

A localized Erdős-Kac theorem

Anup B. Dixit and M. Ram Murty

Dedicated to the memory of Srinivasa Ramanujan

Abstract. Let $\omega_y(n)$ be the number of distinct prime divisors of n not exceeding y . If y_n is an increasing function of n such that $\log y_n = o(\log n)$, we study the distribution of $\omega_{y_n}(n)$ and establish an analog of the Erdős-Kac theorem for this function. En route, we also prove a variant central limit theorem for random variables, which are not necessarily independent, but are well approximated by independent random variables.

Keywords. prime divisors, Erdős-Kac theorem, central limit theorem

2010 Mathematics Subject Classification. 11N25, 11N64, 11K65, 60F05

1. Introduction

Let $\omega(n)$ denote the number of distinct prime divisors of the natural number n . It is an interesting question to study how $\omega(n)$ is distributed as we vary $n \leq x$. The average value of $\omega(n)$ for $n \leq x$ can be computed as

$$\frac{1}{x} \sum_{n \leq x} \omega(n) = \frac{1}{x} \sum_{p \leq x} \sum_{\substack{n \leq x, \\ p|n}} 1 = \frac{1}{x} \sum_{p \leq x} \left[\frac{x}{p} \right] = \log \log x + O(1)$$

as $x \rightarrow \infty$. In 1917, Hardy and Ramanujan [HR1917] proved that the normal order of $\omega(n)$ is $\log \log n$. More precisely, for any $\epsilon > 0$, as $x \rightarrow \infty$, we have

$$\# \left\{ n \leq x \mid n \text{ satisfies } |\omega(n) - \log \log n| > \epsilon \log \log n \right\} = o(x). \quad (1.1)$$

A simplified proof of the Hardy-Ramanujan result was given by Turán [Tu34] in 1934, by considering the second moment of $\omega(n)$. He proved that

$$\sum_{n \leq x} (\omega(n) - \log \log x)^2 \ll x \log \log x. \quad (1.2)$$

Note that in the summand above, $\log \log x$ can be replaced with $\log \log n$, since they are very close in value for all but very small integers, owing to the slow growth of the function $\log \log x$. Thus, one can easily deduce (1.1). In fact, Hardy and Ramanujan [HR1917] showed that if $c(n) \rightarrow \infty$ as $n \rightarrow \infty$, then

$$|\omega(n) - \log \log n| \leq c(n) \sqrt{\log \log n} \quad (1.3)$$

for almost all $n \leq x$, no matter how slowly $c(n)$ grows.

Research of the first author was partially supported by a Coleman Postdoctoral Fellowship. Research of the second author was partially supported by NSREC Discovery grant.

From a probabilistic perspective, inequality (1.2) is reminiscent of Chebycheff's inequality and inequality (1.3) suggested that perhaps ω was a "random variable" on the space of natural numbers with "mean" $\log \log n$ and "standard deviation" $\sqrt{\log \log n}$. So after Turán's paper appeared, M. Kac posed the question of finding the distribution of

$$\frac{\omega(n) - \log \log n}{\sqrt{\log \log n}}, \quad (1.4)$$

as n varies. He suggested that this distribution is perhaps Gaussian. Kac recalls with evident delight "If I remember it correctly I first stated (as a conjecture) the theorem on the normal distribution of the number of prime divisors during a lecture in Princeton in March 1939. Fortunately for me and possibly for Mathematics, Erdős was in the audience, and he immediately perked up. Before the lecture was over he had completed the proof, which I could not have done not having been versed in the number theoretic methods, especially those related to the sieve." This was the magical moment when probabilistic number theory was born (see page 24 of [El79b]).

Shortly after this romantic episode, Erdős and Kac announced their result, which is famously known as the Erdős-Kac theorem [EK40]. They proved that for any real numbers a, b

$$\lim_{x \rightarrow \infty} \frac{1}{x} \#\left\{n \leq x \mid a \leq \frac{\omega(n) - \log \log n}{\sqrt{\log \log n}} \leq b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt.$$

Thus, the quantity in (1.4) has the standard normal distribution. As noted, the original proof of Erdős and Kac used Brun's sieve and the central limit theorem. Alternate proofs of the Erdős-Kac theorem were given later using different methods by Selberg [Sel53], Halberstam [Hal55], Billingsley [Bi69] (using the method of moments which we adopt below) and Shapiro [Sha56].

