RANDOM INVOLUTIONS AND THE NUMBER OF PRIME FACTORS OF AN INTEGER

KIRSTEN WICKELGREN

1. Introduction

Any positive integer n factors uniquely as

$$n = p_1^{e_1} p_2^{e_2} p_3^{e_3} \cdots p_d^{e_d}$$

where $p_1, p_2, \ldots, p_d$ are distinct prime numbers. We can try to understand the behavior of d as n varies. Mathematicians have thought about this for hundreds of years!

An *involution* is a map f such that composing f with itself gives the identity map

$$ff = \mathrm{id}.$$

Let $\mathbb{F}_2$ denote the field with 2 elements, so $\mathbb{F}_2 = \mathbb{Z}/2$.

For each n, there is a beautiful surface called $X_0(n)$ (which is important for many reasons!) and which looks like the surface of a donut with many holes. The number of these holes is called the *genus*, and from the surface $X_0(n)$, one can obtain an involution $\tau(n)$ on $\mathbb{F}_2^{2g(n)}$, where $g(n)$ denotes the genus of $X_0(n)$. (Really, one only obtains this involution up to a certain equivalence called *isomorphism*, but we'll take care of this distinction later.) It is a lovely fact about $X_0(n)$ that for n odd, there are exactly

$$2^{g(n)+2^{d-1}-1} \tag{1}$$

elements of $\mathbb{F}_2^{2g(n)}$ which are fixed by $\tau(n)$. In particular, $g(n)$ and the involution $\tau(n)$ determine d. The reason for this is discussed in Section 6, but for now let's think about some consequences.

$X_0(n)$ is called a *modular curve* and its properties turn out to be extremely important to modern mathematics. The project discussed here is to compute the fixed points of a random involution and compare the number of these to (1). In other words:

1.1. Question. Can we model the number of prime factors of an integer by a random involution?

A positive answer to this question would be great, but even a negative one might tell us something interesting about $X_0(n)$. More specifically, $\tau(n)$ is the involution induced by complex conjugation on a vector space called $H_1(X_0(n), \mathbb{F}_2)$. It is a fact that $\tau(n)$ is determined up to isomorphism by its fixed points (see Theorem 3.6 below), so a negative

Date: Summer 2012.

answer could tell us that $\tau(n)$ is not random in a certain sense. Conversely, a positive answer could tell us that $\tau(n)$ is random.

We need to be more precise about Question 1.1 and what the word "random" means. For instance, for each n, the involution $\tau(n)$ is a particular map, and so of course it is not random at all. Let's say what we really mean, and collect some tools to understand involutions and prime factorizations.

2. How big is a random finite set?

Suppose you want to describe all the ways to draw an equilateral triangle of side length 1 in the plane. This can be done by first choosing where to draw the center and then choosing a direction for the line from a vertex to the center. This direction can be any angle from 0 to 2π, but choosing the angle 0 is no different from choosing the angle $2\pi/3$, because rotating the triangle $2\pi/3$ about its center determines an *automorphism* of the triangle, i.e. a map f from the triangle to itself such that there exists another map g from the triangle to itself with fg and gf equal to the identity map. We can thus describe the ways to draw the triangle as $\mathbb{R}^2 \times \mathbb{R}/(\frac{2}{3}\pi)$.

If instead of an equilateral triangle, you draw a triangle whose sides are all of different lengths, the vertices can be distinguished one from another, giving that the ways to draw the triangle forms the space $\mathbb{R}^2 \times \mathbb{R}/(2\pi)$. The lack of automorphisms increased the size of the space of possibilities.

It is a general principle that one should often count objects weighted by $1/|\operatorname{Aut}|$, where $|\operatorname{Aut}|$ denotes the number of automorphisms. With this principle, we can make sense of the question:

How many finite sets are there?

