A Generalization of the Erdös-Kac Theorem and its Applications

Yu-Ru Liu *

November 8, 2020

Abstract

We axiomatize the main properties of the classical Erdös-Kac Theorem in order to apply it to a general context. We provide applications in the cases of number fields, function fields, and geometrically irreducible varieties over a finite field.
2000 Mathematics Subject Classification. 11N60, 11N80.

1 Introduction.

For $m \in \mathbb{N}$, define $\omega(m)$ to be the number of distinct prime divisors of m. The Turán Theorem is about the second moment of $\omega(m)$. For $x \in \mathbb{Q}$, Turán proved that [12]

$$\sum_{m \leq x} \big(\omega(m) - \log\log x\big)^2 \ll x \log\log x.$$

A direct consequence of this theorem is that

$$\#\Big\{m \leq x, \Big|\frac{\omega(m) - \log\log m}{\sqrt{\log\log m}}\Big| > g_x\Big\} = \mathrm{o}(x),$$

for any sequence $\{g_x\}$ satisfying $g_x \to \infty$ as $x \to \infty$. In particular, it implies a result of Hardy and Ramanujan [5] that the normal order of $\omega(m)$ is $\log\log m$. The idea behind Turán's proof was essentially probabilistic. In 1940, further development of probabilistic ideas led Erdös and Kac [2] to prove a remarkable refinement of the Turán Theorem. They discovered that there exists a Gaussian normal distribution for the quantity

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

More precisely, for $\gamma \in \mathbb{R}$, Erdös-Kac proved that

$$\lim_{x\to\infty} \frac{1}{[x]} \#\Big\{m : \, m \leq x, \, \frac{\omega(m) - \log\log m}{\sqrt{\log\log m}} \leq \gamma\Big\} = G(\gamma) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\gamma} e^{\frac{-t^2}{2}}\, dt,$$

*Research partially supported by NSERC PGSB scholarship .

where $[x]$ is the largest integer $\leq x$.

In their original paper, Erdös and Kac used a technically involved sieve method to obtain this result. In 1955, Halberstam [4] gave a more probabilistically natural approach to this theorem by using the method of 'all moments'. In 1969, by applying the concept of independent random variables, Billingsley [1] provided an elementary proof of the Erdös-Kac Theorem. Thanks to his efforts, we can give a generalization of this Theorem.

Let P be a set of elements with a map

$$N\colon P \to \mathbb{N}\backslash\{1\},\ p \mapsto N(p).$$

Let M be a free abelian monoid generated by elements of P. For each $m \in M$, we write

$$m = \sum_{p\in P} n_p(m)p,$$

with $n_p(m) \in \mathbb{N} \cup \{0\}$ and $n_p(m) = 0$ for all but finitely many p. We extend the map N on M as follows:

$$\begin{aligned} N : M &\longrightarrow \mathbb{N} \\ m = \sum_{p\in P} n_p(m)p &\longmapsto N(m) := \prod_{p\in P} N(p)^{n_p(m)}, \end{aligned}$$

i.e., N is a monoid homomorphism from $(M, +)$ to $(\mathbb{N}, \cdot)$. Let X be a countable subset of $\mathbb{Q}$ that contains the image $\mathrm{Im}(N(M))$ with an extra condition: if $x_1, x_2 \in X$, the fraction x_1/x_2 belongs to X, too. Without loss of generality, we assume $X = \mathbb{Q}$ or $X = \{q^z, z \in \mathbb{Z}\}$ for some $q \in \mathbb{N}$ (see Remark at the end of this section for a more detailed discussion about X).

Given P, M, and X as above, for each (sufficiently large) $x \in X$, we assume that the following two conditions hold: let $m \in M$ and $p \in P$, we have

$$\text{(A)} \sum_{N(m)\leq x} 1 = \kappa x + \mathrm{O}(x^\theta), \quad \text{for some} \quad \kappa > 0 \quad \text{and} \quad 0 \leq \theta < 1.$$

$$\text{(B)} \sum_{N(p)\leq x} 1 = \mathrm{O}\Big(\frac{x}{\log x}\Big).$$

For each $m \in M$, we define

$$\omega(m) = \sum_{\substack{p\in P \\ n_p(m)\geq 1}} 1.$$

It it the number of elements of P that generate m, counted without multiplicity. Given P, M, and X satisfying (A) and (B), the author [9] proved that for $x \in X$, we have

$$\sum_{N(m)\leq x} \big(\omega(m) - \log\log x\big)^2 = \kappa x \log\log x + Cx + \mathrm{O}\Big(\frac{x \log\log x}{\log x}\Big).$$

Here κ is the same constant as in (A) and C is another constant. This result is a generalization of the Turán Theorem. It implies that

$$\#\Big\{m \in M, N(m) \leq x, \Big|\frac{\omega(m) - \log\log N(m)}{\sqrt{\log\log N(m)}}\Big| > g_x\Big\} = \mathrm{o}(x),$$

for any sequence $\{g_x\}$ satisfying $g_x \to \infty$ as $x \to \infty$. In particular, we obtain that the normal order of $\omega(m)$ is $\log\log N(m)$. This result suggests a possible existence of a normal distribution for the quantity

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

This is indeed the case.

Theorem 1 *Given P, M, and X as before, assume they satisfy* (A) *and* (B)*. For $m \in M$, we have*

$$\lim_{x\to\infty} \frac{1}{\#\{m : N(m) \leq x\}} \#\Big\{m : N(m) \leq x, \frac{\omega(m) - \log\log N(m)}{\sqrt{\log\log N(m)}} \leq \gamma\Big\} = G(\gamma).$$

In [9], the author provided the following applications where the general setting can be applied.