In this paper, we are interested in the truncated function, $\omega_y(n)$, which denotes the number of distinct prime divisors of n not exceeding y . We study the behaviour of $\omega_y(n)$ and derive an analog of the Erdős-Kac theorem for it. We first make some observations before stating the main result.

The average order of $\omega_y(n)$ can be computed as before:

$$\frac{1}{x} \sum_{n \leq x} \omega_y(n) = \frac{1}{x} \sum_{n \leq x} \sum_{\substack{p|n, \\ p \leq y}} 1 = \frac{1}{x} \sum_{p \leq y} \left[\frac{x}{p} \right] = \log \log y + O(1)$$

as $x, y \rightarrow \infty$. Also note that

$$\begin{aligned} \frac{1}{x} \sum_{n \leq x} \omega_y^2(n) &= \frac{1}{x} \left(\sum_{\substack{p, q \leq y, \\ p \neq q}} \left[\frac{x}{pq} \right] + \sum_{p \leq y} \left[\frac{x}{p} \right] \right) \\ &= \sum_{p, q \leq y} \frac{1}{pq} - \sum_{p \leq y} \frac{1}{p^2} + \log \log y + O\left(\frac{\pi^2(y)}{x}\right) \\ &= (\log \log y)^2 + O(\log \log y) + O\left(\frac{\pi^2(y)}{x}\right), \end{aligned}$$

where $\pi(y)$ denotes the number of primes $\leq y$. Hence, the second moment is given by

$$\frac{1}{x} \sum_{n \leq x} (\omega_y(n) - \log \log y)^2 = O(\log \log y) + O\left(\frac{\pi^2(y)}{x}\right),$$

as $x, y \rightarrow \infty$. From the above observation, it is reasonable to predict that if y is sufficiently smaller than x , then $\omega_y(n)$ is typically $\log \log y$. More generally, for any multiplicative function $f : \mathbb{N} \rightarrow \mathbb{Q}$, the second moment of $\omega_y(f(n))$ is computed in [MM84a, Theorem 3.1]. We can derive our theorem in a more general setting though we do not do so here for the sake of simplicity of exposition. The general investigation can be found in [MM84a] and [MM84b].

Let Ω be the set of positive integers. P_n be the probability measure placing mass $1/n$ for each $\{1, 2, \dots, n\}$. We prove the following version of the Erdős-Kac theorem for $\omega_y(n)$.

Theorem 1.1. *Let y_n be an increasing sequence of real numbers satisfying $y_n \rightarrow \infty$ as $n \rightarrow \infty$ and suppose*

$$\lim_{n \rightarrow \infty} \frac{\log y_n}{\log n} = 0.$$

Then, for any real numbers a, b ,

$$\lim_{n \rightarrow \infty} P_n \left(m : a \leq \frac{\omega_{y_n}(n) - \log \log y_n}{\sqrt{\log \log y_n}} \leq b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt.$$

Truncated additive functions, such as $\omega_y(n)$, have been previously considered in the literature. The distribution of $\omega_y(n)$ as in Theorem 1.1 appears as a step towards establishing the Erdős-Kac theorem, for instance, see [EK40, Lemma 5], [Sha56, pg.428, formula (10)] and with more restrictive conditions on y_n in [Bi69, formula (21)] and [Hal55, Theorem 2]. The aim of this paper is to highlight the study of *localized* additive functions on account of their independent interest.

2. Preliminaries

The proof of the main theorem relies on the method of moments theorem and the Lyapunov central limit theorem, which are stated below (see pp. 312 and pp. 342 Theorem 30.1 in [Bi79]).