Here's how: two objects A and B are said to be *isomorphic* if there are maps $f : A \to B$ and $g : B \to A$ such that the compositions fg and gf are the identity on B and A respectively. Note that any two sets with the same number of elements are isomorphic. So, up to isomorphism, the finite sets are

$$\emptyset, \{1\}, \{1,2\}, \{1,2,3\} \ldots \{1,2,3,\ldots,k\} \ldots . \tag{2}$$

Note that the automorphisms of the set $\{1,2,3,\ldots,k\}$ are the permutations on k elements, and that there are $k!$ of these. In total, we therefore have

$$\sum_{k=0}^{\infty} \frac{1}{k!} = e$$

finite sets.

2.1. *Example.* The probability that a finite set has size k is $1/(k!e)$.

2.2. *Question.* How big is a random finite set?

2.3. *Question.* Given two finite sets, what is the probability that the larger has at least two more elements that the smaller?

2.4. *Question.* Given two finite sets, how much larger do you expect the larger one to be?

3. Random involutions

How many $\mathbb{F}_2$-vector spaces are there?

For any positive integer m,

$$\mathbb{F}_2^m$$

has the structure of a $\mathbb{F}_2$-*vector space*, i.e. you can add two elements of $\mathbb{F}_2^m$ together, and you can multiply any element of $\mathbb{F}_2^m$ by any element of $\mathbb{F}_2$. Conversely, any $\mathbb{F}_2$-vector space can be written as $\mathbb{F}_2^m$ for some m. This m is unique, and is called the *dimension* of the vector space and the function taking an $\mathbb{F}_2$-vector space to its dimension will be denoted $\dim$. Elements of vector spaces are called *vectors*, and given vectors $v_1, \dots, v_n$, the *span* of $v_1, \dots, v_n$ is the set $\{a_1v_1 + a_2v_2 + \dots a_nv_n : a_i \in \mathbb{F}_2\}$. A nice general reference on vector spaces and their properties is [Art91, Ch 3,4].

The automorphisms of $\mathbb{F}_2^m$ are the $m \times m$ invertible matrices, which form $\mathrm{GL}_m(\mathbb{F}_2)$. (Multiplication of matrices makes GL_m into an object called a *group*.) The number of elements of $\mathrm{GL}_m(\mathbb{F}_2)$ can be computed as follows. The first column can be any non-zero element of $\mathbb{F}_2^m$. There are $2^m - 1$ of these. After choosing the first n-columns, the $n+1$st column can be any element not in the span of the first n-columns. So there are $2^m - 2$ choices for the second column, and more generally, $2^m - 2^n$ choices for the $n+1$st column. Thus

$$|\mathrm{GL}_m \mathbb{F}_2| = \prod_{n=1}^{m} (2^m - 2^{n-1})$$

[Lan02, XIII Ex 15 p 546] and there are

$$\sum_{m=1}^{\infty} \prod_{n=1}^{m} \frac{1}{2^m - 2^{n-1}}$$

$\mathbb{F}_2$-vector spaces.

3.1. *Question.* What's the probability that an $\mathbb{F}_2$-vector space has dimension 3? What's the probability that an $\mathbb{F}_2$-vector space has dimension 5? Compare the sizes of these probabilities e.g. how many times more likely is it that an $\mathbb{F}_2$-vector space has dimension 3? Is it more likely that an $\mathbb{F}_2$-vector space has dimension 100 or 101?

How many $\mathbb{F}_2$-vector spaces with involutions are there?

To compute the number of $\mathbb{F}_2$-vector spaces with involution, we need two things:

(1) A list of the isomorphism classes of $\mathbb{F}_2$-vector spaces with involution.

(2) A computation of the number of automorphisms of each $\mathbb{F}_2$-vector space with involution in the list from (1).

Note that it is these same two things that we needed to compute the number of finite sets and the number of $\mathbb{F}_2$-vector spaces. For instance, the list of isomorphism classes of finite sets is given by equation (2) in Section 2.

