ELLIPTIC CURVES AND SIEVE METHODS

V. Kumar Murty
University of Toronto

Talk at "Contemporary Methods in Cryptography"
January 9-13, 2002
Institute for Pure and Applied Mathematics, UCLA

Let E be an elliptic curve over a finite field $\mathbb{F}$.

In cryptography, we want the order

$$|E(\mathbb{F})|$$

to be prime, or nearly prime (in the sense that it is a prime times a small number).

How likely is this?

In this talk, we will explore conjectural answers as well as provable results.

Several approaches:

(a) We can consider all elliptic curves over $\mathbb{F}$ and compute how many have prime order

(b) We can consider one elliptic curve, say over $\mathbb{Q}$, and determine the primes p for which it has prime order over the finite field $\mathbb{F}_p$ of p elements.

In other words, we can fix the prime p and vary the curve E or we can fix the curve E and vary the prime p.

For the most part, we shall restrict our attention to (b).

We study the quantity

$$\pi(E, x) \ = \ \#\{p \leq x, E(\mathbb{F}_p) \text{has prime order}\}.$$

At the outset, there is a constraint coming from torsion:

$$E(\mathbb{Q})_{\text{tors}} \ \hookrightarrow \ E(\mathbb{F}_p).$$

This implies that the order of $E(\mathbb{F}_p)$ is divisible by the order of the rational torsion.

For example, for the curve

$$y^2 = x(x-1)(x+1)$$

we will have

$$4||E(\mathbb{F}_p)|$$

at all primes of good reduction. Similarly, for the curve

$$y^2 + y = x^3 - x^2$$

we will have

$$5||E(\mathbb{F}_p)|$$

for all primes of good reduction.

So in what follows, we shall suppose that

$$E(\mathbb{Q})_{\text{tors}} = \{\mathcal{O}\}.$$

We have

$$\gcd_p |E(\mathbb{F}_p)| = \mathrm{lcm}_{E' \sim E \text{ over } \mathbb{Q}} |E'(\mathbb{Q})_{\mathrm{tors}}|.$$

Hence, we should actually assume that both sides are equal to 1.

We also suppose that E does not have complex multiplication.

Conjecture of Koblitz (1988):

There exists a constant $C_E > 0$ so that

$$\pi(E, x) \sim C_E \frac{x}{(\log x)^2}.$$

In this conjecture, there are two aspects:

(a) the growth as a function of x

(b) the constant C_E

and both are mysterious.

Heuristics for the growth in x:

Let

$$N_p \;=\; |E(\mathbb{F}_p)| \;=\; p+1-a_p.$$

By Hasse's bound, we have

$$p+1-2\sqrt{p} \;\leq\; N_p \;\leq\; p+1+2\sqrt{p}.$$

If N_p is a random integer in this interval, the probability that it is prime is

$$\sim\; \frac{c}{\log p}$$

for some constant c. Hence, the number in question should be

$$\sum_{p\leq x} \frac{c}{\log p} \;\sim\; \frac{cx}{(\log x)^2}$$

by the Prime Number Theorem.

This has to be adjusted since the N_p are not quite random.

There are two kinds of constraints:

(a) Archimedean: Sato-Tate conjecture

(b) Non-archimedean: ℓ-adic representations

The effect of these constraints is different depending on whether E has or does not have complex multiplication.

Suppose that E does not have complex multiplication, that is

$$\text{End } E \;=\; \mathbb{Z}.$$

Let I be an interval in $[-1,1]$. Then, the Sato-Tate conjecture predicts that

$$\#\{p \le x : \frac{N_p - (p+1)}{2\sqrt{p}} \in I\} \;\sim\; \left(c \int_I \sqrt{1-x^2}dx\right) \pi(x)$$

where c is an explicit constant (not depending on I or E).

For a prime ℓ, let K_ℓ denote the field obtained by adjoining to $\mathbb{Q}$ the coordinates of points in E of order dividing ℓ.

K_ℓ is a Galois extension of $\mathbb{Q}$. Let us denote the Galois group by G_ℓ.

By a theorem of Serre, for large ℓ (depending on the curve),

$$G_\ell \;=\; \mathsf{GL}_2(\mathbb{F}_\ell).$$

The Chebotarev density theorem implies that

$$\#\{p \le x : \; N_p \;\equiv\; 0 \pmod{\ell}\} \;\sim\; \delta(\ell)\pi(x)$$

where $\delta(\ell)$ is the fraction of elements in G_ℓ that have an eigenvalue equal to 1.

