PACIFIC JOURNAL OF MATHEMATICS
Vol. 131, No. 1, 1988

PRIMALITY OF THE NUMBER OF POINTS ON AN ELLIPTIC CURVE OVER A FINITE FIELD

NEAL KOBLITZ

Given a fixed elliptic curve E defined over Q having no rational torsion points, we discuss the probability that the number of points on E mod p is prime as the prime p varies. We give conjectural asymptotic formulas for the number of $p \leq n$ for which this number is prime, both in the case of a complex multiplication and a non-CM curve E. Numerical evidence is given supporting these formulas.

1. Let E be an elliptic curve defined over the field $\mathbf{Q}$ of rational numbers which has no rational torsion points. Motivated by an analogy with a classical question about finite fields (see §2) and by cryptographic applications (where certain public key cryptosystems use an elliptic curve whose group of points mod p has order divisible by a very large prime, see [**6**]), we ask the question: As the prime p varies, what is the probability that the number of points on E mod p is prime? After recalling analogous questions in classical number theory, in §3 we give a conjectural answer to this question in the case of elliptic curves without complex multiplication, and present some numerical evidence supporting the conjecture. In §4 we give a conjectural asymptotic formula in the case of CM curves, and decribe some supporting evidence.

2. In Hardy and Littlewood's paper [**4**] about the Goldbach conjecture and related questions, they give a conjectural asymptotic formula for half the number of twin primes (primes p for which $p + 2$ is prime) less than n:

$$(1) \quad C_2 \frac{n}{(\log n)^2}, \quad \text{where } C_2 = \prod_{\text{primes } l \geq 3} \left(1 - \frac{1}{(l-1)^2}\right) \approx 0.660164.$$

The same heuristics lead to the identical asymptotic formula for a slightly different question (not considered in the Hardy-Littlewood paper): For how many primes $5 \leq p \leq n$ is $(p-1)/2$ prime? It should be recalled, by the way, that, as in the case of twin primes, no one has even been able to prove that there are infinitely many p such that both p and $(p-1)/2$ are prime.

The heuristic argument for the asymptotic formula (1) for the number of prime pairs $(p, (p-1)/2)$, $p \le n$, goes as follows. By the Prime Number Theorem, the probability that a large integer n is prime equals $1/(\log n)$. Given that $n = p$ is prime, if one supposed $(p-1)/2$ to be random, its chance of being prime would be $1/(\log((p-1)/2)) \approx 1/(\log n)$, and so the probability of primality of both $n = p$ and $(p-1)/2$ would be $(\log n)^{-2}$. However, if p is prime, then for each odd prime $l \neq p$ we have $(p-1)/2 \not\equiv (l-1)/2 \pmod{l}$, because l does not divide p. Thus, $(p-1)/2$ can fall in $l-1$ residue classes mod l, of which $l-2$ are nonzero. So the chance that $l \nmid (p-1)/2$ is $(l-2)/(l-1)$ rather than $(l-1)/l$, as would be the case if $(p-1)/2$ were random. Hence, one would expect a correction factor giving the ratio of primality probabilities for $(p-1)/2$ and for a random integer which is equal to the product over l of the ratio of $(l-2)/(l-1)$ to $(l-1)/l$. Since the latter ratio is $1-(l-1)^{-2}$, this correction factor is the constant C_2 in (1). (Note: The same argument applies to the twin prime problem, except that there is an additional correction factor of 2 because automatically $p+2$ is odd, whereas there is only a $1/2$ probability of a random number being odd; thus the formula in (1) is for *half* the number of twin prime pairs.)

As Hardy and Littlewood remark, one expects—and finds—that the asymptotically equivalent formula $C_2 \int_2^n (\log t)^{-2}\, dt$ is in closer agreement with numerical data. In the case of our $(p-1)/2$ problem, one expects still closer agreement if one uses

$$(2) \qquad C_2 \sum_{\substack{\text{primes } p \\ 5 \le p \le n}} \frac{1}{\log((p-1)/2)}.$$

And in fact, for $n = 10^5$ we find that the ratio of $|\{5 \le p \le n \mid (p-1)/2 \text{ is prime}\}|$ to the value predicted by (2) is 1.0043.

Letting $\mathbf{F}_q$ denote the finite field of $q = p^f$ elements, we can rephrase our $(p-1)/2$ question in the form: Does the multiplicative group $\mathbf{F}_p^*$ have any nontrivial subgroups besides $\{\pm 1\}$ and the subgroup of squares, i.e., is $\mathbf{F}_p^*/\{\pm 1\}$ cyclic of prime order? A natural complementary question is: For fixed p and for $f > 1$, how often is $\mathbf{F}_q^*/\mathbf{F}_p^*$ cyclic of prime order, i.e., is $(p^f-1)/(p-1)$ prime? The case $p = 2$ is the famous Mersenne prime problem, and other cases of this primality question have also been investigated (see [1], [13]).

