Distributions of Rationals on the Unit Interval (or, How to (mis)-Count Rationals)

Linas Vepstas <linas@linas.org>

12 October 2004 (revised 9 February 2005)

Abstract

The distribution of rationals on the unit interval is filled with surprises. As a child, one is told that the rationals are distributed "uniformly" on the unit interval. If one considers the entire set $\mathbb{Q}$, then yes, in a certain narrow sense, this is true. But if one considers just subsets, such as the subset of rationals with "small" denominators, then the distribution is far from uniform and full of counter-intuitive surprises, some of which we explore below. This implies that using "intuition" to understand the rationals and, more generally, the real numbers is a dangerous process. Once again, we see the footprints of the set-theoretic representation of the modular group $SL(2,\mathbb{Z})$ at work.

This paper is part of a set of chapters that explore the relationship between the real numbers, the modular group, and fractals.

1 Distributions Of Rationals on the Unit Interval

The entire field of classical calculus and analysis is based on the notion that the real numbers are smoothly and uniformly distributed on the real number line. When one works with a particular representation of the rational numbers, say the dyadic representation, where each rational is represented by a sequence of binary digits, one gets, 'for free', a measure that goes with that representation. In the case of the dyadics, that measure is the idea that all strings of binary digits are uniformly distributed on the unit interval. This statement is so blatantly obvious and taken for granted that it in fact impedes the understanding of measure. But this will be the topic of this chapter.

There are several different ways of representing the rationals (and thier closures), and these are (as we will see shortly) inequivalent. One way is to represent them with p-adic, or base-p expansions of digits. Another way is to represent them as rationals, that is, as ratios of integers. Each of these representations will result in a uniform distribution of reals on the real number line, when one takes the apropriate limit of allowing p-adic strings with an infinite number of digits, or allowing fractions with arbitrarily large denominators. However, if we work with just finite subsets of p-adic expansions, or finite sets of rationals, one finds that the distributions are far from uniform, and are inequivalent to each other. In particular, this implies that the notion of measure on the real number line has a certain kind of ambiguity associated with it.

The next thing that one finds is that the modular group $SL(2,\mathbb{Z})$ becomes manifest, being the symmetry group that connects together the different representations of the rationals. However, insofar as there is no such thing as a 'real number' except as defined by the closure of the rationals, using a specific representation of the rationals, one has that the real numbers themselves have a modular group symmetry, if only because the underlying representations in terms of p-adic expansions and ratios have this symmetry.

We develop the above wild-sounding claim a bit further in later chapters; here, we show one very simple way in which the modular group, and thus Farey Fractions, manifest themselves on the real number line. We do this by (incorrectly) counting rationals, and then wildly scrambling to find the correct way of counting.

1.1 Simple Counting

Lets begin by trying to enumerate the rationals, and seeing how they fall on the real number line. Start by listing all of the fractions with denomintors from 1 to N, and numerators between 0 and the denomintor. Clearly, many of these fractions will be reducible, i.e. the numerator and denominator have common factors, and thus, in this simple-minded enumeration, some rationals are counted multiple times. In particular, we'll count 0 over and over again: it will be in the list as 0/1, 0/2, 0/3 and so on. Likewise, 1 will appear in this list over and over: as 1/1, 2/2, 3/3, etc. We'll have 1/2 also appearing as 2/4, 3/6 and so on. Although this enumeration of the rationals clearly over-counts, it has the advantage of being extremely simple: it is a subset of the rectangular lattice $\mathbb{Z} \times \mathbb{Z}$. Its the canonical grade-school example of how the rationals are enumerable.

How are these rationals p/q distributed on the real number line? In fancy terms, what is the distribution of this lattice on the real number line? Or, what is the measure induced by the projection of the lattice $\mathbb{Z} \times \mathbb{Z}$ onto the real number line? Unfortunately, using words like "measure" implies the taking of a limit to infinity. Lets stick to the simpler language: we want to make a histogram of the rationals. Lets draw some graphs.

The figure 1 shows this enumeration, up to a denominator of K=4000, carved up into N=720 bins, and normalized to unit density. That is, if $n/720 \le p/q < (n+1)/720$, then we assign the fraction p/q to the n'th bin, and so the graph is a histogram. We might expect this graph to have a huge peak at the bin n=360: after all, this bin will hold 1/2 and 2/4 and 3/6 and in general should have a big surfiet coming from the degeneracy at 1/2. One mght expect peaks at 1/3, and 1/4 and etc, but smaller.

