THE ERDŐS-KAC THEOREM

STEVE FAN

Abstract. The celebrated Erdős-Kac theorem states that if $\omega(n)$ denotes the number of distinct prime divisors of a positive integer n, then the distribution of

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

is the standard normal distribution with mean 0 and variance 1. The main objective of this expository note is to present a proof of this result by estimating the moments of the above quantity. Our exposition follows basically the argument of Granville and Soundararajan [5].

1. Introduction

Throughout the paper, let $\mathbb{R}$ denote the field of real numbers, $\mathbb{C}$ the field of complex numbers, $\mathbb{N}_+$ the set of positive integers, and $\mathbb{P}$ the set of prime numbers. For any $x\in\mathbb{R}$, let $\lfloor x\rfloor$ denote the integer part of x. We shall also reserve the letters p and q for primes and denote by $\pi(x)$ the number of primes up to x. For every $n\in\mathbb{N}_+$, let $\omega(n)$ denote the number of distinct prime factors of n. Then we have by a famous theorem of Mertens [7, Theorem 427] that

$$\sum_{n\le x}\omega(n)=\sum_{n\le x}\sum_{p\mid n}1=\sum_{p\le x}\left\lfloor\frac{x}{p}\right\rfloor=x\sum_{p\le x}\frac{1}{p}+O(x)=x\log\log x+O(x).$$

Roughly speaking, this means that $\omega(n)$ is $\log\log x$ on average for $n\le x$. A more precise result obtained by Hardy and Ramanujan [6] states that the normal order of $\omega(n)$ is $\log\log n$. This means that given any $\varepsilon>0$, there exists $x_\varepsilon\ge 2$ such that for all $x\ge x_\varepsilon$, the inequality

$$|\omega(n)-\log\log n|\ge\varepsilon\log\log n$$

holds for at most εx integers $n\le x$. Turán [11] showed that

$$\sum_{n\le x}(\omega(n)-\log\log n)^2=(1+o(1))x\log\log x \tag{1.1}$$

for sufficiently large x. Turán's result immediately implies that the normal order of $\omega(n)$ is $\log\log n$, since $\#\{\sqrt{x}<n\le x\colon |\omega(n)-\log\log n|\ge\varepsilon\log\log n\}$ is bounded above by

$$\begin{aligned}\frac{1}{(\varepsilon\log\log n)^2}\sum_{n\le x}(\omega(n)-\log\log n)^2&\le\frac{(1+o(1))x\log\log x}{\varepsilon^2(\log\log x-\log\log 2)^2}\\&=(1/\varepsilon^2+o(1))\frac{x}{\log\log x}\\&\ll\varepsilon x\end{aligned}$$

for all sufficiently large x. In 1939 Mark Kac gave a lecture at Princeton on the average number of prime factors of a random integer. At the end of his lecture, he described his heuristic for the distribution of $\omega(n)$. He suggested that the distribution of $\omega(n)$ is perhaps normal but had difficulty verifying this. Paul Erdős, who was in the audience and, as Kac recalled (see [8]), half-dozed through most of the lecture, interrupted and announced that he found the solution. This led to the collaboration between Kac and Erdős, who showed that

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

behaves like a normal random variable with mean 0 and variance 1. More precisely, they [4] proved that

$$\lim_{x\to+\infty} \frac{1}{x} \cdot \# \left\{ n \le x \colon \frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \le a \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2}\, dt.$$

It is not hard to see that the above equality holds with $\log\log n$ replaced by $\log\log x$. In [4] they proved a stronger result concerning strongly additive functions (see Theorem 6.1 below). Their original proof uses the central limit theorem and Brun's method from sieve theory. The main objective of this expository note is to present a proof of this result following the argument of Granville and Soundararajan [5]. Our writing is inspired by a lecture on this topic delivered by Lester and Rudnick [9] in 2015 at the University of Montreal.

2. The Second Moment of $\omega(n)$

Turán's result (1.1) tells us that $|\omega(n) - \log\log n|$ is roughly of size $\log\log n$. As a warm-up, we shall prove a refinement of this result. The proof is an adaptation of that given in [11]. We first prove the following lemma.

Lemma 2.1. *For all $x \ge e$ one has*

$$\sum_{p\le x} \frac{(\log p)^k}{p} = \frac{1}{k}(\log x)^k + O((\log x)^{k-1})$$

for all $k \in \mathbb{N}_+$, where the implicit constant in the error term is independent of k.

Proof. A standard result [7, Theorem 425] in prime number theory states that

$$\sum_{p\le x} \frac{\log p}{p} = \log x + O(1).$$

By partial summation we obtain

$$\begin{aligned}\sum_{p\le x} \frac{(\log p)^k}{p} &= (\log x + O(1))(\log x)^{k-1} - (k-1)\int_1^x (\log t + O(1))\frac{(\log t)^{k-2}}{t}\, dt \\ &= \frac{1}{k}(\log x)^k + O((\log x)^{k-1})\end{aligned}$$

for all $k \ge 2$. □

Next, we show

$$\sum_{n=1}^{\infty} \frac{1}{2^n n^2} = \frac{\pi^2}{12} - \frac{(\log 2)^2}{2}.$$

This is a special case of the following lemma [1, Formula 27.7.3, §27.7].

Lemma 2.2. *For every $z \in \mathbb{C}$ with $|z| \leq 1$, define*

$$f(z) := \sum_{n=1}^{\infty} \frac{z^n}{n^2}.$$

Then

$$f(z) + f(1-z) = \frac{\pi^2}{6} - \log z \log(1-z)$$

for all $z \in \mathbb{C}$ with $|z| \leq 1$ and $|1-z| \leq 1$, where log *is the principle branch of the natural logarithm. In particular, we have*

$$\sum_{n=1}^{\infty} \frac{1}{2^n n^2} = \frac{\pi^2}{12} - \frac{(\log 2)^2}{2}.$$

Proof. It is clear that $f(z)$ is continuous on the closed unit disk $|z| \leq 1$ and analytic in the open unit disk $|z| < 1$. Since

$$f'(z) = \sum_{n=1}^{\infty} \frac{z^{n-1}}{n} = -\frac{\log(1-z)}{z},$$

we have

$$f'(z) + \frac{d}{dz}(f(1-z)) = -\frac{\log(1-z)}{z} + \frac{\log z}{1-z} = -\frac{d}{dz}(\log z \log(1-z)).$$

Integrating both sides with respect to z we obtain

$$f(z) + f(1-z) = c - \log z \log(1-z),$$

where $c \in \mathbb{C}$ is some constant. Letting $z \to 0$ and using the fact that $f(1) = \pi^2/6$, we find $c = \pi^2/6$. The second part of the lemma follows on taking $z = 1/2$. □

Now we prove the following refinement of Turán's result (1.1). The proof presented below was found by the author himself, though it may be far from new.

Theorem 2.3. *For sufficiently large x one has*

$$\sum_{n \leq x} (\omega(n) - \log\log x)^2 = x \log\log x + Cx + O\left(\frac{x \log\log x}{\log x}\right),$$

where

$$C = b(2b - 2\log 2 + 1) - \sum_p \frac{1}{p^2} - \left(\frac{\pi^2}{6} - (\log 2)^2\right),$$

$$b = \gamma + \sum_p \left[\log\left(1 - \frac{1}{p}\right) + \frac{1}{p}\right],$$

and $\gamma = 0.5772...$ is Euler's constant.

