Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Number Theory 119 (2006) 155–170
JOURNAL OF Number Theory
www.elsevier.com/locate/jnt

Prime analogues of the Erdős–Kac theorem for elliptic curves

Yu-Ru Liu [1]

Department of Pure Mathematics, University of Waterloo, Waterloo, ON, Canada N2L 3G1

Received 1 May 2005; revised 1 October 2005

Available online 13 December 2005

Communicated by David Goss

Abstract

Let $E/\mathbb{Q}$ be an elliptic curve. For a prime p of good reduction, let $E(\mathbb{F}_p)$ be the set of rational points defined over the finite field $\mathbb{F}_p$. We denote by $\omega(\#E(\mathbb{F}_p))$, the number of distinct prime divisors of $\#E(\mathbb{F}_p)$. We prove that the quantity (assuming the GRH if E is non-CM)

$$\frac{\omega(\#E(\mathbb{F}_p)) - \log\log p}{\sqrt{\log\log p}}$$

distributes normally. This result can be viewed as a "prime analogue" of the Erdős–Kac theorem. We also study the normal distribution of the number of distinct prime factors of the exponent of $E(\mathbb{F}_p)$.

MSC: 11N60; 11G20

Keywords: Prime divisors; Rational points; Elliptic curves

1. Introduction

For $n \in \mathbb{N} := \{1, 2, 3, \ldots\}$, define $\omega(n)$ to be the number of distinct prime divisors of n. The Erdős–Kac theorem [6] is about the existence of a Gaussian normal distribution for the quantity

$$\frac{\omega(n) - \log\log n}{\sqrt{\log\log n}}.$$

E-mail address: yrliu@math.uwaterloo.ca.

[1] Research partially supported by an NSERC discovery grant.

0022-314X/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jnt.2005.10.014

More precisely, for $x, \gamma \in \mathbb{R}$, $x > 1$, Erdős and Kac proved that

$$\lim_{x\to\infty} \frac{1}{x} \#\left\{n \leqslant x\colon\ n \text{ satisfies } \frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \leqslant \gamma\right\} = G(\gamma) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\gamma} e^{\frac{-t^2}{2}}\, dt.$$

The idea behind Erdős–Kac's proof was essentially probabilistic. Further development of probabilistic idea led Kubilius [9] and Shapiro [15] to prove independently a generalization of the Erdős–Kac theorem. Their result is applicable to what are called "strongly additive functions." An interested reader can find a comprehensive treatment of it in the monograph of Elliott [4,5].

Instead of the sequence of natural numbers, we consider only the set of primes now. In 1955, Halberstam [8] proved that for a prime p,

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\left\{p \leqslant x\colon\ p \text{ satisfies } \frac{\omega(p-1) - \log\log p}{\sqrt{\log\log p}} \leqslant \gamma\right\} = G(\gamma),$$

where $\pi(x)$ is the number of primes $p \leqslant x$. This theorem can be viewed as a "prime analogue" of the Erdős–Kac theorem.

Another prime analogue of the Erdős–Kac theorem which can be described as "non-abelian" was discovered by Murty and Murty. Let $\tau(n)$ denote the Ramanujan τ-function. Assuming the GRH (i.e., the Riemann hypothesis for all Dedekind zeta functions of number fields), they proved that [14]

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\left\{p \leqslant x\colon\ p \text{ satisfies } \tau(p) \neq 0 \text{ and } \frac{\omega(\tau(p)) - \log\log p}{\sqrt{\log\log p}} \leqslant \gamma\right\} = G(\gamma).$$

In this paper, we provide another "non-abelian" prime analogue of the Erdős–Kac theorem. Let E be an elliptic curve defined over $\mathbb{Q}$. For a prime p of good reduction, we denote by $E(\mathbb{F}_p)$ the set of rational points defined over the finite field $\mathbb{F}_p$. We prove the theorem.

Theorem 1. *Let $E/\mathbb{Q}$ be an elliptic curve. For $x, \gamma \in \mathbb{R}$, $x > 1$, we have* (*assuming the GRH if E has no complex multiplication* (*non-CM*))

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\left\{p \leqslant x\colon\ p \text{ is of good reduction and } \frac{\omega(\#E(\mathbb{F}_p)) - \log\log p}{\sqrt{\log\log p}} \leqslant \gamma\right\} = G(\gamma).$$

Indeed, the full strength of the GRH is not needed to prove Theorem 1, but any "quasi-RH" is enough (see Remark at the end of Section 4).

It is well known that the group of $\mathbb{F}_p$-rational points $E(\mathbb{F}_p)$ is isomorphic to

$$E(\mathbb{F}_p) \cong (\mathbb{Z}/f_p\mathbb{Z}) \times (\mathbb{Z}/m_p\mathbb{Z}),$$

for unique integers f_p and m_p with $m_p | f_p$. The number f_p is called the *exponent* of $E(\mathbb{F}_p)$ and is the largest possible order of points on $E(\mathbb{F}_p)$. Since $\#E(\mathbb{F}_p) = f_p \cdot m_p$ and $m_p | f_p$, we have

$$\omega(f_p) = \omega\big(\#E(\mathbb{F}_p)\big).$$

Hence, as a direct consequence of Theorem 1, we have:

Theorem 2. *Let $E/\mathbb{Q}$ be an elliptic curve. For $x, \gamma \in \mathbb{R}$, $x > 1$, we have (assuming the GRH if E is non-CM)*

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\left\{p \leqslant x\colon\ p \text{ is of good reduction and } \frac{\omega(f_p) - \log\log p}{\sqrt{\log\log p}} \leqslant \gamma\right\} = G(\gamma).$$

The idea behind our proof is essentially probabilistic. In this paper, we will indeed prove a generalization of the Erdős–Kac theorem which is applicable to the classical Erdős–Kac theorem and all of its prime analogue (see Theorem 3). In Section 2 of this paper, we review some facts in probability theory that are essential for a generalized Erdős–Kac theorem. We state and prove this generalization in Section 3 and apply it in the following sections to show that the quantity

$$\frac{\omega(\#E(\mathbb{F}_p)) - \log\log p}{\sqrt{\log\log p}}$$

distributes normally (assuming the GRH if E is non-CM). In Section 6, we conclude this paper by discussing a possible strategy to remove the GRH assumption for "non-abelian" prime analogues of the Erdős–Kac theorem.

