SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

The Sato-Tate Conjecture

av

Johan Frisk

2012 - No 9

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

The Sato-Tate Conjecture

Johan Frisk

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Torsten Ekedahl/Rikard Bögvad

2012

Abstract

This thesis describes the Sato-Tate conjecture of the number of points on an elliptic curve over finite fields. I will show how this result can be derived using heuristics and numerical experiments. Some theory about elliptic curves will be provided to give a context for the conjecture. Furthermore I will describe applications of some of the methods used in this thesis.

Contents

1 The Sato-Tate conjecture

The problem is to determine the number of solutions N_p to equations of the type:

$$y^2 = x^3 + Ax + B \tag{1}$$
$$4A^3 - 27B^2 \neq 0$$

over a *finite field* F with p elements , where p is a prime number.
The restriction on A and B is a necessary technical condition which will be discussed more thoroughly later. We mention that in the case where $4A^3 - 27B^2 = 0$, the curves take a more simple form and the counterpart to the Sato-Tate conjecture is easy to prove for these curves, see [1].

By a theorem of the German mathematician Helmut Hasse we know that N_p lies in the interval. This is proved in [1]

$$[p + 1 - 2\sqrt{p}, p + 1 + 2\sqrt{p}] \tag{2}$$

Hasse's theorem only gives a lower and an upper bound to N_p. It gives no details about the *distribution* over the interval so we cannot answer questions about how frequently N_p attains a value close to the upper limit or close to p , etc.

Around the year 1960 the American mathematician John Tate and the Japaneese mathematician Mikio Sato independently formulated a hypothesis about the distribution over Hasse's interval, which later came to be known as the *Sato-Tate conjecture*.

The conjecture was made after both heuristical reasoning and numerical experiments. At the time the performance of even the fastest computers in the world was vastly inferior to todays standard computers, so the process of making numerical experiments took a lot of time and had to be done with severe constraints on the size of the fields used in the calculations. See [5] for more details.

This thesis will describe a method of how to derive the Sato-Tate conjecture using both heuristics and an experimental part. The result will be a formulation of the Sato-Tate conjecture with an explicit function describing the distribution of solutions over the interval (2).

1.1 Examples

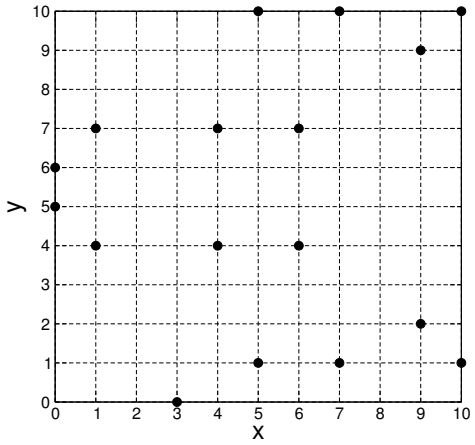


Figure 1: The solutions to $y^2 = x^3 + x + 3 \pmod{11}$

The figure shows the 17 solutions to the elliptic curve $y^2 = x^3 + x + 3$ when taking x and y as integers modulo 11 [1]. Putting $p = 11$ in (2) the interval for possible numbers of solutions is $[6, 18]$. In this case the number of solutions is in the upper end of the allowed interval. If we instead do the arithmetics modulo 19 we will get 20 solutions , a number close to the middle of the interval (2), i.e. with $p = 19$ we get the interval $[11, 27]$.

The distribution conjectured by Sato and Tate would allow us to understand how the number of solutions varies over the interval (2) for fixed values of A and B as p varies over all the primes.

2 Algebra

2.1 Definitions

Definition 1. A **group** $(G, *)$ is a set G equipped with a binary operation $*$ on G such that the following holds:

(i) *Closure:* For all $a, b \in G$ $a * b$ is a uniquely defined element in G

(ii) *Associativity:* $(a * b) * c = a * (b * c)$ for all $a, b, c \in G$

(iii) *Identity:* There exists an element $e \in G$ called the **identity** with the property $e * a = a * e$ for all $a \in G$

(iv) *Inverses:* For each element $a \in G$ there exists an element a^{-1} called the **inverse** of a such that: $a * a^{-1} = a^{-1} * a = e$

[1] See section 2.2 why this is a field

If $a * b = b * a$ for all $a, b \in G$ the group is called **abelian**
If the set G has a finite number of elements the group is called a **finite group.** In this case, the number of elements in G is called the **order** of G, denoted by $|G|$.

Definition 1.1. Let $(G, *)$ be a group , and let H be a subset of G. Then H is called a **subgroup** of G if H itself is a group under the same group operation as in G.

Definition 2. A **field** is a set F on which two binary operations $+$ and $\cdot$ are defined such that:

(i) $(F, +)$ is an abelian group with the identity element 0

(ii) $(F\backslash\{0\}, \cdot)$ is an abelian group with identity element $1 \neq 0$

(iii) The distributive law $a \cdot (b + c) = a \cdot b + a \cdot c$ holds for all $a, b, c \in F$

Note that each element $a \in F$ has an inverse with regards to the $+$ operation and an inverse with regards to the $\cdot$ operator. We separate these by letting $(-a)$ denote the first one and the other by a^{-1}.
A field makes it possible to define the four arithmetic operators $+, -, *, /$ where the operators $-$ and $/$ have the meanings $a - b = a + (-b)$ and $a/b = a * b^{-1}$.

2.2 Basic theorems

To assist the reasoning in future sections we state without proofs some very basic theorems about groups , fields and polynomials.

Proposition 1. In a group $(G, *)$ with $a, b, c \in G$ the following holds:

(a) *Cancellation:* $a * c = b * c \ \Rightarrow \ a = b$

(b) *Unique identity:* $a * b = a \ \Rightarrow \ b = e$

(c) *Unique inverses:* $a * b = e \ \Rightarrow \ b = a^{-1}$

Proposition 2. Let F be a field with the operations $+$ and $\cdot$

(a) For all $a \in F$ we have: $a \cdot 0 = 0$

(b) $a \cdot b = 0 \ \Rightarrow \ a = 0$ or $b = 0$

Theorem 3. Let $f(x)$ be a nonzero polynomial over the field F of degree n and $c \in F$. Then $f(c) = 0$ if and only if $x - c$ is a factor of $f(x)$. That is $f(c) = 0 \ \Leftrightarrow f(x) = (x - c)g(x)$ for some polynomial $g(x)$ over F of degree $n - 1$.

With the definitions in the previous section and these theorems we can prove:

Theorem 4. A polynomial $f(x)$ of degree n with coefficients in a field F has at most n unique roots in F.

2.3 Examples

Proposition 3. Let t be a positive integer and $p > 2$ be a prime number and let $\mathbb{Z}_t$ and $\mathbb{Z}_p$ be the set of possible remainders under division by t and p respectively.

(a) $(\mathbb{Z}_t, +)$ is a group under addition of residues with identity element 0.

(b) $\mathbb{Z}_p$ is a field under addition and multiplication of residues respectively with identities 0 and 1.

We will show this with some simple examples aided by a simple theorem. See [2] for full proofs in a more general context.

Theorem 5. For any integers a and $b > 0$,there exist unique integers q, r such that $a = bq+r, \quad 0 \leq r < b$. q is called the quotient and r is called the remainder of a under division by b.

