THE PROBABILITY THAT THE NUMBER OF POINTS ON AN ELLIPTIC CURVE OVER A FINITE FIELD IS PRIME

STEVEN GALBRAITH AND JAMES MCKEE

ABSTRACT. The paper gives a formula for the probability that a randomly chosen elliptic curve over a finite field has a prime number of points. Two heuristic arguments in support of the formula are given as well as experimental evidence.

The paper also gives a formula for the probability that a randomly chosen elliptic curve over a finite field has kq points where k is a small number and where q is a prime.

1. INTRODUCTION

Cryptographic and computational applications have recently motivated the study of several questions in the theory of elliptic curves over finite fields. For instance, the analysis of the elliptic curve factoring method leads to estimates ([7], [8]) for the probability that the number of points on an elliptic curve is smooth.

In this paper, motivated by the use of elliptic curves in public key cryptosystems, we consider the "opposite" problem. More specifically, we ask the question: What is the probability that a randomly chosen elliptic curve over $\mathbb{F}_p$ has kq points, where k is small and q is prime? Initially we take p to be prime. The minor modifications needed to deal with arbitrary finite fields are considered later.

Koblitz [5] has considered the analogous problem when the elliptic curve E is fixed and where it is the prime p which varies. The paper [5] gives a conjectural formula that the number of primes $p \leq n$ such that $\#E(\mathbb{F}_p)$ is prime is asymptotic to $C_E n/(\log n)^2$ (where C_E is an explicitly given constant depending on E).

To make our question precise, we need to specify what we mean by k being small, and also what we mean by a randomly chosen elliptic curve. The first point is easily dealt with: we fix K and consider $k \leq K$, allowing p to tend to infinity.

Regarding the second point, we use the Weierstrass model (in the case $p > 3$) for elliptic curves:

$$E : y^2 = x^3 + ax + b, \qquad 4a^3 + 27b^2 \neq 0\,.$$

Then whenever we refer to "the probability that an elliptic curve over $\mathbb{F}_p$ has property P", we shall mean "the probability that the elliptic curve defined by

1991 *Mathematics Subject Classification.* 11G20.
The first author thanks the EPSRC for support.

$y^2 = x^3 + ax + b$ over $\mathbb{F}_p$ has property P, given that (a, b) is uniformly distributed in $\mathbb{F}_p^2 - \{(a, b) : 4a^3 + 27b^2 = 0\}$".

The effect of using this model is to count each isomorphism class of curves with weight inversely proportional to the size of the automorphism group. All classes are given equal weight, except for those with j-invariant either 0 or 1728.

We can now state our conjectured answer to the above question, isolating first the special case of a prime number of points.

Conjecture A. *Let P_1 be the probability that a number within $2\sqrt{p}$ of $p+1$ is prime. Then the probability that an elliptic curve over $\mathbb{F}_p$ has a prime number of points is asymptotic to $c_p P_1$ as $p \to \infty$, where*

$$c_p = \frac{2}{3} \prod_{l>2} \left(1 - \frac{1}{(l-1)^2}\right) \prod_{l \mid p-1,\ l>2} \left(1 + \frac{1}{(l+1)(l-2)}\right).$$

Here the products are over all primes *l satisfying the stated conditions.*

Remarks

(1) Note that $\prod_{l>2} \left(1 - \frac{1}{(l-1)^2}\right) \approx 0.6601618$ is the Hardy-Littlewood twin-primes constant.

(2) P_1 can be approximated by

$$\frac{1}{4\sqrt{p}} \int_{p+1-2\sqrt{p}}^{p+1+2\sqrt{p}} \frac{\mathrm{d}t}{\log t} \approx \frac{1}{\log p}.$$

Part of the difficulty of the problem is that we do not know enough about primes in short intervals to say how good an approximation this is.

(3) If all numbers of points between $p+1-2\sqrt{p}$ and $p+1+2\sqrt{p}$ were equally likely, then we would have $c_p = 1$. The expression given for c_p lies between about 0.44 and 0.62, indicating the phenomenon that prime numbers of points are disfavoured.

Conjecture A is interesting in itself, regardless of cryptographic applications. We shall give two heuristic derivations of this same formula, in some detail. We shall also provide experimental evidence which lends strong support to the conjecture. We also briefly indicate the necessary modifications to generalise this to:

Conjecture B. *For $k = 1, 2, 3, \ldots$, let P_k be the probability that a number within $2\sqrt{p}$ of $p+1$ is of the form kq, with q a prime. Then the probability that an elliptic curve over $\mathbb{F}_p$ has k times a prime for its number of points is asymptotic to $c_p f_k P_k$ as $p \to \infty$, where c_p is as in Conjecture A, and f_k is a rational number, defined below, depending only on k and $\gcd(k, p-1)$. For fixed p, f_k is a multiplicative function of k. For a prime power l^t $(t \geq 1)$, we have*

$$f_{l^t} = \frac{l^t \left(r(l^t) - r(l^{t+1})\right)}{1 - r(l)},$$

where

$$r(l^t) = \begin{cases} \frac{1}{l^{t-1}(l-1)} & \text{if } q \not\equiv 1 \pmod{l^u} \\ \\ \frac{l^{v+1}+l^v-1}{l^{t+v-1}(l^2-1)} & \text{if } q \equiv 1 \pmod{l^u} \end{cases}$$

where $u = \lceil t/2 \rceil$, $v = \lfloor t/2 \rfloor$.