Indeed, there is a big upwards spike at 1/2. But there seems to be a big downwards spike just below, at bin 359, seemingly of equal and opposite size. This is the first surprise. Why is there a deficit at bin 359? We also have blips at 1/3, 1/4, 1/5, 1/6, but not at 1/7: something we can hand-wave away by noting that 720 is 6 factorial. (When one attempts 7!=5040 bins, one finds the peak at 1/7 is there, but the one at 1/11 seems to be missing; clearly having the number of bins being divisible by 7 is important.). The other surprising aspect of this picture is the obvious fractal self-similarity of this histogram. The interval between 1/3 and 1/2 seems to reprise the whole. The tallest blip in the middle of this subinterval occurs at 2/5, which is the Farey mediant of 1/2

Figure 1: Distribution of Simple Rationals into 720 Bins
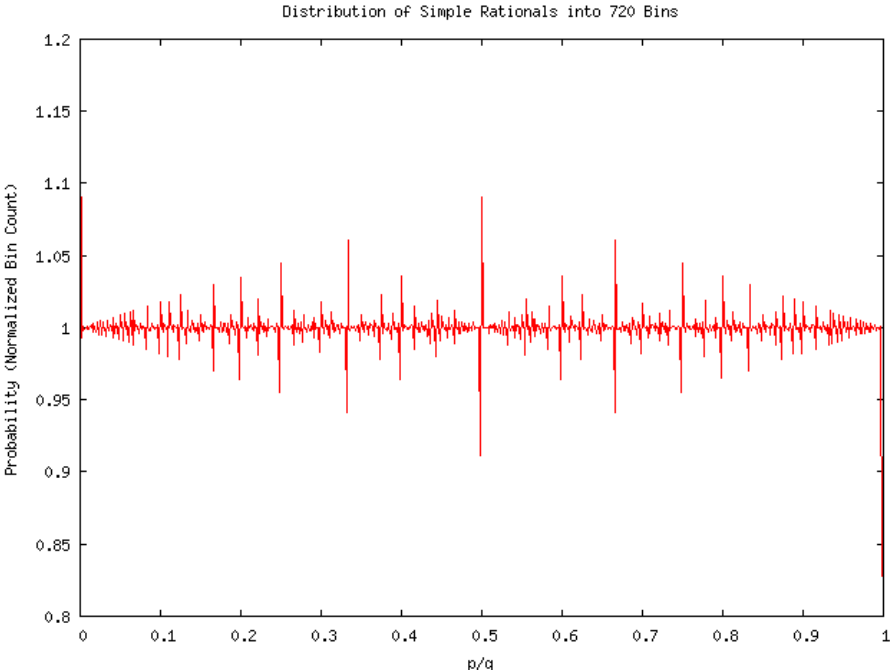
The above is a density graph of the rationals that occur in the simple enumeration, binned into 720 bins, up to a denominator of N=4000. The normalization of the bin count is such that the expected value for each bin is 1.0, as explained in the text.

and 1/3. Why are we getting something that looks like a fractal, when we are just counting rationals? More tanalizingly, why does the fractal involve Farey Fractions?

We suspect that something peculiar happens because the over-counting at 1/2, 2/4'ths etc. falls on exactly the boundary between bins 360 and 359. In fact, any fraction with a denominator that is a multiple of 2,3,4,5, or 6 will have this problem; fractions that have a multiple of 7 in the denominator don't seem to have this problem, perhaps because they are not on a bin boundary. We can validate this idea by binning into 719 bins, noting that 719 is prime. Thus, for the most part, almost all fractions will clearly be in the "middle" of a bin. We expect a flatter graph; the up-down blips should cancel. But it shouldn't be too flat: we still expect a lot of overcounting at 1/2. See below:


Distribution of Simple Rationals into 719 Bins
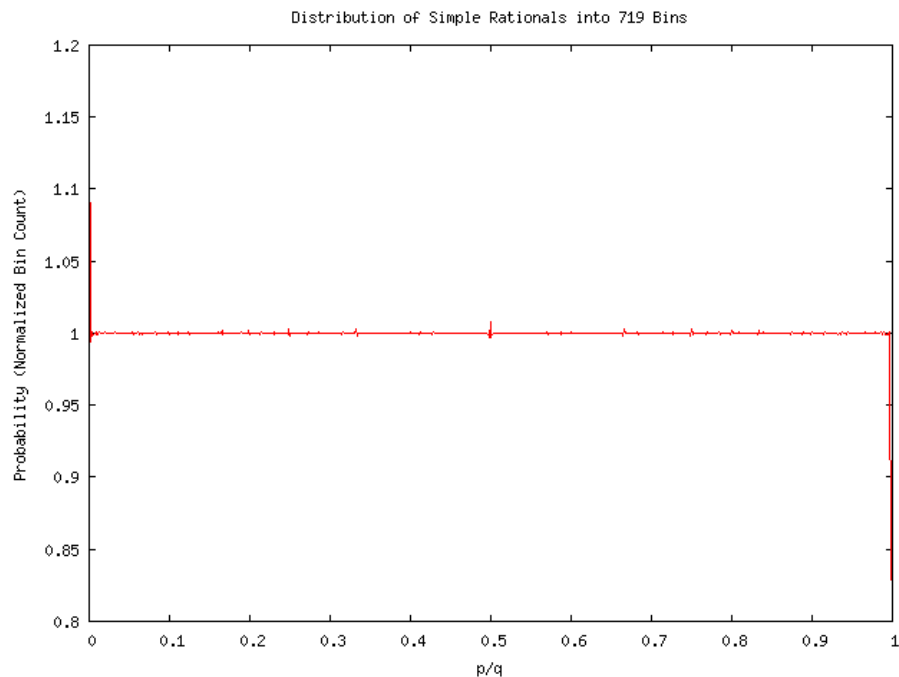
Wow, thats flat! How can this graph possibly be so flat? We should be massively overcounting at 1/2, there should be a big peak there. Maybe its drowned out by the blips at 0 and 1: we are, after all histograming over 8 million fractions, and we expect statistical variations to go as one over the sqaure-root of the sample size. So lets graph the same data, but rescale more appropriately. This is shown below:

