

THE DISTRIBUTION OF THE NUMBER OF POINTS OF ELLIPTIC CURVES ON \mathbb{Z}_p

JAVIER FERNANDEZ

ABSTRACT. In this note we report on an experiment related to the distribution of the number of rational points on elliptic curves over the field with p elements. In particular, we analyze the symmetry that these numbers exhibit. A simple proof is provided and a more elaborate interpretation of the phenomenon is given in terms of isomorphism classes of curves and field extension.

1. THE PROBLEM

In this report we want to discuss a project that was carried through as part of the REU on rational points on elliptic curves that was run at the Department of Mathematics, University of Utah, during the Summer of 2003. Even though we include some statements and proofs, there are no new results in what follows.

The broad objective of this project was to study the distribution of the number of rational points of elliptic curves over the finite field $\mathbb{Z}_p = \mathbb{Z}/(p\mathbb{Z})$. One of the participants, Jenise Smalley, wrote the software and produced the data that led this experiment.

Let $p \geq 5$ be a prime number that we will fix for the rest of this note. All elliptic curves over \mathbb{Z}_p can be written in Weierstrass form

$$C : y^2 = x^3 + bx + c \text{ with } b, c \in \mathbb{Z}_p. \tag{1}$$

We denote by $\#C(\mathbb{Z}_p) = \#\{(x, y) \in \mathbb{Z}_p^2 : y^2 = x^3 + bx + c\}$, the number of rational points on the curve C .

By taking all possible values¹ of b and c in \mathbb{Z}_p we can compute all the possible values of $\#C(\mathbb{Z}_p)$. Figure 1 shows the histogram corresponding to the distribution of these values.

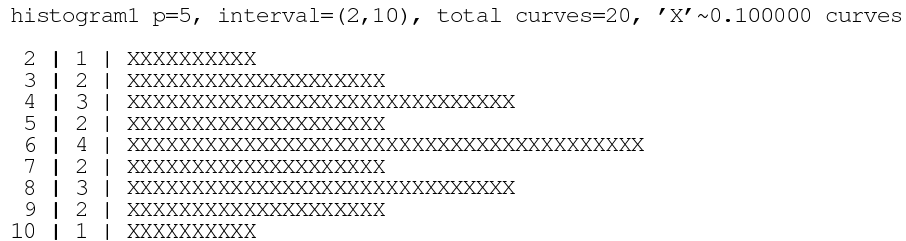


FIGURE 1. Histogram for $p = 5$

Date: June 25, 2003.

¹Actually, we have to discard those values where $\Delta = -4b^3 - 27c^2 = 0$ since these correspond to singular curves

On Figure 1 the first column on the left corresponds to the number of rational points; the second column is the number of curves with the given number of points.

A first observation is that only certain numbers appear as number of points on elliptic curves: all numbers between 2 and 10. Notice that $\#\mathbb{Z}_5^2 = 25$ so the actual range of points is much smaller than it could have been. This has a “simple” explanation, after Hasse’s Theorem that states that for all p and all elliptic curves C

$$|p + 1 - \#C(\mathbb{Z}_p)| < 2\sqrt{p}.$$

Thus, for $p = 5$, $|6 - \#C(\mathbb{Z}_p)| < 4.47$, so that $1.53 < \#C(\mathbb{Z}_p) < 10.47$. The fact that every number in this range is realized by some curve follows from the work of Honda.

Our next observation is that the histogram is symmetric with respect to the number $p + 1 = 6$. We may think that this is a “feature” of the case $p = 5$, but as you can see in Figures 2 and 3, this turns out to be the case in general².

histogram1 p=37, interval=(26,50), total curves=1332, 'X'~3.375000 curves

26		9		XX
27		24		XXXXXXXX
28		60		XXXXXXXXXXXXXXXXXXXX
29		18		XXXXX
30		72		XXXXXXXXXXXXXXXXXXXX
31		54		XXXXXXXXXXXXXXXXXXXX
32		72		XXXXXXXXXXXXXXXXXXXX
33		36		XXXXXXXXXX
34		72		XXXXXXXXXXXXXXXXXXXX
35		54		XXXXXXXXXXXXXXXXXXXX
36		135		XX
37		42		XXXXXXXXXXXX
38		36		XXXXXXXXXX
39		42		XXXXXXXXXXXX
40		135		XX
41		54		XXXXXXXXXXXXXXXXXXXX
42		72		XXXXXXXXXXXXXXXXXXXX
43		36		XXXXXXXXXX
44		72		XXXXXXXXXXXXXXXXXXXX
45		54		XXXXXXXXXXXXXXXXXXXX
46		72		XXXXXXXXXXXXXXXXXXXX
47		18		XXXXX
48		60		XXXXXXXXXXXXXXXXXXXX
49		24		XXXXXX
50		9		XX

FIGURE 2. Histogram for $p = 37$

In what follows we will, first, prove that all histograms are symmetric with respect to $p + 1$. Then, we will explain why this is so.