Proof. Note that

$$\sum_{n \leq x} (\omega(n) - \log\log x)^2 = \sum_{n \leq x} \omega(n)^2 - 2\log\log x \sum_{n \leq x} \omega(n) + \lfloor x \rfloor (\log\log x)^2.$$

Mertens' theorem [7, Theorems 427, 428] says that

$$\sum_{p\le x}\frac{1}{p}=\log\log x+b+O\left(\frac{1}{\log x}\right).$$

Since the number $\pi(x)$ of primes up to x is $O(x/\log x)$ by Chebyshev's estimate [7, Theorem 7], we have

$$\sum_{n\le x}\omega(n)=\sum_{n\le x}\sum_{p|n}1=\sum_{p\le x}\left\lfloor\frac{x}{p}\right\rfloor=x\sum_{p\le x}\frac{1}{p}+O\left(\frac{x}{\log x}\right)=x\log\log x+bx+O\left(\frac{x}{\log x}\right).$$

It follows that

$$\sum_{n\le x}(\omega(n)-\log\log x)^2=\sum_{n\le x}\omega(n)^2-x(\log\log x)^2-2bx\log\log x+O\left(\frac{x\log\log x}{\log x}\right).$$

To prove the theorem, it suffices to show

$$\sum_{n\le x}\omega(n)^2=x(\log\log x)^2+(2b+1)x\log\log x+Cx+O\left(\frac{x\log\log x}{\log x}\right). \tag{2.1}$$

Now we compute

$$\sum_{n\le x}\omega(n)^2=\sum_{n\le x}\sum_{p|n,q|n}1=\sum_{p,q\le x}\sum_{\substack{n\le x\\ p|n,q|n}}1=\sum_{p\le x}\left\lfloor\frac{x}{p}\right\rfloor+\sum_{\substack{pq\le x\\ p\ne q}}\left\lfloor\frac{x}{pq}\right\rfloor.$$

By Chebyshev's estimate [7, Theorem 7], we have

$$\sum_{\substack{pq\le x\\ p\ne q}}1=2\sum_{p<\sqrt{x}}\sum_{p<q\le x/p}1\ll x\sum_{p<\sqrt{x}}\frac{1}{p\log(x/p)}\ll\frac{x}{\log x}\sum_{p<\sqrt{x}}\frac{1}{p}\ll\frac{x\log\log x}{\log x}.$$

It follows that

$$\begin{aligned}\sum_{\substack{pq\le x\\ p\ne q}}\left\lfloor\frac{x}{pq}\right\rfloor&=x\sum_{\substack{pq\le x\\ p\ne q}}\frac{1}{pq}+O\left(\frac{x\log\log x}{\log x}\right)\\&=x\sum_{pq\le x}\frac{1}{pq}-x\sum_{p\le\sqrt{x}}\frac{1}{p^2}+O\left(\frac{x\log\log x}{\log x}\right)\\&=x\sum_{pq\le x}\frac{1}{pq}-x\sum_{p}\frac{1}{p^2}+O\left(\frac{x\log\log x}{\log x}\right).\end{aligned}$$

Thus we obtain

$$\sum_{n\le x}\omega(n)^2=x\log\log x+bx+x\sum_{pq\le x}\frac{1}{pq}-x\sum_{p}\frac{1}{p^2}+O\left(\frac{x\log\log x}{\log x}\right).$$

To prove (2.1), it is hence sufficient to show

$$\sum_{pq\le x}\frac{1}{pq}=(\log\log x)^2+2b\log\log x+2b(b-\log 2)-\left(\frac{\pi^2}{6}-(\log 2)^2\right)+O\left(\frac{\log\log x}{\log x}\right). \tag{2.2}$$

We now write

$$\sum_{pq\le x}\frac{1}{pq}=2\sum_{p\le\sqrt{x}}\frac{1}{p}\sum_{q\le x/p}\frac{1}{q}-\sum_{p,q\le\sqrt{x}}\frac{1}{pq}. \tag{2.3}$$

It is easy to see that

$$\sum_{p,q\le\sqrt{x}}\frac{1}{pq}=(\log\log x)^2+2(b-\log 2)\log\log x+O\left(\frac{\log\log x}{\log x}\right). \tag{2.4}$$

Note that

$$\sum_{p\le\sqrt{x}}\frac{1}{p}\sum_{q\le x/p}\frac{1}{q}=\sum_{p\le\sqrt{x}}\frac{1}{p}\left(\log\log\frac{x}{p}+b+O\left(\frac{1}{\log(x/p)}\right)\right). \tag{2.5}$$

Clearly, we have

$$b\sum_{p\le\sqrt{x}}\frac{1}{p}=b\log\log x+b(b-\log 2)+O\left(\frac{1}{\log x}\right) \tag{2.6}$$

and

$$\sum_{p\le\sqrt{x}}\frac{1}{p\log(x/p)}\le\frac{2}{\log x}\sum_{p\le\sqrt{x}}\frac{1}{p}\ll\frac{\log\log x}{\log x}. \tag{2.7}$$

Finally, we see that

$$\sum_{p\le\sqrt{x}}\frac{1}{p}\log\log\frac{x}{p}=\log\log x\sum_{p\le\sqrt{x}}\frac{1}{p}+\sum_{p\le\sqrt{x}}\frac{1}{p}\log\left(1-\frac{\log p}{\log x}\right) \tag{2.8}$$

and that

$$\log\log x\sum_{p\le\sqrt{x}}\frac{1}{p}=(\log\log x)^2+(b-\log 2)\log\log x+O\left(\frac{\log\log x}{\log x}\right). \tag{2.9}$$

From Lemmas 2.1 and 2.2 it follows that

$$\begin{aligned}\sum_{p\le\sqrt{x}}\frac{1}{p}\log\left(1-\frac{\log p}{\log x}\right)&=-\sum_{k=1}^{\infty}\frac{1}{k(\log x)^k}\sum_{p\le\sqrt{x}}\frac{(\log p)^k}{p}\\&=-\sum_{k=1}^{\infty}\frac{1}{2^k k^2}+O\left(\frac{1}{\log x}\sum_{k=1}^{\infty}\frac{1}{2^{k-1}k}\right)\\&=-\left(\frac{\pi^2}{12}-\frac{(\log 2)^2}{2}\right)+O\left(\frac{1}{\log x}\right).\end{aligned}$$

Combining this estimate with (2.3)–(2.9) gives (2.2). □

3. The Turán-Kubilius Inequality

An arithmetic function $f\colon \mathbb{N}_+ \to \mathbb{C}$ is called additive if $f(mn) = f(m) + f(n)$ for all $m, n \in \mathbb{N}_+$ with $\gcd(m,n) = 1$. Thus $\omega(n)$ is an additive function. In this section, we prove the following inequality of Turán and Kubilius concerning the second moment of $f(n)$. It provides a natural extension of the second moment of $\omega(n)$ discussed in the preceding section. Here we do not pursue the best version of this inequality that is currently known nor its applications to special additive functions. The interested reader is referred to [10, Chapter III.6] for detailed discussions on this topic.