4

Distribution of Simple Rationals into 719 Bins
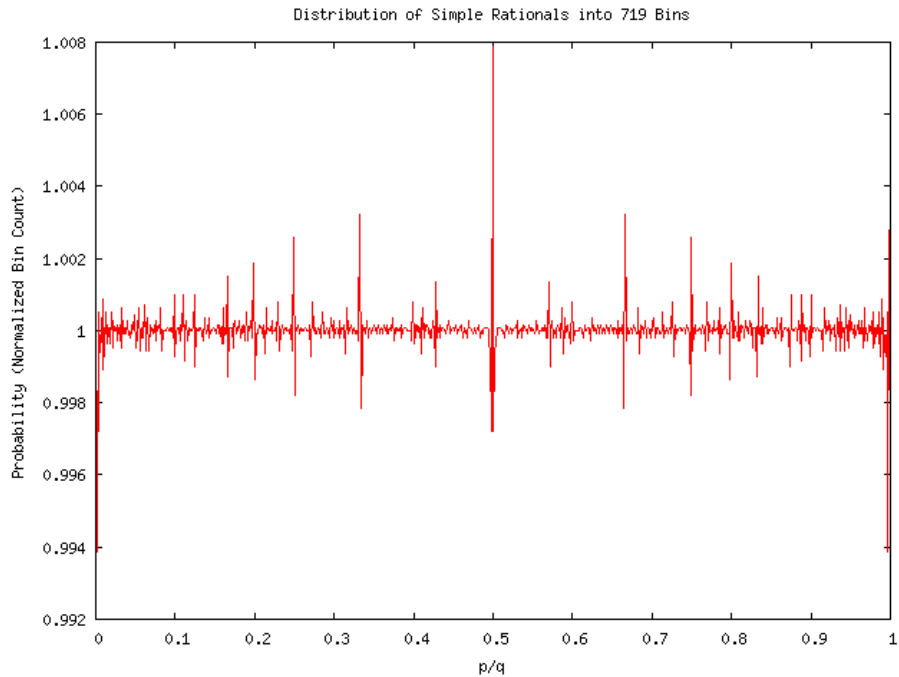
Hmm. Curious. There is indeed a peak at 1/2. But there are also deficits symmetrically arranged at either side. This is still confusing. We might have expected peaks, but no deficits, with the baseline pushed down, to say, 0.999, with all peaks going above, so that the total bin count would still average out to 1.0. But the baseline is at 1.0, and not at 0.999, and so this defies simple intuition. Notice also that the fractal nature is still evident. There are also peaks at 1/3, 1/4, 1/5 and 1/6. But not at 1/7'th. Previously, we explained away the lack of a peak at 1/7'th by arguing about the prime factors of 720; this time, 719 has no prime factors other than itself; thus, this naive argument fails. What do we replace this argument with?

Well, at any rate, lets compare this to the distribution we "should have been using all along", where we eliminate all fractions that are reducible. That is, we should count each rational only once. This mkes a lot more sense, if we are to talk of teh distribution of rationals on the real number line. This is graphed below, again, binned into 719 bins, for all irreducible rationals with denominator less than or equal to 4000:

5

Distribution of Irreducible Rationals into 719 Bins
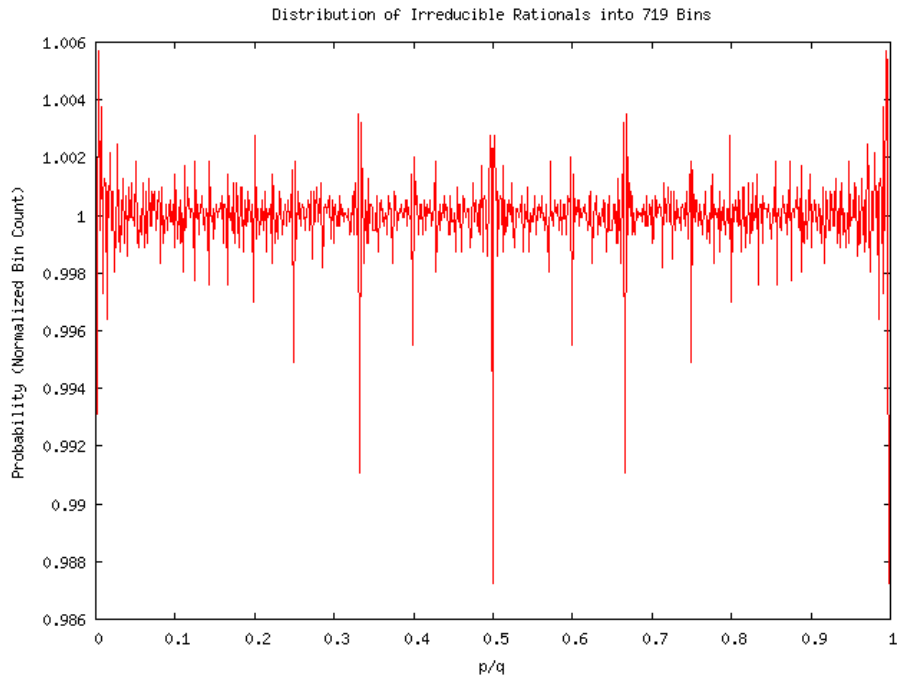
Wow! We no longer have a peak at 1/2. In fact, it sure gives the distinct impression that we are undercounting at 1/2! Holy Banach-Tarski, Batman! What does it mean? Note also the graph is considerably noiser. Compare the scales on the left for a relative measure of the noise. Part, but not all, of the noise is due to the smaller sample size: we are counting fewer fractions: 4863602 are irreducible out of the simple list of 8002000. However, matching the sample sizes does not seem to significantly reduce the small-amplitude noise: qualitatively speaking, the binning of irreducible fractions seems much noisier.

Let us pause for a moment to notice that this noise is not due to some numerical aberation due to the use of floating-point numbers, IEEE or otherwise. The above bincounts are performed using entirely integer math. That is, for every pair of integers p, q, we computed the integer bin number n and the integer remainder $0 \le r < N$ such that $nq = pN + r$ holds as in integer equation, where N was the number of bins. This equation does not have 'rounding error' or 'numerical imprecision'.

Curiously, binning into a non-prime number of bins does seem to reduce the (small-amplitude) noise. Equally curiously, it also seems to erase the prominent features that were occuring ath the Farey Fractions. This is exactly the opposite of the previous experience, where it was bining to a prime that seemed to 'erase' the features. Below is the binning into 720 bins.

Distribution of Irreducible Rationals into 720 Bins
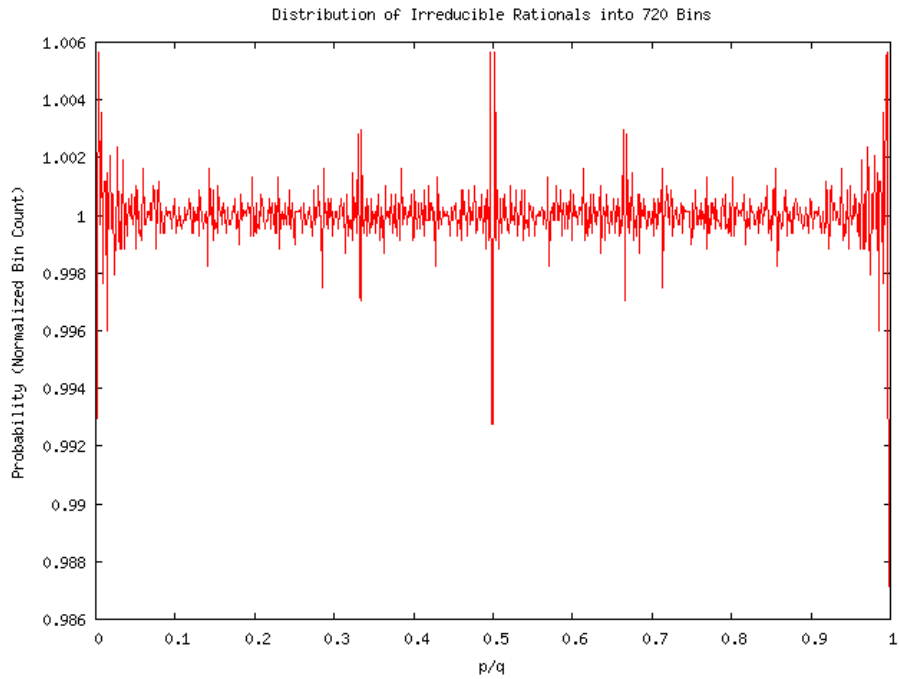
Following the usual laws of statistics and averages, one expects that increasing the sample size reduces the noise. This is true in an absolute sense, but not a relative sense. The graph below shows 720 bins holding all irreducible rationals with denominators less than 16000. The absolute amplitude has been reduced by over a factor of ten compared to the previous graphs; this is not a surprise. We are counting 77809948 irreducible rationals, as opposed to 4863602 before: our sample size is nearly 16 times larger. What is perhaps surprising is that there is relatively far more power in the higher frequencies. There are also still-visible noise peaks near 1/2, 1/3, and 2/3'rds, as well as at 0 and 1.