Consider integer division by t. When we divide any integer by t the remainder will always be a an integer in the set $\mathbb{Z}_t = \{0, 1, \ldots, t-1\}$. If we add any two numbers a, b in this set we will get a unique integer $c = a + b$. By Theorem 5 above this number c will have a unique representation on the form: $c = tq + r$ where $0 \leq r < t$ and thus $r \in \mathbb{Z}_t$. We can then naturally define a group operation $*$ on this set by letting $a * b$ be the remainder of $a + b$ under division by t. This shows that $*$ has the property of closure in Definition 1.
Associativity follows easily because $(a + b) + c = a + (b + c)$ so that the remainder under divison by t will be the same. For the element 0 we have $a * 0 = 0 * a = a + 0 = a$ for all elements in $\mathbb{Z}_t$ by definition of $*$ so that 0 is an identity. 0 is clearly also its own inverse. For the other elements we can order them in pairs $(1, t-1), (2, t-2), \ldots$ so that for every pair (a, b) it holds that $a + b = t$ and $a * b = b * a = 0$ meaning that $b = a^{-1}$ and $a = b^{-1}$ This completes the checklist in Definition 1 and shows that this is a group $(\mathbb{Z}_t, *)$ which we will denote $(\mathbb{Z}_t, +)$ with some slight abuse of notation.

When performing integer division by a prime p we have to show that we have an operation $\times$ that makes the set $\mathbb{Z}_p \backslash \{0\}$ an abelian group with 1 as identity to prove that $\mathbb{Z}_p$ is a field. We define $a \times b$ to be the remainder of the integer product $a \cdot b$ under division with p. This establishes the properties (i)-(iii) in Definition 1 for the same reason as above. For each element $a \in \mathbb{Z}_p \backslash \{0\}$ we have that the *greatest common divisor* $gcd(a, p)$ between a and p is 1 since p is a prime.

Proposition 4. Let a, b be non-zero integers. Then $gcd(a, b) = 1$ if and only if there exists integers m, n such that $ma + nb = 1$

Because of this we can match any $a \in \mathbb{Z}_p \backslash \{0\}$ with an integer m such that $ma + np = 1$. By the uniquness of the numbers q, r by Theorem 5 we have that if $a = bq + r$ then $a - kq = b(q - k) + r$ when $k \in \mathbb{Z}$ so that the numbers a and $a - kq$ have the same reminder under division by b. Under division by p we can then see that $ma = ma + np - np$ has reminder 1. Furthermore, for an integer $o = m + lp$ for some $l \in \mathbb{Z}$ we can write $oa = ma + mlp = ma + (ml)p$. The number oa must then also have the remainder 1 under division by p. We can choose l so that $0 < o = m + lp < p \Rightarrow o \in \mathbb{Z}_p \backslash \{0\}$ by Theorem 5. We have

now found that the pair of elements $a^{-1} = o$ and a are each others inverses with respect to the operation $\times$ because $a^{-1} \times a = a \times a^{-1} = 1$ by definition.

3 Elliptic curves

This section will give a very brief introduction to some basic properties of elliptic curves. For more information and details, see [1], [3] and [4].

As a general explanation to why equations on the form (1) are interesting to study we begin by asserting that every third degree equation in two variables $ax^3 + bx^2y + cxy^2 + dy^3 + ex^2 + fxy + gy^2 + hx + iy + j = 0$ can be rewritten on the form $y^2 = x^3 + ax^2 + bx + c$ through a series of transformations and variable changes. This form is called the *Weierstrass* form. This means that it is sufficient to study curves in Weierstrass form because the understanding of their properties can be transfered back to the original curve.
If we restrict ourselves to equations over finite fields $\mathbb{Z}_p$ with $p > 3$, we can go even further and rewrite the general third degree equation to $y^2 = x^3 + Ax + B$
We can now see that the study of these equations is the next logical step after studying first and second degree equations in the forms: $ab + by + c = 0$ and $ax^2 + bxy + cy^2 + dx + ey + f = 0$.

3.1 Definition

Definition 4. An **elliptic curve** $C(F)$ defined over a field F is the set of solutions (x, y) to the equation

$$y^2 = x^3 + ax^2 + bx + c$$

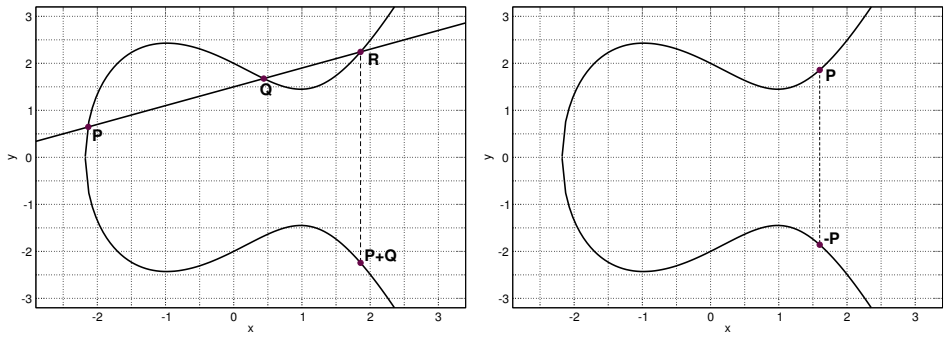
$$4A^3 - 27B^2 \neq 0$$

where $a, b, c \in F$

Note that F does not have to be finite. We will use elliptic curves over the field of real numbers $\mathbb{R}$ in the next section. For this field the property $4A^3 - 27B^2 \neq 0$ ensures that the curve is differentiable for all $x \in \mathbb{R}$.
For a finite field we cannot make the same connection, but in this case $4A^3 - 27B^2 \neq 0$ ensures that $x^3 + Ax + B$ does not have a double root.

3.2 Group structure of elliptic curves

An interesting property of elliptic curves is that it is possible to turn the set of points on an elliptic curve C into an abelian group $E(C)$. This requires an additional special point $\mathcal{O}$, called the *point at infinity*, that will serve as the identity element.
The most simple way to describe the group operation on this set is with a geometrical figure where $A, B, x \in \mathbb{R}$:



(a) Group addition (b) Additive inverse

Figure 2: The group operation

In the most simple case, the addition of two points P and Q is simply done by connecting the points with a line and then finding the third point R where the line crosses the curve. The sum $P + Q$ is then the mirror of R about the x-axis. I.e. if $R = (x, y)$ then $P + Q = -(x, y) = (x, -y)$. This point is also the inverse[2] of R with regards to the new operation. It is easy to see that this operation is commutative by definition because the point R will be the same irrespective of the order of P and Q.
If no such point R exists we define the point at infinity $\mathcal{O}$ as the third intersection point. We can think of $\mathcal{O}$ as a point on the y-axis infinitely far away. By saying this we mean that a line between any point on C and $\mathcal{O}$ is parallel to the y-axis. The mirror point of $\mathcal{O}$ is $\mathcal{O}$ itself. This allows us to make sense of the case $P + (-P) = \mathcal{O}$ in Figure 2.b
The case of $P + \mathcal{O}$ can also be explained by the same figure; if we extend the line between P and $\mathcal{O}$ we get $-P$ as the third intersection point so that $P + \mathcal{O} = -(-P) = P$.