Theorem 3.1 (Turán-Kubilius inequality)**.** *Let $f\colon \mathbb{N}_+ \to \mathbb{C}$ be an additive function. Then*

$$\frac{1}{x}\sum_{n\le x}|f(n)-A_f(x)|^2 \ll B_f(x), \tag{3.1}$$

where

$$A_f(x) = \sum_{p^k\le x}\frac{f(p^k)}{p^k}\left(1-\frac{1}{p}\right),$$
$$B_f(x) = \sum_{p^k\le x}\frac{|f(p^k)|^2}{p^k}.$$

Proof. Let $f^*\colon \mathbb{N}_+ \to \mathbb{C}$ be the additive function defined by

$$f^*(p^k) := \begin{cases} f(p^k) & \text{if } p^k \le \sqrt{x}, \\ 0 & \text{otherwise.}\end{cases}$$

We first prove (3.1) for f^*. To this end, let $\mathbf{n}$ be a random variable chosen uniformly from $\mathbb{N}_+ \cap [1,x]$. For each prime power p^k, define $\mathbf{1}_{p^k}\colon \mathbb{N}_+ \to \{0,1\}$ by

$$\mathbf{1}_{p^k}(n) := \begin{cases} 1 & \text{if } p^k \parallel n, \\ 0 & \text{otherwise.}\end{cases}$$

Then

$$f^*(\mathbf{n}) = \sum_{p^k\le x} f^*(p^k)\mathbf{1}_{p^k}(\mathbf{n}).$$

It is not hard to see that

$$\mathbb{E}[\mathbf{1}_{p^k}(\mathbf{n})] = \frac{\lfloor x/p^k\rfloor - \lfloor x/p^{k+1}\rfloor}{\lfloor x\rfloor} = \frac{x/p^k(1-1/p)+O(1)}{x+O(1)} = \frac{1}{p^k}\left(1-\frac{1}{p}\right) + O\left(x^{-1}\right)$$

and

$$\mathrm{Var}[\mathbf{1}_{p^k}(\mathbf{n})] = \mathbb{E}[\mathbf{1}_{p^k}(\mathbf{n})^2] - \mathbb{E}[\mathbf{1}_{p^k}(\mathbf{n})]^2 \ll \frac{1}{p^k}$$

for all $p^k \le x$. Moreover, if p, q are distinct primes, then

$$\begin{aligned}\mathbb{E}[\mathbf{1}_{p^k}(\mathbf{n})\mathbf{1}_{q^l}(\mathbf{n})] &= \frac{\lfloor x/(p^kq^l)\rfloor - \lfloor x/(p^{k+1}q^l)\rfloor - \lfloor x/(p^kq^{l+1})\rfloor + \lfloor x/(p^{k+1}q^{l+1})\rfloor}{\lfloor x\rfloor} \\ &= \frac{1}{p^kq^l}\left(1-\frac{1}{p}\right)\left(1-\frac{1}{p}\right) + O\left(x^{-1}\right).\end{aligned}$$

Hence

$$\operatorname{Cov}(\mathbf{1}_{p^k}(\mathbf{n}),\mathbf{1}_{q^l}(\mathbf{n})) = \mathbb{E}[\mathbf{1}_{p^k}(\mathbf{n})\mathbf{1}_{q^l}(\mathbf{n})] - \mathbb{E}[\mathbf{1}_{p^k}(\mathbf{n})]\mathbb{E}[\mathbf{1}_{q^l}(\mathbf{n})] = O\left(x^{-1}\right).$$

It follows that

$$\mathbb{E}[f^*(\mathbf{n})] = A_{f^*}(x) + O\left(\frac{1}{x}\sum_{p^k\le\sqrt{x}}|f^*(p^k)|\right)$$

and

$$\begin{aligned}\operatorname{Var}[f^*(\mathbf{n})] &= \sum_{p^k\le x}|f^*(p^k)|^2\operatorname{Var}[\mathbf{1}_{p^k}(\mathbf{n})] + 2\sum_{\substack{p^k,q^l\le x\\ p\neq q}} f^*(p^k)\overline{f^*(q^l)}\operatorname{Cov}(\mathbf{1}_{p^k}(\mathbf{n}),\mathbf{1}_{q^l}(\mathbf{n}))\\ &\ll B_{f^*}(x) + \frac{1}{x}\left(\sum_{p^k\le\sqrt{x}}|f^*(p^k)|\right)^2.\end{aligned}$$

Note that

$$\begin{aligned}\mathbb{E}[|f^*(\mathbf{n}) - A_{f^*}(x)|^2] &\le 2\operatorname{Var}[f^*(\mathbf{n})] + 2\mathbb{E}\left[|\mathbb{E}[f^*(\mathbf{n})] - A_{f^*}(x)|^2\right]\\ &\ll B_{f^*}(x) + \frac{1}{x}\left(\sum_{p^k\le\sqrt{x}}|f^*(p^k)|\right)^2,\end{aligned}$$

where we have used the arithmetic mean-quadratic mean inequality

$$\left|\sum_{i=1}^n a_i\right|^2 \le n\sum_{i=1}^n |a_i|^2 \tag{3.2}$$

for all $a_1,...,a_n\in\mathbb{C}$. Thus we have

$$\frac{1}{x}\sum_{n\le x}|f^*(n) - A_{f^*}(x)|^2 \ll B_{f^*}(x) + \frac{1}{x}\left(\sum_{p^k\le\sqrt{x}}|f^*(p^k)|\right)^2.$$

By (3.2) we have

$$\left(\sum_{p^k\le\sqrt{x}}|f^*(p^k)|\right)^2 \le \sqrt{x}\sum_{p^k\le\sqrt{x}}|f^*(p^k)|^2 \le xB_{f^*}(x).$$

It follows that

$$\frac{1}{x}\sum_{n\le x}|f^*(n) - A_{f^*}(x)|^2 \ll B_{f^*}(x). \tag{3.3}$$

Now we establish (3.1) for f. Suppose that $x\ge 9$. Since (3.2) gives

$$|f(n) - A_f(x)|^2 \le 3\left(|f(n) - f^*(n)|^2 + |f^*(n) - A_{f^*}(x)|^2 + |A_{f^*}(x) - A_f(x)|^2\right),$$

we have

$$\frac{1}{x}\sum_{n\le x}|f(n) - A_f(x)|^2 \ll \frac{1}{x}\sum_{n\le x}|f(n) - f^*(n)|^2 + B_{f^*}(x) + |A_{f^*}(x) - A_f(x)|^2$$

by (3.3). For every positive integer n, denote by $P(n)$ the largest prime power dividing n with the convention that $P(1)=1$. Note that for all $n \le x$ we have

$$f(n)-f^*(n)=\sum_{\substack{p^k>\sqrt{x}\\ p^k\|n}} f(p^k)=\begin{cases} f(P(n)) & \text{if } P(n)>\sqrt{x},\\ 0 & \text{otherwise.}\end{cases}$$

Hence

$$\begin{aligned}\frac{1}{x}\sum_{n\le x}|f(n)-f^*(n)|^2 &= \frac{1}{x}\sum_{\substack{n\le x\\ P(n)>\sqrt{x}}}|f(P(n))|^2\\ &= \frac{1}{x}\sum_{\sqrt{x}<p^k\le x}|f(p^k)|^2\sum_{\substack{n\le x\\ p^k\|n}}1\\ &= \frac{1}{x}\sum_{\sqrt{x}<p^k\le x}|f(p^k)|^2\left(\left\lfloor\frac{x}{p^k}\right\rfloor-\left\lfloor\frac{x}{p^{k+1}}\right\rfloor\right)\\ &\ll \sum_{\sqrt{x}<p^k\le x}\frac{|f(p^k)|^2}{p^k}.\end{aligned}$$

