Elliptic curves: applications and problems

Carl Pomerance, Dartmouth College

IMA Abel Conference, January, 2011

Factoring:

An idea of John Pollard:

Let M_y be the lcm of the integers in $[1, y]$. Suppose p is an odd prime with $p - 1 | M_y$. If n is an integer divisible by p, then

$$p \mid \gcd(2^{M_y} - 1, n).$$

The gcd can be computed in about y arithmetic steps with integers the size of n. Yet p could be much bigger than y, and if so, we could have a cheap way to discover a prime factor of n.

For example, $420 \mid M_7$, so 421 should be very easily discoverable as a prime factor of any number it divides.

To simplify slightly, say a number is y-smooth if all of its prime factors are in $[1, y]$. Then p is easily discoverable as a prime divisor if $p-1$ is y-smooth for a small value of y.

Because of this, cryptographers like to set up the RSA cryptosystem with so-called "safe" primes p of the form $2q+1$, where q is prime. Then $p-1$ is as unsmooth as possible.

Unsolved problem: Are there infinitely many safe primes?

Heuristically there are plenty of them, and this is borne out in practice—so this suffices for the practicing cryptographer.

Just about 27 years ago exactly, Hendrik Lenstra delivered some bad news to cryptographers: no prime is truly safe!

The Pollard $p-1$ factoring method depends on the unit group of $\mathbb{F}_p$ having smooth order. If it doesn't have smooth order, the method fails.

What Lenstra suggested was to replace $\mathbb{F}_p^\times$ with an elliptic curve group $E_{a,b}(\mathbb{F}_p)$ for random choices of a, b. If the curve is nonsingular, we know after Hasse that the order $\#E_{a,b}(\mathbb{F}_p)$ is in the interval $I_p := (p+1-2\sqrt{p},\ p+1+2\sqrt{p})$.

So, one can basically think that one is choosing a random integer in this Hasse interval. In fact Lenstra proved that if there are a fair number of choices of y-smooth integers in I_p, then there is a fair chance of landing upon one such choice, and so discover p as prime factor of some number n.

Helmut Hasse

Hendrik Lenstra

Unsolved problem: Are there a fair number of y-smooth integers in $I_p = (p + 1 - 2\sqrt{p},\ p + 1 + 2\sqrt{p})$?

Heuristically yes, with an optimal value of y as

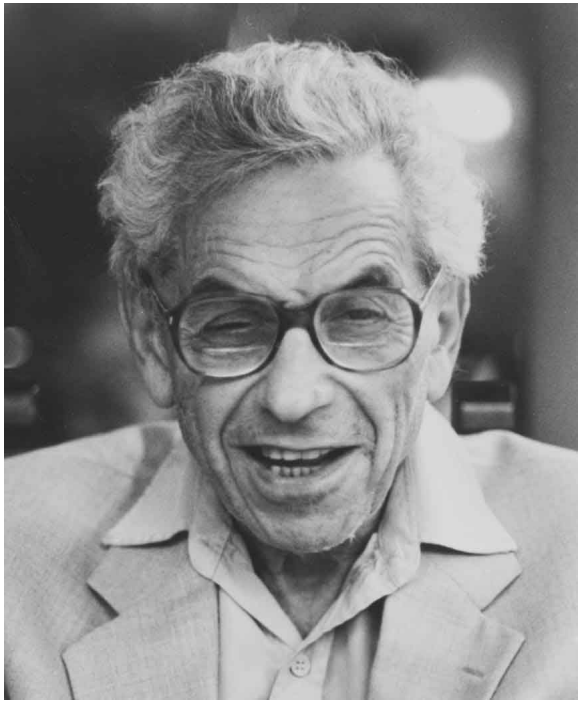
$$\exp\left(\sqrt{(1/2)\log p \log\log p}\right).$$

(A theorem of Canfield, Erdős, & P from 1983 asserts that with $L = \exp(\sqrt{\log p \log\log p})$, the probability that an integer in $[p/2,\ 3p/2]$ is L^α-smooth is about $L^{-1/(2\alpha)}$. If this holds for the smaller interval I_p and one applies the elliptic curve factoring method with $y = L^\alpha$, then the work per choice of curve is about L^α and the expected number of curves is about $L^{1/(2\alpha)}$, for a total of $L^{\alpha+1/(2\alpha)}$ steps. Thus, $\alpha = \sqrt{1/2}$ is optimal.)