Example 1 In the case of rational numbers, let P be the set of primes of $\mathbb{N}$ with the identity map N. Take $M = \mathbb{N}$ and $X = \mathbb{Q}$. Condition (A) is true since

$$\#\{m \in \mathbb{N},\, m \leq x\} = [x] = x + \mathrm{O}(1).$$

Also, Condition (B) is the classical Chebyshev Theorem [11] (p36-37). Hence, by Theorem 1, we recover the classical Erdös-Kac Theorem.

Example 2 Given a number field K, let $\mathcal{O}_K$ be its ring of integer. Let P be the set of prime ideals of $\mathcal{O}_K$ with the standard norm map N, i.e., $\mathfrak{p} \mapsto |\mathcal{O}_K/\mathfrak{p}|$. Let M be the set of ideals and $X = \mathbb{Q}$. Condition (A) is a result of Weber [13]. Also, Condition (B) follows from the classical Chebyshev Theorem and the fact that there are only finitely many prime ideals lying above a rational prime. Thus we have

Corollary 1 *Let $K/\mathbb{Q}$ be a number field and $\mathcal{O}_K$ be its ring of integers. For an ideal $\mathfrak{m}$ of $\mathcal{O}_K$, let $\omega(\mathfrak{m})$ denote the number of distinct prime ideals dividing $\mathfrak{m}$. For $x \in \mathbb{Q}$, we have*

$$\lim_{x\in\infty} \frac{1}{\#\{\mathfrak{m} : |\mathcal{O}_K/\mathfrak{m}| \leq x\}} \#\Big\{\mathfrak{m} : |\mathcal{O}_K/\mathfrak{m}| \leq x, \frac{\omega(\mathfrak{m}) - \log\log(|\mathcal{O}_K/\mathfrak{m}|)}{\sqrt{\log\log(|\mathcal{O}_K/\mathfrak{m}|)}} \leq \gamma\Big\} = G(\gamma).$$

Example 3 Let $\mathbb{F}_q[t]$ be the ring of polynomials of one variable over a finite field $\mathbb{F}_q$. Take P to be the set of monic irreducible polynomials with $p \mapsto q^{\deg p}$, where $\deg p$ is the degree of the polynomial p. Let M be the set of monic polynomials and $X = \{q^z, z \in \mathbb{Z}\}$. Conditions (A) and (B) can be easily derived from the fact that for a fixed $d \in \mathbb{N}$,

$$\#\big\{m \in M, \deg m = d\big\} = q^d.$$

Hence, we have a generalization of the Erdös-Kac Theorem in the case of function fields. Related results about this case can also be found in [14].

Example 4 Let $V/\mathbb{F}_q$ be a geometrically irreducible variety of dimension r over a finite field $\mathbb{F}_q$. Let P be the set of closed points with $p \mapsto (q^r)^{\deg p}$, where $\deg p$ is the length of the corresponding orbit [10] (p259). Take M to be the set of effective 0-cycles and $X = \{(q^r)^z, z \in \mathbb{Z}\}$. Conditions (A) and (B) can be verified by the estimate of Lang-Weil [8] about the number of points of V. Hence, we have

Corollary 2 *Let $V/\mathbb{F}_q$ be a geometrically irreducible variety of dimension r over a finite field $\mathbb{F}_q$. Let P be the set of closed points and M be the set of effective 0-cycles. Let $X = \{(q^r)^z, z \in \mathbb{Z}\}$. For $m \in M$, write $m = \sum_{p \in P} n_p(m)p$. The degree of m is defined by*

$$\deg m = \sum_{p \in P} n_p(m) \deg p,$$

where $\deg p$ is the length of the corresponding orbit of p. Let $\omega(m)$ denote the number of distinct closed points on m. We have

$$\lim_{n \in \infty} \frac{1}{\#\big\{m : \deg m \leq n\big\}} \#\left\{m : \deg m \leq n,\ \frac{\omega(m) - \log(\deg m)}{\sqrt{\log(\deg m)}} \leq \gamma\right\} = G(\gamma).$$

This application can be viewed as the first geometric analogue of the Erdös-Kac Theorem.

Remark The conditions that we impose on the set X give only two choices for it: either X is dense in $\mathbb{R}_0^+ = \big\{r \in \mathbb{R}, r > 0\big\}$ or $X = \big\{q^z, z \in \mathbb{Z}\big\}$ for some $q > 1$. For the purpose of our applications, we take either $X = \mathbb{Q}$ or $X = \big\{q^z, z \in \mathbb{Z}\big\}$ for $q \in \mathbb{N}$. I would like to thank W. Kuo for providing the following theorem.

Theorem 2 (W. Kuo) *Let X be a subset of $\mathbb{R}_0^+$ that satisfies the following two conditions:*

- $\mathrm{Im}N(M) \subset X$, *and*

- *If $x_1, x_2 \in X$, the quotient $x_1/x_2 \in X$.*

Then X is either

- *dense in $\mathbb{R}_0^+$ or*

- *there is a* $q > 1$, *such that* $X = \{q^z | z \in \mathbb{Z}\}$.

In the first case, we say X *is archimedean; the second one is called non-archimedean.*

Proof: Let $p_1 \in P$ such that

$$N(p_1) = \min\{N(p), p \in P\}.$$

We consider the following two cases.