In the definition of an $\mathbb{F}_2$-vector space, "$\mathbb{F}_2$" can be replaced by a more general type of object called a *ring* and the resulting "vector space" is then called a *module* over that ring. (This is discussed in more detail below.) An $\mathbb{F}_2$-vector space with an involution is precisely the same as a module over a certain ring called $\mathbb{F}_2[\mathbb{Z}/2]$. This means that (1) is equivalent to understanding the isomorphism classes of $\mathbb{F}_2[\mathbb{Z}/2]$-modules, and (2) is equivalent to computing the number of automorphisms of the corresponding modules over $\mathbb{F}_2[\mathbb{Z}/2]$. Abstract algebra gives us the answers to (1) and (2). The answer for (1) is contained in Theorem 3.6 and the answer to (2) is Proposition 3.8. (You only need to understand the statements of Theorem 3.6 and Proposition 3.8. You can ignore the arguments used to justify them, and you might want to if this is your first time thinking about modules.)

Here's the definition of $\mathbb{F}_2[\mathbb{Z}/2]$. Consider the elements of $\mathbb{Z}/2$. Let's give them names. The identity element (which is usually denoted 0) will be called $\mathbb{1}$, and τ will denote the non-identity element. The elements of $\mathbb{F}_2$ will be called 0 and 1. Consider the four elements

$$\{0, \mathbb{1}, \tau, \mathbb{1} + \tau\}.$$

These elements could also be written

$$\{a_1 \cdot \mathbb{1} + a_2 \cdot \tau : a_1, a_2 \in \mathbb{F}_2\}$$

where the operations $\cdot$ and $+$ satisfy the natural rules suggested by the notation. For example, $0 \cdot \tau = 0$, and $1 \cdot \tau = \tau$. These four elements form a structure called a *ring*, meaning there is a natural way to add and multiply these elements. Addition and multiplication are defined by the rules

$$(a_1 \cdot \mathbb{1} + a_2 \cdot \tau) + (a_1' \cdot \mathbb{1} + a_2' \cdot \tau) = (a_1 + a_1') \cdot \mathbb{1} + (a_2 + a_2') \cdot \tau$$

$$(a_1 \cdot \mathbb{1} + a_2 \cdot \tau)(a_1' \cdot \mathbb{1} + a_2' \cdot \tau) = (a_1 a_1' + a_2 a_2') \cdot \mathbb{1} + (a_1 a_2' + a_1' a_2) \cdot \tau.$$

The ring formed by these four elements is $\mathbb{F}_2[\mathbb{Z}/2]$.

3.2. *Exercise.* Compute $\tau(\mathbb{1} + \tau)$ and $(\mathbb{1} + \tau)^2$ in $\mathbb{F}_2[\mathbb{Z}/2]$.

3.3. *Example.* Elements of $\mathbb{F}_2[X]$ are polynomials $a_0 + a_1 X + a_2 X^2 + \ldots + a_n X^n$ and polynomials can be added and multiplied. It follows that $\mathbb{F}_2[X]$ is a ring. There is another ring called $\mathbb{F}_2[X]/\langle X^2 \rangle$ which is obtained from $\mathbb{F}_2[X]$ by only remembering $a_0 + a_1 X$. So $\mathbb{F}_2[X]/\langle X^2 \rangle = \{a_0 + a_1 X : a_0, a_1 \in \mathbb{F}_2\}$ and the operations of multiplication and addition come from multiplication and addition of polynomials, except any monomials of degree 2 or higher are erased [Art91, Ch 10]. Convince yourself that by identifying X with $\mathbb{1} + \tau$, we identify the rings $\mathbb{F}_2[X]/\langle X^2 \rangle$ and $\mathbb{F}_2[\mathbb{Z}/2]$.

A *module* over a ring R is a set M such that elements of M can be added together, and elements of M can be multiplied by elements of R. There is also an element 0 of M such that $m+0 = m$ for all $m \in M$, an element $-m$ such that $m+(-m) = 0$, and multiplication by elements of R distributes over addition, i.e. $r(m_1+m_2) = rm_1+rm_2$ for all $m_1, m_2 \in M$ and $r \in R$. A reference on modules is [Art91, Ch 12].