Remarks

(1) P_k can be approximated by

$$\frac{1}{4k\sqrt{p}} \int_{(p+1-2\sqrt{p})/k}^{(p+1+2\sqrt{p})/k} \frac{\mathrm{d}t}{\log t} \approx \frac{1}{k(\log p - \log k)} .$$

(2) The following table gives values of f_k for some small prime powers k, and for prime k.

k	f_k	
1	1	
2	3/2	
3	16/15	if $p \equiv 1 \pmod 3$
	2	if $p \equiv 2 \pmod 3$
4	5/2	if $p \equiv 1 \pmod 4$
	2	if $p \equiv 3 \pmod 4$
5	96/95	if $p \equiv 1 \pmod 5$
	4/3	if $p \not\equiv 1 \pmod 5$
8	9/4	if $p \equiv 1 \pmod 4$
	3	if $p \equiv 3 \pmod 4$
9	22/15	if $p \equiv 1 \pmod 9$
	7/5	if $p \equiv 4, 7 \pmod 9$
	2	if $p \equiv 2 \pmod 3$
16	11/4	if $p \equiv 1 \pmod 8$
	3	if $p \not\equiv 1 \pmod 8$
prime l	$1 + 1/(l^3 - l^2 - l)$	if $p \equiv 1 \pmod l$
	$(l-1)/(l-2)$	if $p \not\equiv 1 \pmod l$

(3) Let Q_k denote the probability that an elliptic curve over $\mathbb{F}_p$ has kq points for some prime q. Then Conjecture A reads:

$$Q_1 \sim c_p P_1, \qquad \text{as } p \to \infty ,$$

and Conjecture B reads:

$$Q_k \sim c_p f_k P_k, \qquad \text{as } p \to \infty .$$

2. Numerical evidence

Evidence for Conjecture A

p	P_1	c_p	c_pP_1	Q_1
1009	0.15748	0.56389	0.08880	0.09366
1019	0.14173	0.44011	0.06238	0.07262
10007	0.10723	0.44011	0.04719	0.04762
10009	0.10723	0.55016	0.05900	0.06025
100003	0.08538	0.56389	0.04814	0.04889
100043	0.08379	0.44011	0.03688	0.03883
199999	0.07714	0.55048	0.04246	0.04339
10000019	0.06238	0.45121	0.02815	0.02766
10000079	0.06218	0.44506	0.02769	0.02708
10000537	0.06174	0.56923	0.03515	0.03497
500000003	0.05042	0.44038	0.02221	0.02215
500000009	0.05042	0.44011	0.02219	0.02213
500000041	0.05045	0.58070	0.02929	0.02925
500000069	0.05046	0.44019	0.02221	0.02206
500000071	0.05046	0.58070	0.02930	0.02934
10000000019	0.04330	0.44014	0.01906	0.01900
10000000033	0.04330	0.55729	0.02413	0.02408
10000000061	0.04330	0.46456	0.02011	0.02009
10000000069	0.04330	0.55014	0.02382	0.02378
10000000097	0.04329	0.44012	0.01905	0.01903
10000000121	0.04330	0.46915	0.02031	0.02037
10000000147	0.04330	0.55013	0.02382	0.02384
10000000259	0.04334	0.44011	0.01907	0.01905
10000000469	0.04333	0.44011	0.01907	0.01908
10000001251	0.04334	0.58083	0.02517	0.02512
10000001551	0.04333	0.60073	0.02603	0.02608
10000050061	0.04335	0.60568	0.02626	0.02626
10223473261	0.04281	0.60930	0.02609	0.02607

The agreement between the final two columns, for primes above 10000, is striking. The last six examples contrast two primes p with $p-1$ twice a prime, followed by four primes p with $p-1$ having small prime factors (making c_p larger).

Evidence for Conjecture B

For a few primes, in varying congruence classes modulo small primes, we computed P_k and Q_k for a range of small k.