1. There is a $p \in P$ such that
$$\frac{\log N(p)}{\log N(p_1)} = \gamma \notin \mathbb{Q}.$$

2. For all $p \in P$,
$$\frac{\log N(p)}{\log N(p_1)} = \frac{m_p}{n_p} \in \mathbb{Q}, \quad m_p, n_p \in \mathbb{N}, (m_p, n_p) = 1.$$

For the first case, we claim that X is dense in $\mathbb{R}_0^+$; its proof is following. By the conditions of X, we know that for $m, n \in \mathbb{N}$,
$$\frac{N(p)^m}{N(p_1)^n} \in X.$$
We shall show that any positive number can be approximated by elements of the form $N(p)^m/N(p_1)^n$. It suffices to show that $\log(N(p)^m/N(p_1)^n)$ is dense in $\mathbb{R}$. We have

$$\log\left(\frac{N(p)^m}{N(p_1)^n}\right) = \log N(p) \cdot \left(m - n \cdot \frac{\log N(p)}{\log N(p_1)}\right) = \log N(p) \cdot (m - n\gamma).$$

Since γ is irrational, the set $\{(m - n\gamma) \,|\, m, n \in \mathbb{Z}\}$ is dense in $\mathbb{R}$. Therefore, X is dense in $\mathbb{R}_0^+$. Now, consider the second case. If we assume first that

$$\overline{\lim_{\substack{p \in P \\ N(p) \to \infty}}} \quad n_p = \infty.$$

Then the set
$$\{N(p_1)^{z/n_p} \,|\, z \in \mathbb{Z},\, p \in P\}$$
is dense in $\mathbb{R}_0^+$ since the set of its log

$$\{z/n_p \,|\, z \in \mathbb{Z},\, p \in P\},$$

is dense in $\mathbb{R}$. Therefore, in this case, X is also dense in $\mathbb{R}_0^+$. On the other hand, if we have
$$\overline{\lim_{\substack{p \in P \\ N(p) \to \infty}}} \quad n_p = M < \infty,$$

Then X contains the set

$$\{N(p_1)^{z/M} \mid z \in \mathbb{Z}\}.$$

If there is any other element of X not containing in the above set, repeat the same argument. We get either X is dense in $\mathbb{R}_0^+$ or X is supported on a power of a positive number.

Moreover, since X is either archimedean or non-archimedean, in either case, Condition (A) indeed imply (B). The case $X = \mathbb{Q}$ is a result of Landau [7] and the case $X = \{q^z, z \in \mathbb{Z}\}$ is proved by Knopfmacher [6] (p76). Since the proof of (A) implies (B) is involved, in the following discussion, we will continue to assume both Conditions (A) and (B) with the understanding that (B) is indeed redundant.

2 Review of probability theory.

In this section, we review some probability theory.

Given a random variable X with a probability measure P. For $t \in \mathbb{R}$, the function F defined by $F(t) = \mathrm{P}\{X \leq t\}$ is the *distribution function* of X. The *expectation* of X is defined by

$$\mathrm{E}\{X\} = \int_{-\infty}^{\infty} t \, dF(t).$$

The *variance* of X measures the difference between X and $\mathrm{E}\{X\}$. It is defined by

$$\mathrm{Var}\{X\} = \mathrm{E}\{(X - \mathrm{E}\{X\})^2\} = \mathrm{E}\{X^2\} - (\mathrm{E}\{X\})^2.$$

Let X and Y be two random variables with the same probability measure P. We have

$$\mathrm{E}\{X + Y\} = \mathrm{E}\{X\} + \mathrm{E}\{Y\}.$$

If X and Y are *independent*, i.e., for all $x \in \mathbb{R}, y \in \mathbb{R}$,

$$\mathrm{P}\{X \leq x, Y \leq y\} = \mathrm{P}\{X \leq x\} \cdot \mathrm{P}\{Y \leq y\},$$

we have

$$\mathrm{E}\{X \cdot Y\} = \mathrm{E}\{X\} \cdot \mathrm{E}\{Y\}$$

and

$$\mathrm{Var}\{X + Y\} = \mathrm{Var}\{X\} + \mathrm{Var}\{Y\}.$$

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

$$\lim_{x \to \infty} \mathrm{P}\{|X_n - \alpha| > \epsilon\} = 0.$$

We denote it by

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

Now, we are in a position to state some facts from probability theory that are needed to prove Theorem 1; most of their proofs can be found in [1] and [3].

Fact 1 *Given a sequence of random variables* $\{X_n\}$*, if*

$$\lim_{n\to\infty} \mathrm{E}\big\{|X_n|\big\} = 0,$$

we have

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

Proof: Fix an $\epsilon > 0$. Since $\lim\limits_{n\to\infty} \mathrm{E}\big\{|X_n|\big\} = 0$, for any $\epsilon_1 > 0$, there exists $N = N(\epsilon_1) \in \mathbb{N}$ such that for all $n > N$, we have

$$\epsilon \cdot \mathrm{P}\big\{|X_n| > \epsilon\big\} \le \int_{-\infty}^{\infty} |t|\, dF_n(t) < \epsilon_1.$$

It implies that

$$\mathrm{P}\big\{|X_n| > \epsilon\big\} < \epsilon_1/\epsilon.$$

By choosing ϵ_1 small enough, the fact follows.