However, rigorously, we cannot even prove that I_p has even one y-smooth number much less as many as suggested by the CEP theorem.

Luckily the numbers we are trying to factor do not know this! They get factored as quickly as we heuristically predict they should.

E. Rodney Canfield



Paul Erdős

Some work-arounds and progress:

In 1992, Lenstra & P gave a rigorous factorization algorithm with the same worst-case complexity that the elliptic curve method is conjectured to have. The algorithm uses quadratic forms of negative discriminant, not elliptic curves. However, it relies on examining many auxiliary numbers, keeping those that are y-smooth, until about y of them have been assembled. (With elliptic curve factoring, one needs just one y-smooth number.) One can use the elliptic curve method to examine these auxiliary numbers for y-smoothness, giving up after a pre-determined amount of effort is expended. This can be used as a subroutine in a rigorous algorithm since we were able to prove that the elliptic curve method *usually* works, and our auxiliary numbers are provably random enough so as not to skew things towards possible exceptional cases.

Soundararajan (2010): Assuming the RH, for each $\epsilon > 0$ there is some number $c(\epsilon)$ such that for all large x, the interval $[x, x + c(\epsilon)\sqrt{x}]$ contains at least one x^{ϵ}-smooth integer.

One can rigorously prove that slightly longer intervals have plenty of smooth numbers. In particular, intervals of the shape $[x, x + x^{3/4}]$ suffice. In a series of papers of Lenstra, Pila, & P from 1993, 2002, and "to appear" (actually, "to be written"), we prove this assertion and give a hyper-elliptic factorization method. This uses Jacobian varieties of hyper-elliptic curves of genus 2, and it stands (or will stand when completed) as the only rigorous method to recognize any given y-smooth number in fewer than y^{ϵ} elementary steps.

K. Soundararajan

Jonathan Pila

Primality testing:

Lucas (ca. 1876): If $a^{p-1} \equiv 1 \pmod{p}$ and $a^{(p-1)/q} \not\equiv 1 \pmod{p}$ for every prime $q \mid p-1$, then p is prime.

For example, $F_n := 2^{2^n} + 1$ is prime if and only if $3^{(F_n-1)/2} \equiv -1 \pmod{F_n}$.

The Lucas method is excellent whenever $p - 1 = \#\mathbb{F}_p^\times$ is easily factorable (basically a smooth number times a prime or prime power). What might one try if $p - 1$ is not easily factorable?

Hmmm...

In his doctoral dissertation from 1983, René Schoof gave a deterministic, polynomial-time algorithm to compute $\#E_{a,b}(\mathbb{F}_p)$. If $p-1$ is not easily factorable, then perhaps $\#E_{a,b}(\mathbb{F}_p)$ is?

In his doctoral dissertation from 1989, Joe Kilian, jointly with his advisor, Shafi Goldwasser, thought of applying the Lucas idea in the elliptic context, using the Schoof algorithm as the key subroutine. Theirs is a random algorithm that expects to rigorously prove primality for prime inputs p in polynomial time, provided each Hasse interval I_p contains as many easily factorable numbers as might be expected. In particular, it should have at least $\sqrt{p}/\log p$ integers of the form $2q$, with q prime. Heuristically, this is true.

If one hits upon a curve with order $2q$, then one can fashion a proof of "if q is prime, then p is prime". Then one can iterate, finding a curve of order $2r$ in I_q, and so on.

René Schoof

Joe Kilian (et al.)

Shafi Goldwasser

D. R. Heath-Brown

Unsolved provlem: Prove that each Hasse interval I_p contains at least $\sqrt{p}/\log p$ integers of the form $2q$ with q prime. Prove that the interval has at least one such number!

Using results of Heath-Brown it is possible to show that most short intervals contain many easily factorable numbers, and as a consequence, most primes can be proved prime in expected polynomial time via the Goldwasser–Kilian algorithm.

In 1992, Adleman & Huang found a way to rigorously remove any possible exceptional set from the Goldwasser & Kilian method. Namely, instead of using elliptic curves, use Jacobian varieties of hyper-elliptic curves of genus 2. Here the analog to the Hasse interval is long enough to guarantee that there are plenty of primes in the interval. Then one has a reduction: "if q is prime, then p is prime," but now $q \approx p^2$. It's hardly a reduction, but one gains *randomness*, and so it is likely we will land outside the Goldwasser–Kilian exceptional set, so that we can then descend using elliptic curves.