In particular, if $G_\ell = \mathrm{GL}_2(\mathbb{F}_\ell)$, then

$$\delta(\ell) = \frac{\ell^2 - 2}{(\ell^2 - 1)(\ell - 1)} = \frac{1}{\ell} + \frac{1}{\ell^2} + \mathbf{O}(\frac{1}{\ell^3}).$$

Thus, we might expect that

$$C_E = C_\infty \cdot \prod_\ell \frac{\delta(\ell)}{\left(1 - \frac{1}{\ell}\right)}.$$

For example, if E is the curve

$$y^2 + y = x^3 - x$$

then

$$\prod_\ell \frac{\delta(\ell)}{\left(1 - \frac{1}{\ell}\right)} \simeq .5052.$$

In the case of complex multiplication, the constants have to be modified. Let K be the field of multiplication. Then, we know that if the prime p does not split in K,

$$N_p \;=\; p+1.$$

Easy problem: How often is $p+1$ prime?

Hard problem: How often is $(p+1)/2$ prime?

Hardy-Littlewood Conjecture: The number of primes $p \leq x$ such that $(p+1)/2$ is also prime is

$$\sim \prod_{\ell \geq 3} \left(1 - \frac{1}{(\ell-1)^2}\right) \frac{x}{(\log x)^2}.$$

Koblitz conjectures that the number of primes $p \leq x$ that split in K and for which $E(\mathbb{F}_p)$ has prime order is

$$\sim C_E \frac{x}{(\log x)^2}$$

where again, C_E is a product of local densities.

In this talk, we use sieve methods and the effective Chebotarev density theorem to prove two results.

This is joint work with S. A. Miri and appeared in Indocrypt 2001.

Theorem 1: Assume the GRH. Let E be an elliptic curve over $\mathbb{Q}$ without CM and satisfying the above hypotheses on torsion. Then

$$\#\{p \leq x : |E(\mathbb{F}_p)| \text{has at most 16 prime divisors}\}$$

is

$$\gg \frac{x}{(\log x)^2}.$$

Theorem 2: Assume the GRH. Then, except for a set of primes p of density zero, $|E(\mathbb{F}_p)|$ has $\sim \log\log p$ prime divisors.

The proof of Theorem 2 uses the normal order method. Let us set

$$\nu(N_p) \;=\; \#\text{distinct prime divisors of } N_p.$$

Then,

$$\sum_{p\le x} \nu(N_p) \;=\; \sum_{\ell\le x} \sum_{\substack{p\le x \\ N_p\equiv 0 \pmod{\ell}}} 1.$$

Now set

$$\pi(x,\ell) \;=\; \sum_{\substack{p\le x \\ N_p\equiv 0 \pmod{\ell}}} 1.$$

By the Chebotarev density theorem,

$$\pi(x,\ell) \;=\; \delta(\ell)\pi(x) \;+\; \mathbf{O}(\ell^3 x^{1/2}\log \ell N x).$$

Hence, with a parameter y to be chosen,

$$\sum_{\ell \le y} \pi(x,\ell) \;=\; \pi(x) \sum_{\ell \le y} \delta(\ell) + \mathbf{O}(y^4 x^{1/2} \log Nxy).$$

Choosing

$$y = x^{\frac{1}{8}-\epsilon}$$

the above is

$$\pi(x) \log\log x \;+\; \mathbf{O}(\pi(x)).$$

Now for the remaining values of ℓ,

$$\sum_{y<\ell\le x} \pi(x,\ell) \;=\; \sum_{p\le x} \sum_{\substack{\ell \mid N_p \\ y<\ell\le x}} 1$$

and this is

$$\ll \pi(x).$$

Hence,

$$\sum_{p\le x} \nu(N_p) \;=\; \pi(x) \log\log x \;+\; \mathbf{O}(\pi(x)).$$

Similarly,

$$\sum_{p \leq x} \nu(N_p)^2 \;=\; \pi(x)(\log\log x)^2 + \mathbf{O}(\pi(x)(\log\log x)).$$

Putting both of the above estimates together, we deduce that

$$\sum_{p \leq x} (\nu(N_p) - \log\log p)^2 \;\ll\; \pi(x)\log\log x.$$

Hence, given $\epsilon > 0$, we have

$$|\nu(N_p) - \log\log p| \;<\; \epsilon \log\log p$$

except possibly for

$$\ll\; \frac{\pi(x)}{\epsilon^2 \log\log x}$$

of the primes $p \leq x$.

The proof of Theorem 1 uses Selberg's lower bound sieve method. The general setup is as follows.

Let

$$f : \mathbb{N} \longrightarrow \mathbb{Z}$$

be a non-zero function.