Fact 2 ([1] p134-135, [3] p247) *Let* $\{X_n\}, \{Y_n\}$*, and* $\{U_n\}$ *be sequences of random variables with the same probability measure* P*. Let* U *be a distribution function. Suppose*

$$X_n \xrightarrow{\ p\ } 1 \qquad \text{and} \qquad Y_n \xrightarrow{\ p\ } 0.$$

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

$$\lim_{x\to\infty} \mathrm{P}\big\{U_n \le \gamma\big\} = U(\gamma)$$

if and only if

$$\lim_{x\to\infty} \mathrm{P}\big\{(X_n U_n + Y_n) \le \gamma\big\} = U(\gamma).$$

We use $G(\gamma)$ to 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 r-th moment of G is defined by

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

Notice that for an odd integer r, we have

$$\begin{aligned}
\int_{\infty}^{\infty} |t|^r\, dG(t) &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} t^r \cdot e^{-t^2/2} dt \\
&= \frac{2}{\sqrt{2\pi}} \int_o^{\infty} (2u)^{(r-1)/2} \cdot e^{-u} du \\
&= \frac{2}{\sqrt{2\pi}} \cdot 2^{(r-1)/2} \cdot \Big(\frac{r-1}{2}\Big)!.
\end{aligned}$$

The last equality holds since $\int_0^\infty t^n e^{-t} dt = n!$. Thus we have

$$\lim_{r\to\infty} \sup \frac{1}{r} \Big(\int_{\infty}^{\infty} |t|^r \, dG(t) \Big)^{1/r} = 0.$$

It follows from [3] (p487) that G is uniquely determined by these moments. Thus we have

Fact 3 ([3] p262-263) *Given a sequences of distribution functions $\{F_n\}$, if for all $r \in \mathbb{N}$,*

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} t^r dF_n(t) = \mu_r,$$

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

$$\lim_{x\to\infty} F_n(\gamma) = G(\gamma).$$

This next fact is an analogue of the Lebesgue Dominated Theorem.

Fact 4 ([3] p244-245) *Let $r \in \mathbb{N}$. Given a sequence of distribution functions $\{F_n\}$, if*

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

and

$$\sup_n \Big\{ \int_{-\infty}^{\infty} |t|^{r+\delta} dF_n(t) \Big\} < \infty, \quad \textit{for some } \delta = \delta(r) > 0,$$

we have

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} t^r dF_n(t) = \mu_r.$$

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

Fact 5 ([3] p256-258) *Let $X_1, X_2, \cdots, X_i, \cdots$ be a sequence of independent random variables and $\mathrm{Im}(X_i)$ is the image of X_i. Suppose*

(1) $\sup_i \big\{ \mathrm{Im}(X_i) \big\} < \infty$.

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

For $n \in \mathbb{N}$, let G_n be the 'normalization' of $X_1, X_2, \cdots, X_n$, i.e.,

$$G_n := \Big(\sum_{i=1}^{n} X_i \Big) \Big/ \Big(\sum_{i=1}^{n} \mathrm{Var}\big\{X_i\big\} \Big)^{\frac{1}{2}}.$$

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

$$\lim_{n\to\infty} \mathrm{P}\big\{G_n \le \gamma\big\} = G(\gamma).$$

3 Technical lemmas.

Given P, M, and X as defined before, assume they satisfy (A) and (B). We need the following two lemmas from [9].

Lemma 1 (Lemma 1(1) [9])

$$\sum_{N(p)\le x} \frac{1}{N(p)^{\alpha}} \ll \frac{x^{1-\alpha}}{\log x} \qquad \textit{if } 0 \le \alpha < 1.$$

Lemma 2 (Lemma 2 [9])

$$\sum_{N(p)\le x} \frac{1}{N(p)} = \log\log x + A + \mathrm{O}\Big(\frac{1}{\log x}\Big),$$

where A is a constant.

For $x \in X$, define

$$M(x) = \big\{m \in M, N(m) \le x\big\}.$$

Let

$$\mathrm{P}_x\big\{m : m\,\text{satisfies some conditions}\big\}$$

denote the quantity

$$\frac{1}{|M(x)|}\,\#\big\{m \in M(x),\, m\,\text{satisfies some conditions}\big\}.$$

Notice that P_x is a probability measure on M. Let f be a function from M to $\mathbb{R}$. The expectation of f with respect to P_x is denoted by

$$\mathrm{E}_x\Big\{m : f(m)\Big\} := \frac{1}{|M(x)|}\sum_{m\in M(x)} f(m).$$

The following lemmas are essential for the proof of Theorem 1. The first one gives an equivalent statement of Theorem 1.

Lemma 3

$$\lim_{x\to\infty} \mathrm{P}_x\Big\{m : \frac{\omega(m) - \log\log N(m)}{\sqrt{\log\log N(m)}} \le \gamma\Big\} = G(\gamma)$$

if and only if

$$\lim_{x\to\infty} \mathrm{P}_x\Big\{m : \frac{\omega(m) - \log\log x}{\sqrt{\log\log x}} \le \gamma\Big\} = G(\gamma).$$

Proof: Since

$$\frac{\omega(m) - \log\log x}{\sqrt{\log\log x}} = \frac{\omega(m) - \log\log N(m)}{\sqrt{\log\log N(m)}} \frac{\sqrt{\log\log N(m)}}{\sqrt{\log\log x}} + \frac{\log\log N(m) - \log\log x}{\sqrt{\log\log x}},$$

by Fact 2, to prove this lemma, it suffices to show that for any $\epsilon > 0$, we have

$$\lim_{x\to\infty} \mathrm{P}_x\Big\{m : \Big|\frac{\sqrt{\log\log N(m)}}{\sqrt{\log\log x}} - 1\Big| > \epsilon\Big\} = 0$$

and

$$\lim_{x\to\infty} \mathrm{P}_x\Big\{m : \Big|\frac{\log\log N(m) - \log\log x}{\sqrt{\log\log x}}\Big| > \epsilon\Big\} = 0.$$