3.2.1 The technical details

The reason for defining the operation like this becomes clearer when making the observation that a straight line seems to intersect an elliptic curve at three points with the exception of a few corner cases. If two points $P = (x_1, y_1), Q = (x_2, y_2)$ on an elliptic curve have $x_1 = x_2$ then the straight line through P and Q is vertical and obviously only crosses the curve at two points. This motivates the introduction of the point at infinity $\mathcal{O}$ which is taken as the third point of intersection and allows us to define inverse elements by y-axis mirroring. This means that we can set $P + Q + R = \mathcal{O}$ when $x_1 \neq x_2$ because by definition of the inverse, the sum of the point R and the point we defined as $P + Q$ is $\mathcal{O}$. In fact, for any three points of intersection between an elliptic curve and a straight line we can define their sum as $\mathcal{O}$. We have covered the cases of $P + Q + R = \mathcal{O}$ in Figure 2.a and $P + (-P) + \mathcal{O} = 0$. The remaining case is the special case of Figure 2.a when $P = Q$. We can set $P + P + Q = \mathcal{O}$ if we take the tangent of the elliptic curve at P and letting Q be the point where the tangent intersects

[2]Note that we denote the inverse of R as $-R$ instead of R^{-1}

the curve again. Then we have $P + P = -Q$, or $P + P + Q = \mathcal{O}$. So far, the examples we have used to show how the group operation for points on an elliptic curve works have been motivated with geometrical reasoning. The figures 2.a and 2.b are drawn for an elliptic curve over the field of real numbers $\mathbb{R}$. Because of this, and the assertion that the line connecting two points on the curve always intersects the curve at a third point, we can easily see that the closure property in Definition 1 holds. From the way we went about defining the point $\mathcal{O}$ we have an element that works as an identity according to Definition 1. We also note that if $P = (x, y)$ is a point on the curve C, the point $-P = (x, -y)$ also has to be on the curve C because $y^2 = x^3 + Ax + B \Leftrightarrow (-y)^2 = x^3 + Ax + B$. This explains why the definition of inverses in this group makes sense.
If we have an elliptic curve over a finite field $\mathbb{F}_p$ we lose the geometrical interpretation. It is no longer clear that we can connect two points $P = (x_1, y_2), Q = (x_2, y_2)$ on C with a line to get a third point R that is also on C. Now, the condition $4A^3 - 27B^2 \neq 0$ is equivalent to the polynomial $x^3 + Ax + B$ having 3 distinct roots x_1, x_2, x_3. By Theorem 3 we have: $x^3 + Ax + b = (x - x_1) \cdot (x - x_2) \cdot (x - x_3)$. We can construct a linear equation: $y = kx + l$ with the property that $y_1 = kx_1 + l$ and $y_2 = kx_2 + l$ which means that $k = \dfrac{y_2 - y_1}{x_2 - x_1}$ and $l = -kx_1 + y_1$. In the case of an elliptic curve over $\mathbb{R}$, this is obviously the straight line connecting P and Q. If we put this equation as y in the equation defining an elliptic curve we get:

$$(kx + l)^2 = x^3 + Ax + B \tag{3}$$

expanding the left hand side gives:

$$x^3 - k^2x^2 + (A - 2kl)x + (B - k^2) = 0$$

We know that x_1 and x_2 are roots to this equation. If we call the third root x_3 and apply theorem 3 we get:

$$x^3 - k^2x^2 + (A - 2kl)x + (B - k^2) = (x - x_1) \cdot (x - x_2) \cdot (x - x_3)$$

expanding the right hand side

$$x^3 - k^2x^2 + (A - 2kl)x + (B - k^2) = x^3 - (x_1 + x_2 + x_3)x^2 + (x_1x_2 + x_1x_3 + x_2x_3)x - x_1x_2x_3$$

in particular we have that the coefficients of the x^2 term must be the same on both sides of the equality:

$$x_1 + x_2 + x_3 = k^2 \Rightarrow x_3 = k^2 - x_1 - x_2 \tag{4}$$

If we now put $x = x_3$ in (3) we see that the point $(x_3, kx_3 + l) = (x_3, kx_3 - kx_1 + y_1)$ is located on the curve C. Since we have closure in any field F it follows that the group operation on points of elliptic curves also has the closure property if we define $P + Q = R$ where $P = (x_1, y_1)$, $Q = (x_2, y_2)$, $R = (x_3, -(kx_3 - kx_1 + y_1))$ and x_3 is calculated as in (4).
We have yet to show that the operation is associative. We have also not shown how to derive an explicit formula for $P + P$.
This formula is just written out in the definition below. We will not derive it in this thesis. See [4] for a complete proof.

Definition 5. Let C be an elliptic curve over a field F defined by $y^2 = x^3 + Ax + B$. Let $P = (x_1, y_2)$ and $Q = (x_2, y_2)$ be points on C separate from the point at infinity $\mathcal{O}$ and with $P + Q \neq \mathcal{O}$. The sum $S = P + Q$ is defined as:

Put $S = (x_3, y_3)$, then
$x_3 = k^2 - x_1 - x_2$,
$y_3 = k(x_1 - x_3) - y_1$
where

$$k = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1} & \text{if } P \neq Q \\ \dfrac{3x_1^2 + A}{2y_1} & \text{if } P = Q \end{cases}$$

Note that k is the slope of a line between P and Q in the first case and the slope of the tangent at P in the second case.

3.3 An example

Below is an example where we show the group operation table for the elliptic curve $y^2 = x^3 + 3x + 7$ over $\mathbb{Z}_{11}$.

The solutions to the equation in $\mathbb{Z}_{11}$ are[3]:

$$(1,0), (5,2), (5,9), (8,2), (8,9), (9,2), (9,9), (10,5), (10,6)$$

If we use the formulas above and introduce the imaginary point at infinity, $\mathcal{O}$, we can make a table for the group operation:

Table 1: Addition table for $y^2 = x^3 + 3x + 7$ over $\mathbb{F}_{11}$

+	$\mathcal{O}$	(1,0)	(5,2)	(5,9)	(8,2)	(8,9)	(9,2)	(9,9)	(10,5)	(10,6)
$\mathcal{O}$	$\mathcal{O}$	(1,0)	(5,2)	(5,9)	(8,2)	(8,9)	(9,2)	(9,9)	(9,10)	(10,6)
(1,0)	(1,0)	$\mathcal{O}$	(8,2)	(8,9)	(5,2)	(5,9)	(10,6)	(10,5)	(9,9)	(9,2)
(5,2)	(5,2)	(8,2)	(10,5)	$\mathcal{O}$	(9,9)	(1,0)	(8,9)	(9,2)	(10,6)	(5,9)
(5,9)	(5,9)	(8,9)	$\mathcal{O}$	(10,6)	(1,0)	(9,2)	(9,9)	(8,2)	(5,2)	(10,5)
(8,2)	(8,2)	(5,2)	(9,9)	(1,0)	(10,5)	$\mathcal{O}$	(5,9)	(10,6)	(9,2)	(8,9)
(8,9)	(8,9)	(5,9)	(1,0)	(9,2)	$\mathcal{O}$	(10,6)	(10,5)	(5,2)	(8,2)	(9,9)
(9,2)	(9,2)	(10,6)	(8,9)	(9,9)	(5,9)	(10,5)	(5,2)	$\mathcal{O}$	(1,0)	(8,2)
(9,9)	(9,9)	(10,5)	(9,2)	(8,2)	(10,6)	(5,2)	$\mathcal{O}$	(5,9)	(8,9)	(1,0)
(10,5)	(10,5)	(9,9)	(10,6)	(5,2)	(9,2)	(8,2)	(1,0)	(8,9)	(5,9)	$\mathcal{O}$
(10,6)	(10,6)	(9,2)	(5,9)	(10,5)	(8,9)	(9,9)	(8,2)	(1,0)	$\mathcal{O}$	(5,2)

Note that the size of the group $E(C)$ is the number of solutions to the equation $y^2 = x^3 + Ax + B$ plus the special point at infinity i.e. $|E(C)| = N_p + 1$
In this example: $|E(C)| = 9 + 1 = 10$.
In the next section we will show how to calculate $|E(C)|$.

[3]See 5.1 for a method to calculate all the solutions

4 A first analysis of the distribution over Hasse's interval

Definition. Let $p > 2$ be a prime and let $x > 0$ be an integer.
x is a **quadratic residue** mod p if there exists $y \in \mathbb{Z}_p$ such that $y^2 \equiv x \pmod{p}$ and a **quadratic non-residue** otherwise.