Leonard Adleman

Ming-deh Huang

In some sense this is all moot following the 2002 deterministic, polynomial-time primality test of Agrawal, Kayal, & Saxena. This test uses the arithmetic of finite fields and is less dependent on analytic number theory than the elliptic curve tests. The fastest deterministic version, due to Lenstra & P, runs in $(\log p)^{6+\epsilon}$ bit operations.

However, in another sense, the elliptic curve methods are alive and kicking. This is in the practical sense of actually proving large primes are really prime. The AKS test can maybe handle numbers of 100 digits, but with elliptic curves, we can handle numbers of 10,000 digits. However, we do not use Schoof's beautiful algorithm, but instead rely on curves with complex multiplication.

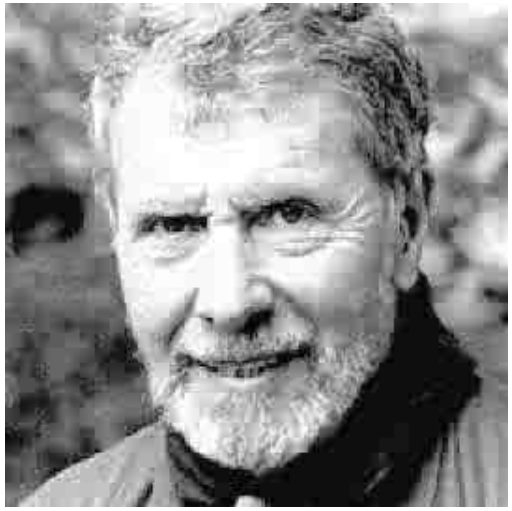
Suppose D is the discriminant of an imaginary quadratic field over $\mathbb{Q}$ and p is a prime which splits in $\mathbb{Q}(\sqrt{D})$ and for which there are integers u, v with

$$4p = u^2 + |D|v^2$$

(asymptotically, $1/h(D)$ primes p which split in $\mathbb{Q}(\sqrt{D})$ have this property, so $1/2h(D)$ in all; these are the primes that split in the Hilbert class field). Then there are elliptic curves over $\mathbb{F}_p$ with group orders

$$p + 1 \pm u.$$

(If $D = -3$ or -4 there are a few more curve orders.) It is fairly easy to find u, v if they exist (by an algorithm of Cornacchia), and somewhat harder to find actual curves with the orders $p + 1 \pm u$. The point being, if the group orders are not useful for us (e.g., easily factorable), we need not construct the actual curves.



Oliver Atkin

François Morain

Jeff Shallit

This then becomes the backbone of the Atkin–Morain elliptic curve primality test. With an improvement of Shallit, the heuristic running time to prove the primality of a prime p is $O((\log p)^{4+\epsilon})$ bit operations. The actual proof produced is shorter by a factor $\log p$.

There is actually one special elliptic curve over $\mathbb{F}_p$ which if we could produce it on demand, we could verify the primality of p in $O((\log p)^{2+\epsilon})$ bit operations. This was shown in [P, 1987] as follows. There is a number $m = 2^k w$ in the Hasse interval I_p with $2^k > 2\sqrt{p}$. Further, there is at least one elliptic curve $E_{a,b}(\mathbb{F}_p)$ with order m. Via this curve (and generators for the 2-Sylow subgroup) one can prove that p is prime in $O((\log p)^{2+\epsilon})$ bit operations.

So, there exist extremely short primality proofs. The rub is in actually finding such a special curve. Naively, some sort of Hensel iteration might be usable?

Unsolved problem: Given some prime p and integer $m \in I_p$, quickly find some elliptic curve $E_{a,b}(\mathbb{F}_p)$ with order m.

(By results of Deuring and Waterhouse, such curves exist.)

This problem is also of interest in cryptography. Many cryptosystems and signature schemes rely on the intractability of the discrete logarithm problem. (This problem: given a cyclic group $G = \langle g \rangle$ and an element $t \in G$, find an integer n with $g^n = t$.)