Example 1: $p = 500000071$, $c_p = 0.5807$.

k	P_k	f_k	$c_p f_k P_k$	Q_k
1	0.05046	1	0.02930	0.02934
2	0.02547	3/2	0.02218	0.02207
3	0.01747	16/15	0.01082	0.01077
4	0.01336	2	0.01552	0.01547
5	0.01093	96/95	0.00642	0.00648
6	0.00928	8/5	0.00862	0.00864
7	0.00814	6/5	0.00567	0.00563
8	0.00708	3	0.01233	0.01236
9	0.00621	7/5	0.00504	0.00500
10	0.00560	144/95	0.00493	0.00488

Example 2: $p = 10000000019$, $c_p = 0.44014$.

k	P_k	f_k	$c_p f_k P_k$	Q_k
1	0.04330	1	0.01906	0.01900
2	0.02244	3/2	0.01482	0.01480
3	0.01516	2	0.01334	0.01342
4	0.01155	2	0.01016	0.01015
5	0.00929	4/3	0.00545	0.00544
6	0.00784	3	0.01036	0.01037
7	0.00674	6/5	0.00356	0.00352
8	0.00607	3	0.00802	0.00804
9	0.00524	2	0.00461	0.00463
10	0.00479	2	0.00422	0.00419

Example 3: $p = 10000000033$, $c_p = 0.55729$.

k	P_k	f_k	$c_p f_k P_k$	Q_k
1	0.04330	1	0.02413	0.02408
2	0.02244	3/2	0.01876	0.01867
3	0.01516	16/15	0.00901	0.00904
4	0.01155	5/2	0.01609	0.01618
5	0.00929	4/3	0.00691	0.00687
6	0.00784	8/5	0.00700	0.00705
7	0.00674	6/5	0.00451	0.00446
8	0.00607	9/4	0.00761	0.00755
9	0.00524	7/5	0.00409	0.00411
10	0.00479	2	0.00534	0.00530

Example 4: $p = 10000000061$, $c_p = 0.46456$.

k	P_k	f_k	$c_p f_k P_k$	Q_k
1	0.04330	1	0.02011	0.02009
2	0.02244	3/2	0.01564	0.01566
3	0.01516	2	0.01409	0.01415
4	0.01155	5/2	0.01341	0.01346
5	0.00929	96/95	0.00436	0.00435
6	0.00784	3	0.01093	0.01096
7	0.00674	6/5	0.00376	0.00375
8	0.00607	9/4	0.00635	0.00634
9	0.00524	2	0.00487	0.00486
10	0.00479	144/95	0.00337	0.00337

Example 5: $p = 10000000069$, $c_p = 0.55014$.

k	P_k	f_k	$c_p f_k P_k$	Q_k
1	0.04330	1	0.02382	0.02378
2	0.02244	3/2	0.01852	0.01851
3	0.01516	16/15	0.00890	0.00891
4	0.01155	5/2	0.01588	0.01603
5	0.00929	4/3	0.00682	0.00679
6	0.00784	8/5	0.00691	0.00687
7	0.00674	6/5	0.00445	0.00441
8	0.00607	9/4	0.00752	0.00748
9	0.00524	7/5	0.00404	0.00403
10	0.00479	2	0.00527	0.00523

Example 6: $p = 10000000097$, $c_p = 0.44012$.

k	P_k	f_k	$c_p f_k P_k$	Q_k
1	0.04329	1	0.01905	0.01903
2	0.02244	3/2	0.01481	0.01487
3	0.01516	2	0.01334	0.01335
4	0.01155	5/2	0.01271	0.01275
5	0.00929	4/3	0.00545	0.00544
6	0.00784	3	0.01035	0.01035
7	0.00674	6/5	0.00356	0.00353
8	0.00607	9/4	0.00601	0.00599
9	0.00524	2	0.00461	0.00460
10	0.00479	2	0.00422	0.00419

(Compare with example 2, for which the value of c_p is almost identical. The two primes are in different congruence classes mod 4, and the only significant difference between the two tables is in the rows for $k = 4$ and $k = 8$.)

Example 7: $p = 10000000121$, $c_p = 0.46915$.

k	P_k	f_k	$c_p f_k P_k$	Q_k
1	0.04330	1	0.02031	0.02037
2	0.02243	3/2	0.01579	0.01578
3	0.01516	2	0.01422	0.01429
4	0.01155	5/2	0.01354	0.01358
5	0.00929	96/95	0.00441	0.00441
6	0.00784	3	0.01103	0.01101
7	0.00674	6/5	0.00379	0.00377
8	0.00607	9/4	0.00641	0.00638
9	0.00524	2	0.00492	0.00493
10	0.00479	144/95	0.00341	0.00340

Example 8: $p = 10000000147$, $c_p = 0.55013$.

k	P_k	f_k	$c_p f_k P_k$	Q_k
1	0.04330	1	0.02382	0.02384
2	0.02243	3/2	0.01851	0.01862
3	0.01516	16/15	0.00890	0.00890
4	0.01155	2	0.01271	0.01268
5	0.00929	4/3	0.00682	0.00684
6	0.00784	8/5	0.00690	0.00689
7	0.00674	6/5	0.00445	0.00438
8	0.00607	3	0.01003	0.00997
9	0.00524	7/5	0.00404	0.00402
10	0.00479	2	0.00527	0.00521

Again the agreement between the last two columns, over a wide variety of values of c_p and f_k, provides support for the Conjecture.