3.4. *The relationship between involutions and $\mathbb{F}_2[\mathbb{Z}/2]$-modules.* An involution f on an $\mathbb{F}_2$-vector space $\mathbb{F}_2^n$ gives $\mathbb{F}_2^n$ the structure of an $\mathbb{F}_2[\mathbb{Z}/2]$-module by defining for all $(a_1\mathbb{1} + a_2\tau) \in \mathbb{F}_1[\mathbb{Z}/2]$ and all v in $\mathbb{F}_2^n$,

$$(a_1\mathbb{1} + a_2\tau)v = a_1v + a_2fv \in \mathbb{F}_2^n.$$

Conversely an $\mathbb{F}_2[\mathbb{Z}/2]$-module is an $\mathbb{F}_2$-vector space with an involution, by letting the vectors of the vector space be the elements of the module and setting the involution equal to τ. Two important examples of $\mathbb{F}_2[\mathbb{Z}/2]$-modules are $\mathbb{F}_2[\mathbb{Z}/2]$ itself, and $\mathbb{F}_2$. In the former case, the underlying vector space is $\mathbb{F}_2^2$ and the involution f is

$$f = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

In the latter case, the underlying vector space is $\mathbb{F}_2$ and the involution is the identity.

3.5. *Exercise.* Write down a matrix corresponding to the involution on $\mathbb{F}_2^{2a+b}$ given by the $\mathbb{F}_2[\mathbb{Z}/2]$-module $\mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b$. For starters, take $a = 2$ and $b = 1$.

3.6. *Theorem. — Any $\mathbb{F}_2[\mathbb{Z}/2]$-module is isomorphic to $\mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b$ for a unique pair of non-negative integers* (a, b).

Since the pair (a, b) is unique, we can consider a and b as functions which take an $\mathbb{F}_2[\mathbb{Z}/2]$ module and return a non-negative integer. We'll adopt this convention, although it's not standard. Note that the dimension of an $\mathbb{F}_2[\mathbb{Z}/2]$-module is

$$\dim = 2a + b.$$

Theorem 3.6 is true because the ring $\mathbb{F}_2[\mathbb{Z}/2]$ is an example of a *principal ideal domain*. See [Lan02, III §7 p. 146] if you are interested. Theorem 3.6 says that the isomorphism classes of involutions on $\mathbb{F}_2^n$ are in bijection with pairs of non-negative integers (a, b) such that $2a + b = n$. In particular, there are $\lceil n/2 \rceil$ of these given by

$$\mathbb{F}_2^n, \mathbb{F}_2[\mathbb{Z}/2] \times \mathbb{F}_2^{n-2}, \mathbb{F}_2[\mathbb{Z}/2]^2 \times \mathbb{F}_2^{n-4} \ldots.$$

3.7. *Exercise.* Let n be an odd integer with d distinct prime factors. Recall that $\tau(n)$ is an involution on $\mathbb{F}_2^{2g(n)}$ with exactly $2^{g(n)+2^{d-1}-1}$ fixed points. Show that $\tau(n)$ is the involution associated to the $\mathbb{F}_2[\mathbb{Z}/2]$-module

$$\mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b$$

for $a = g(n) - 2^{d-1} + 1$ and $b = 2(2^{d-1} - 1)$.