The common element in these questions is that, after dividing by a subgroup that is trivially known to exist, one asks about the likelihood that the quotient group is simple. Thus, an analogous question for elliptic curves is: Given an elliptic curve defined over $\mathbf{Q}$, after one divides by its

Q-torsion subgroup, what is the probability for large p that its reduction modulo p is of prime order?

3. Certain public key cryptographic systems, based on intractability of the discrete logarithm problem, can be implemented using the group of points on an elliptic curve E defined over a finite field $\mathbf{F}_q$ (see [**6**], [**11**]). In that case one wants the cyclic subgroup generated by certain points G to have order divisible by a large prime. One way to accomplish this is to choose E and $\mathbf{F}_q$ so that the group of points has prime order; then the desired condition holds for any point G not the identity (point at infinity).

To find an elliptic curve defined over a finite field having a prime number of points, one method is to choose a fixed large prime p and let the coefficients of E vary in $\mathbf{F}_p$. The probability that the number of points is prime can be estimated using results of Deuring, Waterhouse, Goldwasser, Killian, and Heath-Brown (for details and references, see §5 of [**10**]).

We shall consider a different method. Namely, if we have a fixed elliptic curve E defined over the rational integers $\mathbf{Z}$, we might try to choose a prime finite field $\mathbf{F}_p$ such that $|E \bmod p|$ is prime. One can proceed as follows. Choose a very large odd number M at random, then, with the help of an efficient primality test, find the successive primes p among the sequence $M, M+2, M+4, \ldots$. In each case compute the order of $E \bmod p$. (In principle, one can compute $|E \bmod p|$ using the deterministic polynomial time algorithm of R. Schoof [**12**]; however, Schoof's algorithm is not very practical, and so in practice one might use a probabilistic algorithm. For more discussion of algorithms for computing $|E \bmod p|$, see §4 of [**10**].) Test $|E \bmod p|$ for primality, and stop when it is prime.

For fixed E, what is the probability that $N_p = |E \bmod p|$ is prime? As p varies, N_p is distributed close to p in a uniform way. More precisely, it is known that $(N_p - p - 1)/2\sqrt{p}$ is always between -1 and 1 (Hasse's theorem) and its distribution there as p varies is proportional to the measure $\sqrt{1-x^2}\,dx$ (the Sato-Tate distribution). But it would be wrong to think that N_p behaves like a random number in that range with respect to primality. Namely, using the Chebotarev density theorem as in [**9**], one finds that the probability that N_p is divisible by a fixed prime l is equal *not* to $1/l$ (as for a random integer) but rather, if we let G_l denote the Galois group of l-division points in $\mathrm{GL}(2, \mathbf{Z}/l\mathbf{Z})$, to the ratio

$$|\{g \in G_l \mid g \text{ has eigenvalue } 1\}|/|G_l|.$$

For example, if G_l is all of $\mathrm{GL}(2, \mathbf{Z}/l\mathbf{Z})$, then this ratio is

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

Thus, in that case N_p has a slightly greater chance of being divisible by l. So heuristically, in the case of the "Serre curves" in (4) below, where G_l is always equal to all of $\mathrm{GL}(2, \mathbf{Z}/l\mathbf{Z})$, one would expect that N_p has a lower chance of being prime, by a factor of

$$\prod_{\text{primes } l} \frac{1 - (l^2 - 2)/(l^2 - 1)(l - 1)}{1 - 1/l} = \prod_{\text{primes } l} \left(1 - \frac{l^2 - l - 1}{(l^2 - 1)(l - 1)^2}\right).$$

We first treat curves E without complex multiplication.