Notation. For $x \in \mathbb{R}, x > 0$, let $f(x)$ and $g(x)$ be two functions of x. If $g(x)$ is positive and there exists a constant $C > 0$ such that $|f(x)| \leqslant Cg(x)$, we write either $f(x) \ll g(x)$ or $f(x) = O(g(x))$. If $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$, we write $f(x) = o(g(x))$.

2. Review of probability theory

In this section, we review some probability theory.

Let X be a random variable with a probability measure Pr. Let F be its associated distribution function. Let $\mathrm{E}\{X\}$ and $\mathrm{Var}\{X\}$ be the expectation and variance of X, respectively.

Definition. Given a sequence of random variables $\{X_k\}$ and $\alpha \in \mathbb{R}$, we say $\{X_k\}$ *converges in probability to* α if for any $\epsilon > 0$,

$$\lim_{k\to\infty} \Pr\left\{|X_k - \alpha| > \epsilon\right\} = 0.$$

We denote it by

$$X_k \xrightarrow{p} \alpha.$$

Now, we are in a position to state some results from probability theory that are needed to prove Theorem 1; their proofs can be found in [1,7].

Proposition 1. [1, p. 134] *Given a sequence of random variables $\{X_k\}$, if*

$$\lim_{n\to\infty} \mathrm{E}\left\{|X_k|\right\} = 0,$$

we have

$$X_k \xrightarrow{p} 0.$$

Proposition 2. ([1, pp. 134, 135], [7, p. 247]) *Let* $\{X_k\}$, $\{Y_k\}$, *and* $\{U_k\}$ *be sequences of random variables with the same probability measure* Pr. *Let* U *be a distribution function. Suppose*

$$X_k \xrightarrow{p} 1 \quad \textit{and} \quad Y_k \xrightarrow{p} 0.$$

For all $\gamma \in \mathbb{R}$, *we have*

$$\lim_{k\to\infty} \Pr\{U_k \leqslant \gamma\} = U(\gamma)$$

if and only if

$$\lim_{k\to\infty} \Pr\big\{(X_k U_k + Y_k) \leqslant \gamma\big\} = U(\gamma).$$

For $\gamma \in \mathbb{R}$, let $G(\gamma)$ denote the Gaussian normal distribution, i.e.,

$$G(\gamma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\gamma} e^{\frac{-t^2}{2}}\, dt.$$

For $r \in \mathbb{N}$, the rth moment of G is defined by

$$\mu_r := \int_{-\infty}^{\infty} t^r\, dG(t).$$

The following proposition shows that G is uniquely determined by these moments.

Proposition 3. [7, pp. 262, 263] *Let* $\{F_k\}$ *be a sequence of distribution functions. Suppose for all* $r \in \mathbb{N}$,

$$\lim_{k\to\infty} \int_{-\infty}^{\infty} t^r\, dF_k(t) = \mu_r.$$

Then for all $\gamma \in \mathbb{R}$, *we have*

$$\lim_{k\to\infty} F_k(\gamma) = G(\gamma).$$

This next proposition is an analogue of the Lebesgue Dominated Convergence Theorem.

Proposition 4. [7, pp. 244, 245] *Let* $r \in \mathbb{N}$ *and* $\{F_k\}$ *a sequence of distribution functions. Suppose*

$$\lim_{k\to\infty} F_k(\gamma) = G(\gamma), \quad \textit{for all } \gamma \in \mathbb{R},$$

and

$$\sup_k \left\{ \int_{-\infty}^{\infty} |t|^{r+\delta}\, dF_k(t) \right\} < \infty, \quad \text{for some } \delta = \delta(r) > 0.$$

We have

$$\lim_{k\to\infty} \int_{-\infty}^{\infty} t^r\, dF_k(t) = \mu_r.$$

The next proposition is a special case of the Central Limit Theorem.

Proposition 5. [7, pp. 256–258] *Let $X_1, X_2, \ldots, X_i, \ldots$ be a sequence of independent random variables and* $\mathrm{Im}(X_i)$ *the image of X_i. Suppose*

(1) $\sup_i\{\mathrm{Im}(X_i)\} < \infty$.
(2) $\mathrm{E}\{X_i\} = 0$ *and* $\mathrm{Var}\{X_i\} < \infty$ *for all i.*

For $k \in \mathbb{N}$, let G_k be the "normalization" of $X_1, X_2, \ldots, X_k$, i.e.,

$$G_k := \frac{\sum_{i=1}^k X_i}{\left(\sum_{i=1}^k \mathrm{Var}\{X_i\}\right)^{1/2}}.$$

If $\sum_{i=1}^{\infty} \mathrm{Var}\{X_i\}$ diverges, we have

$$\lim_{k\to\infty} \Pr\{G_k \leqslant \gamma\} = G(\gamma).$$

3. A generalized Erdős–Kac theorem

In this section, we prove a generalization of the Erdős–Kac theorem which can be applied to the classical Erdős–Kac theorem and its prime analogues.