Theorem 2.1. (Method of moments) *Let μ be a probability measure on the line having finite moments*

$$\alpha_k = \int_{-\infty}^{\infty} x^k \mu(dx),$$

for all positive integers k . If the power series

$$\sum_{k=1}^{\infty} \frac{\alpha_k r^k}{k!}$$

has a positive radius of convergence, then μ is the only probability measure with moments $\alpha_1, \alpha_2, \dots$.

Theorem 2.2. (Lyapunov central limit theorem) *For $i \in \mathbb{N}$, let X_i be independent random variables, with mean μ_i and variance σ_i^2 respectively. Denote by $\mathfrak{s}_n^2 = \sum_{i=1}^n \sigma_i^2$. If for some $\delta > 0$, the Lyapunov condition*

$$\lim_{n \rightarrow \infty} \frac{1}{\mathfrak{s}_n^{2+\delta}} \sum_{i=1}^n E \left[|X_i - \mu_i|^{2+\delta} \right] = 0 \tag{2.5}$$

is satisfied, then

$$\frac{1}{\mathfrak{s}_n} \sum_{i=1}^n (X_i - \mu_i) \rightarrow N(0, 1),$$

where $N(0, 1)$ denotes the standard normal distribution, with mean 0 and variance 1.

3. A generalized central limit theorem

Towards the proof of Theorem 1.1, we first establish a central limit theorem for random variables, which are not necessarily independent.

Let f be a positive non-decreasing function on positive integers, such that as n tends to infinity, $f(n)$ tends to infinity and

$$\log f(n) = o(\log n).$$

We prove the following.

Theorem 3.1. *For $i \in \mathbb{N}$, let X_i be independent random variables, taking bounded values and satisfying the Lyapunov condition (2.5) with mean μ_i and variance σ_i^2 . Let Y_i be random variables, not necessarily independent such that*

$$E[X_{i_1} X_{i_2} \cdots X_{i_k}] = E[Y_{i_1} Y_{i_2} \cdots Y_{i_k}] + O\left(\frac{1}{n}\right), \quad (3.6)$$

for $i_j \leq f(n)$ for all $1 \leq j \leq k$. Let

$$s_n^2 = \sum_{i=1}^{f(n)} \sigma_i^2.$$

Then,

$$\frac{1}{s_n} \sum_{i=1}^{f(n)} (Y_i - \mu_i),$$

converges to the standard normal distribution $N(0, 1)$ as n tends to infinity.

Proof. Let $S_n = \sum_{j \leq f(n)} X_j$ and $T_n = \sum_{j \leq f(n)} Y_j$. Denote the mean and variance of S_n as c_n and s_n^2 respectively. As the Lyapunov condition is satisfied for S_n , by Theorem 2.2, we conclude that as n tends to infinity, $(S_n - c_n)/s_n$ converges to the standard normal distribution. Since X_n 's are bounded, the method of moments applies here and from Theorem 2.1 we have that the r -th moment of $(S_n - c_n)/s_n$ converges to the r -th moment of the normal distribution. That is,

$$m_r = \lim_{n \rightarrow \infty} E\left[\left(\frac{S_n - c_n}{s_n}\right)^r\right]$$

for all r , where m_r denotes the r -th moment of the standard normal distribution.

Let d_n and r_n^2 denote the mean and variance of T_n respectively. By condition (3.6), we have

$$c_n = d_n + O(1) \quad \text{and} \quad s_n^2 = r_n^2 + O(1)$$

as n tends to infinity. Hence, to prove Theorem 3.1, it suffices to show that as $n \rightarrow \infty$

$$E\left[\left(\frac{S_n - c_n}{s_n}\right)^r\right] - E\left[\left(\frac{T_n - c_n}{s_n}\right)^r\right] \rightarrow 0 \quad (*)$$

for each r . We have

$$E[S_n^r] = \sum_{u=1}^r \sum' \frac{r!}{r_1! \cdots r_u!} \frac{1}{u!} \sum'' E[X_{i_1} \cdots X_{i_u}], \quad (3.7)$$

where \sum' runs over tuples (r_1, \dots, r_u) satisfying $r_1 + \cdots + r_u = r$ and \sum'' is over tuples (i_1, \dots, i_u) , where i_j 's are distinct and not exceeding $f(n)$.

