Elliptic curve cryptography workshop 2003

Distribution of group orders of abelian varieties

Annegret Weng

Institute for Experimental Mathematics
University Essen-Duisburg

Waterloo, August 11-13, 2003

Overview

- Motivation
- Elliptic curves
- Higher-dimensional abelian varieties
- Open questions

Motivation

Let $(G, \circ)$ be a finite abelian group.

The discrete logarithm problem in G:
Given $g, h \in G$, find an integer $k \in \mathbb{Z}$ such that

$$\underbrace{g \circ \ldots \circ g}_{k \text{ times}} = h$$

if such a k exists.

Let $n = | < g > |$.

In any such group there exists an algorithm solving the discrete logarithm problem in time $O(\sqrt{\ell})$ where ℓ is the largest prime factor of n (Pohlig-Hellmann; Baby-Step-Giant-Step, Pollard-Rho).

In practice, we start with a family $\mathcal{C}$ of groups $(G, \circ)$ and try to find a group whose order is prime or almost prime[1].

For example:

- $\mathcal{C} = \{\mathbb{F}_p^* : p \text{ prime}, s_1 \leq p \leq s_2\}$,
- $\mathcal{C} = \{\mathrm{Cl}(\mathcal{O}_K) : \quad \text{K imag. quadratic number field}\}$,
- $\mathcal{C}_{a,b} = \{E(\mathbb{F}_p) : p \text{ prime}, s_1 \leq p \leq s_2\}$
  where $E : y^2 = x^3 + ax + b$,
- $\mathcal{C}_p = \{E(\mathbb{F}_p) : E \text{ elliptic curve}\}$.

[1] In this talk we use *almost prime* in the sense of *prime up to a small cofactor*. Note that this is not a precise definition but depends on the meaning of *small*.

Given a family $\mathcal{C}$, we might ask the following (theoretical) questions:

1. Can prime group orders occur in $\mathcal{C}$?

2. Can almost prime group orders occur in $\mathcal{C}$?
   Which cofactors can we have in $\mathcal{C}$?

3. What is the probability for a randomly chosen element in $\mathcal{C}$ to have prime or almost prime group order?

Examples:

1. $\mathbb{F}_p^*$, p large[2]: $|\mathbb{F}_p^*| = p - 1$ is never prime.
   When is $q_p = \frac{p-1}{2}$ prime?
   Not known whether there exist infinitely many **Sophie-Germain primes**.

2. $\mathbb{F}_{p^n}^*$: $|\mathbb{F}_{p^n}^*| = p^n - 1$ is divisible by $\Phi_k(p)$ for all $k \mid n$.
   If n prime: What is the probability for $\Phi_n(p)$ to be prime?

3. XTR: Here, $n = 6$. What is the probability that $\Phi_6(p) = p^2 - p + 1$ is prime?

4. Ideal class group Cl_K of a number field K:
   $|Cl_K| = h_K$ class number
   Cohen Lenstra heuristics estimates the probability that $\ell \mid h_K$.

[2]Note that in the first three examples there exist non-generic attacks (Index Calculus). Hence, in practice we choose a subgroup which is much smaller than p.

5. Elliptic curves $\#E(\mathbb{F}_p)$:
   Two possibilities:
   (a) E fixed, p varies: Given a fixed elliptic curve
   $$E : y^2 = x^3 + ax + b$$
   defined over $\mathbb{Q}$ (resp. a number field M), consider the reduction of E modulo different primes. What about $\#E(\mathbb{F}_p)$?
   (prime group orders, heuristics by Koblitz 1988)

   (b) p fixed, coefficients of E vary (heuristics by Galbraith, McKee, 1999).

6. $\#E(\mathbb{F}_{2^n})$: Fix $\mathbb{F}_q = \mathbb{F}_{2^n}$ and consider different elliptic curves.