Finally, we have

$$\begin{aligned}|A_{f^*}(x)-A_f(x)|^2 &= \left|\sum_{\sqrt{x}<p^k\le x}\frac{f(p^k)}{p^k}\left(1-\frac{1}{p}\right)\right|^2\\ &\le \left(\sum_{\sqrt{x}<p^k\le x}\frac{|f(p^k)|^2}{p^k}\right)\left(\sum_{\sqrt{x}<p^k\le x}\frac{1}{p^k}\left(1-\frac{1}{p}\right)^2\right)\\ &\le \left(\sum_{\sqrt{x}<p^k\le x}\frac{|f(p^k)|^2}{p^k}\right)\left(\sum_{\sqrt{x}<p^k\le x}\frac{1}{p^k}\right)\end{aligned}$$

by Cauchy-Schwarz inequality. From Mertens' theorem [7, Theorem 427] it follows that

$$\sum_{\sqrt{x}<p^k\le x}\frac{1}{p^k}=\sum_{\sqrt{x}<p\le x}\frac{1}{p}+O(1)=\log\log x-\log\log\sqrt{x}+O(1)=O(1).$$

Hence

$$|A_{f^*}(x)-A_f(x)|^2 \ll \sum_{\sqrt{x}<p^k\le x}\frac{|f(p^k)|^2}{p^k}.$$

Collecting the above estimates we conclude that

$$\frac{1}{x}\sum_{n\le x}|f(n)-A_f(x)|^2 \ll B_{f^*}(x)+\sum_{\sqrt{x}<p^k\le x}\frac{|f(p^k)|^2}{p^k}\ll B_f(x).$$

This completes the proof. □

4. The Erdős-Kac Theorem: Preliminary Lemmas

The Turán-Kublius inequality provides an upper bound for the second moment of a general additive function $f(n)$. One may wonder about the higher moments of $f(n)$ and the distribution of its values. The Erdős-Kac theorem asserts that under certain conditions, the limit distribution of the normalized values of $f(n)$ is a Gaussian distribution. In this section, we collect some preliminary results needed for the proof of the Erdős-Kac theorem. In what follows, we shall denote by $\tau(n)$ the number of positive divisors of $n \in \mathbb{N}_+$. So if $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ is the prime factorization of n. then

$$\tau(n) = \prod_{i=1}^{r} (1 + \alpha_i).$$

Let $\varphi \colon \mathbb{N}_+ \to \mathbb{N}_+$ denote Euler's totient function, which counts the number of positive integers $a \leq n$ such that $\gcd(a, n) = 1$. Then we have [7, Theorem 62]

$$\varphi(n) = n \prod_{i=1}^{r} \left(1 - \frac{1}{p_i}\right) \tag{4.1}$$

for $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$. Let $\mu \colon \mathbb{N}_+ \to \mathbb{Z}$ denote the Möbius function defined by

$$\mu(n) := \begin{cases} (-1)^r & \text{if } n \text{ is a product of } r \geq 0 \text{ distinct primes,} \\ 0 & \text{otherwise.} \end{cases}$$

It is well known that μ satisfies the following identity [7, Theorem 263]

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, μ is multiplicative. Hence we can rewrite (4.1) as

$$\varphi(n) = n \sum_{d|n} \frac{\mu(d)}{d}.$$

Our first lemma is the following well-known result from elementary number theory.

Lemma 4.1. *Let $m \in \mathbb{N}_+$. Then*

$$\sum_{\substack{n \leq x \\ \gcd(n,m)=1}} 1 = \frac{\varphi(m)}{m} x + O(\tau(m)).$$

Proof. We compute

$$\begin{aligned}\sum_{\substack{n\le x\\ \gcd(n,m)=1}} 1 &= \sum_{n\le x}\sum_{d\mid \gcd(n,m)} \mu(d)\\ &= \sum_{d\mid m}\mu(d)\sum_{\substack{n\le x\\ d\mid n}} 1\\ &= \sum_{d\mid m}\mu(d)\left\lfloor \frac{x}{d}\right\rfloor\\ &= x\sum_{d\mid m}\frac{\mu(d)}{d} + O(\tau(m))\\ &= x\cdot\frac{\varphi(m)}{m} + O(\tau(m)).\end{aligned}$$

This finishes the proof. □

We have defined additive functions in Section 3. An additive function $f\colon \mathbb{N}_+ \to \mathbb{R}$ is called strongly additive if $f(p^\alpha) = f(p)$ for all $p \in \mathbb{P}$ and $\alpha \in \mathbb{N}_+$. Given a strongly additive function f, we define $f_p(n)$ for every $p \in \mathbb{P}$ by

$$f_p(n) := \begin{cases} f(p)(1-1/p) & \text{if } p \mid n,\\ -f(p)/p & \text{otherwise.}\end{cases}$$

For any positive integer $m = p_1^{\alpha_1}\cdots p_r^{\alpha_r}$, we set

$$f_m(n) := \prod_{i=1}^{r}(f_{p_i}(n))^{\alpha_i}.$$

Now we prove the the following lemma.

Lemma 4.2. *Let $f\colon \mathbb{N}_+ \to \mathbb{R}$ be a strongly additive function such that $|f(p)| \le 1$ for all $p \in \mathbb{P}$. Then for $x \ge 1$ and $z = x^{1/\log(B(x)+3)}$, we have*

$$f(n) - \sum_{p\le x}\frac{f(p)}{p} = \sum_{p\le z} f_p(n) + O(\log(B(x)+3)) \tag{4.2}$$

for all $n \le x$, where

$$B(x) := \left(\sum_{p\le x}\frac{f(p)^2}{p}\right)^{1/2}.$$

Proof. For $n \le x$ we have

$$\sum_{p\le x}\frac{f(p)}{p} = -\sum_{p\le x} f_p(n) + \sum_{p\mid n} f(p).$$

Since f is strongly additive, we have

$$f(n) - \sum_{p\le x}\frac{f(p)}{p} = \sum_{p\mid n} f(p) - \sum_{p\le x}\frac{f(p)}{p} = \sum_{p\le x} f_p(n). \tag{4.3}$$

To prove (4.2), it suffices to show

$$\sum_{z<p\leq x} f_p(n) = \sum_{\substack{z<p\leq x\\ p|n}} f(n) - \sum_{z<p\leq x} \frac{f(p)}{p} = O(\log(B(x)+3)).$$

Since $|f(p)| \leq 1$ for all $p \in \mathbb{P}$, we have

$$\sum_{z<p\leq x} \frac{f(p)}{p} = O\left(\sum_{z<p\leq x} \frac{1}{p}\right) = O(\log\log(B(x)+3)).$$

Note that if $z < n \leq x$, then n is not divisible by any product of more than $\lfloor (\log x)/\log z \rfloor = \lfloor \log(B(x)+3) \rfloor$ primes $p > z$. Hence

$$\sum_{\substack{z<p\leq x\\ p|n}} f(n) = O(\log(B(x)+3)).$$

We conclude that

$$\sum_{z<p\leq x} f_p(n) = O(\log(B(x)+3)),$$

as desired. □

Given a strongly additive function f, we define a sequence $\{X(p)\}_{p\in\mathbb{P}}$ of independent random variables with the property that for every $p \in \mathbb{P}$, we have $X(p) = f(p)$ with probability $1/p$ and $X(p) = 0$ with probability $1 - 1/p$. The following lemma connects $f_m(n)$ and $X(p)$.

Lemma 4.3. *Let $f\colon \mathbb{N}_+ \to \mathbb{R}$ be a strongly additive function such that $|f(p)| \leq 1$ for all $p \in \mathbb{P}$. Fix $m = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$. Then*

$$\sum_{n\leq x} f_m(n) = x \cdot \mathbb{E}\left[\prod_{i=1}^{r} \left(X(p_i) - \frac{f(p_i)}{p_i}\right)^{\alpha_i}\right] + O(2^{2r}) \tag{4.4}$$

for all $x \geq 1$.