2. PROVING SYMMETRY

Proposition 1. *Let $p \geq 5$ be a prime number. For every elliptic curve C defined over \mathbb{Z}_p there is another elliptic curve C' defined over \mathbb{Z}_p such that $\#C(\mathbb{Z}_p) + \#C'(\mathbb{Z}_p) = 2(p + 1)$.*

Notice that Proposition 1 proves the symmetry because the numbers of points of $C(\mathbb{Z}_p)$ and $C'(\mathbb{Z}_p)$ are symmetric with respect to $p + 1$.

²You can find histograms for all primes $5 \leq p \leq 293$ at <http://www.math.utah.edu/~jfernand/teaching/elliptic/docs/histograms.pdf>

histogram1 p=89, interval=(72,108), total curves=7832, 'x'~15.400000 curves

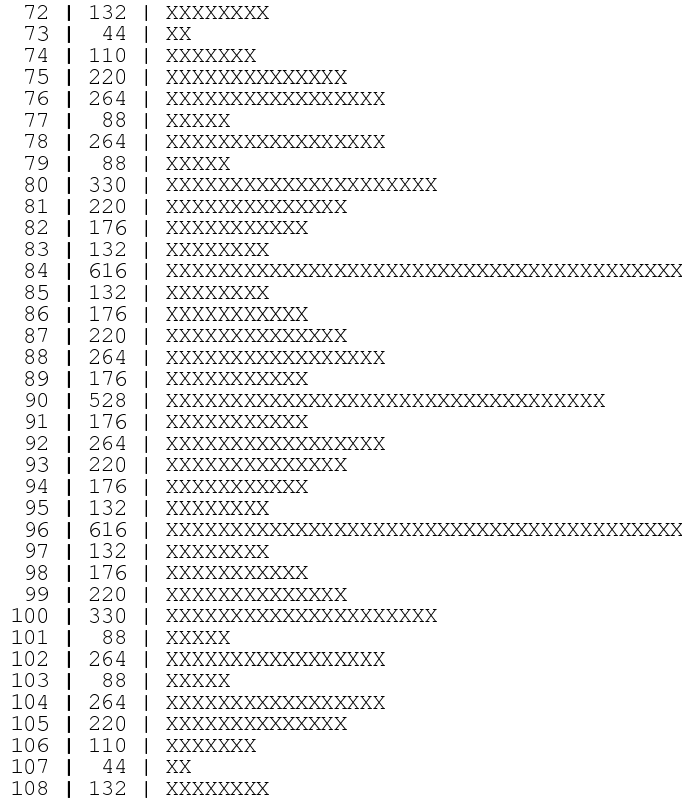


FIGURE 3. Histogram fro $p = 89$

Proof. Our first step is to give a formula for $\#C(\mathbb{Z}_p)$. The idea is very simple: using the expression (1) for C , for each possible value of $x \in \mathbb{Z}_p$, there are three possibilities:

- (1) $x^3 + bx + c = 0$, in which case x produces a single point, $(x, 0)$, on $C(\mathbb{Z}_p)$;
- (2) $x^3 + bx + c$ is a square in \mathbb{Z}_p , in which case x produces two points on $C(\mathbb{Z}_p)$: (x, y_1) and (x, y_2) , where y_1 and y_2 are the two square roots of $x^3 + bx + c$ in \mathbb{Z}_p ;
- (3) $x^3 + bx + c$ is not a square in \mathbb{Z}_p , in which case x produces no point on $C(\mathbb{Z}_p)$.

Additionally, we have to count the point at infinity (the zero in the group), \mathcal{O} .