We now wish to understand the automorphisms of $\mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b$. An automorphism

$$T : \mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b \to \mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b$$

is given by an $(a+b) \times (a+b)$ matrix, separated into blocks of size $a \times a$, $a \times b$, $b \times a$ and $b \times b$

$$T = \left[\begin{array}{cccc|ccc} e_{11} & e_{12} & \dots & e_{1a} & f_{11} & \dots & f_{1b} \\ e_{21} & e_{22} & \dots & e_{2a} & f_{21} & \dots & f_{2b} \\ \vdots & & & \vdots & \vdots & & \vdots \\ e_{a1} & e_{a2} & \dots & e_{aa} & f_{a1} & \dots & f_{ab} \\ \hline c_{11} & c_{12} & \dots & c_{1a} & d_{11} & \dots & d_{1b} \\ \vdots & & & \vdots & \vdots & & \vdots \\ c_{b1} & c_{b2} & \dots & c_{ba} & d_{b1} & \dots & d_{bb} \end{array}\right]$$

with e_{ij}, f_{ij} in $\mathbb{F}_2[\mathbb{Z}/2]$, and c_{ij}, d_{ij} in $\mathbb{F}_2$.

By Example 3.3, we can replace $\mathbb{F}_2[\mathbb{Z}/2]$ with $\mathbb{F}_2[X]/\langle X^2 \rangle$. Since X acts by 0 on $\mathbb{F}_2$, it follows that $Xf_{ij} = 0$, which implies that $f_{ij} = Xf'_{ij}$ for some $f'_{ij} \in \mathbb{F}_2$.

T is an automorphism exactly when the determinant of T in $\mathbb{F}_2[X]/\langle X^2 \rangle$ has its constant coefficient equal to 1, i.e. $\mathrm{Det}\, T = 1 + a_1 X$ for $a_1 \in \mathbb{F}_2$. Viewing $\mathrm{Det}\, T$ as a polynomial in the variable X, we can write this as follows: T is an automorphism if and only if $(\mathrm{Det}\, T)(0) = 1$. Since the determinant is a polynomial in the matrix entries, we are free to consider the matrix entries as polynomials in X and evaluate at 0 before taking the determinant. Thus, f'_{ij} can be any element of $\mathbb{F}_2$ for all i, j and the coefficient of X in all the e_{ij} can also vary freely.

Since the determinant of

$$\left[\begin{array}{c|c} E & 0 \\ \hline C & D \end{array}\right]$$

is $\mathrm{Det}\, E\, \mathrm{Det}\, D$, we have that the automorphisms of $\mathbb{F}_2[X]/\langle X^2 \rangle^a \times \mathbb{F}_2^b$ are in bijection with the matrices

$$\left[\begin{array}{c|c} E + E'X & F'X \\ \hline C & D \end{array}\right]$$

with $E \in \mathrm{GL}_a\, \mathbb{F}_2$, $D \in \mathrm{GL}_b\, \mathbb{F}_2$, $C \in \mathrm{Mat}_{b\times a}\, \mathbb{F}_2$, $E' \in \mathrm{Mat}_a\, \mathbb{F}_2$, and $F' \in \mathrm{Mat}_{a\times b}\, \mathbb{F}_2$. Thus:

3.8. *Proposition. — Let (a, b) be a pair of non-negative integers. The $\mathbb{F}_2[\mathbb{Z}/2]$-module $\mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b$ has exactly*

$$|\mathrm{GL}_a\, \mathbb{F}_2||\mathrm{GL}_b\, \mathbb{F}_2||\mathrm{Mat}_{b\times a}\, \mathbb{F}_2|^2|\mathrm{Mat}_{a\times a}|$$

automorphisms.

3.9. *Exercise.* Express the number of automorphisms of $\mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b$ as a polynomial in a and b.

3.10. *Exercise.* Let (a, b) be a pair of non-negative integers such that $2a + b = n$. What is the probability that an involution on $\mathbb{F}_2^n$ is isomorphic to the involution corresponding to $\mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b$?

3.11. *Remark.* There are finitely many involutions on $\mathbb{F}_2^n$. Call the number of these D. Some finite subset of these will correspond to an $\mathbb{F}_2^n$-module isomorphic to $\mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b$. Call the number of these N. The quotient N/D is equal to

$$\frac{1/|\operatorname{Aut}(a,b)|}{\sum_{a',b'} 1/|\operatorname{Aut}(a',b')|}$$

where $|\operatorname{Aut}(a',b')|$ denotes the number of automorphisms of the $\mathbb{F}_2[\mathbb{Z}/2]$-module $\mathbb{F}_2[\mathbb{Z}/2]^{a'} \times \mathbb{F}_2^{b'}$ and the sum is taken over pairs of non-negative integers (a', b') such that $2a'+b' = n$.