Definition. Let $p > 2$ be a prime. For an integer $a \geq 0$ we define the **Legendre symbol** $\left(\frac{a}{p}\right)$ as:

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } a \equiv 0 \pmod{p} \\ 1 & \text{if a is a quadratic residue mod p} \\ -1 & \text{otherwise} \end{cases} \tag{5}$$

Lemma 1. There are exactly $\frac{p+1}{2}$ distinct quadratic residues in the field $\mathbb{Z}_p$

Proof. If $x^2 = y^2$ then $(x+y)(x-y) = 0$ so that $x = \pm y$. This tells us that each element and its additive inverse correspond to exactly one quadratic residue. By proposition 1 we know that inverses are unique.
The elements in $(\mathbb{Z}_p, +)$ are the $\frac{p-1}{2}$ pairs $x, -x$ s.t. $x \neq -x$ and the element 0. In total: $\frac{p-1}{2} + 1 = \frac{p+1}{2}$ quadratic residues.

□

With these definitions we can write:

$$N_p - 1 = \sum_{x=0}^{p-1} \left(1 + \left(\frac{x^3 + Ax + B}{p}\right)\right) = p + \sum_{x=0}^{p-1} \left(\frac{x^3 + Ax + B}{p}\right) \tag{6}$$

because if y solves $y^2 \equiv x^3 + Ax + B \pmod{p}$ we get *two solutions* (y and $-y$) when $y \neq 0$ and one solution if $y = 0$.
So the problem is to calculate

$$N_p - p - 1 = \sum_{x=0}^{p-1} \left(\frac{x^3 + Ax + B}{p}\right) \tag{7}$$

To make a general prediction about the right hand side of (3) we can view it as a sum of p random variables over the set $\{-1, 0, 1\}$:

$$Y = \sum_{i=0}^{p-1} X_i \tag{8}$$

To decribe the probability functions $p_{X_i}(x)$ we need to determine how the values of the Legendre symbol are distributed depending on the parameters A, B, x and p. There are $\frac{p+1}{2}$ possible values of a quadratic residue in $\mathbb{Z}_p$ and we we assume that the values of $f(x) = x^3 + Ax + B$ for $x = 0, 1 \ldots, p-1$ should in some sense be spread out evenly over $\mathbb{Z}_p$, in particular not be biased towards the quadratic residues.

With these arguments the probability that $f(x)$ is a quadratic residue is:
$P[X_i = 1] = \frac{\frac{p+1}{2}}{p} \approx \frac{1}{2}$ for all i when p is large enough.
Now we have that:

$$P[X_i = -1] + P[X_i = 0] = 1 - P[X = 1] \approx 1 - 0.5 = 0.5 \tag{9}$$

By Theorem 4 we know that the equation $f(x) = 0$ has at most 3 solutions. This implies that $P[X_i = 0] \leq \frac{3}{p}$. When p is large this quantity is of negligible size and thus wont have any impact on the probability function $p_{X_i}(x)$

Motivated by this we suggest the following probability density function:

$$p_X(x) = \begin{cases} 0.5 & \text{if } x = 1 \\ 0.5 & \text{if } x = -1 \\ 0 & \text{if } x = 0 \end{cases} \tag{10}$$

Note that we have omitted the index i simply because we expect the probability that $x^3 + Ax + B$ is a quadratic residue to be independent of x.

To test the asumptions made, and the reasoning above, we want to calculate said probabilities for some polynomials over some fields $\mathbb{Z}_p$.

A simple procedure using the ideas above is to make a list of those values in the interval $\{0, 1, \ldots, p-1\}$ that are quadratic residues modulo p. Then we can calculate $z = f(x)$ for all $x \in \mathbb{Z}_p$ and for each value z check in the list if it is a quadratic residue. After counting the number of residues found, we can divide this number by p to obtain the experimental probability that $f(x)$ is a quadratic residue.

We let a computer program do this given parameters A, B, p. Below is the *pseudocode* for a simple implementation of the procedure described above.

Input: A, B, p
  vector[4] $y[p]$
  vector $squares[p]$
  for $i = 1 \ldots p$ **do**
    $y[i] \leftarrow 0$
    $squares[i] \leftarrow 0$
  end for
  for $i = 0 \ldots p - 1$ **do**
    $index \leftarrow i^2 \pmod p + 1$
    $squares[index] \leftarrow 1$
  end for
  for $x = 0 \ldots p - 1$ **do**
    $index \leftarrow x^3 + Ax + B \pmod p + 1$
    $y[index] \leftarrow y[index] + 1$

[4]This represents a list with p elements with indices in the range $1, 2, \ldots, p$

end for
Output: $\frac{\sum_{i=1}^{p} y[i] \cdot squares[i]}{p}$

With this algorithm we make the following table:

$f(x) = x^3 + Ax + B$	p	$p_X(1)$
$x^3 - 2x + 7$	29741	0.50422
$x^3 + 45x - 22$	29741	0.49823
$x^3 + 45x - 22$	15809	0.49643
$x^3 - 5x + 19$	15809	0.49320
$x^3 + 8x - 13$	5903	0.50178
$x^3 - 4x + 8$	5903	0.50737
$x^3 + 25x + 2$	44971	0.49803

4.1 A heuristic argument

From the table it seems that our reasoning is valid. $P_X(1)$ does indeed seem to be close to 0.5 indepenent of the variables.
For fixed values of A and B we can consider p to be a large random odd prime with a uniform probability distribution and consider the probability density function $p_Y(x)$ where $Y = \sum_{i=0}^{p-1} X_i$
Y is a sum of p numbers ± 1 with equal probability and if we assume the individual X_i's to be independent Y should have a normal distribution around zero with variance $\sqrt{p}$ by the central limit theorem.
We have to be very careful here, because a normally distributed variable could with a certain probability attain a value that is out of the range allowed by Hasses theorem. I.e: $P[|Y| > 2\sqrt{p}] > 0$ if Y has a normal distribution.
This tells us that we must have some kind of dependency structure between the variables X_i, or more explicitly, that the probabilistic model was a bit too simple.
Arguing in an informal way, we would expect a probability distribution that behaves like the normal distribution in the sense that we expect it to be centralized around 0, and show similar variation because of the probabilites above, but always attain a value in the range $[1 - 2\sqrt{p}, 1 + 2\sqrt{p}]$.

Based on these results we can calculate an approximation of $p_Y(x)$ by gathering empirical data for *one polynomial* over many different fields $\mathbb{Z}_p$.

5 Modelling a probability density function

Definition 4. A **cumulative distribution function** for a random variable X is defined as: $F_X := Pr[X \leq x]$, where $-\infty < x < \infty$.

Definition 5. If there exists a function $f_X(x)$ such that $F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx$ then X is said to be a **continuous stochastic variable.**
The function $f_X(x)$ is called the **probability density function** for the variable X.

We will use the abbreviated terms *cdf* and *pdf* for cumultative distrubtion functions and probability density functions respectively.

In this section we will calculate the number of solutions to an equation on the form (1) over 300000 different fields $\mathbb{Z}_p$ where p are large primes. We will also decribe a method to do this effectively. The data will be used to calculate a probability density function by fitting it to a function with the least squares method.

Since we are going to calculate the number of solutions to one equation over many different fields, we have to clarify the reasoning about the probability density function. We clearly expect the variance to depend on p, so therefore we have to normalize the variables in some way.
The result from Hasse's theorem:

$$p + 1 - 2\sqrt{p} \leq N_p \leq p + 1 + 2\sqrt{p} \tag{11}$$

can be written as:

$$-2\sqrt{p} \leq N_p - (p+1) \leq 2\sqrt{p} \tag{12}$$

or

$$-1 \leq \frac{N_p - (p+1)}{2\sqrt{p}} \leq 1 \tag{13}$$

Motivated by this we define $a_p := N_p - (p+1)$ and $c_p := \dfrac{a_p}{2\sqrt{p}} = \dfrac{N_p - (p+1)}{2\sqrt{p}}$
This suggests a nice way to normalize the variables. Given a way to calculate N_p we can do so for a range of different primes and for each value calculate c_p.