Let

$$\mathcal{P} = \{p : p \leq x\}$$

be the set of primes $\leq x$. For an integer d, set

$$\mathcal{P}_d = \{p \in \mathcal{P} : f(p) \equiv 0 \pmod{d}\}$$

We write

$$|\mathcal{P}_d| = \frac{1}{\delta(d)}|\mathcal{P}| + R_d$$

where δ is a multiplicative function, and R_d is the "remainder".

Let $\{\alpha_n\}$ and $\{\lambda_n\}$ be two sequences of real numbers. Suppose that $\lambda_n = 0$ if either n is not squarefree or $n \gg 1$.

Consider the quantity

$$S = \sum_{p \in \mathcal{P}} \left(\sum_{d | f(p)} \alpha_d \right) \left(\sum_{\nu | f(p)} \lambda_\nu \right)^2 .$$

Then

$$S = |\mathcal{P}|\mathcal{S} + \mathbf{O}(\sum_m (\sum_{d|m} \alpha_d)(\sum_{\nu|m} |\lambda_\nu|)^2 R_m).$$

Here,

$$\mathcal{S} \;=\; \sum_{m} \sum_{(m,d)=1} \frac{\mu^2(m)}{\delta_1(m)} \frac{\alpha_d}{\delta_d} \left(\sum_{r|d} \mu(r) \zeta_{rm} \right)^2$$

and

$$\delta_1 = \delta * \mu$$

and

$$\zeta_r \;=\; \mu(r)\delta_1(r) \sum_{\nu} \frac{\lambda_{\nu r}}{\delta(\nu r)}.$$

By Möbius inversion, we can write the λ_ν in terms of ζ_r:

$$\lambda_\nu \;=\; \mu(\nu)\delta(\nu) \sum_{r} \frac{\mu^2(r\nu)}{\delta_1(r\nu)} \zeta_{r\nu}.$$

By appropriate choices of the α_d and the ζ_r, we can get interesting results.

Choose

$$\alpha_d \;=\; \begin{cases} 1 & if\ \ d=1 \\ 0 & if\ \ d>1 \end{cases}$$

and

$$\zeta_r \;=\; \begin{cases} \zeta_1 & if\ \ r<z\ \ is\ \ squarefree \\ 0 & otherwise. \end{cases}$$

Then,

$$\mathcal{S} \;=\; \sum_{m<z} \frac{\mu^2(m)}{\delta_1(m)} \zeta_1^2$$

and

$$\lambda_\nu \;=\; \mu(\nu)\delta(\nu)\zeta_1 \sum_{r<z/\nu} \frac{\mu^2(r\nu)}{\delta_1(r\nu)}.$$

Selberg's theorem gives

$$\sum_{p\in\mathcal{P}}\left(\sum_{\substack{\nu|f(p)\\ \nu<z}}\lambda_\nu\right)^2 = \zeta_1^2\left(\sum_{m<z}\frac{\mu^2(m)}{\delta_1(m)}\right)|\mathcal{P}|$$

$$+\ \mathbf{O}(\sum_m(\sum_{\substack{\nu|m\\ \nu<z}}|\lambda_\nu|)^2R_m).$$

Now choose

$$\alpha_d = \begin{cases} 1 & if\ d\ is\ a\ prime\ < y \\ 0 & otherwise. \end{cases}$$

Then

$$\mathcal{S} = \zeta_1^2 \sum_{m<z} \sum_{\substack{\ell<y \\ \ell \nmid m \\ \ell > z/m}} \frac{\mu^2(m)}{\delta_1(m)} \frac{1}{\delta(\ell)}.$$

Selberg's theorem gives

$$\sum_{p\in\mathcal{P}}\left(\sum_{\substack{d|f(p)\\ d<y\\ d \text{ prime}}} 1\right)\left(\sum_{\substack{\nu|f(p)\\ \nu<z}}\lambda_\nu\right)^2$$

$$= \zeta_1^2\sum_{m<z}\frac{\mu^2(m)}{\delta_1(m)}\{\sum\frac{1}{\delta(\ell)}\}|\mathcal{P}|$$

$$+ \mathbf{O}(\sum_m(\sum_{\substack{d|m\\ d<y\\ \text{prime}}} 1)(\sum_{\substack{\nu|m\\ \nu<z}}|\lambda_\nu|)^2R_m).$$

Hence,

$$\sum_{p\in\mathcal{P}} \left(2 \; - \sum_{\substack{d|f(p)\\ d<y\\ d \text{ prime}}} 1 \right) \left(\sum_{\substack{\nu|f(p)\\ \nu<z}} \lambda_\nu \right)^2$$

$$= \zeta_1^2 \left(\sum_{m<z} \frac{\mu^2(m)}{\delta_1(m)} |\mathcal{P}| \{ 2 - \sum_{\substack{\ell\not|m\\ z/m<\ell<y}} \frac{1}{\delta(\ell)} \} \right) + \text{error.}$$