Let S be an infinite subset of $\mathbb{N}$. For $x \in \mathbb{R}$, $x > 1$, define

$$S(x) = \{n \leqslant x\colon\ n \in S\}.$$

We assume that S satisfies the following cardinality condition, say (C),

$$\left|S\left(x^{1/2}\right)\right| = o\left(\left|S(x)\right|\right), \tag{C}$$

where $|S(x)|$ is the cardinality of $S(x)$. Let f be a map from S to $\mathbb{N}$. For each prime l, we write

$$\frac{1}{|S(x)|} \#\left\{n \in S(x)\colon\ f(n) \text{ is divisible by } l\right\} = \lambda_l(x) + e_l(x),$$

where $\lambda_l := \lambda_l(x)$ can be thought of as a main term (and is usually chosen to be independent of x) and $e_l := e_l(x)$ is an error term. For any u-tuples of distinct primes $(l_1, l_2, \dots, l_u)$, we write

$$\frac{1}{|S(x)|} \#\big\{n \in S(x):\ f(n) \text{ is divisible by } l_1 l_2 \cdots l_u\big\} = \prod_{i=1}^{u} \lambda_{l_i} + e_{l_1 l_2 \cdots l_u}(x).$$

We will use $e_{l_1 l_2 \cdots l_u}$ to abbreviate $e_{l_1 l_2 \cdots l_u}(x)$ below.

Suppose there exist a constant β (independent of x) with $0 < \beta \leqslant 1$ and a function $y = y(x) < x^{\beta}$ such that the following conditions hold:

(1) For each $n \in S(x)$, the number of distinct prime divisors l of $f(n)$ with $l > x^{\beta}$ is bounded uniformly.
(2) $\sum_{y<l\leqslant x^{\beta}} \lambda_l = o((\log\log x)^{1/2})$, where the sum is over primes l.
(3) $\sum_{y<l\leqslant x^{\beta}} |e_l| = o((\log\log x)^{1/2})$.
(4) $\sum_{l\leqslant y} \lambda_l = \log\log x + o((\log\log x)^{1/2})$.
(5) $\sum_{l\leqslant y} \lambda_l^2 = o((\log\log x)^{1/2})$.
(6) For $r \in \mathbb{N}$, let $u = 1, 2, \dots, r$. We have

$$\sum{}'' |e_{l_1 \cdots l_u}| = o\big((\log\log x)^{-r/2}\big),$$

where $\sum''$ extends over all u-tuples of distinct primes $(l_1, l_2, \dots, l_u)$ with $l_i \leqslant y$.

Given S and f satisfying the above conditions, we have the following generalization of the Erdős–Kac theorem.

Theorem 3. *Let S be an infinite subset of $\mathbb{N}$ satisfying condition* (C) *and $f : S \to \mathbb{N}$. Suppose there exist a constant β with $0 < \beta \leqslant 1$ and $y = y(x) < x^{\beta}$ such that conditions* (1)–(6) *hold. Then for $\gamma \in \mathbb{R}$, we have*

$$\lim_{x\to\infty} \frac{1}{|S(x)|} \#\left\{n \in S(x):\ n \text{ satisfies } \frac{\omega(f(n)) - \log\log n}{\sqrt{\log\log n}} \leqslant \gamma\right\} = G(\gamma).$$

We divide our proof of Theorem 3 into Lemmas 1–5 below. For $n \in S$, we define

$$\Pr_{S,x}\{n:\ n \text{ satisfies some conditions}\} := \frac{1}{|S(x)|} \#\big\{n \in S(x):\ n \text{ satisfies some conditions}\big\}.$$

Notice that $\Pr_{S,x}$ is a probability measure on S. Let g be a function from S to $\mathbb{R}$. The expectation of g with respect to $\Pr_{S,x}$ is denoted by

$$\mathrm{E}_{S,x}\big\{g(n)\big\} := \frac{1}{|S(x)|} \sum_{n\in S(x)} g(n).$$

Our goal is to prove that

$$\lim_{x\to\infty} \Pr_{S,x}\left\{n:\ n \text{ satisfies } \frac{\omega(f(n)) - \log\log n}{\sqrt{\log\log n}} \leqslant \gamma\right\} = G(\gamma).$$

The following lemma gives an equivalent statement of Theorem 3. More precisely, it says that we can replace the term $\log\log n$ by $\log\log x$.

Lemma 1. *Let S be an infinite subset of $\mathbb{N}$ satisfying condition* (C) *and $f : S \to \mathbb{N}$. Then for $\gamma \in \mathbb{R}$, we have*

$$\lim_{x\to\infty} \mathrm{Pr}_{S,x}\left\{n\colon n \text{ satisfies } \frac{\omega(f(n)) - \log\log n}{\sqrt{\log\log n}} \leqslant \gamma\right\} = G(\gamma)$$

if and only if

$$\lim_{x\to\infty} \mathrm{Pr}_{S,x}\left\{n\colon n \text{ satisfies } \frac{\omega(f(n)) - \log\log x}{\sqrt{\log\log x}} \leqslant \gamma\right\} = G(\gamma).$$

Proof. Write

$$\frac{\omega(f(n)) - \log\log x}{\sqrt{\log\log x}} = \frac{\omega(f(n)) - \log\log n}{\sqrt{\log\log n}} \cdot \frac{\sqrt{\log\log n}}{\sqrt{\log\log x}} + \frac{\log\log n - \log\log x}{\sqrt{\log\log x}}.$$

Consider those integers $n \in S(x)$ with $n > x^{1/2}$. Given $\epsilon > 0$, if we have

$$\frac{\sqrt{\log\log n}}{\sqrt{\log\log x}} < 1 - \epsilon,$$

it follows that

$$\log\log x < \frac{\log 2}{\epsilon(2-\epsilon)}.$$

Hence, for x large enough, we have

$$\mathrm{Pr}_{S,x}\left\{n\colon n \text{ satisfies } \left|\frac{\sqrt{\log\log n}}{\sqrt{\log\log x}} - 1\right| > \epsilon\right\} \leqslant \mathrm{Pr}_{S,x}\left\{n\colon n \leqslant x^{1/2}\right\} = o(1).$$

The last equality follows from condition (C). Thus we have

$$\frac{\sqrt{\log\log n}}{\sqrt{\log\log x}} \xrightarrow{p} 1.$$

Similarly, we can prove

$$\frac{\log\log n - \log\log x}{\sqrt{\log\log x}} \xrightarrow{p} 0.$$

By Proposition 2, we obtain the equivalence of the statements in the lemma. □

Remark. This lemma is the only place where condition (C) is applied. Notice that the lemma still holds if the exponent $1/2$ is replaced by any constant between 0 and 1.