5.1 Counting N_p for an equation in $\mathbb{Z}_p$

The set: $S = \{(x, y) \mid x, y \in \mathbb{Z}_p\}$ contains p^2 elements , including all possible solutions to the equation:

$$y^2 = x^3 + Ax + B \tag{14}$$

A naive algorithm could calculate N_p in $O(p^2)$ time by testing for each element in S if it solves the equation.
A faster algorithm would allow us to get a larger dataset in the same amount of time.
We can speed things up by calculating the $\dfrac{n+1}{2}$ values of y^2 in $\mathbb{Z}_p$ and for each value setting the corresponding element in a list to 1, leaving the untouched values to the initial value 0[5] . This requires $O(p)$ memory and $O(p)$ time.
We then proceed by evaluating the right hand side of (9) for $x = 0, \ldots, x-1$ to $z = x^3 + Ax + B$ and checking in the precalculated list if element number z is 1 or 0. In the first case we know that z is a quadratic residue. If $z = 0$ then $(x, 0)$ is a solution to (9), if $z \neq 0$ then there exists two numbers y and $-y$ so that $y^2 = (-y)^2 = z$ giving us the two solutions (x, y) and $(x, -y)$.
In the second case that element number z is 0 there is no possible value y that would make (x, y) a solution to (9).
This suggests the following algorithm with a time complexity of $O(p)$ and

[5]See the algorithm in section 4

memory $O(p)$ which is a significant improvement over the naive algorithm:

Input: A, B, p
  vector $squares[p]$
  $N_p \leftarrow 0$
  for $i = 1 \ldots p$ **do**
    $squares[i] \leftarrow 0$
  end for
  for $y = 0 \ldots p-1$ **do**
    $index \leftarrow y^2 \ (mod\ p)$
    $squares[index + 1] \leftarrow 1$
  end for
  for $x = 0 \ldots p-1$ **do**
    $z \leftarrow x^3 + Ax + B \ (mod\ p)$
    if $squares[z + 1] = 1$ **then**
      if $z = 0$ **then**
        $N_p \leftarrow N_p + 1$
      else
        $N_p \leftarrow N_p + 2$
      end if
    end if
  end for
Output: N_p

5.2 The number of solutions for $y^2 = x^3 + 2x + 3$

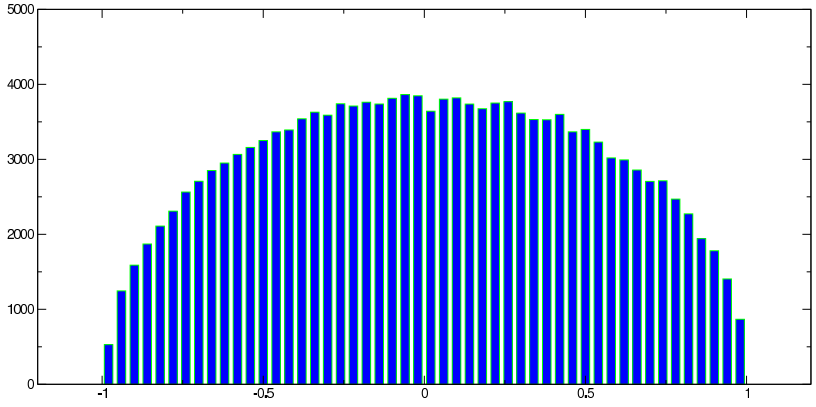
We use a program that implements the algorithm above to count the number of solutions to the equation

$$y^2 = x^3 + 2x + 3 \pmod{p} \tag{15}$$

The program generates a list of 300000 positive integers N_p over the fields $\mathbb{Z}_p$ where p are the first 300000 primes $p \geq 44773$. The numbers are then normalized to a set of values c_p according to section 5. We call this set L. We view the data as 300000 outcomes of a random variable X from the distribution we want to determine. Call the cumulative distribution function $F_X(x)$ and the probability density function $f_X(x)$.

The data is then plotted as a histogram with 100 bins:

Figure 3



The histogram is constructed so that $\sum_{i=1}^{100} m(i) = 300000$ where $m(i) = |b_i|$ is the number of elements in the i:th bin defined as:

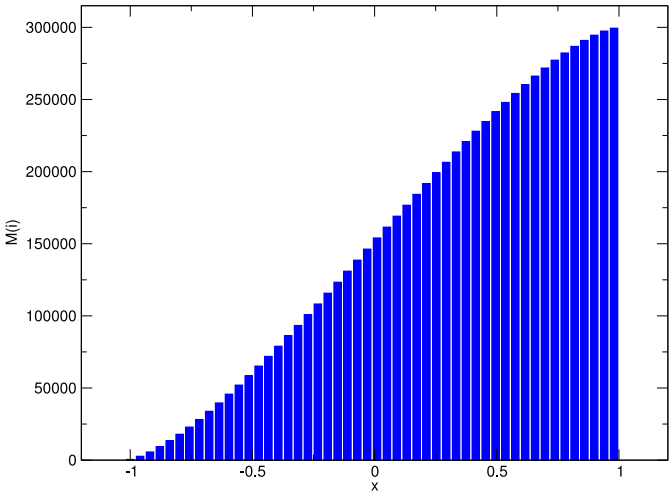
$$b_i = \{x \in L | -1 + (i-1)h < x \leq -1 + ih\} \tag{16}$$

$$\text{where } h = \frac{1-(-1)}{100} = \frac{2}{100}$$

We can also make a cumulative histogram. Consider the function:

$$M(i) = \sum_{j=1}^{i} m(j) \tag{17}$$

Plotting for $i = 1, \ldots, 100$ gives us:



To relate these figures to the functions f_X and F_X we begin with noticing that the probability for a value c_p to be placed in a bin to the left of or in the i:th bin is :

$$pr(i) = \frac{M(i)}{300000} \tag{18}$$

Now , for all x_i on the form $x_i = -1 + ih$, $i = 1, \ldots, 100$ we find that

$$\tilde{F_X}(x_i) = Pr(X \leq x_i) = \frac{M(i)}{300000} \tag{19}$$

is an estimation of $F_X(x_i)$ at a discrete set of points.

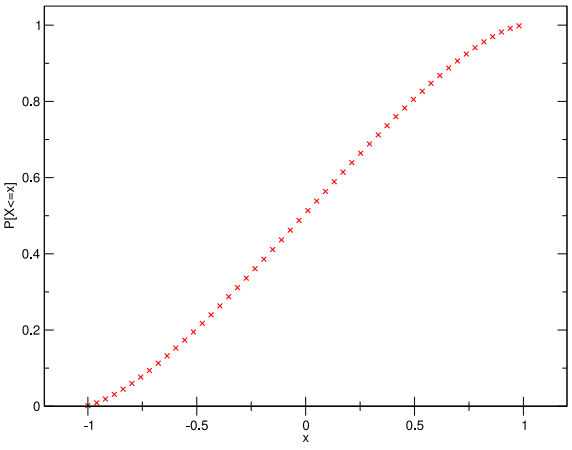


Figure 4: Empirical cdf

The functions f_X and F_X are closely related in the sense that:

$$f_X(x) = \frac{d}{dx}F_X(x) = \lim_{a\to 0}\frac{F_X(x+a) - F_X(x)}{a} \tag{20}$$

Since we have an approximation $\tilde{F}_X(x)$ we can use it to approximate the derivative of $F_X(x)$ at the points $x_1, \ldots, x_{100}$

$$\frac{F_X(x_i+a) - F_X(x_i)}{a} \approx \frac{\tilde{F}_X(x_i+a) - \tilde{F}_X(x_i)}{a} \tag{21}$$

By assigning $a = h = \frac{2}{100}$ we have that

$$\frac{d}{dx}F_X(x_i) \approx \frac{\tilde{F}_X(x_i+h) - \tilde{F}_X(x_i)}{h} = \frac{\tilde{F}_X(x_{i+1}) - \tilde{F}_X(x_i)}{h} = \frac{\frac{M(i+1)}{300000} - \frac{M(i)}{300000}}{h} = \frac{\frac{M(i+1)-M(i)}{300000}}{h} = \frac{\frac{m(i+1)}{300000}}{h} = \frac{m(i+1)}{300000}\cdot\frac{100}{2} = \frac{m(i+1)}{6000} \tag{22}$$