Distribution of Irreducible Rationals into 720 Bins
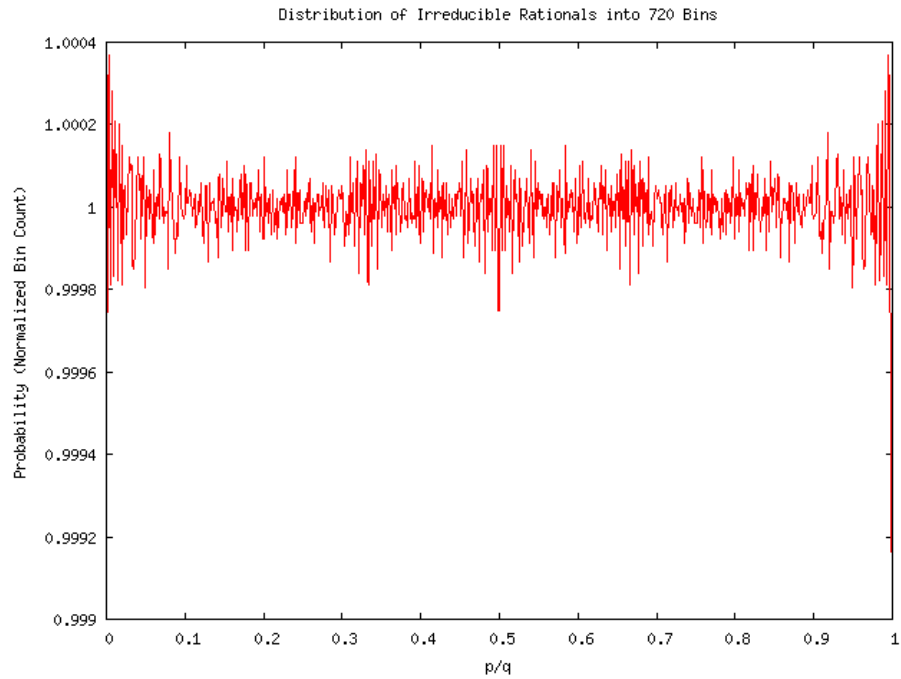
Let reiterate that the noise in this figue is not due to floating-point errors or numerical imprecision. Its really there, deeply embedded in the rationals. As we count more and more rationals, and bin them into a fixed number of bins, then we will expect that the mean deviation about the norm of 1.0 to shrink and shrink, as some power law. It is in this sense that we can say that the rationals are uniformly distributed on the real-number line: greater sample sizes seemingly leads to more uniform distributions, albeit with strangely behaved variances. But even this statement is less than conclusive, because it hides a terrible scale invarience. We have one more nasty histogram to demonstrate.

8

Distribution of Irreducible Fractions into 2880 Bins
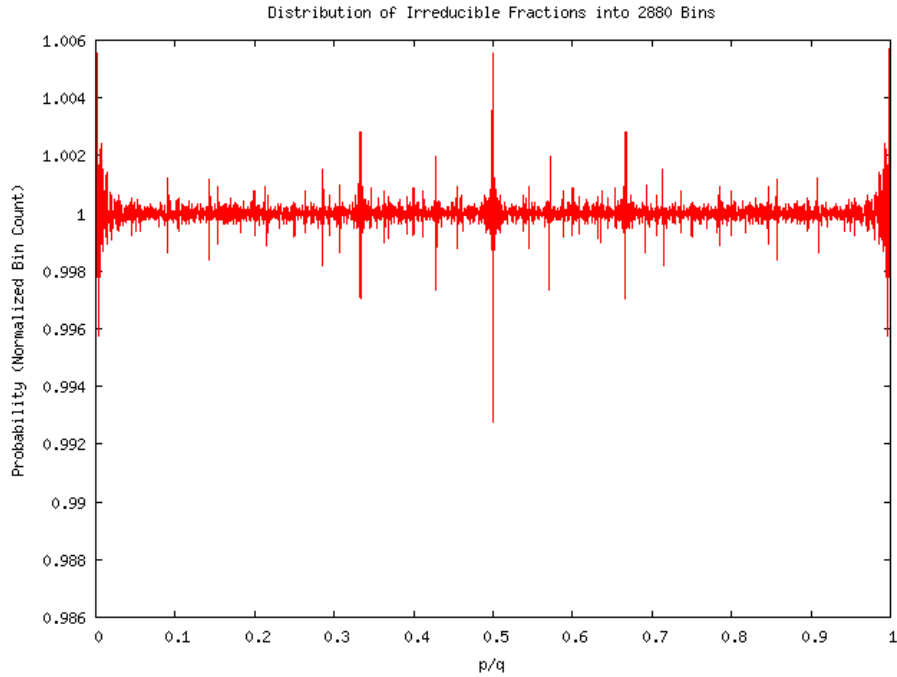
This one shows irreducible fractions with denominators less than 16000, which, as we've mentioned, represents a sample size almost 16 times larger than the first sets of graphs. We bin these into four times as many bins: 2880=4x720. Compare the normalized scale on the vertical axis to the corresponding picture for the smaller sample size and smaller number of bins. The vertical scales are identical, and the sizes of the peaks are identical. Each bin, on average, holds four times as many rationals (16 times as many rationals, 4 times as many bins). We've increased our sample size, but the features are not 'washing out': they are staying constant in size, and are becoming more distinct and well-defined.