Proof. Note that $|f_m(n)| \leq 1$ for all $n \in \mathbb{N}_+$. Let $a := p_1 \cdots p_r$. Observe that if $\gcd(n,a) = d$, then $f_m(n) = f_m(d)$. By Lemma 4.1 we have

$$\begin{aligned}
\sum_{n\leq x} f_m(n) &= \sum_{d|a} f_m(d) \sum_{\substack{n\leq x\\ \gcd(n,a)=d}} 1 \\
&= \sum_{d|a} f_m(d) \left(\frac{\varphi(a/d)}{a/d} \cdot \frac{x}{d} + O(\tau(a/d))\right) \\
&= x \sum_{d|a} f_m(d) \frac{\varphi(a/d)}{a} + O(\tau(a)^2) \\
&= x \sum_{d|a} f_m(d) \frac{\varphi(a/d)}{a} + O(2^{2r}).
\end{aligned}$$

To prove (4.4), it is sufficient to show

$$\sum_{d|a} f_m(d)\frac{\varphi(a/d)}{a} = \mathbb{E}\left[\prod_{i=1}^{r}\left(X(p_i)-\frac{f(p_i)}{p_i}\right)^{\alpha_i}\right]. \tag{4.5}$$

It is easy to see that

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^{r}\left(X(p_i)-\frac{f(p_i)}{p_i}\right)^{\alpha_i}\right] &= \prod_{i=1}^{r}\mathbb{E}\left[\left(X(p_i)-\frac{f(p_i)}{p_i}\right)^{\alpha_i}\right]\\ &= \prod_{i=1}^{r}\left[\frac{1}{p_i}\left(f(p_i)-\frac{f(p_i)}{p_i}\right)^{\alpha_i}+\left(1-\frac{1}{p_i}\right)\left(\frac{-f(p_i)}{p_i}\right)^{\alpha_i}\right].\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\sum_{d|a} f_m(d)\frac{\varphi(a/d)}{a} &= \sum_{d|a}\frac{\varphi(a/d)}{a}\prod_{i=1}^{r}(f_{p_i}(d))^{\alpha_i}\\ &= \sum_{d|a}\frac{\varphi(a/d)}{a}\prod_{\substack{i=1\\ p_i|d}}^{r}\left(f(p_i)-\frac{f(p_i)}{p_i}\right)^{\alpha_i}\prod_{\substack{i=1\\ p_i|(a/d)}}^{r}\left(-\frac{f(p_i)}{p_i}\right)^{\alpha_i}\\ &= \sum_{d|a}\prod_{\substack{i=1\\ p_i|d}}^{r}\frac{1}{p_i}\left(f(p_i)-\frac{f(p_i)}{p_i}\right)^{\alpha_i}\prod_{\substack{i=1\\ p_i|(a/d)}}^{r}\left(1-\frac{1}{p_i}\right)\left(-\frac{f(p_i)}{p_i}\right)^{\alpha_i}\\ &= \prod_{i=1}^{r}\left[\frac{1}{p_i}\left(f(p_i)-\frac{f(p_i)}{p_i}\right)^{\alpha_i}+\left(1-\frac{1}{p_i}\right)\left(\frac{-f(p_i)}{p_i}\right)^{\alpha_i}\right].\end{aligned}$$

This proves (4.5). □

By (4.3) we have

$$\sum_{n\le x}\left(f(n)-\sum_{p\le x}\frac{f(p)}{p}\right)^{k} = \sum_{n\le x}\left(\sum_{p\le x} f_p(n)\right)^{k}$$

for all $k\in\mathbb{N}_+$. Now we relate this quantity to $X(p)$.

Lemma 4.4. *Let $f\colon \mathbb{N}_+\to\mathbb{R}$ be a strongly additive function such that $|f(p)|\le 1$ for all $p\in\mathbb{P}$. For $k\in\mathbb{N}_+$ and $x, z\ge 1$, we have*

$$\sum_{n\le x}\left(\sum_{p\le z} f_p(n)\right)^{k} = x\cdot\mathbb{E}\left[\left(\sum_{p\le z}\left(X(p)-\frac{f(p)}{p}\right)\right)^{k}\right] + O(2^{2k}\pi(z)^k).$$

Proof. Note that

$$\sum_{n\le x}\left(\sum_{p\le z} f_p(n)\right)^{k} = \sum_{p_1,\ldots,p_k\le z}\sum_{n\le x} f_{p_1\cdots p_k}(n).$$

By Lemma 4.3 we have

$$\sum_{n\le x} f_{p_1\cdots p_k}(n) = x\cdot \mathbb{E}\left[\prod_{i=1}^k\left(X(p_i)-\frac{f(p_i)}{p_i}\right)\right] + O(2^{2k}).$$

It follows that

$$\begin{aligned}\sum_{n\le x}\left(\sum_{p\le z} f_p(n)\right)^k &= \sum_{p_1,\dots,p_k\le z}\left(x\cdot\mathbb{E}\left[\prod_{i=1}^k\left(X(p_i)-\frac{f(p_i)}{p_i}\right)\right]+O(2^{2k})\right)\\ &= x\cdot\mathbb{E}\left[\sum_{p_1,\dots,p_k\le z}\prod_{i=1}^k\left(X(p_i)-\frac{f(p_i)}{p_i}\right)\right]+O(2^{2k}\pi(z)^k)\\ &= x\cdot\mathbb{E}\left[\left(\sum_{p\le z}\left(X(p)-\frac{f(p)}{p}\right)\right)^k\right]+O(2^{2k}\pi(z)^k).\end{aligned}$$

This completes the proof. □

5. Computing Moments

Let $f\colon \mathbb{N}_+\to\mathbb{R}$ be a strongly additive function such that $|f(p)|\le 1$ for all $p\in\mathbb{P}$. Suppose further that the series

$$\sum_p \frac{f(p)^2}{p}$$

diverges. In this section we compute $\mathbb{E}[Y_N^k]$ for all $k\in\mathbb{N}_+$, where

$$Y_N=\sum_{p\le N}\left(X(p)-\frac{f(p)}{p}\right)$$

and $N\in\mathbb{N}_+$ is sufficiently large. Consider the moment generating function $M_{Y_N}(z):=\mathbb{E}[e^{zY_N}]$, where $z\in\mathbb{C}$. Then $M_{Y_N}(z)$ is an entire function and $\mathbb{E}[Y_N^k]=M_{Y_N}^{(k)}(0)$. Note that

$$M_{Y_N}(z)=\mathbb{E}\left[\exp\left(z\sum_{p\le N}\left(X(p)-\frac{f(p)}{p}\right)\right)\right]=\prod_{p\le N}\mathbb{E}\left[\exp\left(z\left(X(p)-\frac{f(p)}{p}\right)\right)\right]$$

and that

$$\left|X(p)-\frac{f(p)}{p}\right|\le|f(p)|\left(1-\frac{1}{p}\right)<1.$$

For each $p\le z$, we have

$$\exp\left(z\left(X(p)-\frac{f(p)}{p}\right)\right)=1+z\left(X(p)-\frac{f(p)}{p}\right)+\frac{z^2}{2}\left(X(p)-\frac{f(p)}{p}\right)^2+R(z)$$

for $|z|\le 1/2$, where

$$R(z)=O\left(\left|X(p)-\frac{f(p)}{p}\right|^3|z|^3\right).$$

Observe that

$$\mathbb{E}\left[X(p)-\frac{f(p)}{p}\right]=\frac{f(p)}{p}\left(1-\frac{1}{p}\right)-\left(1-\frac{1}{p}\right)\frac{f(p)}{p}=0,$$

$$\mathbb{E}\left[\left(X(p)-\frac{f(p)}{p}\right)^2\right]=\frac{f(p)^2}{p}\left(1-\frac{1}{p}\right)^2+\left(1-\frac{1}{p}\right)\frac{f(p)^2}{p^2}=\frac{f(p)^2}{p}\left(1-\frac{1}{p}\right),$$

$$\mathbb{E}\left[\left|X(p)-\frac{f(p)}{p}\right|^3\right]=\frac{|f(p)|^3}{p}\left(1-\frac{1}{p}\right)^3+\left(1-\frac{1}{p}\right)\frac{|f(p)|^3}{p^3}=O\left(\frac{|f(p)|^3}{p}\right).$$