The next figure shows a plot of this function:

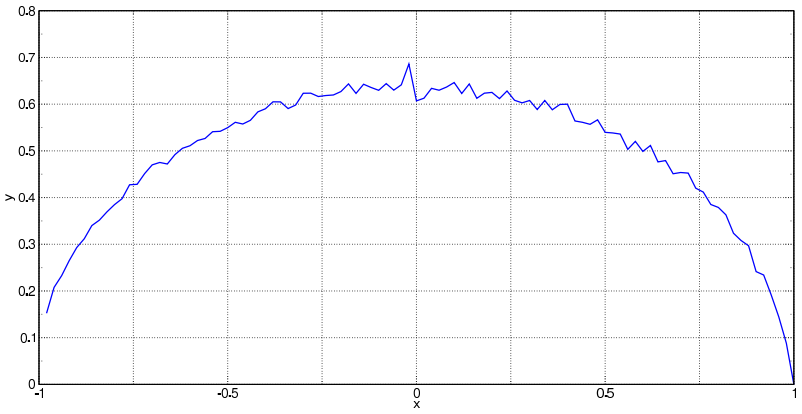


Figure 5

5.3 Modelling a density function

To conclude the previous section we can say that the numbers c_p indeed seem to converge towards a distribution. To fit the data to a function we need to make an analysis of what the distribution looks like, to be able to choose a suitable model function.
Since all the values c_p are in the range $[-1, 1]$ we can associate each c_p with an angle $\theta_p \in [0, \pi]$ such that $\cos\theta_p = c_p$ and instead try to fit a function to the set of the values θ_p.
By doing so we make the restrictions on c_p implicit in our model.
In figure 3 we have an approximation of the function $f_X(x)$ we want to determine. Apart from the choppiness in the graph it looks like a quite well behaved

function. Given that we constructed the graph by a numerical approximation of the derivative of the graph in figure 2, we expect that the choppiness in the graph is caused by numerical issues. To motivate this further, we expect the graph in figure 2 to have a smoother derivative than this.
If we disregard the choppiness (by refering to Figure 3) , the function appears to be an *even* function. We integrate this observation into our model.

Motivated by all this we try to fit the data to a Fourier series. Assuming the function is even we can directly try with a cosine series:

$$f_\Theta(\theta) = \sum_{n=0}^{N} a_n \cos n\theta \tag{23}$$

Because of the fact that there were some issues with determining the approximation of $f_X(x)$ we can expect them to cause further problems with the numerical accuracy when we try to fit the data to the cosine series. However , if we instead try to fit the curve in figure 2 to a function we can expect a more accurate result. This means that we will try to fit the datapoints $\{x_1, \tilde{F}(x_1)), \ldots, (x_{100}, \tilde{F}(x_{100}))\}$ to the primitive function $F_\Theta(\theta) = \int_0^\theta f_\Theta(\tilde{\theta})\, d\tilde{\theta}$

We integrate (23) to get the cdf:

$$F_\Theta(\theta) = \int_0^\theta \sum_{n=0}^{N} a_n \cos n\tilde{\theta}\, d\tilde{\theta} = \left[a_0\tilde{\theta}\right]_0^\theta + a_n \sum_{n=1}^{N} \left[\frac{\sin n\tilde{\theta}}{n}\right]_0^\theta =$$

$$= a_0\theta + \sum_{n=1}^{N} \frac{a_n}{n} \sin n\theta$$

We must have that

$$F_\Theta(\pi) = 1 \;\Rightarrow\; a_0\pi + 0 + \ldots = 1 \;\Rightarrow\; a_0 = \frac{1}{\pi}$$

so that the general form of the series we want to fit our function to is:

$$F_\Theta(\theta) = \frac{\theta}{\pi} + \sum_{n=1}^{N} b_n \sin n\theta \tag{24}$$

5.4 Fitting the data to our model

The set $\{x_1, \tilde{F}(x_1)), \ldots, (x_{100}, \tilde{F}(x_{100}))\}$ cannot directly be used with the model (24). For each value x_i we need to calculate its corresponding angle $\theta_i \in [0, \pi]$. The correct way to do this is with the bijection $\theta_i = \pi - \cos^{-1}(x_i)$ because we want F_Θ to be an increasing function. We use $N = 12$ to only include 12 coefficients $b_1, \ldots, b_{12}$ in the model.
We view the data as a vector $\mathbf{x} = [\theta_1, \ldots, \theta_{100}]^T$ and a vector $\mathbf{y} = \left[\tilde{F}(x_1), \ldots, \tilde{F}(x_{100})\right]^T$

and the matrix:

$$X = \begin{bmatrix} \sin(1 \cdot \theta_1) & \sin(2 \cdot \theta_1) & \dots & \sin(12 \cdot \theta_1) \\ \sin(1 \cdot \theta_2) & \sin(2 \cdot \theta_2) & \dots & \sin(12 \cdot \theta_2) \\ \vdots & \vdots & \ddots & \vdots \\ \sin(1 \cdot \theta_{100}) & \sin(2 \cdot \theta_{100}) & \dots & \sin(12 \cdot \theta_{100}) \end{bmatrix}$$

Our model can now be written in matrix form:

$$\frac{1}{\pi}\mathbf{x} + X\mathbf{b} = \mathbf{y} \tag{25}$$

Form the new vector: $\mathbf{z} = \mathbf{y} - \frac{1}{\pi}\mathbf{x}$
The problem is now to calculate the vector $\mathbf{b}$ that satisfies:

$$X\mathbf{b} = \mathbf{z} \tag{26}$$

Since the system is overdetermined we know that there is no solution but that the vector $\mathbf{b}$ that minimizes the norm

$$\|\mathbf{z} - X\mathbf{b}\|^2 \tag{27}$$

is given by the normal equations:

$$(X^T X)\mathbf{b} = X^T\mathbf{z} \tag{28}$$

Solving (28) gives:

$$b = \begin{bmatrix} 0.00019087087 \\ -0.1594348064 \\ 0.00015471255 \\ 0.00013657114 \\ 0.00011571785 \\ -0.0000360403 \\ -0.0002155770 \\ 0.00005559181 \\ -0.0000210021 \\ 0.00007706562 \\ -0.0001150259 \\ -0.0000778209 \end{bmatrix}$$

All the coefficients are very small except b_2. If we ignore the other numbers we get:

$$F_\Theta(\theta) = \frac{\theta}{\pi} - 0.1594348064 \sin(2\theta) = \frac{1}{\pi}(\theta - 0.5008792165 \sin(2\theta)) \tag{29}$$

As the final step we guess that the function we are looking for is:

$$F_\Theta(\theta) = \frac{1}{\pi}\left(\theta - \frac{\sin(2\theta)}{2}\right) \tag{30}$$

We plot this function together with the data from figure 2 marked as red circles:

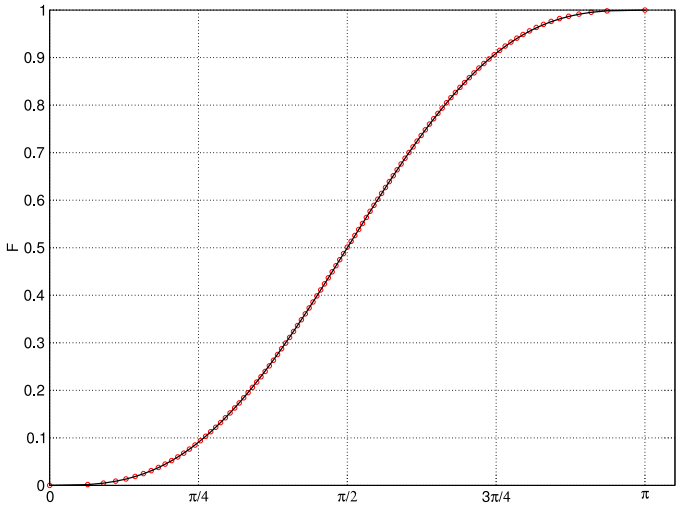


Figure 6

This would give a probability density function:

$$\frac{dF_\Theta}{d\theta} = \frac{1}{\pi}\left(1 - \cos 2\theta\right) = \frac{1}{\pi}(1 - (\cos^2\theta - \sin^2\theta)) = \frac{2}{\pi}\sin^2\theta \tag{31}$$