3. First derivation of Conjecture A

Given t with $|t| < 2\sqrt{p}$, the probability that an elliptic curve over $\mathbb{F}_p$ has exactly $p + 1 - t$ points is

$$H(t^2 - 4p)/2p\,,$$

where $H(\Delta)$ is the Kronecker/Hurwitz class number, counting all equivalence classes of binary quadratic forms with discriminant Δ, with forms proportional to $x^2 + y^2$ being counted with weight 1/2, and forms proportional to x^2+xy+y^2 being counted with weight 1/3. We have the analytic class number formula

$$H(t^2 - 4p) = \frac{\sqrt{4p - t^2}}{\pi} \prod_l \left\{ \left(1 - \left(\frac{t^2 - 4p}{l} \right) \Big/ l \right)^{-1} \psi_3(l) \right\} .$$

Here, as ever, the product is over primes l, and $\left(\frac{t^2-4p}{l}\right)$ is the Legendre/Kronecker symbol. The function $\psi_3(l)$ is the contribution to McKee's $\psi_2(l)$ (see [8]) from the

prime l: if m is maximal such that both (i) $l^{2m} | t^2 - 4p$, and (ii) $(t^2 - 4p)/l^{2m} \equiv 0$ or $1 \pmod 4$, then

$$\psi_3(l) = \frac{l - l^{-m}}{l - 1}, \quad \frac{l}{l-1}, \quad \frac{l - 2/(l^m + l^{m-1})}{l - 1}$$

according as $\left(\frac{(t^2-4p)/l^{2m}}{l}\right)$ is 0, 1 or -1.

Our plan is to consider the average value of each factor

$$\left(1 - \left(\frac{t^2 - 4p}{l}\right) \Big/ l\right)^{-1} \psi_3(l)\,,$$

first over all t, and then over all t such that $p + 1 - t$ is prime. By averaging over "all t", we shall mean all t between 1 and l^r, where we shall let r tend to infinity.

We make three assumptions:

(i) averaging over all t, in the above sense, gives a value which is asymptotically the same as if we average over $|t| < 2\sqrt{p}$;

(ii) the average value of $\prod_l \left\{ \left(1 - \left(\frac{t^2-4p}{l}\right) \Big/ l\right)^{-1} \psi_3(l) \right\}$ is the product of the average values of each factor $\left(1 - \left(\frac{t^2-4p}{l}\right) \Big/ l\right)^{-1} \psi_3(l)$;

(iii) the average of $\left(1 - \left(\frac{t^2-4p}{l}\right) \Big/ l\right)^{-1} \psi_3(l)$ over all t such that $p + 1 - t$ is prime equals the average over all t such that $\gcd(l, p + 1 - t) = 1$.

The reasonableness of these assumptions is reflected in the goodness of fit of the model to the numerical data in the previous section. The first assumption is reasonable if l is small compared to $\sqrt{p}$, and is false if l is too large, but the product formula for $H(t^2 - 4p)$ is dominated by the relatively small primes. Indeed for this reason we can safely exclude $l = p$ from the ensuing discussion, and always think of l as small compared to $\sqrt{p}$. The second assumption asserts the independence of distinct primes. The third assumption claims that the prime l cannot distinguish between primes and numbers which are prime to l.

The purpose of computing first the average over all t is twofold: it encompasses most of the hard work for the case of interest, and it provides a comforting check that the average is 1 for each l, lending support to assumption (ii).

We recall the character sum (see, for example, exercise 8 of chapter 5 in [4])

$$\sum_{t=1}^{l} \left(\frac{t^2 - 4p}{l}\right) = -1$$

if l does not divide $4p$. This tells us how many times the Legendre symbol is 1 or -1, given that it is zero for $1 + \left(\frac{p}{l}\right)$ values of $t \pmod l$. The prime 2 requires special treatment, as usual. Note that since p is odd, $t^2 - 4p \equiv 5 \pmod 8$ whenever t is odd, so that $\left(\frac{t^2-4p}{2}\right) = -1$ whenever t is odd.