Conjecture A. Let E be an elliptic curve of discriminant Δ defined over $\mathbf{Z}$ which is not $\mathbf{Q}$-isogenous to a curve with nontrivial $\mathbf{Q}$-torsion and which does not have complex multiplication. Then

$$\left|\{\text{primes } p \le n,\ p \nmid \Delta \mid |E \bmod p| \text{ is prime}\}\right|$$

is asymptotic to

$$C\frac{n}{(\log n)^2}, \tag{3}$$

where C is a positive constant of the form $C = \prod a(l)$ which depends on E. Here $a(l)$ is the ratio of $1 - |\{g \in G_l \mid g \text{ has eigenvalue } 1\}|/|G_l|$ to $1 - 1/l$, where G_l denotes the Galois group over $\mathbf{Q}$ of the field of l-division points; for all but finitely many primes l

$$a(l) = 1 - \frac{l^2 - l - 1}{(l^2 - 1)(l - 1)^2}.$$

We tested this conjecture numerically using the same three curves that Lang and Trotter used to test their conjecture about primitive points in [**9**], namely, the curves

$$A\colon y^2 + y = x^3 - x, \quad B\colon y^2 + y = x^3 + x^2, \quad C\colon y^2 + xy + y = x^3 - x^2 \tag{4}$$

of conductor 37, 43, 53, respectively. These curves have no $\mathbf{Q}$-torsion, and the Galois group of l-division points is always the full general linear group mod l.

In the case of the curves (4), we have

$$C = \prod_l \left(1 - \frac{l^2 - l - 1}{(l^2 - 1)(l - 1)^2}\right) \approx 0.5052.$$

Table I shows the function of n in the conjecture

$$f(n) = \left|\{\text{primes } p \leq n, p \nmid \Delta \mid |E \bmod p| \text{ is prime}\}\right|$$

for n at intervals of 2000 up to 30000, for the three curves in (4). We compare with the predicted value, where instead of (3) we use the asymptotically equivalent formula $C\Sigma_{\text{primes } p \leq n, p \nmid \Delta}(\log p)^{-1}$.

TABLE I

Number of $p \leq n$ such that $|E \bmod p|$ is prime

n	predicted value	Curve A	Curve B	Curve C
2000	26	29	30	23
4000	42	43	42	38
6000	55	54	51	52
8000	68	61	62	61
10000	80	74	77	75
12000	92	84	87	86
14000	103	91	103	91
16000	114	97	113	105
18000	125	111	123	117
20000	135	121	131	125
22000	145	131	141	143
24000	155	146	156	154
26000	165	160	165	166
28000	175	169	176	181
30000	184	179	183	194

4. We now suppose that E is an elliptic curve of discriminant Δ defined over $\mathbf{Z}$ which has trivial $\mathbf{Q}$-torsion (more precisely, is not $\mathbf{Q}$-isogenous to a curve with nontrivial $\mathbf{Q}$-torsion) and has complex multiplication by the ring of integers O_K of an imaginary quadratic field K (necessarily of class number 1). Let $N_p = |E \bmod p|$ for $p \nmid \Delta$. First, if p remains prime in K, then $N_p = p + 1$ is never prime for $p > 2$; but one can ask for $(p + 1)/2$ to be prime. This question—and the more general question of factorization of $N_{p^f} = p^f + 1$—is analogous to the $(p - 1)/2$ and $(p^f - 1)/(p - 1)$ problem mentioned in §2. The same heuristics apply, giving the conjectural asymptotic formula (1), except that one must insert

a factor of $\frac{1}{2}$ (because we are asking only for p which remain prime in K for which $(p+1)/2$ is prime), and, in the case when K has discriminant $-l$, one must also change the l-term in C_2 to $1-(l+1)/(l-1)^2$. (Note that when $K=\mathbf{Q}(\sqrt{-3})$, this makes $C_2=0$, because the 3-term is zero; in addition, C_2 must be replaced by 0 if $K=\mathbf{Q}(i)$, since in that case as well $(p+1)/2$ is even whenever p remains prime in K; thus, one should really ask about primality of $(p+1)/w$, where w is the number of roots of unity in K.) Alternately, one has the estimate (2) with $(p-1)/2$ replaced by $(p+1)/2$ and the summation taken only over p which remain prime in K.

Now we consider the primes p which split in K. In investigations of similar questions, such as the primitive point conjecture of Lang-Trotter, it has turned out to be natural to restrict one's attention to the class of p which split in K (see [3]).

Conjecture B. Let E be an elliptic curve of discriminant Δ defined over $\mathbf{Z}$ which is not $\mathbf{Q}$-isogenous to a curve with nontrivial $\mathbf{Q}$-torsion and which has complex multiplication by the ring of integers O_K of an imaginary quadratic field K. Then

$$\left|\{\text{primes } p \le n,\ p \nmid \Delta,\ p \text{ splits in } K \mid |E \bmod p| \text{ is prime}\}\right|$$

is asymptotic to

$$C\frac{n}{2(\log n)^2}, \tag{5}$$

where C is a positive constant of the form $C=\prod a(l)$ which depends on E. Here $a(l)$ is again $(1-|\{g\in G_l \mid g \text{ has eigenvalue } 1\}|/|G_l|)/(1-1/l)$, where here G_l denotes the Galois group over K of the field of l-division points; for all but finitely many primes l

$$a(l)=1-\chi(l)\frac{l^2-l-1}{(l-\chi(l))(l-1)^2}, \tag{6}$$

where χ is the quadratic character corresponding to the field K.