3.12. *Exercise.* What is the expected number of fixed points of an involution on $\mathbb{F}_2^n$?

3.13. *Exercise.* How many $\mathbb{F}_2$-vector spaces with involution are there?

3.14. *Exercise.* Fix a positive integer n. What is the probability that a randomly chosen $\mathbb{F}_2$-vector space with involution will have dimension n? In other words, what is the probability that a random $\mathbb{F}_2$-vector space with involution is isomorphic to some $\mathbb{F}_2[\mathbb{Z}/2]^a \times \mathbb{F}_2^b$ with $2a + b = n$?

3.15. *Exercise.* Fix a positive integer N. What is the expected value of $a - \dim$ for a random involution on a vector space of dimension $> N$? Note that taking $N = -1$ gives the expected value of $a - \dim$ for a random involution. The expected value of a function is the sum over the possibilities of the probability of the possibility times the value of the function i.e.

$$\text{Expectedvalue}(f) = \sum_{\text{possibilities } a} \text{Probability}(a) f(a).$$

4. Some information on the distribution of primes

Mertens' 2nd theorem states that

$$|\sum_{p \leq n} \frac{1}{p} - \log\log n - \mathbb{M}| \leq \frac{4}{\log(n+1)} + \frac{2}{n \log n} \tag{3}$$

where $\log$ denotes the logarithm base e and $\mathbb{M}$ denote's the Meissel-Mertens constant which is about

$$.2614972\ldots.$$

In particular,

$$\lim_{n\to\infty} (\sum_{p\leq n} \frac{1}{p} - \log\log n) = \mathbb{M}.$$

Let $d(n)$ denote the number of distinct prime factors of n. Let $\pi(n)$ denote the number of primes $\leq n$. We approximate $\sum_{n=N+1}^{M} d(n)$. For a prime $p \leq M$, the number of n in

$[N+1, M]$ which are divisible by p is $(M-N)/p + r_p$, where $r \in [0,1)$. Thus

$$(M-N)\sum_{p\leq M}\frac{1}{p} \leq \sum_{n=N}^{M} d(n) \leq (M-N)\sum_{p\leq M}\frac{1}{p} + \pi(M) \tag{4}$$

4.1. *Exercise.* Combine (4) and (3) to obtain bounds for $\sum_{n=N}^{M} d(n)$ that do not contain sums over all primes $\leq M$.

4.2. *Exercise.* Obtain bounds for $\sum_{n>N}^{M} d(n)$ where n only runs over the odd integers contained in $(N, M]$.

The Prime Number Theorem says that $\lim_{n\to\infty} \frac{\pi(n)}{n/\log n} = 1$. (The difference $\pi(n) - \frac{n}{\log n}$ is related to the Riemann hypothesis. The Riemann hypothesis is perhaps the most famous unsolved problem in math.) More generally, let $\pi(n, k)$ denote the number of positive integers $\leq n$ which such that $d = k$. Let

$$G_k(n) = \frac{n}{\log n}\frac{(\log\log n)^{k-1}}{(k-1)!}.$$

Then

$$\lim_{n\to\infty} \frac{\pi(n,k)}{G_k(n)} = 1.$$

The Hardy-Ramanujan inequality gives that there exist constants c_1 and c_2 such that

$$\pi(n,k) \leq c_1 \frac{n}{\log n}\frac{(\log_2 n + c_2)^{k-1}}{(k-1)!}$$

for all $n \geq 3$ and all $k \geq 1$.

See [HT88] for more information on $\pi(n, k)$.