For $y = y(x)$, define

$$\omega_y(n) = \#\{l \leqslant y\colon\ l \text{ is a prime and } l \mid n\}.$$

It is a truncation function of $\omega(n)$. We can restate Theorem 3 in terms of ω_y instead of ω.

Lemma 2. *Let S be an infinite subset of $\mathbb{N}$ and $f : S \to \mathbb{N}$. Suppose there exist a constant β with $0 < \beta \leqslant 1$ and $y = y(x) < x^{\beta}$ such that conditions* (1) *to* (3) *hold. Then for $\gamma \in \mathbb{R}$, we have*

$$\lim_{x\to\infty} \Pr_{S,x}\left\{n\colon\ n \textit{ satisfies } \frac{\omega(f(n)) - \log\log x}{\sqrt{\log\log x}} \leqslant \gamma\right\} = G(\gamma)$$

if and only if

$$\lim_{x\to\infty} \Pr_{S,x}\left\{n\colon\ n \textit{ satisfies } \frac{\omega_y(f(n)) - \log\log x}{\sqrt{\log\log x}} \leqslant \gamma\right\} = G(\gamma).$$

Proof. Since

$$\frac{\omega_y(f(n)) - \log\log x}{\sqrt{\log\log x}} = \frac{\omega(f(n)) - \log\log x}{\sqrt{\log\log x}} + \frac{\omega_y(f(n)) - \omega(f(n))}{\sqrt{\log\log x}},$$

by Propositions 1 and 2, it suffices to prove that

$$\lim_{x\to\infty} \mathrm{E}_{S,x}\left\{\left|\frac{\omega(f(n)) - \omega_y(f(n))}{\sqrt{\log\log x}}\right|\right\} = 0.$$

Consider

$$\sum_{n\in S(x)} \big|\omega\big(f(n)\big) - \omega_y\big(f(n)\big)\big| = \sum_{y<l\leqslant x^{\beta}} \sum_{\substack{n\in S(x)\\ l\mid f(n)}} 1 + \sum_{n\in S(x)} \sum_{\substack{l>x^{\beta}\\ l\mid f(n)}} 1.$$

By condition (1), the second sum is bounded by

$$\sum_{n\in S(x)} \sum_{\substack{l>x^{\beta}\\ l\mid f(n)}} 1 = O\big(\big|S(x)\big|\big).$$

By conditions (2) and (3), we have

$$\sum_{y<l\leqslant x^{\beta}} \sum_{\substack{n\in S(x)\\ l\mid f(n)}} 1 = \sum_{y<l\leqslant x^{\beta}} \big|S(x)\big|(\lambda_l + e_l) = o\big(\big|S(x)\big|(\log\log x)^{1/2}\big).$$

Hence, we have

$$\sum_{n\in S(x)} \big|\omega\big(f(n)\big) - \omega_y\big(f(n)\big)\big| = o\big(\big|S(x)\big|(\log\log x)^{1/2}\big).$$

It follows that as $x \to \infty$,

$$\mathrm{E}_{S,x}\left\{\left|\frac{\omega(f(n)) - \omega_y(f(n))}{\sqrt{\log\log x}}\right|\right\} = \frac{o(|S(x)|(\log\log x)^{1/2})}{|S(x)|(\log\log x)^{1/2}} = o(1).$$

Thus Lemma 2 follows. □

From Lemmas 1 and 2, we see that to prove Theorem 3, it suffices to prove

$$\lim_{x\to\infty} \mathrm{Pr}_{S,x}\left\{n\colon n \text{ satisfies } \frac{\omega_y(f(n)) - \log\log x}{\sqrt{\log\log x}} \leqslant \gamma\right\} = G(\gamma).$$

The ω_y function can be associated to a sum of the following independent random variables. For a prime l, define an independent random variables X_l by

$$\Pr\{X_l = 1\} = \lambda_l \quad \text{and} \quad \Pr\{X_l = 0\} = 1 - \lambda_l.$$

For $y = y(x)$, let S_y be a random variable defined by

$$S_y := \sum_{l\leqslant y} X_l.$$

By conditions (4) and (5), we have

$$\mathrm{E}\{S_y\} = \sum_{l\leqslant y} \lambda_l = \log\log x + o(\log\log x)^{1/2},$$

$$\mathrm{Var}\{S_y\} = \sum_{l\leqslant y} \lambda_l(1-\lambda_l) = \log\log x + o(\log\log x)^{1/2}.$$

The terms $\log\log x$ in Lemma 2 can be replaced by $\mathrm{E}\{S_y\}$ and $\mathrm{Var}\{S_y\}$.