1.1.1 Some Notes about Histogramming

In light of the fact that the above graphs have some surprising features, we take a moment to try to be precise about what we mean when we say "histogram" and "normalize".

Lets go back to the first figure. The total number of rationals in the histogram is $K(K+1)/2 = 4000 \times 4001/2 = 8002000$, a little over eight million: a decent sample size. Each bin will have some count C_n of these rationals. We want to talk in statistical terms, so we normalize the bin count as $D_n = NC_n/(K(K+1)/2)$, so that the average value or expected value of D_n is 1.0. That is, we have, by definition,

$$\sum_{n=0}^{N} D_n = N \tag{1}$$

The act of bining a rational p/q requires a division; that is, in order to determine if

9

$n/N \le p/q < (n+1)/N$, a division is unavoidable. However, we can avoid numerical imprecision by sticking to integer division; using floating point here potentially casts a cloud over any results. With integer division, we are looking for n such that $nq \le Np < (n+1)q$; performing this computation requires no rounding or truncation. The largest such integers we are likely to encounter in the previous sections are $2880 \times 16000 \approx 50M$, for which ordinary 32-bit math is perfectly adequate; there is no danger of overflow. If one wanted to go deeper, one could use arbitrary precision libraries; for example, the Gnu Bignum Library, GMP, is freely available. But the point here is that to see these effects, one does not need to work with numbers so large that arbitrary precision math libraries would be required.

1.2 Some Properties of Rational Numbers

So what is it about the rational numbers that makes them behave like this? Lets review some basic properties.

We can envision an arbitrary fraction m/n made out of the integers m and n as corresponding to a point $[m,n]$ on a square lattice. This lattice is generated by the vectors $e_1 = [1,0]$ and $e_2 = [0,1]$: these are the vectors that point along the x and y axes. Every point on the lattice can be represented by the vector $me_1 + ne_2 = [m,n]$ for some integers m and n. This grid is a useful way to think about rationals: by looking out onto this grid, we can "see" all of the rationals, all at once.

Theorem: The lattice $\Lambda = \{[m,n] : m,n \in \mathbb{Z}\}$ is a group under addition. We recall the definition of a group: a group is closed under addition: for $[m,n] \in \Lambda$ and $[p,q] \in \Lambda$ one has $[m+p,n+q] \in \Lambda$. A group has an identity element, which, when added to any other group element, gives that element. For Λ the identity is $[0,0]$. Finally, for every element in the group, the inverse is also in the group. In other words, $[m,n] + [-m,-n] = [0,0]$ and $[-m,-n] \in \Lambda$.

Theorem: The generators e_1 and e_2 generate the lattice. That is, $\Lambda = \{me_1 + ne_2 : m,n \in \mathbb{Z}\}$.

Theorem A lattice point $\omega = me_1 + ne_2 \in \Lambda$ is visible from the origin if and only if $\gcd(m,n) = 1$. By "visible" we mean that if one stood at the origin, and looked out on a field of pegs located at the grid corners, a given peg would not be behind another peg. Here, gcd is the "greatest common divisor", and so the statement is that a peg is visible if and only if the fraction m/n cannot be reduced.

Note that e_1 and e_2 are not the only possible generators. For example, $\omega_1 = [7,4]$ and $\omega_2 = [5,3]$ also generate the lattice. That is, every point in the lattice can be written as $p\omega_1 + q\omega_2$ for some integers p and q. That is, given *any* integers m,n then there exist some integers p,q such that $me_1 + ne_2 = p\omega_1 + q\omega_2$. There are an infinite number of such possible generators. The rest of this section attempts to describe this set of generators.

Theorem: (Apostol Thm 1.1) Two vectors ω_1 and ω_2 generate the lattice if and only if the parallelogram formed by 0, ω_1, $\omega_1 + \omega_2$ and ω_2 does not contain any lattice

points in its interior, or on its boundary. Such a parallelogram is called a *cell* or a *fundamental region*.

The above theorem is not entirely obvious, and it is a good excercise to try to prove it. Note that as a corrolary, we have that both ω_1 and ω_2 are visible from the origin (there would be lattice points on the boundary, if they weren't). In other words, all generators are visible: all generators can be represented by a pair of irreducible fractions. However, not all pairs of fractions generate the lattice, as the next theorem shows.

Theorem: (Apostol Thm 1.2) Let $\omega_1 = ae_1 + ce_2$ and $\omega_2 = be_1 + de_2$ for some integers a, b, c, d. Then ω_1 and ω_2 generate the lattice if and only if $ad - bc = \pm 1$.