5. Project description

The goal is to compare random involutions with $\tau(n)$ for a lot of odd values of n.

The genus $g(n)$ of $X_0(n)$ has a rather involved formula, which is given below as (5). However, the dominant term should be $\frac{n}{12}$. Consider the following model or approximation of $\tau(n)$.

5.1. Model. $\tau(n)$ is similar to a random involution on an $\mathbb{F}_2$-vector space of dimension $\geq \frac{n}{12}$.

5.2. *Remark.* A random involution is much more likely to have small dimension than large dimension (see Exercise 3.14.) However $g(n)$ gets arbitrarily large as n gets large.

Model 5.1 is forcing the random involution to have its underlying vector space be reasonably large.

5.3. Project. Test 5.1 as follows. Recall that the value of $a - \frac{1}{2}\dim$ for $\tau(n)$ is $-2^{d-1}+1$ for n odd by Exercise 3.7. Recall as well that we have estimates for $\sum_{n>N}^{M} d(n)$ where n runs over the odd integers from $N+1$ to M by Exercise 4.2 . For n odd $N < n \leq M$, compute the expected value of $a - \dim$ of a random involution of dimension $\geq \frac{n}{12}$, and call the result $E(n)$. This was Exercise 3.15. Compute $S(N,M) = \sum_{N<n}^{M} \log_2(1-E(n))$, where this sum runs over odd n. Compare $S(N,M)$ to the estimates obtained for $\sum_{n>N}^{M} d(n)$ in Exercise 4.2 by giving bounds for

$$\frac{S(N,M)}{\sum_{n>N}^{M} d(n)}.$$

Can you compute the limit

$$\lim_{N\to\infty}\lim_{M\to\infty}\frac{S(N,M)}{\sum_{n>N}^{M} d(n)}?$$

Or

$$\lim_{N\to\infty}\frac{S(N,2N)}{\sum_{n>N}^{2N} d(n)}?$$

We can also try to be more precise about $g(n)$. The genus $g(n)$ of $X_0(n)$ is given by the formula:

$$g(n) = 1 + \frac{\mu}{12} - \frac{\nu_2}{4} - \frac{\nu_3}{3} - \frac{\nu_\infty}{2} \tag{5}$$

where for n odd

$$\begin{aligned}
\mu &= n\prod_{p|n}(1+p^{-1}) \\
\nu_2 &= \prod_{p|n}(1+(\frac{-1}{p})) \\
\nu_3 &= \begin{cases} 0 \text{ if } 9|n, \\ \prod_{p|n}(1+(\frac{-3}{p})) \text{ otherwise.}\end{cases} \\
\nu_\infty &= \sum_{d|n}\varphi((d,n/d))
\end{aligned}$$

Here φ is Euler's function, $\varphi(1)=1$, and $(\frac{-}{p})$ is the quadratic residue symbol, so

$$(\frac{-1}{p}) = \begin{cases} 1 \text{ if } p \equiv 1 \mod (4), \\ -1 \text{ if } p \equiv 3 \mod (4), \end{cases}$$

$$(\frac{-3}{p}) = \begin{cases} 0 \text{ if } p = 3, \\ 1 \text{ if } p \equiv 1 \mod (3), \\ -1 \text{ if } p \equiv 2 \mod (3), \end{cases}$$

See [Shi71, Prop 1.40, Prop 1.43].

5.4. Refined project. Find lower bounds for $g(n)$ and replace $\frac{n}{12}$ by these bounds in 5.3. How does $S(N, M)$ change as you adjust these bounds?

It would be nice to know if a randomly chosen $\tau(n)$ is similar to a randomly chosen involution. This seems hard to me, but here are possible projects along these lines:

5.5. *Project.* Choose a random $\mathbb{F}_2[\mathbb{Z}/2]$ module. Let E_1 be the expected value of $a - \mathrm{dim}$. Choose a random non-negative integer, where random means that the probability of choosing k is $1/(k!e)$. Try to compute the expected number E_2 of prime factors d of k. Compare E_1 and $2^{-E_2+1} - 1$. Interpret the comparison using Exercise 3.7.