Consider $m \in M$ with $x^{1/2} < N(m) \le x$. If we have

$$\frac{\sqrt{\log\log N(m)}}{\sqrt{\log\log x}} < 1 - \epsilon,$$

it follows that

$$(\log\log x - \log 2)^{1/2} < (\log\log N(m))^{1/2} < (1-\epsilon)(\log\log x)^{1/2}.$$

Taking square on both sides, we get

$$\frac{1}{(1-\epsilon)^2}(\log\log x - \log 2) < \log\log x.$$

It follows that

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

Similarly, for $m \in M$ with $x^{1/2} < N(m) \le x$, if we have

$$\frac{\log\log x - \log\log N(m)}{\sqrt{\log\log x}} > \epsilon,$$

it implies that

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

Hence, there exists $x(\epsilon) \in \mathbb{R}$ such that for all $x \ge x(\epsilon)$, we have

$$\mathrm{P}_x\Big\{m : \Big|\frac{\sqrt{\log\log N(m)}}{\sqrt{\log\log x}} - 1\Big| > \epsilon\Big\} \le \mathrm{P}_x\{m : N(m) \le x^{1/2}\}$$

and

$$\mathrm{P}_x\Big\{m : \Big|\frac{\log\log N(m) - \log\log x}{\sqrt{\log\log x}}\Big| > \epsilon\Big\} \le \mathrm{P}_x\{m : N(m) \le x^{1/2}\}.$$

Applying Condition (A), we have

$$\begin{aligned}
\mathrm{P}_x\big\{m : N(m) \le x^{1/2}\big\} &= \frac{1}{|M(x)|} \cdot |M(x^{1/2})| \\
&= \frac{\kappa x^{1/2} + \mathrm{O}(x^{\theta/2})}{\kappa x + \mathrm{O}(x^{\theta})} \\
&\longrightarrow 0,
\end{aligned}$$

as $x \to \infty$. Hence, we obtain the equivalence of the statements in the lemma.

For $x \in X$, define

$$y = x^{1/\log\log x}.$$

For $m \in M$, define

$$\omega_y(m) = \sum_{\substack{p \in P \\ n_p(m) \ge 1 \\ N(p) \le y}} 1.$$

It is a truncation function of $\omega(m)$. Notice that we have

$$y = \mathrm{o}\big(x^{\epsilon}\big) \quad \text{for any } \epsilon > 0.$$

By Lemma 2, we have

$$\sum_{y < N(p) \le x} \frac{1}{N(p)} \ll \log\log\log x = \mathrm{o}\big((\log\log x)^{1/2}\big).$$

We have another equivalent formulation of the Erdös-Kac Theorem in terms of ω_y.

Lemma 4

$$\lim_{x \to \infty} \mathrm{P}_x\Big\{m : \frac{\omega(m) - \log\log x}{\sqrt{\log\log x}} \le \gamma\Big\} = G(\gamma)$$

if and only if

$$\lim_{x \to \infty} \mathrm{P}_x\Big\{m : \frac{\omega_y(m) - \log\log x}{\sqrt{\log\log x}} \le \gamma\Big\} = G(\gamma).$$

Proof: Since

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

by Facts 1 and 2, if we have

$$\lim_{x \to \infty} \mathrm{E}_x\Big\{m : \Big|\frac{\omega(m) - \omega_y(m)}{\sqrt{\log\log x}}\Big|\Big\} = 0,$$

the lemma follows. Consider

$$\begin{aligned}\sum_{N(m)\le x} \bigl|\omega(m)-\omega_y(m)\bigr| &= \sum_{y<N(p)\le x} \sum_{\substack{N(m)\le x\\ n_p(m)\ge 1}} 1\\ &= \sum_{y<N(p)\le x} \biggl(\frac{\kappa x}{N(p)} + \mathrm{O}\Bigl(\frac{x^\theta}{N(p)^\theta}\Bigr)\biggr)\\ &= \mathrm{o}\bigl(\kappa x(\log\log x)^{1/2}\bigr) + \mathrm{O}\bigl(x\bigr).\end{aligned}$$

The last equality follows from the remark before Lemma 4 and Lemma 1. Hence, we have

$$\mathrm{E}_x\Bigl\{m : \Bigl|\frac{\omega(m)-\omega_y(m)}{\sqrt{\log\log x}}\Bigr|\Bigr\} = \frac{\mathrm{o}\bigl(x(\log\log x)^{1/2}\bigr)}{\bigl(\kappa x + \mathrm{O}(x^\theta)\bigr)\bigl(\log\log x\bigr)^{1/2}} \longrightarrow 0,$$

as $x \to \infty$. Thus Lemma 4 follows.