Note that the infinite product of the $a(l)$ in (6) converges conditionally to a nonzero limit, as one can see by comparison with the Euler product for the Dirichlet L-series value $L(1,\chi)=\prod(1-\chi(l)/l)^{-1}$.

Since we have restricted attention to rational primes of degree 1 in K, in applying the Chebotarev density theorem we must work with the Galois group over K of the l-division points. Since E has complex multiplication

by O_K, this Galois group $G_l \subset \mathrm{GL}(2, \mathbf{Z}/l\mathbf{Z})$ is isomorphic to a subgroup of the group of units of O_K/lO_K. This group of units is isomorphic as a group to either $\mathbf{F}_{l^2}^*$ (if l remains prime) or $\mathbf{F}_l^* \oplus \mathbf{F}_l^*$ (if l splits) or $\mathbf{F}_l^* \oplus \mathbf{Z}/l\mathbf{Z}$ (if l ramifies). Meanwhile, the number of elements with eigenvalue 1 equals, respectively: $1, 2l-3, l$. Thus, if G_l is the largest it can be (as is the case for all but finitely many l), the factor $a(l)$ is respectively:

$$\left(1 - \frac{1}{l^2-1}\right)\Big/\left(1 - \frac{1}{l}\right), \quad \left(1 - \frac{2l-3}{(l-1)^2}\right)\Big/\left(1 - \frac{1}{l}\right),$$
$$\left(1 - \frac{1}{l-1}\right)\Big/\left(1 - \frac{1}{l}\right).$$

This equals (6) in the first two cases, and $a(l) = 1 - (l-1)^{-2}$ when l ramifies.

To test Conjecture B, we chose two families of CM-curves for which the fields of l-division points are probably always largest possible and linearly disjoint:

$$D_j\colon\ y^2 - y = x^3 - (j+1)/4; \tag{7}$$

$$E_j\colon\ y^2 = j(x^3 - x^2 - 7x + 41/4), \tag{8}$$

where j is a prime that remains prime in the CM-field K. The curves D_j and E_j have complex multiplication by the ring of integers in $\mathbf{Q}(\sqrt{-3})$ and $\mathbf{Q}(\sqrt{-11})$, respectively; they have discriminant $\Delta = 27j^2$ and $121j^2$, respectively. In (7) we let j run through the first 50 primes $\equiv -1 \bmod 12$; in (8) we let j run through the first 50 primes for which $(j/11) = -1$. In each case let $f_j(n)$ denote the number of primes $p \le n$ such that $(p/3) = 1$ (resp. $(p/11) = 1$) and $|D_j \bmod p|$ (resp. $|E_j \bmod p|$) is prime. In Table II we give the *average* of $f_j(n)$ for the 50 values of j for n at intervals of 2000 up to 10000. We compare with the predicted values

$$C \sum_{\text{primes } p \le n,\ \chi(p)=1} (\log p)^{-1},$$

where

$$C = a(l_0) \prod_{\chi(l)\neq 0} \left(1 - \chi(l)(l^2 - l - 1)(l - \chi(l))^{-1}(l-1)^{-2}\right).$$

Here $l_0 = 3$ and 11, respectively, denotes the one ramified value of l; for this value we have $a(l_0) = 1 - (l_0 - 1)^{-2} = 0.75$ and 0.99, respectively.

TABLE II

Average value of $f_j(n)$ in the CM-case

n	Curves D_j predicted	Curves D_j actual	Curves E_j predicted	Curves E_j actual
2000	24.3	23.0	10.9	13.8
4000	39.8	39.8	17.9	22.4
6000	53.2	51.8	24.1	25.6
8000	65.9	67.0	30.1	28.4
10000	78.8	79.3	35.7	35.2

As in Table I, the agreement between the actual incidence of primality and the predictions of the conjecture is as good as can reasonably be expected. (Of course, this numerical evidence supports Conjecture B only in some "average sense," since in order to obtain a large sample we averaged over families of curves.)

REMARK. Here are two directions relevant to cryptographic applications in which the primality question considered above can be generalized: (1) In the case when E has a nontrivial $\mathbf{Q}$-torsion subgroup E_t, what is the probability that $|(E \bmod p)/(E_t \bmod p)|$ is prime?, and (2) Given E and also a bound B (perhaps depending on the size of p), what is the probability that $E \bmod p$ has a subgroup of prime order with index less than B?