5.6. *Project.* Choose a random non-negative integer, where random means that the probability of choosing k is $1/(k!e)$. Try to compute the expected number of prime factors d of k . Try to compute the expected value of $2^{g(k)+2^{d(k)-1}-1}$. Compare this to the expected number of fixed points of a random $\mathbb{F}_2$-vector space with involution.

There is an operation called direct sum on modules and denoted by $\oplus$. In fact, taking the direct sum of finitely many modules is the same as taking their product. It would be nice if the involution $\tau(N, M) = \oplus_{n=N+1}^{M} \tau(n)$ on $\oplus_{n=N+1}^{M} \mathbb{F}_2^{2g(n)}$ could be modeled by random involutions. This again seems hard to me, but hopefully I'm wrong, so here is a project about this. (Note that since these direct sums are finite, you are welcome to replace $\oplus$ with the product. The reason for this notation is that if we let $M \to \infty$, the product and direct sum will be different, and I think $\oplus$ is more appropriate, because it would correspond to H_1 of $\coprod X_0(n)$.)

5.7. *Project.* If possible, compare $\tau(N, M)$ with the direct sum of randomly chosen involutions on $\mathbb{F}_2^{2g(n)}$ as n ranges over odd integers between $N + 1$ and M.

6. Notes on the project

It is a result of Ogg, Akbas, and Singerman that for n odd, the real points of $X_0(n)$ form exactly 2^{d-1} circles, where d is the number of distinct prime factors of n. See [AS92, p. 6] [Sno11, Prop 1.2.1]. (The reason for this result is that for each primary part of the marked cyclic subgroup of an elliptic curve with level structure, complex conjugation can either act trivially or not.) It follows that

$$H_1(X_0(n), \mathbb{Z}) = \mathbb{Z}[\mathbb{Z}/2]^a \oplus \mathbb{Z}(1)^{b'} \oplus \mathbb{Z}^{b'},$$

with $b' = 2^{d-1} - 1$, for instance by [Wic12, Prop 3.1]. Thus $a = g(n) - 2^{d-1} + 1$.

As pointed out to me by Andrew Snowden, if the real structure on $X_0(n)$ is twisted in a certain way, the number of connected components of real points is related to the size of the class groups of certain number fields. So randomness in those class groups is related to randomness in the corresponding involution. It's fun to recall the Cohen-Lenstra heuristics in this context.

This project is partially based on joint work with Andrew Snowden.

References

[Art91] Michael Artin, *Algebra*, Prentice Hall Inc., Englewood Cliffs, NJ, 1991. MR 1129886 (92g:00001)

[AS92] M. Akbas and D. Singerman, *Symmetries of modular surfaces*, Discrete groups and geometry (Birmingham, 1991), London Math. Soc. Lecture Note Ser., vol. 173, Cambridge Univ. Press, Cambridge, 1992, pp. 1–9. MR 1196910 (93j:20099)

[HT88] Adolf Hildebrand and Gérald Tenenbaum, *On the number of prime factors of an integer*, Duke Math. J. **56** (1988), no. 3, 471–501. MR 948530 (89k:11084)

[Lan02] Serge Lang, *Algebra*, third ed., Graduate Texts in Mathematics, vol. 211, Springer-Verlag, New York, 2002. MR 1878556 (2003e:00003)

[Shi71] Goro Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publications of the Mathematical Society of Japan, No. 11. Iwanami Shoten, Publishers, Tokyo, 1971, Kanô Memorial Lectures, No. 1. MR 0314766 (47 #3318)

[Sno11] Andrew Snowden, *Real components of modular curves*, arXiv:1108.3131v1, 2011.

[Wic12] Kirsten Wickelgren, *2-nilpotent real section conjecture*, arXiv:1006.0265v2, 2012.

E-mail address: kwickelg@math.harvard.edu