The discrete logarithm problem is highly dependent on the form in which the cyclic group is presented. For example, both $\mathbb{Z}/100\mathbb{Z}$ and $\mathbb{F}_{101}^{\times}$ are cyclic of order 100, but it is completely trivial (via Euclid) to compute discrete logs in the former group, and less trivial in the latter. (Both groups are generated by the element 3. Try to find the discrete log of 17 in each group and you will see what I am saying.)

Discrete log cryptosystems were first proposed for groups in the family $\mathbb{F}_q^\times$, where q is a prime power and where $q-1$ has a very large prime factor (say $q-1$ is prime or twice a prime). We have since developed sub-exponential discrete log algorithms for such groups, causing cryptographers to use expensively large values of q.

Neal Koblitz

Victor Miller

An alternative (Koblitz, Miller): use elliptic curve groups $E_{a,b}(\mathbb{F}_q)$ with order divisible by a very large prime, or better yet, prime order. Here we essentially only have generic meet-in-the-middle discrete log algorithms which take about $\sqrt{q}$ steps. So, because of our inability to come up with anything better to solve discrete logs, elliptic curve cryptography is a very viable and competitive platform.

To set up such a system, one needs a curve. Often it is nice to have some special underlying prime p or prime power q, so as to make the elliptic arithmetic somewhat more friendly (e.g., p is a Mersenne prime or q is a power of 2). This then raises the spectre of some of our unsolved problems: must there be a prime in the Hasse interval? How do we find a curve with such an order?

Cryptographers happily ignore the problem of whether there are primes in the Hasse interval, since heuristically (and so far in practice) there are plenty of them. Finding one such curve can then be accomplished via the (also unproved, but heuristic) methods of Atkin, Morain, & Shallit.

Unsolved problem: Find a fast way to compute discrete logarithms in an elliptic curve group, or prove that the problem is as hard as computing discrete logs in a generic group.

Dan Gordon

In the summer of 1986, just after he received his PhD at UCSD, Dan Gordon also thought of the Atkin–Morain idea, but was just a tad late. A disappointment for a new PhD, but he did manage to salvage a new idea, *elliptic pseudoprimes*.

We have seen that if the prime p that splits in an imaginary quadratic number field (and in fact, splits in the Hilbert class field for the quadratic field), then we can say something about certain elliptic curves over $\mathbb{F}_p$. On the other hand, suppose we have an elliptic curve $E_{a,b}(\mathbb{Q})$ which has complex multiplication by an order in the imaginary quadratic field $\mathbb{Q}(\sqrt{D})$ and p is a prime of good reduction which remains inert in the field. Then $\#E_{a,b}(\mathbb{F}_p) = p + 1$.

This conclusion about certain primes then might be applied to numbers n for which primality is unknown, so creating a test which hopefully weeds out all or most composites. We call such a procedure a probable prime test. For example, if $2^n \equiv 2 \pmod{n}$, we say n is a base-2 Fermat "probable prime". (A composite probable prime is called a pseudoprime.)

How might we develop the CM fact above into a probable prime test?

Gordon did this via division polynomials. Say we have an elliptic curve $E_{a,b}(\mathbb{Q})$. Let

$$\psi_0 =, \quad \psi_1 = 1, \quad \psi_2 = 2y, \quad \psi_3 = 3x^4 + 6ax^2 + 12bx - a^2,$$

$$\psi_4 = 4y(x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - 8b^2 - a^3)$$

with the recursion

$$\psi_{2k+1} = \psi_k^3\psi_{k+2} - \psi_{k+1}^3\psi_{k-1}, \quad 2y\psi_{2k} = \psi_k(\psi_{k+2}\psi_{k-1}^2 - \psi_{k-2}\psi_{k+1}^2).$$

If p is a prime of good reduction and $(x_1, y_1) \in E_{a,b}(\mathbb{F}_p)$, with $y_1 \not\equiv 0 \pmod{p}$, then for an integer $m > 2$, $[m]P = \mathcal{O}$ if and only if $\psi_m(x_1, y_1) \equiv 0 \pmod{p}$.

Now suppose that E has CM by an order in $\mathbb{Q}(\sqrt{D})$ ($D < 0$, class number 1), $P = (x_1, y_1)$ is a (rational) point on E of infinite order, and n is an odd natural number coprime to the discriminant of E, coprime to y_1, and with $(\Delta/n) = -1$. Then n is an elliptic probable prime with respect to E and P if $\psi_{n+1}(x_1, y_1) \equiv 0 \pmod{n}$. If n is also composite, it is an elliptic pseudoprime with respect to E and P.