A third question is the analog of the $(p^f - 1)/(p - 1) = |\mathbf{F}_{p^f}^*/\mathbf{F}_p^*|$ problem. Namely, if E is an elliptic curve defined over $\mathbf{F}_p$, then, by Weil's theorem, there is a quadratic imaginary algebraic integer α of norm p such that $N_{p^f} = |E(\mathbf{F}_{p^f})|$ is equal to $p^f + 1 - \alpha^f - \bar{\alpha}^f = \mathbf{N}(\alpha^f - 1)$ (where $\mathbf{N}(x) = x \cdot \bar{x}$ denotes the norm). Thus, a question complementary to primality of $N_p = \mathbf{N}(\alpha - 1)$ (as p and hence α varies, for fixed E defined over $\mathbf{Z}$) is the question of primality of

$$\left|E(\mathbf{F}_{p^f})/E(\mathbf{F}_p)\right| = \mathbf{N}\big((\alpha^f - 1)/(\alpha - 1)\big)$$

s f varies (for fixed E and p).

EXAMPLE. When $p = 2$ or 3, it is possible for an elliptic curve E defined over $\mathbf{F}_p$ to have only one $\mathbf{F}_p$-point (the point at infinity). Namely, let E be given by the equation $y^2 + y = x^3 - x + 1$. In that case the

number N_{p^f} of $\mathbf{F}_{p^f}$-points on E is

$$N_{2^f} = \mathbf{N}\big((1+i)^f - 1\big); \quad N_{3^f} = \mathbf{N}\big((1+\omega)^f - 1\big),$$

$$\text{where } \omega = \frac{1}{2} + \frac{\sqrt{3}}{2}i,$$

and it makes sense to ask whether N_{p^f} is prime. Clearly, this is possible only if f is prime. The N_{p^f} are analogs of the Mersenne numbers; in **[16]** the numbers $(1+i)^f - 1$ are called "complex Mersenne numbers."

References

[1] J. Brillhart, D. H. Lehmer, J. L. Selfridge, B. Tuckerman and S. S. Wagstaff, Jr., *Factorization of* $b^n \pm 1$, $b = 2,3,5,6,7,10,11,12$ *up to high powers*, Amer. Math. Soc., 1983.

[2] W. Bosma, *Primality testing using elliptic curves*, Report 85–12, Mathematisch Instituut, Universiteit van Amsterdam, 1985.

[3] R. Gupta and M. R. P. Murty, *Primitive points on elliptic curves*, Compositio Math., **58** (1986), 13–44.

[4] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio numerorum'; III: on the expression of a number as a sum of primes*, Acta Math., **44** (1923), 1–70.

[5] N. Koblitz, *Introduction to Elliptic Curves and Modular Forms*, Springer-Verlag, New York, 1984.

[6] _____, *Elliptic curve cryptosystems*, Math. of Computation, **48** (1987), 203–209.

[7] S. Lang, *Elliptic Curves: Diophantine Analysis*, Springer-Verlag, New York, 1978.

[8] S. Lang and J. Tate, eds., *The Collected Papers of Emil Artin*, Addison-Wesley, Reading, Mass., 1965.

[9] S. Lang and H. Trotter, *Primitive points on elliptic curves*, Bull. Amer. Math. Soc., **83** (1977), 289–292.

[10] H. W. Lenstra, Jr., *Elliptic curves and number-theoretic algorithms*, Report 86–19, Mathematisch Instituut, Universiteit van Amsterdam, 1986.

[11] V. S. Miller, *Use of elliptic curves in cryptography*, Abstracts for Crypto '85.

[12] R. Schoof, *Elliptic curves over finite fields and the computation of square roots* mod p, Math. of Computation, **44** (1985), 483–494 and 175–182.

[13] E. Seah and H. C. Williams, *Some primes of the form* $(a^n - 1)/(a - 1)$, Math. of Comput., **33** (1979), 1337–1342.

[14] J.-P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Inventiones Math., **15** (1972), 259–331.

[15] D. Shanks, *Solved and Unsolved Problems in Number Theory*, 3rd ed., Chelsea Publ. Co., New York, 1985.

[16] R. Spira, *The complex sum of divisors*, Amer. Math. Monthly, **68** (1961), 120–124.

Received August 21, 1986 and in revised form April 9, 1987.

UNIVERSITY OF WASHINGTON
SEATTLE, WA 98195