Using the Legendre symbol

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & \text{if } a \text{ is a square in } \mathbb{Z}_p \\ 0, & \text{if } a \text{ is } 0 \text{ in } \mathbb{Z}_p \\ -1, & \text{if } a \text{ is not a square in } \mathbb{Z}_p \end{cases}$$

we can write

$$\#C(\mathbb{Z}_p) = 1 + \sum_{x \in \mathbb{Z}_p} \left(1 + \left(\frac{x^3 + bx + c}{p}\right)\right).$$

Fix a non-square number $k \in \mathbb{Z}_p$. Next, for the given curve C we define the elliptic curve C' :

$$C' : k \cdot y^2 = x^3 + bx + c \text{ or, in Weierstrass form, } y^2 = x^3 + k^2b + k^3c.$$

It is immediate that C' is an elliptic curve defined over \mathbb{Z}_p .

Similarly to what we did above we can count the points on C' as follows. For each x in \mathbb{Z}_p there are three possibilities —here we will use the expression $k \cdot y^2 = x^3 + bx + c$ for C' —:

- (1) $x^3 + bx + c = 0$, in which case x produces a single point on $C(\mathbb{Z}_p)$;
- (2) $x^3 + bx + c$ is a square in \mathbb{Z}_p , in which case x produces no points on $C(\mathbb{Z}_p)$, since otherwise k would be a square in \mathbb{Z}_p ;
- (3) $x^3 + bx + c$ is not a square in \mathbb{Z}_p , in which case $k^{-1} \cdot (x^3 + bx + c)$ is a square³ and it has two square roots y_1, y_2 so that x contributes the two points (x, y_1) and (x, y_2) .

We can rewrite these conditions using the Legendre symbol:

$$\#C'(\mathbb{Z}_p) = 1 + \sum_{x \in \mathbb{Z}_p} \left(1 - \left(\frac{x^3 + bx + c}{p} \right) \right).$$

All together, we have

$$\begin{aligned} \#C(\mathbb{Z}_p) + \#C'(\mathbb{Z}_p) &= 1 + \sum_{x \in \mathbb{Z}_p} \left(1 + \left(\frac{x^3 + bx + c}{p} \right) \right) + \\ &\quad + 1 + \sum_{x \in \mathbb{Z}_p} \left(1 - \left(\frac{x^3 + bx + c}{p} \right) \right) \\ &= 2 + 2 \sum_{x \in \mathbb{Z}_p} 1 = 2 + 2p, \end{aligned}$$

as wanted. □

Notice that in the previous Proposition we can replace p by $q = p^r$ with $r \in \mathbb{N}$ and so obtain, with the same proof, the following result.

Proposition 2. *Let $q = p^r$ with $p \geq 5$ prime and $r \in \mathbb{N}$. For every elliptic curve C defined over \mathbb{Z}_q there is another elliptic curve C' defined over \mathbb{Z}_q such that $\#C(\mathbb{Z}_q) + \#C'(\mathbb{Z}_q) = 2(q + 1)$.*

That is, the symmetry holds over arbitrary finite fields.

3. TRYING TO UNDERSTAND

The first time that we noticed the symmetry of the histograms we didn't have any idea of where it was coming from and, much less, how to give a proof of this "experimental fact".

The first idea, suggested by Jim Carlson, was to look at the isomorphism classes of curves. In other words: instead of looking at all the possible curves at the same time, consider only those curves that are isomorphic to a given one. If this smaller set of curves still exhibits the same symmetry then, perhaps, it will be easier to understand the phenomenon here.

³This follows from the fact that, since k is a non-square so is k^{-1} and the product of two non-squares in \mathbb{Z}_p is a square (exercise).

Before going on, we have to understand what we mean by isomorphic curves. It is a simple but lengthy exercise to see that two elliptic curves in Weierstrass form

$$C : y^2 = x^3 + bx + c \text{ and } C' : y^2 = x^3 + b'x + c' \text{ with } a, b, a', b' \in \mathbb{Z}_p$$

are isomorphic if (and only if) there is $\alpha \in F$ such that $b' = \alpha^4 b$ and $c' = \alpha^6 c$. In this case the isomorphism $m_\alpha : C \rightarrow C'$ is given by $m_\alpha(x, y) = (\alpha^2 x, \alpha^3 y)$. One aspect is unclear here: what is F ? F is any field extension⁴ of \mathbb{Z}_p . We say then that C and C' are isomorphic over F . If $\alpha \in \mathbb{Z}_p$ the two curves are isomorphic in “the usual sense” and, in particular, points on both curves are identified by m_α , so, both curves have the same number of elements. But, if $\alpha \notin \mathbb{Z}_p$ the two curves are seen to be “the same” only after we consider points (x, y) with $x, y \in F$. An example of this kind of problem will come in a moment.