For $p \in P$, define the independent random variables X_p by

$$\mathrm{P}\{X_p = 1\} = \frac{1}{N(p)}$$

and

$$\mathrm{P}\{X_p = 0\} = 1 - \frac{1}{N(p)}.$$

Define a new random variable S_y by

$$S_y := \sum_{\substack{p\in P\\ N(p)\le y}} X_p.$$

By Lemma 2 and the choice of y, we have

$$\mathrm{E}\bigl\{S_y\bigr\} = \sum_{N(p)\le y} \frac{1}{N(p)} = \log\log x + \mathrm{o}\bigl(\log\log x\bigr)^{1/2},$$

$$\mathrm{Var}\bigl\{S_y\bigr\} = \sum_{N(p)\le y} \frac{1}{N(p)}\biggl(1 - \frac{1}{N(p)}\biggr) = \log\log x + \mathrm{o}\bigl(\log\log x\bigr)^{1/2}.$$

We have another equivalent formulation of Theorem 1.

Lemma 5

$$\lim_{x\to\infty} \mathrm{P}_x\Bigl\{m : \frac{\omega_y(m)-\log\log x}{\sqrt{\log\log x}} \le \gamma\Bigr\} = G(\gamma)$$

if and only if

$$\lim_{x\to\infty} \mathrm{P}_x\Bigl\{m : \frac{\omega_y(m)-\mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}} \le \gamma\Bigr\} = G(\gamma).$$

Proof: Write

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

Since

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

we have

$$\frac{\sqrt{\log\log x}}{\sqrt{\mathrm{Var}\{S_y\}}} \xrightarrow{p} 1.$$

Also, since

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

it follows that

$$\lim_{x\to\infty} \mathrm{E}_x\Big\{m : \Big|\frac{\mathrm{E}\{S_y\} - \log\log x}{\sqrt{\mathrm{Var}\{S_y\}}}\Big|\Big\} = 0.$$

By Facts 1 and 2, the lemma follows.

Now, for $p \in P$, define a random variable $\delta_p : M \to \mathbb{R}$ by

$$\delta_p(m) := \begin{cases} 1 & \text{if } n_p(m) \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, we can write

$$\omega_y(m) = \sum_{\substack{p\in P \\ N(p)\leq y}} \delta_p(m).$$

Notice that for a fixed $p \in P$ and $x \in X$, by Condition (A), we have

$$\begin{aligned} \mathrm{P}_x\big\{m : \delta_p(m) = 1\big\} &= \frac{1}{|M(x)|} \cdot \Big|M\Big(\frac{x}{N(p)}\Big)\Big| \\ &= \frac{1}{\kappa x + \mathrm{O}(x^\theta)} \Big(\frac{\kappa x}{N(p)} + \mathrm{O}\Big(\frac{x^\theta}{N(p)^\theta}\Big)\Big) \\ &= \frac{1}{N(p)} + \mathrm{O}\big(x^{\theta-1}\big). \end{aligned}$$

Since the expectations of random variables X_p and δ_p are close, the sum S_y is a good approximation of ω_y. Indeed, the r-th moments of their normalizations are equal as $x \to \infty$.

Lemma 6 *Let $r \in \mathbb{N}$. We have*

$$\lim_{x\to\infty} \Big|\mathrm{E}_x\Big\{\Big(\frac{\omega_y(m) - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}\Big)^r\Big\} - \mathrm{E}\Big\{\Big(\frac{S_y - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}\Big)^r\Big\}\Big| = 0.$$

Proof: For $0 \le k \le r$, write

$$\mathrm{E}\{S_y^k\} = \sum_{u=1}^{k} \sum{}' \frac{k!}{k_1! \cdots k_u!} \sum{}'' \mathrm{E}\{X_{p1}^{k_1} \cdots X_{p_u}^{k_u}\}.$$

Here $\sum'$ extends over all u-tuples $(k_1, k_2, \cdots, k_u)$ of positive integers such that $k_1 + k_2 + \cdots + k_u = k$ and $\sum''$ extends over all u-tuples $(p_1, p_2, \cdots, p_u)$ of elements P such that $N(p_i) \le y$ for all i and $p_i \ne p_j$ if $i \ne j$, regardless of their orders. Since each X_{p_i} takes values 0 or 1 and the X_{p_i}'s are independent, we have

$$\mathrm{E}\{X_{p_1} \cdots X_{p_u}\} = \frac{1}{N(p_1) \cdots N(p_u)}.$$

Similarly, we have

$$\mathrm{E}_x\{\omega_n^k\} = \sum_{u=1}^{k} \sum{}' \frac{k!}{k_1! \cdots k_u!} \sum{}'' \mathrm{E}_x\{\delta_{p1}^{k_1} \cdots \delta_{p_u}^{k_u}\},$$

with the same $\sum'$ and $\sum''$ as above. By Condition(A), we have

$$\begin{aligned}
\mathrm{E}_x\{\delta_{p1} \cdots \delta_{p_u}\} &= \frac{1}{|M(x)|} \cdot \left| M\Big(\frac{x}{N(p_1) \cdots N(p_u)}\Big) \right| \\
&= \frac{1}{\kappa x + \mathrm{O}(x)} \Big(\frac{\kappa x}{N(p_1) \cdots N(p_u)} + \mathrm{O}\Big(\frac{x^\theta}{N(p_1)^\theta \cdots N(p_u)^\theta} \Big) \Big) \\
&= \frac{1}{N(p_1) \cdots N(p_u)} + \mathrm{O}\big(x^{\theta-1}\big).
\end{aligned}$$

Hence, we have

$$\Big| \mathrm{E}_x\{\omega_y^k\} - \mathrm{E}\{S_y^k\} \Big| \ll x^{\theta-1} \Big(\sum_{N(p) \le y} 1 \Big)^k \le y^k \cdot x^{\theta-1}.$$