Lemma 3. *Let S be an infinite subset of $\mathbb{N}$ and $f : S \to \mathbb{N}$. Suppose there exist a constant β with $0 < \beta \leqslant 1$ and $y = y(x) < x^{\beta}$ such that conditions* (4) *and* (5) *hold. Then for $\gamma \in \mathbb{R}$, we have*

$$\lim_{x\to\infty} \mathrm{Pr}_{S,x}\left\{n\colon n \textit{ satisfies } \frac{\omega_y(f(n)) - \log\log x}{\sqrt{\log\log x}} \leqslant \gamma\right\} = G(\gamma)$$

if and only if

$$\lim_{x\to\infty} \mathrm{Pr}_{S,x}\left\{n\colon n \textit{ satisfies } \frac{\omega_y(f(n)) - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}} \leqslant \gamma\right\} = G(\gamma).$$

Proof. Write

$$\frac{\omega_y(f(n)) - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}} = \frac{\omega_y(f(n)) - \log\log x}{\sqrt{\log\log x}} \cdot \frac{\sqrt{\log\log x}}{\sqrt{\mathrm{Var}\{S_y\}}} + \frac{\log\log x - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}.$$

The above computations of $\mathrm{E}\{X\}$ and $\mathrm{Var}\{X\}$ imply that

$$\frac{\sqrt{\log\log x}}{\sqrt{\mathrm{Var}\{S_y\}}} \xrightarrow{p} 1 \quad \text{and} \quad \frac{\log\log x - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}} \xrightarrow{p} 0.$$

By Proposition 2, the lemma follows. □

Now, for a prime l, let $\delta_l : \mathbb{N} \to \{0, 1\}$ be a random variable defined by

$$\delta_l(n) := \begin{cases} 1 & \text{if } l \mid n, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, we can write

$$\omega_y\big(f(n)\big) = \sum_{l \leqslant y} \delta_l\big(f(n)\big).$$

Notice that

$$\mathrm{Pr}_{S,x}\big\{n\colon n \text{ satisfies } \delta_l\big(f(n)\big) = 1\big\} = \lambda_l + e_l(x).$$

Hence the expectations of random variables X_l and δ_l are close. Thus, the sum S_y of X_l is a good approximation of the sum ω_y of δ_l. Indeed, the rth moments of their normalizations are equal as $x \to \infty$.

Lemma 4. *Let S be an infinite subset of $\mathbb{N}$ and $f : S \to \mathbb{N}$. Suppose there exist a constant β with $0 < \beta \leqslant 1$ and $y = y(x) < x^{\beta}$ such that condition* (6) *holds. Then for $r \in \mathbb{N}$, we have*

$$\lim_{x\to\infty} \left| \mathrm{E}_{S,x}\left\{ \left(\frac{\omega_y(f(n)) - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}} \right)^r \right\} - \mathrm{E}\left\{ \left(\frac{S_y - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}} \right)^r \right\} \right| = 0.$$

Proof. For $0 \leqslant k \leqslant r$, write

$$\mathrm{E}\big\{S_y^k\big\} = \sum_{u=1}^{k} {\sum}' \frac{k!}{k_1! \cdots k_u!} {\sum}'' \mathrm{E}\big\{X_{l_1}^{k_1} \cdots X_{l_u}^{k_u}\big\},$$

where $\sum'$ extends over all u-tuples $(k_1, k_2, \ldots, k_u)$ of positive integers such that $k_1 + k_2 + \cdots + k_u = k$ and $\sum''$ extends over all u-tuples of distinct primes $(l_1, l_2, \ldots, l_u)$ with $l_i \leqslant y$. Since each X_{l_i} takes values 0 or 1 and the X_{l_i} are independent, we have

$$\mathrm{E}\big\{X_{l_1}^{k_1} \cdots X_{l_u}^{k_u}\big\} = \prod_{i=1}^{u} \lambda_{l_i}.$$

Similarly, if we abbreviate $\omega_y(f(n))$ and $\delta_l(f(n))$ by ω_y and δ_l, respectively, we have

$$\mathrm{E}_{S,x}\big\{\omega_y^k\big\} = \sum_{u=1}^{k} {\sum}' \frac{k!}{k_1! \cdots k_u!} {\sum}'' \mathrm{E}_{S,x}\big\{\delta_{l_1}^{k_1} \cdots \delta_{l_u}^{k_u}\big\},$$

with the same $\sum'$ and $\sum''$ as above. Notice that by definitions of λ_l and e_l, we have

$$\left|\mathrm{E}\left\{X_{l_1}^{k_1}\cdots X_{l_u}^{k_u}\right\} - \mathrm{E}_{S,x}\left\{\delta_{l_1}^{k_1}\cdots\delta_{l_u}^{k_u}\right\}\right| = |e_{l_1 l_2\cdots l_u}|.$$

Write

$$\mathrm{E}\left\{\left(S_y - \mathrm{E}\{S_y\}\right)^r\right\} = \sum_{k=0}^{r}\binom{r}{k}\mathrm{E}\left\{S_y^k\right\}\mathrm{E}\{S_y\}^{r-k}$$

and

$$\mathrm{E}_{S,x}\left\{\left(\omega_y - \mathrm{E}\{S_y\}\right)^r\right\} = \sum_{k=0}^{r}\binom{r}{k}\mathrm{E}_{S,x}\left\{\omega_y^k\right\}\mathrm{E}\{S_y\}^{r-k}.$$

Since

$$\mathrm{E}\{S_y\} = \log\log x + o(\log\log x)^{1/2},$$

by condition (6), we have

$$\begin{aligned}&\left|\mathrm{E}_{S,x}\left\{\left(\omega_y - \mathrm{E}\{S_y\}\right)^r\right\} - \mathrm{E}\left\{\left(S_y - \mathrm{E}\{S_y\}\right)^r\right\}\right|\\ &\quad\ll \sum_{k=0}^{r}\binom{r}{k}\left(\sum_{u=1}^{k}\sum{}'\frac{k!}{k_1!k_2!\cdots k_u!}\sum{}''|e_{l_1\cdots l_u}|\,(\log\log x)^{r-k}\right)\\ &\quad= o\left((\log\log x)^{r/2}\right).\end{aligned}$$

Notice that

$$\mathrm{Var}\{S_y\} = \log\log x + o(\log\log x)^{1/2}.$$

Thus the lemma follows. □

Following the same argument as in [11, Lemma 7], we have the following lemma which is about the rth moment of S_y.