This implies

$$\mathbb{E}\left[\exp\left(z\left(X(p)-\frac{f(p)}{p}\right)\right)\right]=1+\frac{f(p)^2}{2p}\left(1-\frac{1}{p}\right)z^2+O\left(\frac{|f(p)|^3}{p}|z|^3\right).$$

for $|z|\le 1/2$. It follows that for sufficiently large N we have

$$\begin{aligned}M_{Y_N}(z)&=\prod_{p\le N}\left[1+\frac{f(p)^2}{2p}\left(1-\frac{1}{p}\right)z^2+O\left(\frac{|f(p)|^3}{p}|z|^3\right)\right]\\&=\prod_{p\le N}\exp\left(\frac{f(p)^2}{2p}\left(1-\frac{1}{p}\right)z^2+O\left(\frac{|f(p)|^3}{p}|z|^3\right)\right)\\&=\exp\left(\frac{z^2}{2}\sum_{p\le N}\frac{f(p)^2}{p}\left(1-\frac{1}{p}\right)+O\left(|z|^3\sum_{p\le N}\frac{|f(p)|^3}{p}\right)\right)\\&=\exp\left(\frac{z^2}{2}\sum_{p\le N}\frac{f(p)^2}{p}+O\left(|z|^2+|z|^3\sum_{p\le N}\frac{|f(p)|^3}{p}\right)\right)\\&=\exp\left(\frac{z^2}{2}\sum_{p\le N}\frac{f(p)^2}{p}\right)\left(1+O\left(\left(\sum_{p\le N}\frac{f(p)^2}{p}\right)^{-1}+|z|\right)\right)\end{aligned}$$

for $|z|\le 1/2$. Let $q\in\mathbb{P}$ be an arbitrary prime with $f(q)\ne 0$ and let

$$Z_N:=Y_N\left(\sum_{p\le N}\frac{f(p)^2}{p}\right)^{-1/2}$$

for $N\ge q$. Then the moment generating function $M_{Z_N}(z)$ of Z_N is given by

$$M_{Z_N}(z)=M_{Y_N}\left(z\left(\sum_{p\le N}\frac{f(p)^2}{p}\right)^{-1/2}\right).$$

Thus we have

$$M_{Z_N}(z)=e^{z^2/2}\left(1+O\left(\left(\sum_{p\le N}\frac{f(p)^2}{p}\right)^{-1}+|z|\left(\sum_{p\le N}\frac{f(p)^2}{p}\right)^{-1/2}\right)\right)$$

for $|z| \le \delta$, where

$$\delta := \frac{1}{2}\left(\sum_{p\le q}\frac{f(p)^2}{p}\right)^{1/2}.$$

We conclude that $M_{Z_N}(z)$ converges to $e^{z^2/2}$ uniformly for $|z| \le \delta$. It follows that

$$\lim_{N\to\infty} M_{Z_N}^{(k)}(0) = \frac{d^k}{dz^k}(e^{z^2/2})_{z=0} = \frac{d^k}{dz^k}\left(\sum_{n=0}^{\infty}\frac{z^{2n}}{(2n)!!}\right)_{z=0} = \mu_k,$$

where

$$\mu_k = \begin{cases} k!/(k)!! & \text{if } 2 \mid k,\\ 0 & \text{if } 2 \nmid k.\end{cases}$$

Therefore, we have

$$\mathbb{E}[Y_N^k] = M_{Y_N}^{(k)}(0) = \left(\sum_{p\le N}\frac{f(p)^2}{p}\right)^{k/2} M_{Z_N}^{(k)}(0) = (1+o(1))\left(\sum_{p\le N}\frac{f(p)^2}{p}\right)^{k/2}\mu_k$$

as $N\to\infty$. Combining this estimate with Lemma 4.4 we obtain the following result.

Corollary 5.1. *Let $f\colon \mathbb{N}_+ \to \mathbb{R}$ be a strongly additive function such that $|f(p)| \le 1$ for all $p \in \mathbb{P}$. For $k \in \mathbb{N}_+$ and $x, z \ge 1$, we have*

$$\sum_{n\le x}\left(\sum_{p\le z} f_p(n)\right)^k = (1+o(1))\left(\sum_{p\le z}\frac{f(p)^2}{p}\right)^{k/2}\mu_k x + O(2^{2k}\pi(z)^k).$$

6. Proof of the Erdős-Kac Theorem

Now we are in a position to prove the following theorem due to Erdős and Kac [4] concerning the distribution of values of strongly additive functions.

Theorem 6.1 (Erdős-Kac, 1940)**.** *Let $f\colon \mathbb{N}_+ \to \mathbb{R}$ be a strongly additive function such that $|f(p)| \le 1$ for all $p \in \mathbb{P}$. Put*

$$A(x) := \sum_{p\le x}\frac{f(p)}{p},$$

$$B(x) := \left(\sum_{p\le x}\frac{f(p)^2}{p}\right)^{1/2},$$

where $x \ge 1$. Suppose that $B(x)\to\infty$ as $x\to\infty$. Then for every $k\in\mathbb{N}_+$,

$$\sum_{n\le x}(f(n) - A(x))^k = (1+o(1))(B(x))^k\mu_k x, \tag{6.1}$$

or equivalently,

$$\sum_{n\le x}(f(n) - A(n))^k = (1+o(1))(B(x))^k\mu_k x. \tag{6.2}$$

As a consequence, we have

$$\lim_{x\to+\infty}\frac{1}{x}\cdot\#\{n\le x\colon f(n)\le A(x)+aB(x)\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a}e^{-t^2/2}\,dt \tag{6.3}$$

for any $a\in\mathbb{R}$.

Proof. Set $z:=x^{1/\log(B(x)+3)}$. Then we have by Lemma 4.2 that

$$f(n)-A(x)=\sum_{p\le z}f_p(n)+S(x)$$

for all $n\le x$, where $S(x)=O(\log(B(x)+3))$. It follows that

$$\sum_{n\le x}(f(n)-A(x))^k=\sum_{n\le x}\sum_{l=0}^{k}\binom{k}{l}\left(\sum_{p\le z}f_p(n)\right)^l(S(x))^{k-l}. \tag{6.4}$$

Note that

$$\begin{aligned}\sum_{n\le x}\sum_{l=0}^{k-1}\binom{k}{l}\left(\sum_{p\le z}f_p(n)\right)^l(S(x))^{k-l}&\le\sum_{l=0}^{k-1}\binom{k}{l}|S(x)|^{k-l}\max_{0\le l\le k-1}\sum_{n\le x}\left|\sum_{p\le z}f_p(n)\right|^l\\&\ll|S(x)|^k\max_{0\le l\le k-1}\sum_{n\le x}\left|\sum_{p\le z}f_p(n)\right|^l\\&\ll(\log(B(x)+3))^k\max_{0\le l\le k-1}\sum_{n\le x}\left|\sum_{p\le z}f_p(n)\right|^l.\end{aligned}$$