Similarly, we get

$$E[T_n^r] = \sum_{u=1}^r \sum' \frac{r!}{r_1! \cdots r_u!} \frac{1}{u!} \sum'' E[Y_{i_1} \cdots Y_{i_u}], \quad (3.8)$$

where \sum' and \sum'' are as in (3.7). By (3.6), the summands in (3.7) and (3.8) differ by $O(1/n)$. Hence,

$$\left| E[S_n^r] - E[T_n^r] \right| \ll \frac{1}{n} \left(\sum_{j \leq f(n)} 1 \right)^r = O\left(\frac{f(n)^r}{n}\right).$$

Now we have

$$E[(S_n - c_n)^r] = \sum_{k=0}^r \binom{r}{k} E[S_n^k] (-c_n)^{r-k}.$$

Similarly,

$$E[(T_n - c_n)^r] = \sum_{k=0}^r \binom{r}{k} E[T_n^k] (-c_n)^{r-k}.$$

Comparing these expressions, we get

$$\left| E[(S_n - c_n)^r] - E[(T_n - c_n)^r] \right| \ll \sum_{k=0}^r \binom{r}{k} \frac{f(n)^k}{n} c_n^{r-k} = \frac{(f(n) + c_n)^r}{n}.$$

Since X_i 's take bounded values, we have $c_n = O(f(n))$. Using the condition $\log f(n) = o(\log n)$, we conclude that

$$\lim_{n \rightarrow \infty} \frac{(f(n) + c_n)^r}{n} = 0.$$

Dividing by s_n^r , we see that (*) follows.

4. Proof of Theorem 1.1

Our method of proof follows Billingsley [Bi69]. For a prime p , let

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

Then,

$$\omega_y(m) = \sum_{p \leq y} \delta_p(m).$$

If p_1, p_2, \dots, p_u are a set of distinct primes $\leq y$, then

$$P_n \left[m \mid \delta_{p_1}(m) = \cdots = \delta_{p_u}(m) = 1 \right] = \frac{1}{n} \left[\frac{n}{p_1 p_2 \cdots p_u} \right],$$

where $[x]$ denotes the greater integer $\leq x$. This shows that under P_n , δ_{p_i} 's behave like independent random variables asymptotically, but are not independent random variables. For a function f of positive integers, define

$$E_n[f] = \frac{1}{n} \sum_{m=1}^n f(m).$$

For all primes p , let X_p be independent random variables taking values $\{0, 1\}$, satisfying

$$P[X_p = 1] = \frac{1}{p} \quad \text{and} \quad P[X_p = 0] = 1 - \frac{1}{p}.$$

If p_1, \dots, p_u are distinct, then we have

$$P\left[X_{p_1} = \dots = X_{p_u} = 1\right] = \frac{1}{p_1 p_2 \dots p_u}.$$

Let $S_n = \sum_{p \leq y_n} X_p$. The mean and variance of S_n are given by

$$c_n = \sum_{p \leq y_n} \frac{1}{p} = \log \log y_n + O(1)$$

and

$$s_n^2 = \sum_{p \leq y_n} \frac{1}{p} \left(1 - \frac{1}{p}\right) = \log \log y_n + O(1).$$

Since X_p 's are independent, we have

$$E[X_{p_1} \dots X_{p_u}] = \frac{1}{p_1 \dots p_u}. \quad (4.9)$$

Also,

$$E_n[\delta_{p_1} \dots \delta_{p_u}] = \frac{1}{n} \left[\frac{n}{p_1 \dots p_u} \right]. \quad (4.10)$$

Hence,

$$E[X_{p_1} \dots X_{p_u}] - E_n[\delta_{p_1} \dots \delta_{p_u}] = \frac{1}{n} \left(\frac{n}{p_1 \dots p_u} - \left[\frac{n}{p_1 \dots p_u} \right] \right) = O\left(\frac{1}{n}\right)$$

for all $p_i \leq y_n$. The proof now follows from Theorem 3.1.

As an application of this result, we deduce the following corollary.