To get the cumulative density function $f_X(x)$ we have to differentiate:

$$\frac{dF_\Theta}{dx} = \frac{dF_\Theta}{d\theta}\frac{d\theta}{dx} \tag{32}$$

where

$$\theta(x) = \pi - \cos^{-1}(x) \tag{33}$$

with

$$\frac{d\theta}{dx} = \frac{1}{\sqrt{1-x^2}} \tag{34}$$

so that

$$f_X(x) = \frac{2}{\pi}\sin^2\theta(x)\frac{d\theta}{dx} \tag{35}$$

now using the trigonometric angle subtraction formula for sine gives:

$$\sin\theta(x) = \sin\left(\pi - \cos^{-1}(x)\right) = \sin\pi \cdot \cos\left(\cos^{-1}(x)\right) - \sin\left(\cos^{-1}(x)\right)\cdot\cos\pi = \\ = 0 \cdot x - \sin\left(\cos^{-1}(x)\right)\cdot(-1) = \sin\left(\cos^{-1}(x)\right)$$

using the Pythagorean trigonometric identity we can write:

$$\sin^2\theta(x) = \sin^2\left(\cos^{-1}(x)\right) = 1 - \cos^2\left(\cos^{-1}(x)\right) = 1 - x^2 \tag{36}$$

and finally:

$$f_X(x) = \frac{2}{\pi}\sin^2\theta(x)\frac{dx}{d\theta} = \frac{2}{\pi}\left(1-x^2\right)\cdot\frac{1}{\sqrt{1-x^2}} = \frac{2}{\pi}\sqrt{1-x^2} \tag{37}$$

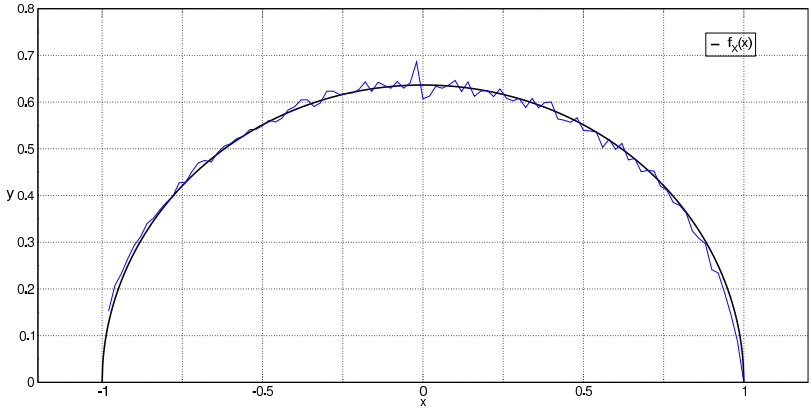


Figure 7

5.5 Testing the model

As we can see the function $f_x(x)$ seems to model the data quite well. Plotting the difference between $f_x(x)$ and the empirical density function from figure (5) at the discrete set of points gives a measure of how good an approximation our model is.

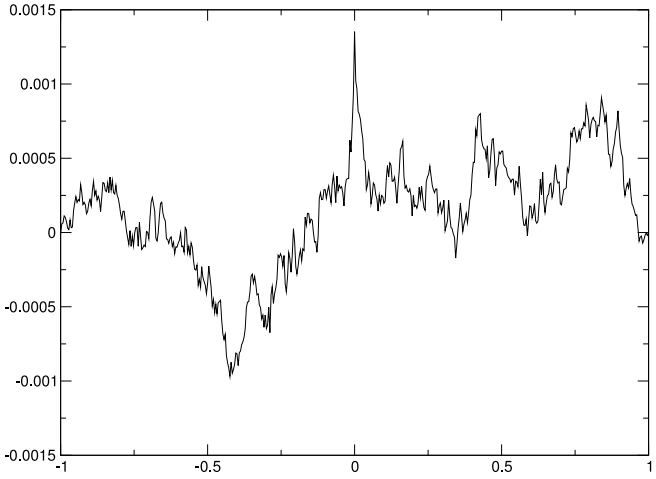


Figure 8

6 The conjectured result

If we repeated the process described in the previous chapter for another equation on the form (1) we would get very similar results. This insight backed by heurestic reasoning made the mathematicians in question conjecture the following result.

The Sato-Tate conjecture. Let C be an elliptic curve over a finite field $\mathbb{Z}_p$ As p varies over the range of primes, the number $a_p = \frac{N_p - (p+1)}{2\sqrt{p}}$ will be randomly distributed over the interval $[-1, 1]$ with the probability density function $f_X(x) = \frac{2}{\pi}\sqrt{1-x^2}$

The conjecture is sometimes defined for the angles θ_p with the *pdf* (29) instead. The conjecture can be made more general to also apply to so called *modular curves* in addition to the elliptic curves. This is beyond the scope of this thesis though. See [7] for more information.

The distribution function $f_X(x) = \frac{2}{\pi}\sqrt{1-x^2}$ is sometimes also refered to as the *Wigner semicircle distribution*. It appears among other things as the distribution of eigenvalues for random $N \times N$ matrices with entries chosen from a standard normal distribution. In this context this distribution has a relation to the natural measure on the 2-dimensional unitary group through "the action of Frobenius", see [9].

6.1 Proving the Sato-Tate conjecture

In March 2006, a major breakthrough in the process of proving the Sato-Tate conjecture was presented by Richard Taylor at Harvard University together with Laurent Clotzel,Nicholas Shepherd-Barron and Michael Harris. The proof apply to wide classes of elliptic curves satisfying some technical conditions. A general proof for elliptic curves over all fields is still missing, see [6] for an overview of this result.

7 Related topics

Elliptic curves are a very important field of interest in cryptography. Elliptic curves over finite fields $\mathbb{F}_p$ can be used to implement a cryptographic system. Cryptographers use theory of mathematics to find so called *one way functions*. These functions have the property that it is easy to calculate $y = F(x)$ for a value x but hard to invert the function in order to calculate $x = F^{-1}(y)$ unless you know a (secret) key value. Such a function can be used to encrypt a value x as $y = F(x)$. The encryption is safe if it is hard to get back x from y.
One way to implement this is by using the group of points on an elliptic curve C, a point P on this curve and a secret positive integer n. We use n to calculate a point

$$Q = P^n = nP = \underbrace{P + P + \ldots + P}_{n \text{ additions}} \tag{38}$$

For this calculation to be fast even for very large values of n a technique called *double and add* can be employed. This is done by writing n in base 2

$$n = d_0 + 2d_1 + 4d_2 + \ldots + 2^m d_m \Rightarrow$$

$$nP = (d_0 + 2d_1 + 4d_2 + \ldots + 2^m d_m)P = d_0P + 2d_1P + 4d_2P + \ldots + 2^m d_m P$$

The calculation of nP can then be done in $m = log_2(n)$ steps by starting with $Q = \mathcal{O}$ and $i = 0$ and for each binary digit d_i calculate $2^iP = 2^{i-1}P + 2^{i-1}P$ if $i > 0$ and then add this number to the cumulative sum Q if $d_i = 1$.