We recognize $ad - bc$ as the determinant of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The set of all matrices with determinant equal to $+1$ or -1 is called $SL(2, \mathbb{Z})$, the modular group. Thus, the set of generators of the lattice correspond to elements of the group $SL(2, \mathbb{Z})$.

Theorem: If $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z})$ then $\gcd(a, b) = 1 = \gcd(b, d) = \gcd(c, d) = \gcd(a, c)$.
That is, the fractions given by the rows and columns are all visible from the origin. But we knew that already.

Note that the matrices in $SL(2, \mathbb{Z})$ act on the lattice by simple multiplication: for any point ω in the lattice, the product $A\omega$ is another point in the lattice.

Theorem: If ω is visible, then $A\omega$ is visible as well, for any $A \in SL(2, \mathbb{Z})$. In other words, the action of the modular group on the lattice never mixes visible points with invisible ones. In other words, if ω is an irreducible fraction, then so is $A\omega$; and if ω is reducible, then so is $A\omega$.

Theorem: (Topology) Elements of $SL(2, \mathbb{Z})$ can be paramterized by $\mathbb{Q} \times \mathbb{Z} \times \mathbb{Z}_2$; equivalently, the elements of the modular group can be thought of as a collection of a certain special set of intervals on the real number line.

Proof: We start by freely picking any $a/c \in \mathbb{Q}$ (understanding that we've picked so that a/c is irreducible). For good luck, we pick so that both a and c are positive; we return to negative values later. Then $ad - bc = \pm 1$ implies that $b = (ad \mp 1)/c$. But we can't pick d freely; only certain special values of d result in b being an integer. Mini-theorem: there exists an integer $d \in \{1, 2, ..., c\}$ such that b is an integer. Call this integer d_0. Than another mini-theorem: the resulting b, which we'll call b_0, belongs to the set $\{1, 2, ..., a\}$. So we now have $ad_0 + b_0c = \pm 1$. Next we note that for any $n \in \mathbb{Z}$, the fraction

$$\frac{b_n}{d_n} = \frac{b_0 + na}{d_0 + nc} \tag{2}$$

solves $ad_n + b_nc = \pm 1$. Thus, we've picked freely a number from $\mathbb{Q}$ and another number from $\mathbb{Z}$, and so we've almost proven the paramterization. We have one

bit of remaining freedom, and that is to pick *a* or *c* to be negative: all other sign changes can be eliminated. Finally, note that the fractions a/c and b/d represent an interval on the real number line. One endpoint of the interval can be picked freely; but the other can only be choosen from a limited (but infinite) set.

What have we learned from this excercise? A new way to visualize rationals. In grade school, one traditionally learns to think of rationals as being somehow laid out evenly on the real number line. Maybe we even realize that there is a grid involved: and the grid is comfortingly square and uniform. But in fact, the the irreducible rationals are anything but square and uniform. If we look out onto the grid of pegs, we see some that are very far away, while others are hidden by nearby pegs. If we look off in the direction $\tan\theta = m/n$, the distance $\sqrt{m^2+n^2}$ to the first visible peg at $[m,n]$ seems to be a completely unpredictable and indeed a very chaotic function of θ.

Next, we've learned that the symmetries of a square grid are hyperbolic. Of course, everyone knows that square grids have a translational symmetry; we didn't even mention that. Square grids don't have a rotational symmetry, except for rotations by exactly 90 degrees. But only a few seem to know about the "special relativity" of a square lattice. Just like "real" special relativity, there is a strange squashing and shrinking of lengths while a "cell" or "fundamental region" is squashed. Worse, this group $SL(2,\mathbb{Z})$, known as the *modular group*, is implicated in a wide variety of hyperbolic goings-on. It is a symmetry group of surfaces with constant negative curvature (the Poincare upper half-plane). All sorts of interesting chaotic phenomena happen on hyprbolic surfaces: geodesics diverge from each other, and are thus said to have positive Lyapunov exponent, and the like. The Riemann zeta function, and its chaotic layout of zeros (never mind the chaotic layout of the prime numbers) are closely related. In general, whenever one sees something hyperbolic, one sees chaos. And here we are, staring at rational numbers and seeing something hyperbolic.

It is also worth noting that the square grid, while being a cross-product $\mathbb{Z} \times \mathbb{Z}$ of integers, is not a free product. By this we mean that there are multiple paths from the origin to any given point on the grid: thus, to get to $[1,1]$, we can go right first, and then up, or up first, and then right. Thus the grid is actually a quotient space of a free group. (XXX need to expand on this free vs. quotient thing).

To conclude, we've learned the following: the set of rationals $\mathbb{Q}$ consists entirely of the set of points on the grid that are visible from the origin. The entire set of rationals can be generated from just a pair of rationals a/c and b/d, as long as $ad - bc = \pm 1$. By "generated" we mean that every rational number can be written in the form

$$\frac{am+bn}{cm+dn} \tag{3}$$

where m, n are integers with $\gcd(m,n) = 1$. Of course, this sounds a little dumb, because if $\gcd(m,n) = 1$, then every rational can already be written as m/n. The point here is that the last is a special case of the previous, with $a/c = 1/0$ and $b/d = 0/1$. This is the broadest such generalization of this form.