The j -invariant of the curve C given by (1) is

$$j(C) = \frac{48^3 b^3}{4b^3 + 27c^2}.$$

It is easy to see that this number is unchanged if we replace C by an isomorphic curve C' . Conversely, two curves that have the same j -invariant are isomorphic over some extension of \mathbb{Z}_p .

Example 3. Consider the following curves over \mathbb{Z}_5 :

$$C_1 : y^2 = x^3 + x + 2,$$

$$C_2 : y^2 = x^3 + x + 3,$$

$$C_3 : y^2 = x^3 + 4x + 1.$$

It is easy to check that their j -invariant (mod 5) is 1 so that they are isomorphic in some extension of \mathbb{Z}_5 . Let's make explicit the isomorphisms.

Start with C_1 and C_2 . We saw above that the isomorphism is given by m_α for some α and that it mapped b to $\alpha^4 b$ and c to $\alpha^6 c$. For these two curves we have

$$\frac{3}{1} = \frac{\alpha^6 \cdot 2}{\alpha^2 \cdot 1} = \alpha^2 2$$

so that $3 = 2\alpha^2$, or, $\alpha^2 = 4$ (remember that we are working over \mathbb{Z}_5). Thus we can take $\alpha = 2$ and the isomorphism is given by $m_2(x, y) = (2^2 x, 2^3 y) = (4x, 3y)$, that is clearly defined over \mathbb{Z}_5 .

Now consider the curves C_1 and C_3 . The same argument as above leads to $\alpha^2 = 2$ and we see that 2 is not a square in \mathbb{Z}_5 , so that α is not in \mathbb{Z}_5 . So we “extend” \mathbb{Z}_5 by adding a square root of 2, that is, we consider $F = \mathbb{Z}_5[\sqrt{2}]$. In F we can take $\alpha = \sqrt{2}$ and consider $m_{\sqrt{2}} : C_1(F) \rightarrow C_3(F)$ given by $m_{\sqrt{2}}(x, y) = (\sqrt{2}^2 x, \sqrt{2}^3 y) = (2x, 2\sqrt{2}y)$. Notice that if you apply the isomorphism to a point in $C_1(\mathbb{Z}_5)$, like $(1, 2)$, we get a point in $C_3(F)$, $m_{\sqrt{2}}(1, 2) = (2, 4\sqrt{2})$ but is not in $C_3(\mathbb{Z}_5)$. Therefore, C_1 and C_3 are isomorphic over F but not over \mathbb{Z}_5 .

It is easy to check that $\#C_1(\mathbb{Z}_5) = \#C_2(\mathbb{Z}_5) = 4$ (they had to agree because the curves are isomorphic over \mathbb{Z}_5), but $\#C_3(\mathbb{Z}_5) = 8$, showing that C_1 and C_3 are not isomorphic over \mathbb{Z}_5 . A nice exercise is to check that the cardinality of all three curves over F —where they are isomorphic—is 32. More about this in Section 4.

⁴A field extension is a field containing the given field as a subfield. For example, \mathbb{R} is a field extension of \mathbb{Q} and \mathbb{C} is a field extension of both. A field extension of \mathbb{Z}_5 is, for instance, the field $\mathbb{Z}_5[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}_5\}$.

Now that we understand the notion of isomorphism and being warned that some subtleties are involved —field extensions— we can try to understand how all the elliptic curves in one isomorphism class (that is, having the same j -invariant) are related to each other. Say that C is given by (1), and take $t \in \mathbb{Z}_p^* = \mathbb{Z}_p - \{0\}$.

- If t is a square, $m_{\sqrt{t}}$ is an isomorphism (over \mathbb{Z}_p) between C and some other curve C_t . In this case, $\#C(\mathbb{Z}_p) = \#C_t(\mathbb{Z}_p)$.
- If t is not a square we consider the extension $F = \mathbb{Z}_p[\sqrt{t}]$ and, again, have an isomorphism between C and some other curve C_t , only that this time the isomorphism is defined over F , so that $\#C(F) = \#C_t(F)$ but in general $\#C(\mathbb{Z}_p) \neq \#C_t(\mathbb{Z}_p)$. In fact, since t is not a square in \mathbb{Z}_p we can use the argument used in the proof of Proposition 1 to conclude that $\#C(\mathbb{Z}_p) + \#C_t(\mathbb{Z}_p) = 2(p+1)$.