Write

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

and

$$\mathrm{E}_x\big\{\big(\omega_y - \mathrm{E}\{S_y\}\big)^r\big\} = \sum_{k=0}^{r} \binom{r}{k} \mathrm{E}_x\{\omega_y^k\} \cdot \mathrm{E}\{S_y\}^{r-k}.$$

Their difference is

$$\begin{aligned}
\Big| \mathrm{E}_x\big\{\big(\omega_y - \mathrm{E}\{S_y\}\big)^r\big\} - \mathrm{E}\big\{\big(S_y - \mathrm{E}\{S_y\}\big)^r\big\} \Big| &\ll \sum_{k=0}^{r} \binom{r}{k} y^k \cdot x^{\theta-1} \cdot \mathrm{E}\{S_y\}^{r-k} \\
&= x^{\theta-1} \big(y + \mathrm{E}\{S_y\}\big)^r.
\end{aligned}$$

Notice that

$$\mathrm{E}\big\{S_y\big\} = \sum_{N(p)\le y} \frac{1}{N(p)} \le \sum_{N(m)\le y} 1 \ll y.$$

Since for any $\epsilon > 0$.

$$y = \mathrm{o}\big(x^\epsilon\big),$$

we have

$$\Big|\mathrm{E}_x\big\{(\omega_y - \mathrm{E}\big\{S_y\big\})^r\big\} - \mathrm{E}\big\{\big(S_y - \mathrm{E}\big\{S_y\big\}\big)^r\big\}\Big| \longrightarrow 0,$$

as $x \to \infty$. Thus the lemma holds.

The following lemma is about the r-th moment of S_y.

Lemma 7 *For $r \in \mathbb{N}$,*

$$\sup_x \Big|\mathrm{E}\Big\{\Big(\frac{S_y - \mathrm{E}\big\{S_y\big\}}{\sqrt{\mathrm{Var}\big\{S_y\big\}}}\Big)^r\Big\}\Big| < \infty.$$

Proof: Define $Y_p = X_p - \frac{1}{N(p)}$. We have

$$\mathrm{E}\big\{\big(S_y - \mathrm{E}\big\{S_y\big\}\big)^r\big\} = \sum_{u=1}^{r} \sum{}' \frac{r!}{r_1! \cdots r_u!} \sum{}'' \mathrm{E}\{Y_{p1}^{r_1} \cdots Y_{p_u}^{r_u}\},$$

where $\sum'$ and $\sum''$ are defined as in Lemma 6 except replacing k by r. Since $\mathrm{E}\big\{Y_p\big\} = 0$, without loss of generality, we can assume $r_i \ge 2$. Since $|Y_p| \le 1$ and $r_i \le 2$, we have

$$\big|\mathrm{E}\big\{Y_{p_i}^{r_i}\big\}\big| \le \mathrm{E}\big\{Y_{p_i}^2\big\}.$$

Hence, we have

$$\begin{aligned}
\mathrm{E}\big\{\big(S_y - \mathrm{E}\big\{S_y\big\}\big)^r\big\} &\le \sum_{u=1}^{r} \sum{}' \frac{r!}{r_1! \cdots r_u!} \sum{}'' \mathrm{E}\big\{Y_{p_1}^2 \cdots Y_{p_u}^2\big\} \\
&\le \sum_{u=1}^{r} \sum{}' \frac{r!}{r_1! \cdots r_u!} \Big(\sum_{N(p)\le y} \mathrm{E}\big\{Y_p^2\big\}\Big)^u \\
&\le \sum_{u=1}^{r} \sum{}' \frac{r!}{r_1! \cdots r_u!}\ \mathrm{Var}\big\{S_y\big\}^u \\
&\le \sum_{u=1}^{r} \sum{}' \frac{r!}{r_1! \cdots r_u!}\ \mathrm{Var}\big\{S_y\big\}^{r/2}.
\end{aligned}$$

The last inequality holds because $2u \le r$. Hence, we obtain

$$\mathrm{E}\Big\{\Big(\frac{S_y - \mathrm{E}\big\{S_y\big\}}{\sqrt{\mathrm{Var}\big\{S_y\big\}}}\Big)^r\Big\} \le \sum_{u=1}^{r} \sum{}' \frac{r!}{r_1! \cdots r_u!} < \infty.$$

Thus Lemma 7 follows.

4 Proof of Theorem 1.

We are now equipped to embark on the proof of Theorem 1. Given P, M, and X as before, assume they satisfy Conditions (A) and (B). For $m \in M$, we shall show that the quantity

$$\frac{\omega(m) - \log\log N(m)}{\sqrt{\log\log N(m)}}$$

distributes normally. By the equivalent statements of Lemmas 3, 4, and 5, to prove Theorem 1, it suffices to show

$$\lim_{x\to\infty} \mathrm{P}_x\Big\{m : \frac{\omega_y(m) - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}} \le \gamma\Big\} = G(\gamma).$$

The distribution function F_x respect to P_x is defined by

$$F_x(\gamma) := \mathrm{P}_x\Big\{m : \frac{\omega_y(m) - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}} \le \gamma\Big\}.$$