Lemma 5. *Let S be an infinite subset of $\mathbb{N}$. Then for $r \in \mathbb{N}$, we have*

$$\sup_y\left|\mathrm{E}\left\{\left(\frac{S_y - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}\right)^r\right\}\right| < \infty.$$

Combine Lemmas 1–5. Applying Propositions 3–5, using the same argument as in [11, Theorem 1], we conclude that: if S is an infinite subset of $\mathbb{N}$ satisfying condition (C) and $f : S \to \mathbb{N}$ satisfying conditions (1)–(6), for $\gamma \in \mathbb{R}$, we have

$$\lim_{x\to\infty}\frac{1}{|S(x)|}\#\left\{n \in S(x)\colon\ n \text{ satisfies } \frac{\omega(f(n)) - \log\log n}{\sqrt{\log\log n}} \leqslant \gamma\right\} = G(\gamma).$$

Thus we obtain Theorem 3.

Although it seems difficult at first to check all conditions in Theorem 3, in most cases, they can be verified very easily. For example, let $S = \mathbb{N}$, f the identity map, and $\beta = 1$. Then conditions (C) and (1) follow. If we take $\lambda_l = \frac{1}{l}$, we can bound $|e_l|$ and $|e_{l_1 l_2 \cdots l_u}|$ by $O(\frac{1}{x})$. By choosing $y = x^{1/\log\log x}$, conditions (2) and (4) follow from the classical Mertens theorem [12] and the series in condition (5) is convergent. Also, we have

$$\sum_{y<l\leqslant x} |e_l| \ll x \cdot \frac{1}{x} \ll 1 \quad \text{and} \quad \sum{}'' |e_{l_1 l_2 \cdots l_u}| \ll y^u \cdot \frac{1}{x} \ll x^{\epsilon-1},$$

where $\epsilon \to 0$ as $x \to 0$. Hence, all conditions are satisfied and we recover from Theorem 3 the classical Erdős–Kac theorem.

In the following sections, we consider only the set of rational primes. The following corollary is a direct consequence of Theorem 3.

Corollary 1. *Let $S \subseteq \mathbb{N}$ be a set of all but finitely many primes. For a map $f : S \to \mathbb{N}$, suppose there exists a constant $\delta > 0$ such that $f(p) \leqslant p^{\delta}$ for all $p \in S$. Define λ_l, e_l, $e_{l_1 l_2 \cdots l_u}$, y, and β as before. Assuming they satisfy conditions* (2)–(6)*, for $\gamma \in \mathbb{R}$, we have*

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\left\{ p \in S(x):\ p \text{ satisfies } \frac{\omega(f(p)) - \log\log p}{\sqrt{\log\log p}} \leqslant \gamma \right\} = G(\gamma).$$

4. Elliptic curves without complex multiplication

Let $E/\mathbb{Q}$ be an elliptic curve. For a prime p of good reduction, let $E(\mathbb{F}_p)$ be the set of rational points of E defined over a finite field $\mathbb{F}_p$. If E is a non-CM elliptic curve, assuming the GRH, Miri and Murty [13] proved that the normal order of $\omega(\#E(\mathbb{F}_p))$ is $\log\log p$. For a CM elliptic curve, the author [10] showed that the same conclusion holds unconditionally. These results suggest a possible existence of a normal distribution for the quantity

$$\frac{\omega(\#E(\mathbb{F}_p)) - \log\log p}{\sqrt{\log\log p}}.$$

In this section, we prove Theorem 1 in the case of elliptic curves without complex multiplication. Let $E/\mathbb{Q}$ be a non-CM elliptic curve and $S \subseteq \mathbb{N}$ the set of primes of good reduction. Let f be a map from S to $\mathbb{N}$ defined by $p \mapsto \#E(\mathbb{F}_p)$. Notice that $\#E(\mathbb{F}_p) \leqslant p + 1 + 2\sqrt{p} \leqslant p^3$. In [13], we have seen that for all but finitely many primes l, assuming the GRH, we have

$$\#\big\{ p \leqslant x:\ \#E(\mathbb{F}_p) \text{ is divisible by } l \big\} = \frac{(l^3 - 2l)}{(l^2-1)(l^2-l)} \operatorname{li} x + \mathrm{O}\big(l^3 x^{1/2} \log l^4 N x\big),$$

where N is the conductor of E and $\operatorname{li} x = \int_2^x \frac{dt}{\log t}$. We define

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

Let $y = x^{1/\log\log x}$ and $0 < \beta \leqslant 1$ (a choice of β will be made later). Since

$$\sum_{l:\ \text{prime}} \left|\lambda_l - \frac{1}{l}\right| \ll 1,$$

Conditions (2) and (4) follow from the Mertens theorem [12] and the series in (5) is convergent.

Consider

$$\sum_{y<l\leqslant x^\beta} |e_l| \ll \frac{1}{\pi(x)} \sum_{y<l\leqslant x^\beta} l^3 x^{1/2} \log l^4 Nx \ll \frac{1}{\pi(x)} \sum_{y<l\leqslant x^\beta} x^{(1/2+\epsilon)} l^4 \ll x^{(\epsilon-1/2+5\beta)}.$$

By taking $\beta = 1/11$, condition (3) follows. Consider the product of primes $l_1 l_2 \cdots l_u$. Using the same argument as the one for e_l, for all but finitely many $l_1 l_2 \cdots l_u$, assuming the GRH, we have

$$|e_{l_1 l_2 \cdots l_u}| \ll x^{(\epsilon-1/2)} l_1^4 l_2^4 \cdots l_u^4.$$

For $r \in \mathbb{N}$, let $\sum''$ be the sum over all u-tuples of distinct primes $(l_1, l_2, \ldots, l_u)$ with $l_i \leqslant y$. We have

$$\sum{}'' |\, e_{l_1 \cdots l_u}| \ll x^{\epsilon-1/2} \cdot \left(\sum_{l\leqslant y} l^4\right)^u \ll x^{\epsilon-1/2} \cdot y^{5u} \ll x^{\epsilon-1/2}.$$