Case 1: l odd, $\left(\frac{p}{l}\right) = -1$

Euler factor, $\left(1 - \left(\frac{t^2-4p}{l}\right)\Big/ l\right)^{-1}$	Probability of having this Euler factor	Expected value of $\psi_3(l)$, given this Euler factor
$\left(1-\frac{1}{l}\right)^{-1}$	$(l-1)/2l$	1
$\left(1+\frac{1}{l}\right)^{-1}$	$(l+1)/2l$	1
1	0	Not applicable

This is the simplest case, for $\psi_3(l)$ is always 1, and the average of $\left(1 - \left(\frac{t^2-4p}{l}\right)\Big/ l\right)^{-1} \psi_3(l)$ over all t is

$$\frac{l-1}{2l}(1-1/l)^{-1} + \frac{l+1}{2l}(1+1/l)^{-1} = 1\,.$$

Case 2: l odd, $\left(\frac{p}{l}\right) = +1$

Euler factor	Probability of having this Euler factor	Expected value of $\psi_3(l)$, given this Euler factor
$\left(1-\frac{1}{l}\right)^{-1}$	$(l-3)/2l$	1
$\left(1+\frac{1}{l}\right)^{-1}$	$(l-1)/2l$	1
1	$2/l$	$l^2/(l^2-1)$

For the two values of $t \pmod l$ for which l divides t^2-4p, the Euler factor is 1, but the ψ_3 contribution is non-trivial, on average. Given $l|t^2-4p$, we have $l^r \| t^2-4p$ with probability $(l-1)/l^r$ (assuming that t runs between 1 and l^s with $s \to \infty$). Let m be maximal such that $l^{2m}|t^2-4p$: $m = \lfloor r/2 \rfloor$. If r is odd, then $\left(\frac{(t^2-4p)/l^{2m}}{l}\right) = 0$. If r is even, then $\left(\frac{(t^2-4p)/l^{2m}}{l}\right) = \pm 1$, with each sign being equally likely. Adding over all r (splitting into $r = 2s-1$ and $r = 2s$), the average value of $\psi_3(l)$, given that $l|t^2-4p$, is

$$(l-1)\sum_{s=1}^{\infty}\left\{\frac{l-l^{-(s-1)}}{l^{2s-1}(l-1)} + \frac{1}{2l^{2s}}\left(\frac{l}{l-1} + \frac{l-2/(l^s+l^{s-1})}{l-1}\right)\right\} = l^2/(l^2-1)\,,$$

after summing the series.

Finally, for this case, we compute the average value of $\left(1 - \left(\frac{t^2-4p}{l}\right)\Big/ l\right)^{-1} \psi_3(l)$ for such l, averaging over all t:

$$\frac{l-1}{2l}\cdot\frac{l}{l+1} + \frac{l-3}{2l}\cdot\frac{l}{l-1} + \frac{2}{l}\cdot\frac{l^2}{l^2-1} = 1\,.$$

Case 3: $l = 2$

Euler factor	Probability of having this Euler factor	Expected value of $\psi_3(l)$, given this Euler factor
2	0	Not applicable
2/3	1/2	1
1	1/2	4/3

There are three subcases, according as $p \equiv 3 \pmod 4$, $p \equiv 5 \pmod 8$, or $p \equiv 1 \pmod 8$. This last is the most difficult, so we go through the details, leaving the other cases as exercises.

Suppose, then, that $p \equiv 1 \pmod 8$. The only figure that needs explanation is the "4/3" when t is even. If $t \equiv 0 \pmod 4$, then $\psi_3(2) = 1$: we cannot remove a factor of 4 from the discriminant. If $t \equiv 2 \pmod 4$, then $t^2 - 4p \equiv 0 \pmod{32}$, and $2^r \| t^2 - 4p$ with probability $16/2^r$ ($r \geq 5$). If r is odd, then with m as in the definition of ψ_3 we have

$$m = (r-3)/2, \quad \left(\frac{(t^2-4p)/2^{2m}}{2} \right) = 0 \, .$$

If r is even, then the pair $(m, \left(\frac{(t^2-4p)/2^{2m}}{2} \right))$ takes values $(r/2-1, 0)$, $(r/2, 1)$, $(r/2, -1)$ with probabilities 1/2, 1/4, 1/4 respectively. Hence, summing over all $r \geq 5$, splitting into $r = 2s$ and $r = 2s-1$, we get that the expected value of $\psi_3(2)$, conditional on t being even, is

$$\frac{1}{2} \cdot 1 + \frac{1}{2} \sum_{s=3}^{\infty} \left\{ \frac{16}{2^{2s-1}} (2 - 2^{-(s-2)}) + \frac{16}{2^s} \left\{ \frac{1}{2} (2 - 2^{-(s-1)}) + \frac{1}{4} \cdot 2 + \frac{1}{4} (2 - \frac{1}{3.2^{s-1}}) \right\} \right\} ,$$

which sums to 4/3.

The cases $p \equiv 5 \pmod 8$ and $p \equiv 3 \pmod 4$ also yield 4/3, but much more simply, with no infinite sums.

Averaging $\left(1 - \left(\frac{t^2-4p}{l} \right) \Big/ l \right)^{-1} \psi_3(l)$ over all t we get

$$\frac{1}{2} \cdot \frac{2}{3} \cdot 1 + \frac{1}{2} \cdot 1 \cdot \frac{4}{3} = 1 \, .$$

Now we move to averaging over $p + 1 - t \not\equiv 0 \pmod l$.

The hard work is done: we simply delete one value of $t \pmod l$ from the above tables.