Are elliptic pseudoprimes rare with respect to primes?

Let $N_{E,P}(x)$ denote the number of elliptic pseudoprimes with respect to E and P.

A flurry of results:

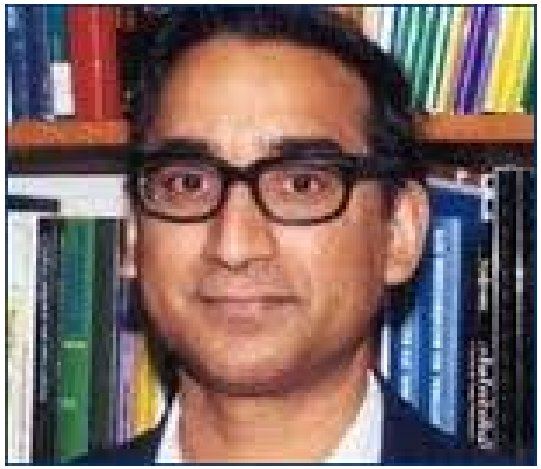
Gordon (1989): Assuming GRH, $N_{E,P}(x) \leq \dfrac{x \log\log x}{(\log x)^2}$. And for certain E, P, $N_{E,P} > \sqrt{\log x}/\log\log x$.

Miyamoto & Murty (1989): Unconditionally,
$N_{E,P}(x) \leq \dfrac{x(\log\log x)^{7/2}}{(\log x)^{3/2}}$.

Balasubramanian & Murty (1990): Unconditionally,
$N_{E,P}(x) \leq x^{1-c\sqrt{(\log\log x)/\log x}}$.

Gordon & P (1991): Unconditionally,
$N_{E,P}(x) \leq x^{1-(\log\log\log x)/(3\log\log x)}$.

R. Balasubramanian



M. Ram Murty

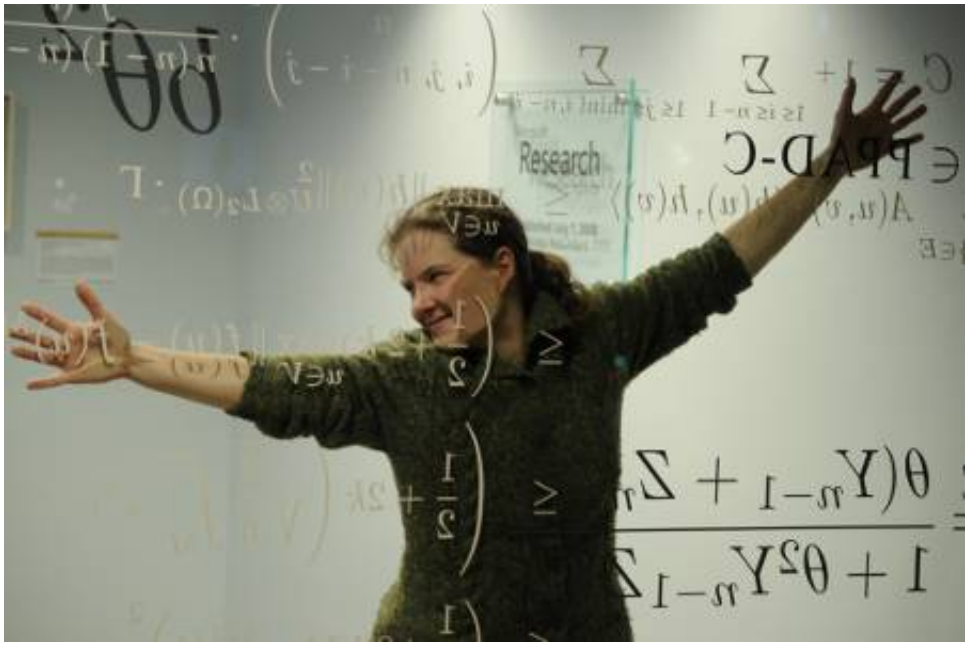
There have been some recent papers, as by Siguna Müller (2010), but the above counts remain unchanged since 1991.

Unsolved problem: Can one do better on the lower bound for $N_{E,P}(x)$?