The security of this system relies on the fact that it is hard to calculate the integer n given the two points P and Q. This is called the *Discrete logarithm problem*. The number n is defined as the discrete logarithm of Q with respect to P. We could of course use another group than the group of points on an elliptic curve, but the reason for this choice is that there is no known algorithm for calculating n fast enough. The best known algorithm is of time complexity $O(\sqrt{n})$ for a general elliptic curve. This quickly becomes infeasible to do as the size of the group C is in the order of a few hundred decimal digits.

One can show that the set of points $\langle P \rangle = \{\mathcal{O}, P, 2P, 3P, \ldots\}$ is a subgroup of $E(C)$. We say that P is a generator of the subgroup $\langle P \rangle$. The relevant measure of security, i.e. how hard it is to solve the discrete logarithm problem in $E(C)$, is the size of the subgroup $\langle P \rangle$. The following theorem by *Lagrange* relates the size of the subgroup to the order of $E(C)$.

Theorem 6. Let H be a subgroup of the finite group G, then $|H|$ is a divisor of $|G|$.

Proof. See [2] □

For a group $E(C)$ of points on an elliptic curve C and a point $P \in E(C)$ we define the *cofactor*

$$h = \frac{|E(C)|}{|\langle P \rangle|}$$

For cryptographic purposes a small cofactor $h \leq 4$ is required. This means that we have to find a group $E(C)$ of order $|E(C)|$ prime or near prime. In particular, a fast method for counting the number of points on an elliptic curve is needed to certify that an elliptic curve is safe for cryptographic use. The method we used in section 6.1 is far too slow for this purpose since the minimal recommended field F_p for use has $p > 2^{224}$.

7.1 Counting points on elliptic curves

The method for counting points $|E(C)|$ devised in 6.1 is of time complexity $O(p)$ for a finite field $\mathbb{F}_p$ since calculations had to be done for each $x \in \mathbb{F}_p$. If we assume that each of the p required calculations can be done in $\frac{1}{3 \cdot 10^9}$ seconds we would require about 10^{50} years to calculate the number of points on an elliptic curve over a field $\mathbb{F}_p$ if $p \approx 2^{224}$. Given that a fast computer works at a frequency $\approx 3 \cdot 10^9$ Hz , the expected time is more than likely an

underestimation.

By using the group $E(C)$ and some theory we can construct a method that is much faster. By Hasse's theorem we know that $|E(C)|$ is a number in the interval $[p+1-2\sqrt{p}, p+1+2\sqrt{p}]$ so that $|E(C)|$ is one of $4\sqrt{p}$ possible numbers. For a subgroup $\langle P \rangle$ one can show that for any element $a \in \langle P \rangle$ it holds that $a^N = \mathcal{O}$ where $N = |\langle P \rangle|$. By Theorem 6 we know that N divides $|E(C)|$ so that for any $a \in E(C)$ we have $a^M = (a^N)^{\frac{M}{N}} = \mathcal{O}^{\frac{M}{N}} = \mathcal{O}$ with $M = |E(C)|$.

If we can find a point $R \in E(C)$ and a unique integer $m \in [p - 2\sqrt{p}, p + 2\sqrt{p}]$ so that $R^m = \mathcal{O}$, it must then hold that $m = |E(C)|$.

A concrete way to find such a unique number m in Hasse's interval is to pick a point $P \in E(C)$ and then calculate mP for $p - 2\sqrt{p} < m < p + 2\sqrt{p}$. If we only find one number m with the property $a^m = \mathcal{O}$ we are done. If we find several numbers we choose another point on the curve and try again. We are only interested in calculating mP for $m >= \lceil p - 2\sqrt{p} \rceil$ and $m \leq \lfloor p + 2\sqrt{p} \rfloor$. To do this we let $l = \lceil p - 2\sqrt{p} \rceil$ and calculate lP with the double and add method. We then proceed to calculate $(l+1)P, (l+2)P, \ldots, (\lfloor p+2\sqrt{p} \rfloor)P$ by starting with lP and then in each step adding P. This requires $4\sqrt{p}$ additions. Calculating lP requires $O(\log_2 l) = O(\log_2 p) < O(\sqrt{p})$ additions so that the time needed to perform the $4\sqrt{p}$ additions will dominate over calculating lP. This method requires $O(\sqrt{p})$ operations which is a huge improvement over the algorithm in 6.1.

We can give an example of this procedure by working with the curve C: $y^2 = x^3+3x+7$ over $\mathbb{F}_{11}$. The group operation for C is listed in Table 1. The interval for $|E(C)|$ is $[6, 18]$.

If we pick $P = (5, 2)$ as the starting point for the algorithm above we get: $6P = 2P + 4P = 2(P + 2P)$ now we calculate $2P = (10, 5)$, $P + 2P = (10, 6)$ and finally $6P = (10, 6) + (10, 6) = (5, 2)$. When we calculate $7P, 8P, \ldots, 18P$ we find that $10P = 15P = \mathcal{O}$ so the algorithm failed for this choice of P.
With $P = (8, 2)$ we find that the only $m \in [6, 12]$ such that $P^m = \mathcal{O}$ is $m = 10$ which is the correct size of $E(C)$.

The time complexity can be improved to $O(\sqrt[4]{p})$ with some optimization. This is called the *Baby-step Giant-step* algorithm. See [4] for a full description.

7.2 A conjecture about the primality of $|E(C)|$

A question of great interest in the study of Elliptic Curve Cryptography is how hard it is to find parameters A, B, p so that the elliptic curve $C : y^2 = x^3+Ax+B$ over $\mathbb{F}_p$ has a group of points $E(C)$ of prime order, i.e. N_p+1 is a prime number. Such a group would have a cofactor $h = 1$ by Lagrage's theorem since the only possible sizes for a subgroup on the form $\langle P \rangle$ is 1 or $|E(C)|$.
For the security of elliptic curves to scale well with p it must be possible to find curves with groups of prime order reasonably fast.
There is no proof that this will always be possible, but there is a conjectured result from 1988 by Neil Koblitz [8]:

Conjecture (Koblitz). Let $C : y^2 = x^3 + Ax + B$ be an elliptic curve over a finite field $\mathbb{F}_p$ with $\nabla = 4A^3 - 27B^2$. Then

$$|\{p \in \mathbb{P},\ p \leq n, p \nmid \nabla \ \mid\ |E(C) \mod p| \in \mathbb{P}\}|$$

is asymptotic to

$$D\frac{n}{\log^2 n}$$

where D is a positive constant depending on C.

The article [8] uses heuristic arguments backed by numerical results to come up with this result similar to the methods used in this thesis.

8 Acknowledgements

First, I would like to thank my supervisor Torsten Ekedahl, who unfortunately passed away last year, for introducing me to this very interesting subject.
I would like to thank Rikard Bøgvad for helping me to finish this thesis.
Also, a huge thank you goes out to those who helped me in the process of writing this thesis by pointing out errors and suggesting things I could improve upon.

9 Bibliography

References

[1] Joseph H. Silverman, *The arithmetic of elliptic curves*
Second Edition , Springer (2009)

[2] J.A. Beachy and W.D. Blair, *Abstract Algebra*
Waveland Pr, Inc, 2006.

[3] Joseph H. Silverman and John Tate, *Rational Points on Elliptic Curves*
Springer-Verlag, New York, 1992.

[4] Jeffrey Hoffstein, Jill Piper and Joseph H. Silverman,
An introduction to Mathematical Cryptography
Springer, 2008

[5] http://www2.math.ou.edu/ rschmidt/satotate/page5.html

[6] http://www.cirm.univ-mrs.fr/videos/2006/exposes/17w2/Harris.pdf

[7] http://www.institut.math.jussieu.fr/projets/fa/bpFiles/Introduction.pdf

[8] http://www.ams.org/mathscinet-getitem?mr=89h:11023

[9] http://www.math.su.se/ teke/