It follows from Cauchy-Schwarz inequality and Corollary 5.1 that

$$\sum_{n\le x}\left|\sum_{p\le z}f_p(n)\right|^l\le\sqrt{x}\left(\sum_{n\le x}\left(\sum_{p\le z}f_p(n)\right)^{2l}\right)^{1/2}\ll\sqrt{\mu_l}(B(z))^l x.$$

Hence

$$\sum_{n\le x}\sum_{l=0}^{k-1}\binom{k}{l}\left(\sum_{p\le z}f_p(n)\right)^l(S(x))^{k-l}\ll\sqrt{\mu_k}(B(x))^{k-1}x(\log(B(x)+3))^k. \tag{6.5}$$

Note that

$$0\le B(x)-B(z)\le\left(\sum_{z<p\le x}\frac{1}{p}\right)^{1/2}\ll(\log\log(B(x)+3))^{1/2}.$$

Thus we have

$$(B(z))^k=(B(x))^k+O\left(k(B(x))^{k-1}(\log\log(B(x)+3))^{1/2}\right).$$

By Corollary 5.1 we have

$$\sum_{n\le x}\left(\sum_{p\le x}f_p(n)\right)^k=(1+o(1))(B(z))^k m_k x+O(2^{2k}\pi(z)^k)=(1+o(1))(B(x))^k\mu_k x. \tag{6.6}$$

It follows from (6.4)–(6.6) that

$$\sum_{n\le x}(f(n)-A(x))^k = (1+o(1))(B(x))^k\mu_k x.$$

Next, we establish the equivalence of (6.1) and (6.2). Indeed, we have

$$\sum_{n\le z}(f(n)-A(n))^k \ll (\log\log z)^k z \ll (\log\log x)^k x^{1/\log(B(x)+3)} = o(x).$$

Similarly, we have

$$\sum_{n\le z}(f(n)-A(x))^k = o(x).$$

Note that

$$\sum_{z<n\le x}(f(n)-A(n))^k = \sum_{l=0}^{k}\binom{k}{l}\sum_{z<n\le x}(A(x)-A(n))^{k-l}(f(n)-A(x))^l.$$

By Cauchy-Schwarz inequality and (6.1) we have

$$\sum_{z<n\le x}|f(n)-A(x)|^l \le \sqrt{x}\left(\sum_{z<n\le x}(f(n)-A(x))^{2l}\right)^{1/2} \ll (B(x))^l x$$

for $0\le l<k$. It follows that

$$\begin{aligned}\sum_{z<n\le x}(A(x)-A(n))^{k-l}(f(n)-A(x))^l &\ll |A(x)-A(z)|^{k-l}\sum_{z<n\le x}|f(n)-A(x)|^l\\ &\ll (\log\log(B(x)+3))^{k-l}(B(x))^l x\\ &= o((B(x))^k x)\end{aligned}$$

for $0\le l<k$. Thus we have

$$\begin{aligned}\sum_{n\le x}(f(n)-A(n))^k &= \sum_{z<n\le x}(f(n)-A(n))^k + o((B(x))^k x)\\ &= \sum_{z<n\le x}(f(n)-A(x))^k + o((B(x))^k x)\\ &= \sum_{n\le x}(f(n)-A(x))^k + o((B(x))^k x)\\ &= (1+o(1))(B(x))^k m_k x.\end{aligned}$$

This shows that (6.1) implies (6.2). The converse can be proved in a similar way.

Now we prove (6.3). Let

$$F_N(w) := \frac{1}{N}\cdot\#\{n\le N\colon f(n)\le A(N)+wB(N)\}$$

for all $N\in\mathbb{N}_+$ with $B(N)>0$. Then $F_N(w)$ is the probability distribution function of some random variable W_N. It follows from (6.1) that

$$\mathbb{E}[W_N^k] = \int_{-\infty}^{+\infty} w^k\,dF_N(w) = \frac{1}{N}\sum_{n\le N}\left(\frac{f(n)-A(N)}{B(N)}\right)^k = (1+o(1))\mu_k.$$

This shows that $\mathbb{E}[W_N^k]$ converges to μ_k, which is the same as the kth moment of a normal random variable with mean 0 and variance 1. Since the normal distribution is completely determined by all its moments, we have by a well-known theorem [2, Theorem 30.2] in probability theory that W_N converges in distribution to a normal random variable W with mean 0 and variance 1. This establishes (6.3). $\square$

Remark. It is not hard to show that in (6.3) one can replace $A(x), B(x)$ by $A(n), B(n)$, respectively. Indeed, we see that $A(n) - A(x) = o(B(x))$ and $B(n) - B(x) = o(B(x))$ for $z < n \leq x$. Given any $\varepsilon > 0$, we have

$$A(x) + (a-\varepsilon)B(x) < A(n) + aB(n) < A(x) + (a+\varepsilon)B(x)$$

for $z < n \leq x$ with x sufficiently large. Replacing a by $a \pm \varepsilon$ we obtain

$$A(n) + (a-\varepsilon)B(n) < A(x) + aB(x) < A(n) + (a+\varepsilon)B(n)$$

for $z < n \leq x$ with x sufficiently large. These are sufficient for proving our claim.

Applying Theorem 6.1 to $f(n) = \omega(n)$ we obtain the Erdős-Kac theorem for $\omega(n)$.

Corollary 6.2 (Erdős-Kac, 1940)**.** *For any $a \in \mathbb{R}$ we have*

$$\lim_{x\to+\infty} \frac{1}{x} \cdot \# \left\{ n \leq x \colon \frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \leq a \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2}\, dt.$$

Let $\Omega(n)$ denote the total number of prime divisors of $n \in \mathbb{N}_+$ with multiplicity. Explicitly, we have $\Omega(n) = \sum_{i=1}^{r} \alpha_i$ for $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$. It is easy to see that $\Omega(n)$ is additive but not strongly additive. Thus Theorem 6.1 is not applicable to $\Omega(n)$. Nevertheless, we can derive the following result directly from Corollary 6.2.

Corollary 6.3. *For any $a \in \mathbb{R}$ we have*

$$\lim_{x\to+\infty} \frac{1}{x} \cdot \# \left\{ n \leq x \colon \frac{\Omega(n) - \log\log n}{\sqrt{\log\log n}} \leq a \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2}\, dt. \tag{6.7}$$

Proof. We first compute the average value of $\Omega(n)$ for $n \leq x$. It is easy to see that

$$\sum_{n\leq x} \Omega(n) = \sum_{n\leq x} \sum_{p^\alpha | n} 1 = \sum_{n\leq x} \omega(n) + \sum_{\substack{p^\alpha \leq x \\ \alpha\geq 2}} \left\lfloor \frac{x}{p^\alpha} \right\rfloor = \sum_{n\leq x} \omega(n) + x \sum_{\substack{p^\alpha \leq x \\ \alpha\geq 2}} \frac{1}{p^\alpha} + O(T(x)),$$

where

$$T(x) = \sum_{\substack{p^\alpha \leq x \\ \alpha\geq 2}} 1 \leq \sum_{p\leq\sqrt{x}} \frac{\log x}{\log p} \leq \frac{\log x}{\log 2}\, \pi(\sqrt{x}) \ll \sqrt{x}.$$