This procedure explains how one curve is associated to several other curves in its isomorphism class. In principle, one curve C_t is constructed for each value of $t \in \mathbb{Z}_p^* = \mathbb{Z}_p - \{0\}$, but it could happen that we obtain the same curve C_t for different values of t . Lets look into this problem more closely.

If we start from C with coefficients (b, c) , application of $m_{\sqrt{t}}$ generates a curve C_t with coefficients (t^2b, t^3c) so that if $bc \neq 0$, and $(t_1^2b, t_1^3c) = (t_2^2b, t_2^3c)$ we conclude that $t_1 = t_2$ and we obtained a different curve for each value of $t \in \mathbb{Z}_p^*$. These are all the curves in the isomorphism class.

But suppose that we start from the curve $(b, 0)$ —whose j -invariant is congruent to $\frac{48^3}{4} = 27648 \pmod{p}$. Then we obtain the curve $(t^2b, 0)$ and, if $(t_1^2b, 0) = (t_2^2b, 0)$ we can only conclude that $t_1 = \pm t_2$. Thus, we only obtain half of the curves in the isomorphism class! In any case, the number of solutions of these curves are still paired as described above. To obtain the other half of the curves we can choose one of the curves that we didn't obtain before and repeat the process to obtain all curves in this isomorphism class⁵.

Finally, if we start from the curve $(0, c)$ —that has j -invariant 0— we see that $(0, t_1^3c) = (0, t_2^3c)$ that only says that $t_1 = t_2u$, where u is a cubic root of 1 in \mathbb{Z}_p : for different values of p this can have only one solution (if $p \equiv 5 \pmod{6}$) or three different solutions (if $p \equiv 1 \pmod{6}$). So in this case it may again happen that to cover all the isomorphic curves we have to add to the family obtained from the initial curve some additional curves. But, again, in each family the symmetry in the cardinality remains valid⁶.

All together, this argument sheds some light on how the number of solutions of different curves with the same j -invariant (and so isomorphic over some field extension of \mathbb{Z}_p) are distributed.

Remark 4. The construction mapping C to C' used in the proof of Proposition 1 appears naturally as the isomorphism $m_{\sqrt{t}}$ for t non-square in \mathbb{Z}_p .

Example 5. Consider the case of $p = 5$, whose histogram is shown in Figure 1. The j -invariant $\pmod{5}$ ranges from 0 to 4.

The “generic case” corresponding to curves with coefficients b and c with $bc \neq 0$ is as follows: pick one curve, and choose $t = 1, 2, 3, 4$. For $t = 1, 4$ (the squares in

⁵Still, one may wonder, all these curves are isomorphic in some extension field. What is the extension required? A short computation shows that an extension of order 4 is required. For example, in the context of Example 3, $y^2 = x^3 + x$ and $y^2 = x^3 + 3x$ are isomorphic over $\mathbb{Z}_5[\sqrt[4]{3}]$.

⁶In this case we may need to consider extensions of order 6 of \mathbb{Z}_p to realize the isomorphism.

\mathbb{Z}_5) $m_{\sqrt{t}}$ produces a curve that is isomorphic over \mathbb{Z}_5 and thus has the same number of rational points over \mathbb{Z}_5 . For $t = 2, 3$ (the non-squares in \mathbb{Z}_5) $m_{\sqrt{t}}$ produces a curve that is isomorphic over a quadratic extension of \mathbb{Z}_5 ; in this case, the number of points is such that $\#C(\mathbb{Z}_5) + \#(m_{\sqrt{t}}C)(\mathbb{Z}_5) = 12$. Thus for the “generic case” in each isomorphism class there are 4 curves, two with the same number of points and two with the “symmetric” number.

Next consider the case when $b = 0$, that is the $j = 0$ case. Starting from the curve with $c = 1$, since every $w \in \mathbb{Z}_5$ is a cube we can take $t = \sqrt[3]{w}$ and the curve $(b, c) = (0, 1)$ is mapped to $(0, w)$ under $m_{\sqrt{t}}$. Since 1 and 4 are squares in \mathbb{Z}_5 , $(0, 1)$ and $(0, 3)$ have the same number of points over \mathbb{Z}_5 , while $(0, 2)$ and $(0, 4)$ have the symmetric number. In any case, all these curves have the same number of points, 6.