Case 1: l odd, $\left(\frac{p}{l} \right) = -1$

Euler factor	Probability of having this Euler factor	Expected value of $\psi_3(l)$, given this Euler factor
$\left(1 - \frac{1}{l}\right)^{-1}$	$(l-3)/2(l-1)$	1
$\left(1 + \frac{1}{l}\right)^{-1}$	$(l+1)/2(l-1)$	1
1	0	Not applicable

The average value of $\left(1 - \left(\frac{t^2-4p}{l} \right) \Big/ l \right)^{-1} \psi_3(l)$ is

$$\frac{l-3}{2(l-1)} \cdot (1 - 1/l)^{-1} + \frac{l+1}{2(l-1)} \cdot (1 + 1/l)^{-1} = 1 - \frac{1}{(l-1)^2} \, .$$

Case 2: l odd, $\left(\frac{p}{l} \right) = 1$, $l \not| p - 1$

Euler factor	Probability of having this Euler factor	Expected value of $\psi_3(l)$, given this Euler factor
$\left(1 - \frac{1}{l}\right)^{-1}$	$(l-5)/2(l-1)$	1
$\left(1 + \frac{1}{l}\right)^{-1}$	1/2	1
1	$2/(l-1)$	$l^2/(l^2-1)$

The average value of $\left(1 - \left(\frac{t^2-4p}{l}\right) \Big/ l\right)^{-1} \psi_3(l)$ is

$$\frac{l-5}{2(l-1)} \cdot (1-1/l)^{-1} + \frac{1}{2} \cdot (1+1/l)^{-1} + \frac{2}{l-1} \cdot \frac{l^2}{l^2-1} = 1 - \frac{1}{(l-1)^2},$$

as in Case 1.

Case 3: l odd, $l \mid p-1$

Euler factor	Probability of having this Euler factor	Expected value of $\psi_3(l)$, given this Euler factor
$\left(1-\frac{1}{l}\right)^{-1}$	$(l-3)/2(l-1)$	1
$\left(1+\frac{1}{l}\right)^{-1}$	$1/2$	1
1	$1/(l-1)$	$l^2/(l^2-1)$

The average value of $\left(1 - \left(\frac{t^2-4p}{l}\right) \Big/ l\right)^{-1} \psi_3(l)$ is

$$\frac{l-3}{2(l-1)} \cdot (1-1/l)^{-1} + \frac{1}{2} \cdot (1+1/l)^{-1} + \frac{1}{l-1} \cdot \frac{l^2}{l^2-1} = 1 - \frac{1}{(l-1)^2(l+1)}.$$

Case 4: $l = 2$

Mercifully this is trivial when we condition on $p + 1 - t \not\equiv 0 \pmod 2$:
$\left(1 - \left(\frac{t^2-4p}{2}\right) \Big/ 2\right)^{-1} \psi_3(2)$ is always $2/3$.
Multiplying over all primes l, we get our expression for c_p:

$$c_p = \frac{2}{3} \prod_{l>2} \left(1 - \frac{1}{(l-1)^2}\right) \prod_{l \mid p-1,\ l>2} \left(1 + \frac{1}{(l+1)(l-2)}\right).$$

4. Second derivation of Conjecture A

Consider the following non-equality:

$$P_1 \neq \prod_{l \leq \sqrt{p}} \left(1 - \frac{1}{l}\right).$$

Indeed Mertens' theorem (Theorem 429 in [2]) tells us that $\prod_{l \leq \sqrt{p}} \left(1 - \frac{1}{l}\right) \sim 2e^{-\gamma}/\log p$ as $p \to \infty$, where γ is Euler's constant, whereas one expects that $P_1 \sim 1/\log p$. If we replace $1 - 1/l$ by the probability that the number of points on an elliptic curve over $\mathbb{F}_p$ is not divisible by l, then we would get another non-equality:

$$Q_1 \neq \prod_{l \not| p-1,\ l \leq \sqrt{p}} \left(1 - \frac{1}{l-1}\right) \prod_{l \mid p-1} \left(1 - \frac{l}{l^2-1}\right).$$

where the factors $1 - 1/(l-1)$ and $1 - l/(l^2-1)$ come from Proposition (1.14) of [7], ignoring the error terms.

Assuming that these two non-equalities are "equally incorrect" asymptotically, which is by no means implausible, their ratio should give us c_p. And it does!

This provides a much quicker way of deriving the formula for c_p, albeit a less honest one.

5. Derivation of Conjecture B

We shall use this second heuristic method to derive Conjecture B. The same results could presumably be recovered from the analytic class number formula, but not so quickly.