7. Koblitz curves (subfield curves): E elliptic curve defined over a small field $\mathbb{F}_q$. Consider the DL problem in $\#E(\mathbb{F}_{q^n})$ for some n (n prime).
   What is the distribution of
   $$\frac{\#E(\mathbb{F}_{q^n})}{\#E(\mathbb{F}_q)}?$$
   (for fixed E and q, varying n, heuristics by Koblitz, 2001)

8. Hyperelliptic curves $\#J_C(\mathbb{F}_q)$: Similar to elliptic curves

From now on we restrict to $E(\mathbb{F}_p)$.

Statistics on the distribution of the group order gives a good feeling what happens in practice.

Take 10000 random curves over a prime field $\mathbb{F}_p$, $p \sim 2^{40}$:
We find 214 prime group orders and 864 group orders which are prime up to a cofactor ≤ 10.
Take 10000 prime fields $\mathbb{F}_p$, $p \sim 2^{40}$ and a fixed elliptic curve:
We find 187 prime group orders and 798 group orders which are prime up to a cofactor ≤ 10.
Below 10000 random numbers of size $\sim 2^{40}$ we find 372 prime numbers and 1102 number which a prime up to a cofactor ≤ 10.

Suppose we want to construct the curve using **complex multiplication**!

Let $D < 0$ the discriminant of the imaginary quadratic field. E.g. $E : y^2 = x^3 - x$, $\mathrm{End}(E) \simeq \mathbb{Z} + i\mathbb{Z}$, $D = -4$.

Compare $D = -67$ and $D = -11$:

S	$p \leq S$ p splits in $\mathbb{Q}(\sqrt{-11})$ $\|E \bmod p\|$ prime	$p \leq S$ p splits in $\mathbb{Q}(\sqrt{-67})$ $\|E \bmod p\|$ prime
10^3	8,5	43
10^4	35,5	193
10^5	220,5	1050
10^6	1427,5	6954

Many discriminants can not lead to a prime group order!

Suppose we want to construct a cryptographically secure elliptic curve over $\mathbb{F}_p$. We have two possibilies:

1. The complex multiplication method (complexity depends on the discriminant and the class number)

2. Choose random curves, count the number of points until a good curve is found.

Questions: What class number can we choose? When does the second approach becomes faster?

To answer these questions we need information about the distribution of group orders.

Consider the case E fixed over $\mathbb{Q}$ and reduce modulo different primes.

How often is has the reduction $\overline{E} := E \mod p$ prime group order?

What is the asymptotic behaviour of

$$\pi_E(x) = \{p \leq x : p \text{ prime}, N_p = \#\overline{E}(\mathbb{F}_p) \text{ prime}\}?$$

The twin prime problem: Are there infinitely many prime pairs $(p, p+2)$?

Solution not known!

<u>Our problem is very similar:</u>

Given an elliptic curve defined over $\mathbb{Q}$.

We are looking for two primes

- p for the field of definition $\mathbb{F}_p$ and
- $q = \#\overline{E}(\mathbb{F}_p)$ with $\overline{E} := E \mod p$.

By the Theorem of Hasse, p and q are of the same size.

Hardy and Littlewood had a heuristics for the number of twin prime pairs.

This can be applied to elliptic curves (Koblitz, 1988):

$\pi(n)$ =number of primes smaller than n
By the prime number theorem,

$$\pi(n) \sim \frac{n}{\log n}.$$

Hence, the probability of a number of size n to be prime is approximately

$$\frac{1}{\log n}.$$

We expect that the number of pairs (p, q), $p \leq n$, p, q of the same size is roughly

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

But the prime number theorem assume that a number is divisible by ℓ with probabiliy $\frac{1}{\ell}$!

For group orders of elliptic curves this is wrong!

For every ℓ we have to insert a correction term

$$a_\ell = \frac{1 - p_\ell}{1 - \frac{1}{\ell}}$$

(p_ℓ - probability that $\#E(\mathbb{F}_p)$ is divisible by ℓ).