The last inequality holds since $y = o(x^\epsilon)$ for any $\epsilon > 0$. Combine all the above results. Applying Corollary 1, we conclude that under the GRH, for a non-CM elliptic curve E, we have

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\left\{ p \leqslant x:\ p \text{ is of good reduction and } \frac{\omega(\#E(\mathbb{F}_p)) - \log\log p}{\sqrt{\log\log p}} \leqslant \gamma \right\} = G(\gamma).$$

Remark. Let $L/\mathbb{Q}$ be a number field and $\zeta_L(s)$ the Dedekind zeta function of L. A generalized Riemann Hypothesis states that $\zeta_L(s)$ has no zero in the region $\mathrm{Re}(s) > 1/2$. For some $\delta \in \mathbb{R}$ with $1/2 < \delta < 1$, we assume a weaker condition that $\zeta_L(s)$ has no zero in the region $\mathrm{Re}(s) > \delta$. It is called the δ-quasi-Riemann Hypothesis. Assuming the δ-quasi-Riemann Hypothesis, the term $x^{1/2}$ appearing as a part of error term for $\#\{p \leqslant x:\ \#E(\mathbb{F}_p)$ is divisible by $l\}$ is replaced by x^δ. Choosing $0 < \theta < (1-\delta)/10$ in the above proof, it follows that the δ-quasi-Riemann Hypothesis is sufficient to prove Theorem 1 in the case of non-CM elliptic curves.

5. Elliptic curves with complex multiplication

In this section, we prove Theorem 1 in the case of CM elliptic curves. Let $E/\mathbb{Q}$ be an elliptic curve with complex multiplication. Let S and f be defined as in Section 4. For each prime l, we write

$$\frac{1}{\pi(x)} \#\left\{ p \in S(x):\ \#E(\mathbb{F}_p) \text{ is divisible by } l \right\} = \frac{1}{\varphi(l)} + e_l,$$

where $\varphi(l)$ is the Euler function and

$$e_l = \frac{1}{\pi(x)}\left[\#\{p \in S(x):\ \#E(\mathbb{F}_p) \text{ is divisible by } l\} - \frac{\pi(x)}{\varphi(l)}\right].$$

Let $y = x^{1/\log\log x}$ and $0 < \beta \leqslant 1$ (a choice of β will be made later). Choosing $\lambda_l = 1/\varphi(l)$, conditions (2), (4), and (5) are satisfied.

Consider

$$\sum_{y<l\leqslant x^\beta} |e_l| = \frac{1}{\pi(x)} \sum_{y<l\leqslant x^\beta}\left[\#\{p \in S(x):\ \#E(\mathbb{F}_p) \text{ is divisible by } l\} - \frac{\pi(x)}{\varphi(l)}\right].$$

We subdivide the elements of S into two classes according to whether $p \in S$ is supersingular or $p \in S$ is ordinary. If p is supersingular, by Deuring's lemma [3] and the Bombieri–Vinogradov theorem [2,16], it was proved in [10] that

$$\sum_{y<l\leqslant x^\beta} \#\{p \in S(x):\ p \text{ is supersingular and } l \mid \#E(\mathbb{F}_p)\} = \sum_{y<l\leqslant x^\beta} \frac{\operatorname{li} x}{2\varphi(l)} + \mathrm{O}\big(x(\log x)^{-A}\big),$$

for any constant $A > 1$, provided the constant $\beta < 1/2$. If p is ordinary, by Wilson's result [17] on an analogue of the Bombieri–Vinogradov theorem in number fields, it was proved in [10] that if $2\beta < 1/3$,

$$\sum_{y<l\leqslant x^\beta} \#\{p \in S(x):\ p \text{ is ordinary and } l \mid \#E(\mathbb{F}_p)\} = \sum_{y<l\leqslant x^\beta} \frac{\operatorname{li} x}{2\varphi(l)} + \mathrm{O}\big(x(\log x)^{-A}\big).$$

Combine all the above results. By choosing $\beta = 1/7$, we have

$$\sum_{y<l\leqslant x^\beta} |e_l| \ll \frac{1}{\pi(x)}\big(x(\log x)^{-A}\big) \ll 1.$$

Thus condition (3) follows.

For distinct primes $l_1, l_2, \ldots, l_u$, we have

$$e_{l_1 l_2 \cdots l_u} = \frac{1}{\pi(x)}\#\{p \leqslant x:\ \#E(\mathbb{F}_p) \text{ is divisible by } l_1 l_2 \cdots l_u\} - \frac{1}{\varphi(l_1 l_2 \cdots l_u)}.$$

Following the same arguments as in the verification of condition (3), the Bombieri–Vinogradov theorem and Wilson's theorem imply that

$$\sum{}'' \#\{p \leqslant x:\ \#E(\mathbb{F}_p) \text{ is divisible by } l_1 l_2 \cdots l_u\} = \frac{\operatorname{li} x}{\varphi(l_1 l_2 \cdots l_u)} + \mathrm{O}\big(x(\log x)^{-A}\big),$$

for any constant $A > 1$. Hence, for all $r \in \mathbb{N}$, we have

$$\sum{}'' |e_{l_1 \cdots l_u}| \ll \frac{1}{\pi(x)}\big(x(\log x)^{-A}\big) \ll 1.$$

Thus condition (6) follows. Applying Corollary 1, we conclude that for a CM elliptic curve $E/\mathbb{Q}$, we have

$$\lim_{x\to\infty}\frac{1}{\pi(x)}\#\left\{p\leqslant x\colon\ p \text{ is of good reduction and } \frac{\omega(\#E(\mathbb{F}_p))-\log\log p}{\sqrt{\log\log p}}\leqslant\gamma\right\}=G(\gamma).$$

Thus, we complete the proof of Theorem 1.