One oddity that we should notice is the superficial resemblance to Farey addition: given two rational numbers a/c and b/d, we add them not as normal numbers, but

instead combining the numerator and denominator. As we will see, Farey fractions and the modular group are intimately intertwined.

Homework: prove all of the above teorems.

1.3 Orbits of the Modular Group

The symmetries of the histograms are given by $SL(2,\mathbb{Z})$, a fact that we develop in later chapters. (XXX see the other pages on this website for now). Just to provide a taste of what is to come, here's a picture of the orbit of a vector under the action of the group elements of the dyadic representation of the modular group:
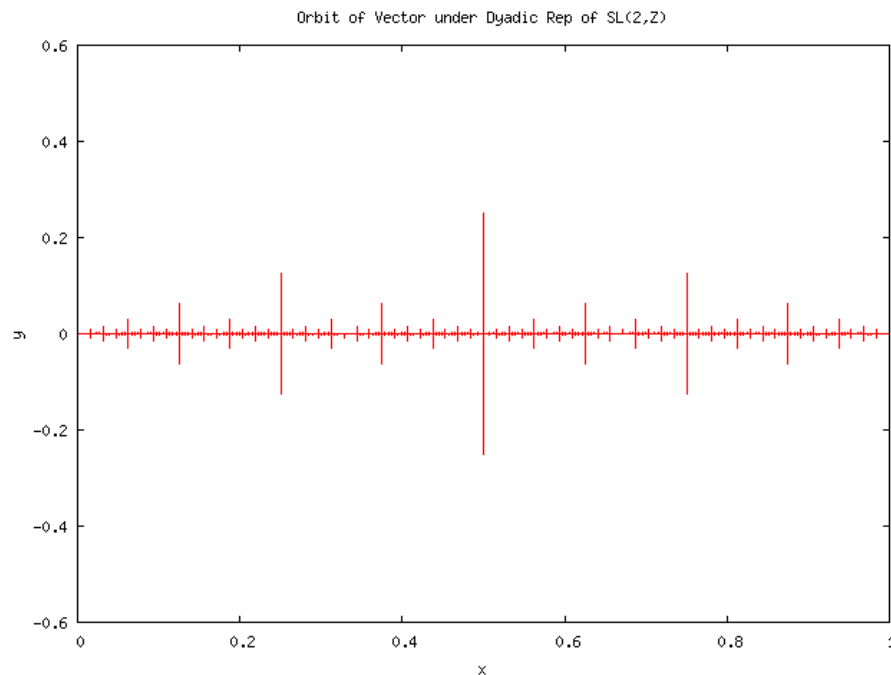
That is, we consider how the vector $(x,y) = (1,0)$ transforms under the group elements generated by

$$g_D = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \text{ and } r_D = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}$$

where we can write a general group element as $\gamma = g^{a_1} r g^{a_2} r g^{a_3} r...rg^{a_N}$. Lets avoid some confusion: the dyadic representation is *not* the canonical rep of $SL(2,\mathbb{Z})$; it is a different rep that is isomorphic; we establish this elsewhere.

In this representation, the only naturally occuring numbers are of the form $p/2^n$, and so the main sequence of the peaks are rooted at 1/2, 1/4, 1/8 etc. To get to the peaks occuring at the Farey numbers, we need to work through the Minkowksi Question mark function, which provides the isomorphism between the Farey Numbers and the

Dyadics. (This is done in the next chapter). (XXXX we really need to re-write this section so it doesn't have to allude to the 'other stuff').

As to the origin of the (white) noise, a better perspective can be gotten on the chapter on continued fraction gaps.

1.4 Conclusion

Write me. Introduce the next chapter.

This is kind-of a to-do list.

It sure would be nice to develop a generalized theory that can work with these peculiar results, and in particular, giving insight into what's happening near 1/2 and giving a quantitative description of the spectra near 1/3 and 2/3, etc. We want to graph the mean-square distribution as a function of sample size. We want to perform a frequency analysis (fourrier transform) and get the power spectrum.

$$f(\tau) = \sum_n c(n) \exp(2\pi i n \tau)$$

We want to explore to what extent the power spectrum has the approximate scaling relationship of a modular form. (We expect this relationship because the fractal self-similarity should manifest itself in the Fourrier spectrum as well, as a scaling relationship. This is not merely "1/f" noise, its more than that.)

When we deal with a finite number of bins, we cannot, of course, get the full symmetry of the modular group. For a finite number of bins, we expect to see the action of only some finite subgroup (or subset) of the modular group. What is that subgroup (subset)? What are its properties?

We also have a deeper question: we will also need to explain why the modular group shows up when one is counting rationals; we will do this in the next chapter, where we discuss the alternate representations of the reals. Its almost impossible to avoid.