Set $C := \prod a_\ell$ (product should converge!). Let $N_p = \#E(\mathbb{F}_p)$ and let

$$\pi_E(x) = \{p \leq x : p \text{ prime}, N_p \text{ prime}\}.$$

Asymptotically we expect

$$\pi_E(x) \sim C \frac{n}{\log^2 n}.$$

How do we get the probability p_ℓ?

- $E[\ell]$ ℓ-torsion points over $\overline{\mathbb{Q}}$.
  $\mathbb{Q}(E[\ell])/\mathbb{Q}$ is Galois.

- $G_\ell := Gal(\mathbb{Q}(E[\ell])/\mathbb{Q})$ operates on
  $$E[\ell] \simeq (\mathbb{Z}/\ell\mathbb{Z})^2,$$
  more precisely
  $$G_\ell \subseteq Gl_2(\mathbb{F}_\ell).$$

We have
$$p_\ell = \frac{|\{g \in G_\ell : g \text{ eigenvalue one}\}|}{|G_\ell|}.$$

Why?

Let $p \neq \ell$ be a prime such that E has good reduction modulo p. $\Rightarrow p$ in $\mathbb{Q}(E[\ell])$ unramified.
For every prime ideal $\mathfrak{p}$ over p, the Artin symbol

$$\sigma_{\mathfrak{p}} = \left(\frac{\mathbb{Q}(E[\ell])/\mathbb{Q}}{\mathfrak{p}}\right)$$

is well-defined.

Theorem 1. *The group order $E \mod p$ is divisible by ℓ if $\sigma_{\mathfrak{p}}$ for some $\mathfrak{p}$ (for all $\mathfrak{p}$ over p) as element in $Gl(2, \mathbb{F}_\ell)$ has eigenvalue 1.*

Proof. $\sigma_{\mathfrak{p}}$ has eigenvalue one $\Leftrightarrow$
$\sigma_{\mathfrak{p}}$ leaves a 1-dimensional subspace of $E[\ell]$ fixed. $\Leftrightarrow$ Frobenius leaves a 1-dimensional subspace of $E[\ell]$ fixed $\Leftrightarrow$ All elements of a 1-dimensional subspace of $E[\ell]$ are defined over $\mathbb{F}_p$. $\square$

Theorem of Chebotarev:

The Frobenius elements are equally distributed in $Gal(E(\mathbb{Q}[\ell])|\mathbb{Q})$.

Example 1. (non-CM)

$$E : y^2 + y = x^3 - x,$$

We have $Gal(\mathbb{Q}(E[\ell])/\mathbb{Q}) \simeq Gl(2, \mathbb{F}_\ell)$ for all ℓ (Serre) and

$$p_\ell = \frac{\ell^2 - 2}{(\ell - 1)(\ell^2 - 1)}.$$

Example 2. (CM) Let E be an elliptic curve defined over $\mathbb{Q}$ with complex multiplication by $\mathcal{O}_K$, $K = \mathbb{Q}(\sqrt{-D})$ e.g. $y^2 = x^3 - x$, $\mathcal{O}_K = \mathbb{Z} + i\mathbb{Z}$.

We have $Gal(\mathbb{Q}(E[\ell])/\mathbb{Q}) \simeq (\mathcal{O}_K/\ell\mathcal{O}_K)^*$ and

$$p_\ell = \begin{cases} \frac{2\ell-1}{(l-1)^2}, & \text{if } \ell \text{ splits} \\ \frac{1}{\ell^2-1}, & \text{if } \ell \text{ inert} \\ \frac{1}{\ell-1}, & \text{if } \ell \text{ ramifies.} \end{cases}$$

Consider $\ell = 2$: We have $\#E(\mathbb{F}_p)$ is always divisible by 2 if ℓ splits or ramifies.

If the small primes tend to be inert in K, we get more prime group orders!