6. Conclusion

In the proof of the existence of a normal distribution, the main difference between the sequence of natural numbers and the sequence of primes is the estimate of $|e_{l_1l_2\cdots l_u}|$ appearing in conditions (3) (i.e., $u=1$) and (6). In the case of natural numbers, we have

$$|e_{l_1l_2\cdots l_u}|\ll\frac{1}{x},$$

for products of distinct primes $l_1l_2\cdots l_u$. However, in the case of primes, we do not have such a good control of $|e_{l_1l_2\cdots l_u}|$. There are two strategies that we can apply. One is to assume the GRH and get an estimate for $|e_{l_1l_2\cdots l_u}|$. For example, in the proof of Theorem 1 for non-CM elliptic curves, assuming the GRH, we have

$$|e_{l_1l_2\cdots l_u}|\ll x^{(\epsilon-1/2)}l_1^4l_2^4\cdots l_u^4.$$

Another method to verify conditions (3) and (6) is to get an average result for $|e_{l_1l_2\cdots l_u}|$ over u-tuples of distinct primes $(l_1,l_2,\ldots,l_u)$ with $l_i\leqslant y$. The case of CM elliptic curves is one of these types where we have

$$\sum{}''\left[\#\big\{p\leqslant x\colon\ \#E(\mathbb{F}_p) \text{ is divisible by } l_1l_2\cdots l_u\big\}-\frac{\pi(x)}{\varphi(l_1l_2\cdots l_u)}\right]\ll x(\log x)^{-A},$$

for some $A>1$. Thus we omit the assumption of the GRH. Following the philosophy of the second approach, we see that in the case of a non-CM elliptic curve $E/\mathbb{Q}$, to remove the condition of the GRH, it suffices to prove an analogous result of Bombieri–Vinogradov for E. More precisely, if we have

$$\sum{}''\left[\#\big\{p\leqslant x\colon\ \#E(\mathbb{F}_p) \text{ is divisible by } l_1l_2\cdots l_u\big\}-\pi(x)\prod_{i=1}^{u}\lambda_{l_i}\right]\ll x(\log x)^{-A},$$

we can obtain Theorem 1 without the GRH. The same principle can be applied to all previous prime analogues of the Erdős–Kac theorem that were obtained under the GRH. However, it seems that the recent techniques in analytic number theory are not able to tackle such a "non-abelian" analogue of the Bombieri–Vinogradov theorem. It is certainly a project worth to be investigated.

Acknowledgments

This paper is part of my PhD thesis at Harvard. I thank Professor B. Mazur and Professor R. Murty for their comments about this paper. I also thank the referee for his her valuable suggestions.

References

[1] P. Billingsley, On the central limit theorem for the prime divisor functions, Amer. Math. Monthly 76 (1969) 132–139.
[2] E. Bombieri, On the large sieve, Mathematika 12 (1965) 201–225.
[3] M. Deuring, Die typen der muptiplikatorenringe elliptischer funktionenkörper, Abh. Math. Sem. Hansischen Univ. 14 (1941) 197–272.
[4] P.D.T.A. Elliott, Probabilistic Number Theory I. Mean Value Theorem, Grundlehren Math. Wiss., vol. 239, Springer, New York, 1979.
[5] P.D.T.A. Elliott, Probabilistic Number Theory II. Central Limit Theorems, Grundlehren Math. Wiss., vol. 240, Springer, New York, 1979.
[6] P. Erdős, M. Kac, The Gaussian law of errors in the theory of additive number theoretic functions, Amer. J. Math. 62 (1940) 738–742.
[7] W. Feller, An Introduction to Probability Theory and Its Applications, vol. II, Wiley, New York, 1966.
[8] H. Halberstam, On the distribution of additive number theoretic functions, I–III, J. London Math. Soc. 30 (1955) 43–53, 31 (1956) 1–14, 15–27.
[9] J. Kubilius, Probabilistic Methods in Number Theory, Transl. Math. Monogr., vol. 11, Amer. Math. Soc., Providence, RI, 1964.
[10] Y.-R. Liu, Prime divisors of number of rational points on elliptic curves with complex multiplication, Bull. London Math. Soc. 37 (2005) 658–664.
[11] Y.-R. Liu, A generalization of the Erdős–Kac theorem and its applications, Canad. Math. Bull. 47 (2004) 589–606.
[12] F. Mertens, Ein Beitrag zur analytischen Zahlentheorie, J. Reine Angew. Math. 78 (1874) 46–62.
[13] S.A. Miri, V.K. Murty, An application of sieve methods to elliptic curves, in: Progress in Cryptology—INDOCRYPT, in: Lecture Notes in Comput. Sci., vol. 2247, Springer, Berlin, 2001, pp. 91–98.
[14] M.R. Murty, V.K. Murty, An analogue of the Erdős–Kac theorem for Fourier coefficients of modular forms, Indian J. Pure Appl. Math. 15 (1984) 1090–1101.
[15] H. Shapiro, Distribution functions of additive arithmetic functions, Proc. Natl. Acad. Sci. USA 42 (1956) 426–430.
[16] A.I. Vinogradov, On the density hypothesis for Dirichlet L-functions, Izv. Akad. Nauk SSSR Ser. Math. 29 (1965) 903–934, 30 (1966) 719–720.
[17] R.J. Wilson, The large sieve in algebraic number fields, Mathematika 16 (1969) 189–204.