Note that

$$\sum_{p>\sqrt{x}} \sum_{\alpha=2}^{\infty} \frac{1}{p^\alpha} = \sum_{p>\sqrt{x}} \frac{1}{p(p-1)} < \sum_{n>\sqrt{x}} \frac{1}{n(n-1)} \ll x^{-1/2},$$

$$\sum_{p\leq\sqrt{x}} \sum_{\alpha>\log_p x} \frac{1}{p^\alpha} \leq \frac{1}{x} \sum_{p\leq\sqrt{x}} \frac{p}{p-1} \leq \frac{1}{x} \sum_{p\leq\sqrt{x}} 2 \leq \frac{2}{\sqrt{x}},$$

where $\log_p x := \log x/\log p$. So we have

$$\sum_{\substack{p^\alpha > x\\ \alpha\geq 2}} \frac{1}{p^\alpha} = \sum_{p\leq\sqrt{x}}\sum_{\alpha>\log_p x}\frac{1}{p^\alpha} + \sum_{p>\sqrt{x}}\sum_{\alpha=2}^{\infty}\frac{1}{p^\alpha} = O(x^{-1/2}).$$

It follows that

$$\sum_{\substack{p^\alpha \leq x\\ \alpha\geq 2}} \frac{1}{p^\alpha} = \sum_{p}\sum_{\alpha=2}^{\infty}\frac{1}{p^\alpha} - \sum_{\substack{p^\alpha > x\\ \alpha\geq 2}}\frac{1}{p^\alpha} = \sum_p \frac{1}{p(p-1)} + O(x^{-1/2}).$$

Thus we conclude

$$\sum_{n\leq x}\Omega(n) = x\log\log x + cx + O\left(\frac{x}{\log x}\right),$$

where

$$c = b + \sum_p \frac{1}{p(p-1)} > 0.$$

This shows that the average value of $\Omega(n)$ for $n\leq x$ is $\log\log x$ and that

$$\sum_{n\leq x}(\Omega(n)-\omega(n)) = O(x).$$

Hence for any $\varepsilon>0$, the set

$$E_\varepsilon := \left\{n\leq x\colon \Omega(n)-\omega(n) > \varepsilon\sqrt{\log\log x}\right\}$$

is of size

$$\#E_\varepsilon = O\left(\frac{x}{\varepsilon\sqrt{\log\log x}}\right).$$

Now fix $a\in\mathbb{R}$. On the one hand, we clearly have

$$\#\left\{n\leq x\colon \frac{\Omega(n)-\log\log x}{\sqrt{\log\log x}}\leq a\right\} \leq \#\left\{n\leq x\colon \frac{\omega(n)-\log\log x}{\sqrt{\log\log x}}\leq a\right\}.$$

On the other hand, we have

$$\frac{\Omega(n)-\log\log x}{\sqrt{\log\log x}} = \frac{\Omega(n)-\omega(n)}{\sqrt{\log\log x}} + \frac{\omega(n)-\log\log x}{\sqrt{\log\log x}} \leq \frac{\omega(n)-\log\log x}{\sqrt{\log\log x}} + \varepsilon$$

for all $n\in[1,x]\setminus E_\varepsilon$. This implies

$$\#\left\{n\in[1,x]\setminus E_\varepsilon\colon \frac{\Omega(n)-\log\log x}{\sqrt{\log\log x}}\leq a\right\} \geq \#\left\{n\in[1,x]\setminus E_\varepsilon\colon \frac{\omega(n)-\log\log x}{\sqrt{\log\log x}}\leq a-\varepsilon\right\}.$$

By Corollary 6.2 we have

$$\varliminf_{x\to+\infty}\frac{1}{x}\cdot\#\left\{n\leq x\colon \frac{\Omega(n)-\log\log x}{\sqrt{\log\log x}}\geq a\right\} \geq \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a-\varepsilon} e^{-t^2/2}\,dt,$$

$$\varlimsup_{x\to+\infty}\frac{1}{x}\cdot\#\left\{n\leq x\colon \frac{\Omega(n)-\log\log x}{\sqrt{\log\log x}}\leq a\right\} \leq \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-t^2/2}\,dt.$$

Letting $\varepsilon \to 0$ we obtain

$$\lim_{x\to+\infty} \frac{1}{x} \cdot \# \left\{ n \le x \colon \frac{\Omega(n) - \log\log x}{\sqrt{\log\log x}} \le a \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2}\, dt,$$

which is equivalent to (6.7). □

We close our discussion with two more applications of Theorem 6.1. Consider first

$$f(n) := \sum_{\substack{p\mid n\\ p\ge 3}} (-1)^{\frac{p-1}{2}}.$$

Equivalently, we have

$$f(n) = \# \{p \mid n \colon p \equiv 1 \pmod 4\} - \# \{p \mid n \colon p \equiv 3 \pmod 4\}.$$

It is clear that f satisfies the conditions of Theorem 6.1 with $A(x) \to A$ as $x \to \infty$ and

$$B(x)^2 = \log\log x + b - \frac{1}{2} + O\left(\frac{1}{\log x}\right),$$

where the series

$$A := \sum_{p\ge 3} (-1)^{\frac{p-1}{2}} \frac{1}{p}$$

is convergent (see [3, §7]). We conclude that

$$\lim_{x\to+\infty} \frac{1}{x} \cdot \# \left\{ n \le x \colon \frac{f(n) - A}{\sqrt{\log\log n}} \le a \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2}\, dt,$$

or equivalently,

$$\lim_{x\to+\infty} \frac{1}{x} \cdot \# \left\{ n \le x \colon \frac{f(n)}{\sqrt{\log\log n}} \le a \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2}\, dt.$$

For another example, consider the function $\log d(n)/\log 2$, which is identically 1 on primes. It is easy to see that

$$\sum_{p\le x} \frac{\log d(p)/\log 2}{p} = \log\log x + b + O\left(\frac{1}{\log x}\right),$$

$$\sum_{p\le x} \frac{(\log d(p)/\log 2)^2}{p} = \log\log x + b + O\left(\frac{1}{\log x}\right).$$

It follows from Theorem 6.1 that

$$\lim_{x\to+\infty} \frac{1}{x} \cdot \# \left\{ n \le x \colon d(n) \le 2^{a\sqrt{\log\log n}} (\log n)^{\log 2} \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2}\, dt.$$

References

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, 1972.

[2] P. Billingsley, *Probability and Measure*, 3rd. ed., John Wiley & Sons, Hoboken, 1995.

[3] H. Davenport, *Multiplicative Number Theory*, 3rd. ed., Grad. Texts in Math., vol. 74, Springer-Verlag, New York, 2000. Revised and with a preface by H. L. Montgomery.

[4] P. Erdős and M. Kac, *The Gaussian law of errors in the theory of additive number theoretic functions*, Am. J. Math. **62** (1940), 738–742.

[5] A. Granville and K. Soundararajan, *Sieving and the Erdős-Kac theorem*, Equidistribution in Number Theory, An Introduction, NATO Sci. Ser. II Math. Phys. Chem., vol. 237, Springer, Dordrecht, 2007, pp. 15–27.

[6] G. H. Hardy and S. Ramanujan, *The normal number of prime factors of a number n*, Quart. J. Pure. Appl. Math. **48** (1917), 76–97.

[7] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 6th. ed., Oxford Univ. Press, Oxford, 2008. Revised by D. R. Heath-Brown and J. H. Silverman; With a forward by A. J. Wiles.

[8] P. Hoffman, *The Man Who Loved Only Numbers: The Story of Paul Erdős and the Search for Mathematical Truth*, Hyperion, New York, 1998.

[9] S. Lester and Z. Rudnick, *Normal order of $\omega(n)$ and the Erdős-Kac theorem*, Sieve theory and its applications, University of Montreal, 2015.

[10] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, 3rd. ed., Grad. Stud. Math., vol. 163, Amer. Math. Soc., Providence, R.I., 2015.

[11] P. Turán, *On a theorem of Hardy and Ramanujan*, J. Lond. Math. Soc. **9** (1934), 274–276.

Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA
Email address: steve.fan.gr@dartmouth.edu