What is the probability P_k that the group order is of the form k·prime? (A.W. 2002)

- a_{ℓ^e} - prob. that $\#\overline{E}(\mathbb{F}_p)$ is divisible by ℓ^e but not ℓ^{e+1}.
- Set
$$c_{k,x} = \prod_{\ell^e || k} a_{\ell^e} \prod_{\ell \text{ prime}, \ell \nmid k, \ell \leq x} (1 - p_\ell)$$
and
$$a_{k,x} = \prod_{\ell^e || k} \frac{\ell - 1}{\ell^{e+1}} \prod_{\ell \text{ prime}, \ell \nmid k, \ell \leq x} (1 - \frac{1}{\ell}).$$
- $Q_k := \lim_{x \to \infty} \frac{c_{k,x}}{a_{k,x}}$

A number of size p equals k·prime with prob. approx.
$$\frac{1}{k(\log p - logk)}.$$
Given an intervall $[2^n, 2^n + c]$ we expect that
$$Q_k \sum_{p \in [2^n, 2^n + c]} \frac{1}{k(\log p - \log k)}$$
group orders are of the form $k \cdot prime$.

Experimental results on non-CM curves

Curve of Example 1						
k	$n = 10^4$		$n = 10^5$		$n = 10^6$	
	#curves	heuristics	#curves	heuristics	#curves	heuristics
1	366	362.78	3584	3521.51	32128	32052.04
2	304	286.34	2747	2775.25	25138	25147.52
3	211	194.50	1916	1883.27	17024	17017.11
4	234	226.63	2173	2192.80	19721	19771.91
5	94	102.49	1012	991.11	8978	8921.37
6	127	154.20	1472	1490.46	13213	13397.23
7	63	70.28	684	678.99	6005	6095.17
8	162	139.93	1364	1351.44	12219	12119.77
9	77	80.36	766	775.82	7035	6950.77
10	64	81.45	788	768.13	7010	7036.91

Experimental results on CM curves

Discriminant $D = -11$						
	$n = 10^4$		$n = 10^5$		$n = 10^6$	
	#curves	heuristics	#curves	heuristics	#curves	heuristics
1	311	323.49	2173	2178.2	27761	27634.14
2	0	0	0	0	0	0
3	480	468.09	3082	3066.2	39560.5	39504.18
4	134.5	134.66	886.5	875.2	11315.5	11324.60
5	193	194.97	1225	1259.1	16366.5	16350.47
6-8	0	0	0	0	0	0
9	339	341.15	2177	2164.9	28378	28385.60
10	0	0	0	0	0	0

Galbraith, McKee: Fix a prime p and consider how often a randomly chosen curve over $\mathbb{F}_p$ has almost prime group order.

Heuristics. P_k - prob. that a number in $[p+1-2\sqrt{p}, p+1+2\sqrt{p}]$ is k·prime. The prob. that $E(\mathbb{F}_p)$ has order k·prime is asymptotic to $c_p f_k P_k$ where

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

and

$$f_1 = 1, f_{\ell^t} = \frac{\ell^t(r(\ell^t) - r(\ell^{t+1}))}{(1 - r(\ell))}$$

with $r(\ell^t) = \begin{cases} \frac{1}{\ell^{t-1}(\ell-1)} & \text{if } p \not\equiv 1 \mod \ell^u \\ \frac{\ell^{v+1}+\ell^v-1}{\ell^{t+v-1}(\ell^2-1)} & \text{if } p \equiv 1 \mod \ell^u \end{cases}$

where $u = \lceil \frac{t}{2} \rceil$ and $v = \lfloor \frac{t}{2} \rfloor$.

Their Idea: For each number in the Hasse interval for p, consider how often it is obtained as the group order of an elliptic curve using Kronecker's class number.