Here is a related problem recently considered by Joseph Silverman and Katherine Stange. Given a non-singular elliptic curve E over $\mathbb{Q}$ and a rational point P of infinite order, consider the elliptic divisibility sequence D_n as defined above. Silverman & Stange (2010) study the algebraic structure of the numbers n with $n \mid D_n$, following the lead of Chris Smyth and others who studied the analogous problem for the Fibonacci sequence and for more general Lucas sequences. However elliptic divisibility sequences do not obey a linear recurrence.

And even for linear recurrences, though there is a fairly large literature on terms divisible by their subscripts, little seems to have been discussed statistically.

Joseph Silverman

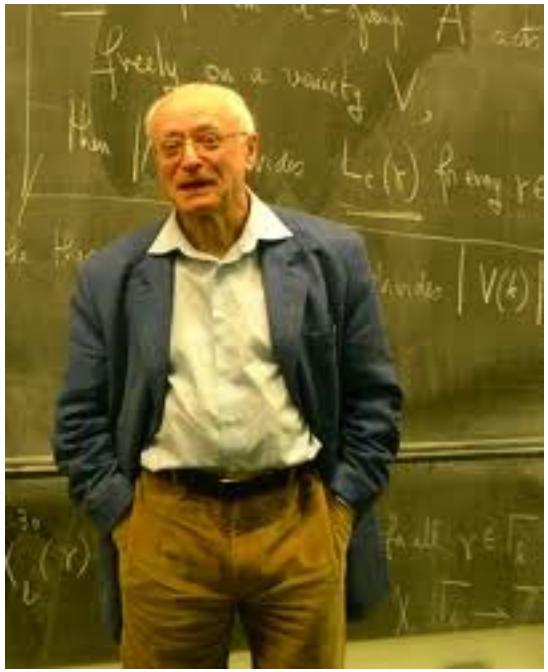


Katherine Stange

In a 2011(!) preprint, González, Luca, P, & Shparlinski showed that for a Lucas sequence (u_n) with characteristic polynomial $f(x)$ satisfying $|f(0)| = 1$, the number of integers $n \in [1, x]$ with $n \mid u_n$ is bounded between x^{c_1} and $x^{1-c_2\sqrt{(\log\log x)/\log x}}$. (The upper bound holds without the requirement that $|f(0)| = 1$.

Avram Gottschlich has just recently achieved a similar upper bound for the count of $n \in [1, x]$ with $n \mid D_n$. He was able to use somewhat similar techniques as in the above result for Lucas sequences to show that the count is bounded above by a function of the shape $x^{1-c\sqrt{(\log\log x)/\log x}}$, but only under the assumption that either the curve is CM or the GRH holds. Unconditionally for non-CM curves he has the count at most $x/(\log x)^{4/3+o(1)}$ using a result of Serre (1981) on *anomalous* primes (primes p with $p \mid D_p$).

Avram Gottschlich



Jean-Pierre Serre

Ranks:

Our last topic concerns ranks of elliptic curves. It is known after Mordell and Weil that the rank of the elliptic curve group for an elliptic curve over a global field is finite. It is a folk conjecture that this rank can be arbitrarily large.

Over $\mathbb{Q}$, the current record (Noam Elkies – 2006) has rank at least 28:

$$y^2 + xy + y = x^3 - x^2 - ax + b$$

where

$a =$

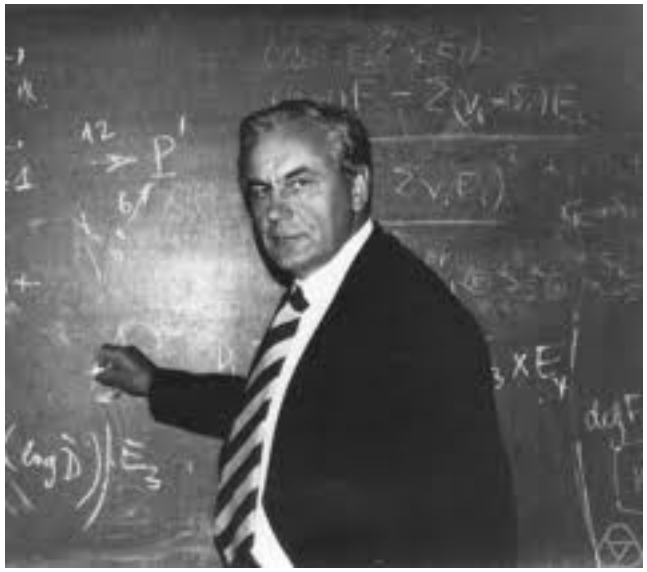
20067762415575526585033208209338542750930230312178956502

and

$b =$

3448161179503055646703298569039072037485594359319180361266008296291939448732243429.