Notice that the r-th moment of F_x is equal to

$$\begin{aligned}
&\int_{-\infty}^{\infty} t^r dF_x(t)\\
&= \sum_{t=-\infty}^{\infty} \left\{ \lim_{u\to\infty} \sum_{i=1}^{u} (t+i/u)^r \Big(F_x(t+i/u) - F_x(t+(i-1)/u)\Big)\right\}\\
&= \sum_{t=-\infty}^{\infty} \left\{ \lim_{u\to\infty} \sum_{i=1}^{u} (t+i/u)^r \mathrm{P}_x\Big\{m : (t+(i-1)/u) < \Big(\frac{\omega_y(m) - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}\Big) \le (t+i/u)\Big\}\right\}\\
&= \frac{1}{\#\{m : N(m) \le x\}} \sum_{N(m)\le x} \Big(\frac{\omega_y(m) - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}\Big)^r\\
&= \mathrm{E}_x\Big\{\Big(\frac{\omega_y(m) - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}\Big)^r\Big\}.
\end{aligned}$$

Hence, to prove

$$\lim_{x\to\infty} F_x(\gamma) = G(\gamma),$$

by Fact 3, it suffices to show that for all $r \in \mathbb{N}$,

$$\lim_{x\to\infty} \mathrm{E}_x\Big\{\Big(\frac{\omega_y(m) - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}\Big)^r\Big\} = \mu_r.$$

By Lemma 6, we see that the last equality holds if

$$\lim_{x\to\infty} \mathrm{E}\Big\{\Big(\frac{S_y - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}\Big)^r\Big\} = \mu_r.$$

Define a new random variable $G_y = G_{y(x)}$ on M by

$$G_y := \frac{S_y - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}.$$

Applying Fact 5, the Central Limit Theorem implies that

$$\lim_{x\to\infty} G_y = G.$$

Also, Lemma 7 implies that for each $r \in \mathbb{N}$, there exists $\delta = \delta(r) > 0$ such that

$$\sup_x \int_{-\infty}^{\infty} |t|^{r+\delta} dG_y(t) < \infty.$$

By Fact 4, we have

$$\lim_{x\to\infty} \mathrm{E}\Big\{\Big(\frac{S_y - \mathrm{E}\{S_y\}}{\sqrt{\mathrm{Var}\{S_y\}}}\Big)^r\Big\} = \mu_r;$$

thus

$$\lim_{x\to\infty} F_x(\gamma) = G(\gamma)$$

follows. Hence, we obtain Theorem 1, i.e., a generalization of the Erdös-Kac Theorem holds in this general setting.

Remark For $m \in M$, we define

$$\Omega(m) = \sum_{\substack{p\in P \\ n_p(m)\geq 1}} n_p(m),$$

the number of generators of m, counted with multiplicity. Applying the same method as in the classical case, we can also obtain generalizations of the Turán Theorem and the Erdös-Kac Theorem for $\Omega(m)$ in our general setting.

Conclusion The Erdös-Kac Theorem is a refinement of the Turán Theorem. When we compare these two, we naturally think that the latter is 'more difficult' than the former. However, when we put these two theorems in a general context, they both require only Conditions (A) and (B). Thus we conclude that these two results are of 'the same difficulty'.

Acknowledgement This paper is part of my Ph.D. thesis at Harvard. I would like to thank my thesis advisor, Prof. B. Mazur, for many important suggestions about this work. He has inspired me in so many ways and it is a great privilege to be his student. I also would like thank Prof. R. Murty for introducing to me probabilistic number theory and for being always on my side. His encouragement and unlimited support are the main sources for me to complete my Ph.D. work. Finally, I would like to thank M. Agarwal, S. Mohit, C.Y. Kao and W. Kuo for their comments about this paper. Especially, I owe my thanks to Satya for many useful discussions related to this work.

References

[1] P. Billingsley, *On the central limit theorem for the prime divisor functions*, Amer. Math. Monthly, 76 (1969), 132-139.

[2] P. Erdös & M. Kac, *The Gaussian law of errors in the theory of additive number theoretic functions*, Amer. J. Math., 62 (1940), 738-742.

[3] W. Feller, *An introduction to probability theory and its applications*, Vol. II, Wiley (1966).

[4] H. Halberstam, *On the distribution of additive number theoretic functions*, I., II., & III., J. London Math. Soc., 30 (1955), 43-53; 31 (1956), 1-14, 15-27.

[5] G.H. Hardy & S. Ramanujan, *The normal number of prime factors of a number n*, Quar. J. Pure. Appl. Math., 48 (1917), 76-97.

[6] J. Knopfmacher, *Analytic arithmetic of algebraic function fields*, Lecture Notes in Pure and Applied Math., 50 (1979).

[7] E. Landau, *Neuer Beweis des Primzahlsatzes und Beweis des Primidealsatzes*, Math. Annal, 56 (1903), 645-670.

[8] S. Lang & A. Weil, *Number of points of varieties in finite fields*, Am. J. of Math. 76 (1954), 819-827.

[9] Y.-R. Liu, *A generalization of the Turán Theorem and its applications*, to appear in Canadian Math. Bulletin.

[10] D. Lorenzini, *An invitation to arithmetic geometry*, Grad. Stud. in Math. Vol. 9, AMS (1996).

[11] M.R. Murty, *Problems in analytic number theory*, Springer Verlag (2001).

[12] P. Turán, *On a theorem of Hardy and Ramanujan*, J. London Math. Soc. 9 (1934), 274-276.

[13] H. Weber, *Über Zahlengruppen in algebraischen Körpern*, Math. Ann., Vol 49 (1897), 83-100.

[14] W.-B. Zhang, *probabilistic number theory in additive arithmetic semigroup I*, Analytic number theory, Vol 2 (1995), 839-885.

Departments of Mathematics, Harvard University, Cambridge, MA, USA 02138

Email: `yrliu@math.harvard.edu`