Comparison fixed prime-fixed curve

The table lists constants $c_p f_K$ resp. Q_k:

k	Fixed prime field $\mathbb{F}_p$ $10^{20}+151$	$10^{20}+129$	Non-CM	CM D=-163	D=-899
1	0.4646	0.5517	0.5052	3.5262	0.2725
2	0.6969	0.8276	0.7578	0	0
3	0.9292	0.5885	0.7484	0	1.0907
4	0.9292	1.3793	1.1368	5.2894	0.4088
5	0.4695	0.7356	0.6312	0	0.7267
6	1.3938	0.8828	1.1227	0	0
7	0.5575	0.6621	0.5894	0	0.6540
8	1.3938	1.2414	1.3263	0	0
9	0.9292	0.7724	0.8483	4.0367	2.1801
10	0.7042	1.1035	0.9468	0	0

Because of the analogy to the twin prime problem, a proof of results on prime group orders would be a real break-through!

Twin primes (J.R. Chen, 1983): There exist infinitely many primes p such that $p+2$ has at most two divisors. Proof uses sieving techniques from analytic number theory.

Theorem 2. *(S.A. Miri, K.V. Murty, Indocrypt 2001) Assume GRH.*
If E with $End(E) \simeq \mathbb{Z}$ and $E(\mathbb{Q})_{tor}$ has a trivial torsion group over $\mathbb{Q}$, then $\Omega(N_p) \leq 16$ for $\gg \frac{N}{(\log N)^2}$ primes $p \leq N$.

Theorem 3. *(J. Steuding, A.W. 2002)*
E elliptic curve over $\mathbb{Q}$, $E' \sim E$ trivial $\mathbb{Q}$-torsion group. Assume GRH, then:

i) If E does not have CM, then

$$\#\{N_p : p \leq N, \nu(N_p) \leq 5\} \geq C_1 \frac{N}{(\log N)^2};$$

($C_1(E)$ pos. comp. constant; $\Omega(N_p) \leq 8$).

ii) If E has CM with $\mathcal{O}$, χ corr. quadratic character, then

$$\#\{N_p : p \leq N, \chi(p) = 1, \Omega(N_p) \leq 3\} \geq C_2 \frac{N}{(\log N)^2};$$

where $C_2(E)$ is pos. computable constant.

Conclusions

- To prove something (of practical importance) seems to be difficult, but we can give good heuristics to support our experimental datas.

- The group orders of elliptic curves are less often prime number than randomly chosen curves.

- If we only ask for almost prime numbers, the difference between random numbers and group orders of elliptic curves seems less dramatic.

- If we fix the prime p, the distribution depends on p.

- But CM curves can be better!

- We can easily modify the heuristics to other situations. E.g. If we select the curve by an early-abort strategy, so we can assume that the group order is never divisible by some prime ℓ.

Let A be an (principally polarized) abelian variety of dimension $g \geq 2$.
E.g. Jacobian of a hyperelliptic curve

It is difficult to prove anything!

Can try to modify the Koblitz heuristic argument to this case.
Again we are looking for two primes:

- p for $\mathbb{F}_p$ and
- $q = \#A(\mathbb{F}_p)$.

If p is of size approximately n, then $q = |A(\mathbb{F}_p)|$ is of size n^g.

Let

$$a_\ell = \frac{1 - p_\ell}{1 - 1/\ell}$$

where p_ℓ is the probability that ℓ divides $\#A(\mathbb{F}_p)$.

Set $C := \prod a_\ell$. The asymptotic value of the number of prime orders should be

$$C \frac{n}{g \log^2 n}.$$

More difficult to determine
$G_\ell =$Gal$(\mathbb{Q}(A[\ell])/\mathbb{Q})$ and p_ℓ!