Over $\mathbb{F}_p(t)$, Igor Shafarevich and John Tate showed in 1967 that ranks of elliptic curves can be arbitrarily large.



Igor Shafarevich

John Tate

The curves exhibited by Shafarevich & Tate are *isotrivial*, meaning the j-invariants are in $\mathbb{F}_p$. In 2002, Douglas Ulmer exhibited a family of curves with large rank over $\mathbb{F}_p(t)$ whose j-invariants are not in $\mathbb{F}_p$.

In particular, Ulmer considered curves over $\mathbb{F}_q$ in the family

$$E_d: \quad y^2 + xy = x^3 - t^d,$$

where d divides some number of the form $p^n + 1$ (where p is the characteristic of $\mathbb{F}_q$). Ulmer showed that the Birch & Swinnerton-Dyer conjecture holds for such curves E_d, they are not isotrivial, and he gave a formula for the rank, showing it is unbounded.

In the case of $q = p$ and $d = p^n + 1$, the rank of $y^2 + xy = x^3 - t^d$ over $\mathbb{F}_p(t)$ is within 4 of $\log(p^d)/\log(d^2)$. This expression tends to infinity with n, and compares very nicely with the universal upper bound of Brumer (1992):

$$\frac{\log(p^d)}{\log(d^2)}\left(1 + O\left(\frac{\log p}{\log d}\right)\right).$$

Douglas Ulmer

Armand Brumer

In his paper, Ulmer gave the exact rank for curves in his family: Let

$$I_q(d) = \sum_{m|d} \frac{\varphi(m)}{\ell_q(m)},$$

where $\ell_q(m)$ is the order of q in $(\mathbb{Z}/m\mathbb{Z})^\times$. (Recall that $d \mid p^n + 1$ for some n so that d and its divisors are coprime to q, a power of p.) From Ulmer's exact formula, we have that

$$I_q(d) - 4 \leq R_q(d) \leq I_q(d),$$

where $R_q(d)$ is the rank of E_d.

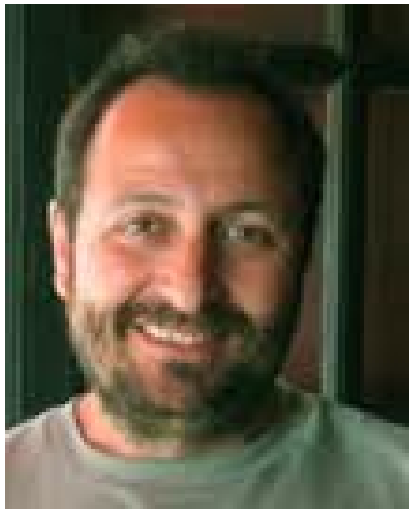
Brumer has shown that on average, ranks of elliptic curves over $\mathbb{F}_q(t)$ are bounded above by 2.3. One might then ask about the ranks of the general curves in Ulmer's family.

In 2010, P & Shparlinski showed a few statistical results about the curves E_d. Fix the prime p. On average, the rank of E_d is greater than d^α, where $\alpha > 1/2$ is a constant, and on average smaller than $d^{1-(\log\log\log d)/(2\log\log d)}$.

(For the upper bound, we assume that d is restricted to numbers which divide $p^n + 1$ for some n.)

Also, we show that for each $\epsilon > 0$, on a set of integers d of asymptotic density 1 (depending on ϵ, p), the rank exceeds $(\log d)^{(1/3-\epsilon)\log\log\log d}$.

We use the methods in a 1991 paper of Erdős, P, & Schmutz.



Igor Shparlinski

Eric Schmutz

Henri Darmon

The EPS paper deals with statistical properties of the universal-exponent function $\lambda(n)$, the order of the largest cyclic subgroup of $(\mathbb{Z}/n\mathbb{Z})^{\times}$.

It is hopeful that some improvements can be made here, including achieving a tight formula for the normal order of the rank for those d dividing $p^n + 1$ for some n. There are other curve families as well (such as some due to Darmon) that might be attacked.

THANK YOU!