Finally we consider the case of $c = 0$. Starting from the curve $(b, c) = (1, 0)$ we choose $t = 2$ and obtain the curve $(4, 0)$, but since t is not a square both curves have symmetric number of points. The other values of $t \in \mathbb{Z}_5^*$ don't produce any new curves. So we pick another of the curves in the isomorphism class: $(2, 0)$. Choosing $t = 2$ produces the curve $(3, 0)$ that has the symmetric number of points. Notice that if we wanted to find an isomorphism between the curve $(1, 0)$ and $(2, 0)$ we have to choose $t = \sqrt{2}$, so that, eventually, the isomorphism will be defined over $\mathbb{Z}_5[\sqrt[4]{2}]$.

4. COUNTING POINTS OVER QUADRATIC EXTENSIONS OF \mathbb{Z}_p

As we noted in Section 3, for an elliptic curve C defined over \mathbb{Z}_p , we may have to consider points over an extension F of \mathbb{Z}_p . In this section we want to show how we can find $\#C(\mathbb{Z}_{p^2})$, knowing $\#C(\mathbb{Z}_p)$.

First we need some notation: let $N_m = \#C(\mathbb{Z}_{p^m})$. The zeta function of C is defined by

$$\zeta_C(s) = \exp\left(\sum_{m=1}^{\infty} \frac{N_m}{m} p^{-ms}\right).$$

The following result is part of a theorem due to A. Weil.

Theorem 6. $\zeta_C(s)$ is a rational function with poles at $s = 0$ and $s = 1$. The zeroes of ζ_C are located on the line $\text{Re}(s) = \frac{1}{2}$. Furthermore,

$$\zeta_C(s) = \frac{1 - Tp^{-s} + p^{1-2s}}{(1 - p^{-s})(1 - p^{1-s})} \quad (2)$$

for some $T \in \mathbb{Z}_p$.

Remark 7. The previous Theorem is part of a more general statement that has been proved by Weil not only for elliptic curves but also for curves of higher genus. This result was, in turn, extended to higher dimensional algebraic varieties by the work of a number of people, including Dwork, Artin, Grothendieck and Deligne.

Using the definition of ζ_C , and taking logarithm on both sides of (2) we obtain

$$\sum_{m=1}^{\infty} \frac{N_m}{m} p^{-ms} = \log(1 - Tp^{-s} + p^{1-2s}) - \log(1 - p^{-s}) - \log(1 - p^{1-s})$$

Since p^{-s} is small for $\text{Re}(s) \gg 0$ we can expand the logarithms using the formula $\log(1 - h) = -h - \frac{1}{2}h^2 - \dots$ and order the result by order of vanishing at $s = \infty$

(that is, in powers of p^{-s}):

$$\begin{aligned} N_1 p^{-s} + \frac{N_2}{2} p^{-2s} + \dots &= -(Tp^{-s} - pp^{-2s}) - \frac{1}{2}(Tp^{-s} - p^{-2s})^2 - \\ &\quad - (-p^{-s} - \frac{1}{2}p^{-2s}) - (-pp^{-s} - \frac{p^2}{2}p^{-2s}) + \dots \\ &= (-T + 1 + p)p^{-s} + \frac{1}{2}(2p - T^2 + 1 + p^2)p^{-2s} + \dots \end{aligned}$$

and we conclude that

$$\begin{aligned} N_1 &= -T + 1 + p \\ N_2 &= N_1(2(p + 1) - N_1). \end{aligned}$$

This last expression allows us to compute N_2 given N_1 . For example, the curves C_1 and C_3 of Example 3 have $N_1(C_1) = 4$ and $N_1(C_3) = 8$. Thus, applying the previous formula we get $N_2(C_1) = N_2(C_2) = 32$. Notice that we expected these numbers to be the same because both curves are isomorphic over \mathbb{Z}_{p^2} .

Remark 8. The process described above can be carried through to any order giving formulas for all the N_m in terms of N_1 . Hence, we conclude that two varieties having the same N_1 have the same zeta function.

E-mail address: jfernand@math.utah.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF UTAH, SALT LAKE CITY, UT 84112