If A has **no** complex multiplication, i.e. $End(A) \simeq \mathbb{Z}$, and $\dim(A) = g = 2, 6$ or odd, then by Serre

$$Gal(\mathbb{Q}(A[\ell])/\mathbb{Q}) \simeq GSp(2g, \ell)$$

for all but finitely many ℓ.

$$GSp(2, \mathbb{F}_\ell) = \{M \in Gl(2g, \mathbb{F}_\ell) : M^t J M = \alpha J, \alpha \in \mathbb{F}_l^*\}$$

with $J = \begin{pmatrix} 0 & I_g \\ -I_g & 0 \end{pmatrix}$.

Again,

$$p_\ell = \frac{|\{g \in G_\ell : g \text{ eigenvalue one}\}|}{|G_\ell|}.$$

Count the number of matrices in $\mathrm{GSp}(2g,\ell)$ with eigenvalue one (using an idea of Washington for $\mathrm{Gl}(n,\ell)$). We find:

	$\#M_g(1)$
$g=2$ (cf. Kuhlmann)	$\ell^{10}-3\ell^{8}+\ell^{7}-\ell^{6}+\ell^{5}+4\ell^{4}$
$g=3$	$\ell^{21}-3\ell^{19}+\ell^{18}-\ell^{17}+2\ell^{16}+2\ell^{15}$ $-\ell^{14}+3\ell^{13}-3\ell^{12}-2\ell^{10}-6\ell^{9}$
$g=5$	$\ell^{36}-3\ell^{34}+\ell^{33}-\ell^{32}+2\ell^{31}+2\ell^{30}-$ $\ell^{29}+2\ell^{28}-4\ell^{27}+4\ell^{26}-2\ell^{25}$ $-5\ell^{24}+3\ell^{21}-2\ell^{20}+$ $6\ell^{19}+\ell^{18}+3\ell^{17}+8\ell^{16}$

Let A be an abelian variety with $\mathrm{End}(A) \simeq \mathcal{O}_K$ where $\mathcal{O}_K$ is the ring of integers in a CM field.
Here, the Galois group G_ℓ=Gal$(\mathbb{Q}(A[\ell])/\mathbb{Q})$ is abelian. If ℓ unramified, it is contained in a maximal torus of $\mathrm{GSp}(2g, \ell)$.

For $g = 2$ we can determine G_ℓ and p_ℓ (A.W. 2001). By class field theory, we can show that G_ℓ is a maximal torus for all but finitely many ℓ . For p_ℓ we find

$\frac{4\ell^2-14\ell+13}{(\ell-1)^3}$ ℓ splits completely,

$\frac{2l+1}{(l+1)^2(l-1)}$ ℓ splits only in $K_0/\mathbb{Q}$,

$\frac{2l-3}{(l-1)^2(l+1)}$ ℓ splits only in K/K_0,

$\frac{2l^2-5}{(l-1)^2(l+1)}$ ℓ splits in three prime ideals,

$\frac{1}{(l^2+1)(l-1)}$ ℓ inert.

We can set up a similar heuristics.

	CM-field $\mathbb{Q}(i\sqrt{8+3\frac{-1+\sqrt{13}}{2}})$	
n	real value	Prediction
10^3	1,5	1,6
10^4	9,3	9,1
10^5	58	55,3
10^6	369	369,0
	CM-field $\mathbb{Q}(i\sqrt{17+5\frac{-1+\sqrt{29}}{2}})$	
n	real value	Prediction
10^3	7	6,0
10^4	29,5	32,3
10^5	194	195,1
10^6	1312,5	1313,4

Probability for a prime group order is really high!

Some open questions:

- Consider the case where $\mathbb{F}_q$ is fixed in the case of hyperelliptic curve of genus 2/higher-dimension abelian varieties![3]

- Find unconditional results not assuming GRH!

- Prove the twin prime problem! Any such proof would probably give some hint how to prove all the problem mentioned in this talk![4]

[3] Some work in this direction has been done by Adleman-Huang in *Primality testing using abelian varieties, LNM 1512*

[4] The twin prime problem is very closely related to the Goldbach conjecture. For the proof of the Goldbach conjecture one can earn 1 Million Dollar!