If we can ensure that the error is negligible compared to the main term and that

$$2 - \sum_{\substack{\ell\not|m\\ z/m<\ell<y}} \frac{1}{\delta(\ell)} \; > \; 0$$

for some choice of parameters y and z, it will follow that for many primes,

$$2 \; - \sum_{\substack{d|f(p)\\ d<y\\ d \text{ prime}}} 1 \; > \; 0.$$

For these primes, $f(p)$ has at most 1 prime divisor $< y$.

Apply this with $f(p) = N_p$. In this case, for $\ell \gg 1$,

$$\delta(\ell) \;=\; \ell \;+\; \mathbf{O}(1).$$

Also, δ is a multiplicative function.

We see that

$$\sum_{\substack{\ell \nmid m \\ z/m < \ell < y}} \frac{1}{\delta(\ell)} \;\sim\; \sum_{\substack{\ell \nmid m \\ z/m < \ell < y}} \frac{1}{\ell}.$$

By elementary estimates from prime number theory,

$$\sum_{\ell \le t} \frac{1}{\ell} \;=\; \log(1+\log t) + c + \mathbf{O}(\frac{1}{\log t}).$$

Hence,

$$\sum_{z/m<\ell<y} \frac{1}{\ell} \;\sim\; \log(\frac{\log y}{\log z}) \;+\; \log(\frac{\log z}{1+\log z/m}).$$

Putting this into the above, we find that the essential requirement on y and z is that

$$1 \;-\; \log\left(\frac{\log y}{\log z}\right) \;>\; 0.$$

What about the error terms?

Return to the distribution law

$$|\mathcal{P}_d| \;=\; \frac{1}{\delta(d)}|\mathcal{P}| \;+\; R_d$$

where

$\mathcal{P}_d \;=\; \{p \in \mathcal{P} : N_p \;=\; p+1-a_p \;\equiv 0 \pmod{d}\}.$

The Chebotarev density theorem gives

$$R_d \;\ll\; d^3 x^{1/2} \log dNx$$

provided we assume the GRH.

Inserting these estimates into the expression for the error term, we find that we have to majorize

$$\sum_{m<yz^2} \mathbf{d}(m)^3 m^3 x^{1/2} \log(mNx) \ll x^{1/2}(yz^2)^{4+\epsilon}.$$

This is

$$\ll x^{1-\epsilon}$$

provided

$$yz^2 < x^{1/8-\epsilon}.$$

This is satisfied if we choose

$$y = x^{\frac{1}{16}+\epsilon}$$

and

$$z = x^{\frac{1}{32}-\epsilon}.$$

This choice also satisfies the other constraint, namely

$$1 - \log\left(\frac{\log y}{\log z}\right) \;=\; 1 \;-\; \log(2+\epsilon') \;>\; 0.$$

Hence, N_p has at most one prime divisor $< y = x^{\frac{1}{16}+\epsilon}$ and at most 15 prime divisors $> y$.

In other words, N_p has at most 16 prime divisors.

This can be refined. By choosing the α_d differently, we can show that for some $a < \frac{1}{16} + \epsilon$, there are at least

$$\gg \frac{x}{(\log x)^2}$$

primes $p \leq x$ for which N_p has all of its prime divisors $> p^a$.

Also, by working harder with the Chebotarev density theorem, we can prove the stronger estimate

$$R_d \ll dx^{1/2} \log dNx$$

and this will allow us to reduce 16 to a somewhat smaller number.

Further remarks:

a) In the non-CM case, without the GRH, we do not get a bounded number of prime factors with this method.

b) If we begin with a CM elliptic curve, it is possible that the GRH can be eliminated.

c) Related conjecture of Galbraith and McKee: fix p and vary the curve E over $\mathbb{F}_p$. They conjecture that the probability that such a curve has a prime number of points is asymptotic to

$$c_p P_1$$

as $p \longrightarrow \infty$ where

$$c_p = \frac{2}{3} \prod_{\ell > 2} \left(1 - \frac{1}{(\ell-1)^2}\right) \prod_{\substack{\ell | p-1 \\ \ell > 2}} \left(1 + \frac{1}{(\ell+1)(\ell-2)}\right)$$

and P_1 is the probability that a number in the interval

$$(p + 1 - 2\sqrt{p}, p + 1 + 2\sqrt{p})$$

is prime. Note that

$$0.41 < c_p < 0.62$$