Suppose that $k = p_1^{\alpha_1} \dots p_m^{\alpha_m}$, with $p_1, \dots, p_m$ distinct primes. Then, asymptotically,

$$P_k \neq \prod_{i=1}^{m} \left(\frac{1}{p_i^{\alpha_i}} - \frac{1}{p_1^{\alpha_i+1}} \right) \prod_{l \leq \sqrt{p},\ l \notin \{p_1,\dots,p_m\}} \left(1 - \frac{1}{l}\right).$$

For the analogous non-equality for elliptic curves, we need Howe's extension of Lenstra's probabilities to cover all small divisors of the number of points (see [3]): the probability that the number of points is divisible by l^t tends to $r(l^t)$ (as defined in Conjecture B) as $p \to \infty$. Hence our second desired non-equality is

$$Q_k \neq \prod_{i=1}^{m} \left(r(p_i^{\alpha_i}) - r(p_i^{\alpha_i+1}) \right) \prod_{l \leq \sqrt{p},\ l \notin \{p_1,\dots,p_m\}} (1 - r(l)) .$$

Comparing the ratio of these non-equalities with c_p gives the stated formula for f_k.

6. Arbitrary finite fields

In this section we consider elliptic curves over non-prime finite fields $\mathbb{F}_q$. Note that Howe's work is valid over any finite field, of any characteristic, always counting classes of curves with weight inversely proportional to the size of the automorphism group. Hence our second heuristic approach readily extends to this more general setting.

The case of most interest to cryptographers is when the field has 2^n elements and when the elliptic curves under consideration are non-supersingular. We give the full details only for this case. In this situation, the number of points is always even and so never prime. Nevertheless, it is still interesting to consider the probability that the number of points is k times a prime for small even values k.

All supersingular curves in characteristic 2 have j-invariant zero. We use the normal form

$$E : y^2 + xy = x^3 + ax^2 + b, \qquad b \neq 0$$

for ordinary elliptic curves over $\mathbb{F}_{2^n}$ (in this case $j(E) = 1/b$). The probability we are studying is over uniformly distributed choices $(a, b) \in \mathbb{F}_{2^n} \times \mathbb{F}_{2^n}^*$. It follows that there are $2^{n+1} - 2$ $\mathbb{F}_{2^n}$-isomorphism classes of ordinary elliptic curves over $\mathbb{F}_{2^n}$.

With $p = 2^n$, we have that $r(2^t) = 1/2^{t-1}$, which is precisely the probability that an even number is divisible by 2^t. For k odd, and $t \geq 1$, let $P'_{2^t k}$ be the probability that an even number within $2\sqrt{2^n}$ of $2^n + 1$ is $2^t k$ times a prime. Then we expect that the probability that an elliptic curve over $\mathbb{F}_{2^n}$ has $2^t k$ times a prime for its number of points will be asymptotic to

$$P'_{2^t k} f_k \prod_{l>2} \left(1 - \frac{1}{(l-1)^2}\right) \prod_{l | 2^n - 1} \left(1 + \frac{1}{(l+1)(l-2)}\right),$$

as $n \to \infty$.

We provide some experimental data to support this conjecture. In the tables below, the column labelled $Q'_{2^t k}$ is the experimentally observed probability.

Example 9: $q = 2^{27}$, $c_q = 0.67679$.

$2^t k$	$P'_{2^t k}$	f_k	$c_q f_k P'_{2^t k}$	$Q'_{2^t k}$
2	0.05558	1	0.03762	0.03793
4	0.03021	1	0.02044	0.02091
6	0.02037	2	0.02757	0.02726
8	0.01497	1	0.01014	0.01030
10	0.01191	4/3	0.01075	0.01084
12	0.00928	2	0.01472	0.01458
14	0.00928	288/287	0.00630	0.00622
16	0.00785	1	0.00531	0.00511
18	0.00703	2	0.00952	0.00993

Example 10: $q = 2^{28}$, $c_q = 0.87289$.

$2^t k$	$P'_{2^t k}$	f_k	$c_q f_k P'_{2^t k}$	$Q'_{2^t k}$
2	0.05499	1	0.04800	0.04805
4	0.02844	1	0.02483	0.02525
6	0.01877	16/15	0.01747	0.01733
8	0.01519	1	0.01327	0.01328
10	0.01147	96/95	0.01012	0.01023
12	0.00986	16/15	0.00918	0.00904
14	0.00824	6/5	0.00863	0.00864
16	0.00745	1	0.00650	0.00635
18	0.00671	7/5	0.00820	0.00826

We see that the results agree fairly closely with the theoretical prediction.

7. Application to Cryptography

In cryptography it is often desirable to use elliptic curves which have been chosen at random, so that there can be no suspicion that any special properties of the curve are being exploited. In this situation it is useful to know how many random curves will need to be chosen, on average, until a curve with a prime (or "nearly prime") number of points is found. The results of this paper give a very precise estimate of this number:

$$\frac{1}{\text{expected number of curves}} = \sum_{k=1}^{K} Q_k \approx c_p \sum_{k=1}^{K} \frac{f_k}{k(\log p - \log k)}.$$

It is worth noting that the benefits of taking $K > 1$ are greater than they would be if the numbers of points were uniformly distributed, since prime numbers of points are a little rarer than prime numbers. As an example, consider $p = 2^{200} + 235$ (the first 201-bit prime). The expected number of random trials before finding an

elliptic curve with a prime number of points is 291, while the expected number of trials for the cases $K = 5, 10$ and 20 is 127, 99 and 80 respectively.