Corollary 4.1. *Suppose $\log \log y_n = (\log \log n)(1 + o(1))$. With probability 1, almost all the prime factors of n are less than y_n . In particular, we can apply this to $y_n = n^{1/\log \log n}$.*

5. Concluding remarks

The fact that for additive functions f for which an Erdős-Kac type theorem holds, that a localized Erdős-Kac also holds, as in Theorem 1.1, is known and is due to Kubilius, Barban and A.I. Vinogradov (see page 123 of [El79a]). Our contribution here is a new proof of this fact adapting the method of Billingsley [Bi69]. This method is quite versatile and it has wider applications in a very general setting discussed in [MMP]. The relation between distribution of f_{y_n} under the measure P_n and that of $\sum_{p < y_n} X_p$, for independent random variables X_p , taking values $f(p)$ and 0 with probability $1/p$ and $1 - 1/p$ respectively, is captured in the Kubilius Fundamental lemma alluded to above. This lemma states that

$$P_n(m : f_y(m) \leq z) = P\left(\sum_{p \leq y} X_p \leq z\right) + O\left(\exp\left(-c \frac{\log n}{\log y}\right)\right) + n^{-\delta}.$$

From this, one can also derive explicit error terms.

The idea of deriving a central limit theorem in the context of not necessarily independent random variables is also not new. For instance, the notion of α -mixing discussed in [Bi79, p. 363] relaxes the notion of independence. A central limit theorem can be derived in such a context. However, our treatment offers an alternate approach to a variation of the central limit theorem suitable for other applications.

Acknowledgement. We thank Siddhi Pathak and the referee for detailed comments on an earlier version of the paper.

References

- [Bi69] P. Billingsley, On the central limit theorem for the prime divisor functions, *Amer. Math. Monthly*, **76** (1969), 132–139.
- [Bi79] P. Billingsley, Probability and measure, *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, New York-Chichester-Brisbane, (1979).
- [El79a] P.D.T.A. Elliot, Probabilistic number theory. I, *Grundlehren der Mathematischen Wissenschaften [Fundamental principles of mathematical science]*, Vol. 239, Springer-Verlag, New York-Berlin, (1979).
- [El79b] P.D.T.A. Elliot, Probabilistic number theory. II, *Grundlehren der Mathematischen Wissenschaften [Fundamental principles of mathematical science]*, Vol. 239, Springer-Verlag, New York-Berlin, (1979).
- [EK40] P. Erdős, M. Kac, The Gaussian law of errors in the theory of additive number theoretic functions, *Amer. J. Math.*, **62** (1940), 738–742.
- [Hal55] H. Halberstam, On the distribution of additive number theoretic functions, *J. London Math. Soc.*, **30** (1955), 43–53.
- [HR1917] G.H. Hardy, S. Ramanujan, The normal number of prime factors of a number n , *Quart. J. Pure. Appl. Math.*, **48** (1917), 323–339.
- [MM84a] M.R. Murty, V.K. Murty, Prime divisors of Fourier coefficients of modular forms, *Duke Math. J.*, **51** no. 1, (1984), 57–76.
- [MM84b] 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** no. 10, (1984), 1090–1101.
- [MMP] M.R. Murty, V.K. Murty, S. Pujahari An all-purpose Erdős-Kac theorem, *to appear*.
- [Sel53] A. Selberg, Note on a paper by L.G. Sathe, *J. Indian Math. Soc.*, **17** (1953), 83–141.
- [Sha56] H.N. Shapiro, Distribution functions of additive arithmetic functions, *Proc. Nat. Acad. Sci. U.S.A.*, **42** (1956), 426–430.
- [Tu34] P. Turán, On a theorem of Hardy and Ramanujan, *J. London Math. Soc.*, **9** (1934), 274–276.

Anup B. Dixit

Institute of Mathematical Sciences
CIT Campus Taramani
Chennai – 600113, INDIA
e-mail: anupdixit@imsc.res.in

M. Ram Murty

Department of Mathematics and Statistics
Jeffery Hall, Queen’s University, Kingston
Canada, ON K7L 3N6
e-mail: murty@mast.queensu.ca