A comparison of the results of this paper with those of [5] shows that roughly the same amount of work is needed to find an elliptic curve suitable for cryptography whether we fix the field $\mathbb{F}_q$ and select random E, or we fix E and vary the field.

If one insists on seeking a prime number of points, then it helps a little to choose p such that $p-1$ is divisible by some small primes, say $p \equiv 1 \pmod{105}$, so that c_p is larger. On the other hand, if a prime l divides $p-1$, then f_l is smaller, and with $K = 10$ there is little variation in $\sum_{k=1}^{K} Q_k$ for primes of the same size.

In cryptography there is some concern about the use of elliptic curves whose endomorphism ring has small class number, such as those elliptic curves which are generated using the CM method (see [1], [9], [6]). This is one of the reasons why it is recommended that elliptic curves should be chosen at random. We will now indicate that the class numbers corresponding to the randomly chosen elliptic curves used in cryptography are actually somewhat smaller than might be expected.

The analytic class number formula shows that $H(t^2-4p)$ is roughly $\frac{1}{\pi}\sqrt{4p-t^2}$ with the correction factor $\prod_l \left(1 - (\frac{t^2-4p}{l})l^{-1}\right)^{-1} \psi_3(l)$. In Section 3 we have given an analysis of these correction factors. Taking the asymptotic average over all t gives a factor of 1. However, over those t such that the number of points is a prime we have seen that the correction factor is the number c_p, which lies in the range $0.44 \leq c_p \leq 0.62$ (and usually is closer to 0.44). Similarly, for curves whose number of points is kq, the correction factor is $c_p f_k$.

We give one example to illustrate this phenomenon. The table gives the expected value of the class number for a randomly chosen elliptic curve modulo $10^{10} + 19$ over those curves having kq points where $1 \leq k \leq K$.

K	Expected class number
1	25125
5	38218
10	46376
15	49881
∞	77857

This shows that the expected class number for curves with a prime number of points is, over this field, about a third of the expected value over all elliptic curves.

From another point of view, these considerations show that elliptic curves with smooth order are likely to have endomorphism rings of large class number. This phenomenon has been noted previously as it explains why the ECM factoring method works slightly better than might be expected.

In conclusion, we see that it is not necessarily advisable to insist on using elliptic curves with a prime number of points for cryptography. Relaxing the condition means that far fewer random trials are required to find a suitable curve and that the endomorphism ring of the resulting curve is likely to have larger class number. Of course, having found a curve with kq points, where k is small and q is prime,

one should work entirely within the subgroup of order q to avoid small subgroup attacks.

8. Acknowledgements

The authors are grateful to Neal Koblitz and Nigel Smart for their comments on an early version of this paper.

References

[1] A. O. L. Atkin, F. Morain, 'Finding suitable curves for the elliptic curve method of factorization', *Math. Comp.*, **60**, No. 201, (1993), pp. 399–405
[2] G.H. Hardy and E.M. Wright, 'An introduction to the theory of numbers', 5th edition, Oxford University Press, (1979)
[3] E. Howe, 'On the group orders of elliptic curves over finite fields', *Compositio Mathematica*, **85**, (1993), pp. 229–247
[4] K. Ireland and M. Rosen, 'A classical introduction to modern number theory', 2nd edition, Springer Graduate Texts in Mathematics 84, (1990)
[5] N. Koblitz, 'Primality of the number of points on an elliptic curve over a finite field', *Pacific J. Math.*, **131**, No. 1, (1988), pp. 157–165
[6] G.-J. Lay, H. G. Zimmer, 'Constructing elliptic curves with given group order over large finite fields', in *ANTS-I*, L. M. Adleman ed., Springer LNCS 877, (1994), pp. 250–263
[7] H.W. Lenstra, Jr., 'Factoring integers with elliptic curves', *Annals of Mathematics*, **126**, (1987), pp. 649–673
[8] J.F. McKee, 'Subtleties in the distribution of the numbers of points on elliptic curves over a finite prime field', *Journal of the London Mathematical Society*, **(2) 59**, (1999), pp. 448–460
[9] F. Morain, 'Building cyclic elliptic curves modulo large primes', in *EUROCRYPT '91*, Springer LNCS 547, (1991), pp. 328–336

Royal Holloway University of London, Egham, Surrey TW20 0EX; Pembroke College, Oxford OX1 1DW
E-mail address: s.galbraith@rhbnc.ac.uk, mckee@maths.ox.ac.uk