Canad. J. Math. Vol. **56** (2), 2004 pp. 356–372

Non-Abelian Generalizations of the Erdős-Kac Theorem

M. Ram Murty and Filip Saidak

Abstract. Let a be a natural number greater than 1. Let $f_a(n)$ be the order of a mod n. Denote by $\omega(n)$ the number of distinct prime factors of n. Assuming a weak form of the generalised Riemann hypothesis, we prove the following conjecture of Erdös and Pomerance:

The number of $n \le x$ coprime to a satisfying

$$\alpha \le \frac{\omega(f_a(n)) - (\log\log n)^2/2}{(\log\log n)^{3/2}/\sqrt{3}} \le \beta$$

is asymptotic to $\left(\frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-t^2/2}dt\right)\frac{x\phi(a)}{a}$, *as x tends to infinity.*

1 Introduction

Let $\omega(n)$ and $\Omega(n)$ denote the number of prime factors of n, counted without multiplicity and with multiplicity, respectively. In 1917, Hardy and Ramanujan [7] proved that the normal order of $\omega(n)$ and $\Omega(n)$ is $\log\log n$. This means that given any $\epsilon > 0$, the number of $n \le x$ failing to satisfy the inequality

$$|f(n) - \log\log n| < \epsilon \log\log n,$$

with $f = \omega$ or Ω, is $o(x)$ as $x \to \infty$. Subsequently, in 1934, Turán [23] gave a simple proof of this fact by showing that

$$\sum_{n\le x}(\omega(n) - \log\log n)^2 \ll x\log\log x. \tag{1}$$

Recently, Saidak [19] improved Turán's theorem by proving the asymptotic formula

$$\sum_{n\le x}\big(\omega(n) - \log\log n\big)^2 = x\log\log x + Cx + O\Big(\frac{x\log\log x}{\log x}\Big),$$

for some constant C. The underlying ideas behind these theorems form the foundations of probabilistic number theory. Indeed, in 1940, Erdős and Kac [4] proved that the quantity

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

Received by the editors May 15, 2002; revised August 26, 2002.
The first author was partially supported by an NSERC grant.
AMS subject classification: 11K36, 11K99.
Keywords: Turán's theorem, Erdős-Kac theorem, Chebotarev density theorem, Erdős-Pomerance conjecture.

is distributed 'normally'. More precisely, they proved that for any $\alpha < \beta$,

$$\#\left\{ n \leq x : \alpha \leq \frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \leq \beta \right\} \sim \Phi(\alpha, \beta)x, \tag{2}$$

as $x \to \infty$, where

$$\Phi(\alpha, \beta) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-t^2/2}\, dt.$$

This theorem opened the way for a more general theory by Kubilius [8] and Shapiro [22] applicable to a wider class of what are called "strongly additive" (see definition below) functions. In the early 60s and subsequently the 70s, the theory was refined by many authors and one can find a comprehensive treatment of it in the monograph of Elliott [2].

For instance, a "prime" analog of the theorem of Turán can be proved using the methods indicated in [14]:

$$\sum_{p \leq x} \big(\omega(p-1) - \log\log p \big)^2 \ll \pi(x) \log\log x, \tag{3}$$

where the summation is over primes $p \leq x$ and $\pi(x)$ denotes the number of primes $p \leq x$. The normal order of $\omega(p-1)$ was first determined in 1935 by Erdős [3] by more complicated methods. The corresponding version of the Erdős-Kac theorem was discovered by Halberstam [6] in 1955, who proved that for all $\alpha < \beta$,

$$\#\left\{ p \leq x : \alpha \leq \frac{\omega(p-1) - \log\log p}{\sqrt{\log\log p}} \leq \beta \right\} \sim \Phi(\alpha, \beta)\pi(x), \tag{4}$$

as $x \to \infty$.

A new type of "Erdős-Kac" theorem, which can be described as "non-abelian", was discovered by Kumar and Ram Murty [14] in the early 1980's. A special case of their theorem will illustrate our meaning.

Let $\tau(n)$ denote the Ramanujan τ-function. Assuming a generalized Riemann hypothesis (to be made more precise below), they proved that

$$\#\left\{ p \leq x : \alpha \leq \frac{\omega\big(\tau(p)\big) - \log\log p}{\sqrt{\log\log p}} \leq \beta \right\} \sim \Phi(\alpha, \beta)\pi(x),$$

and assuming $\tau(n)$ never vanishes (Lehmer's conjecture [11])

$$\#\left\{ n \leq x : \alpha \leq \frac{\omega\big(\tau(n)\big) - \frac{1}{2}(\log\log n)^2}{\frac{1}{\sqrt{3}}(\log\log n)^{3/2}} \leq \beta \right\} \sim \Phi(\alpha, \beta)x,$$

as $x \to \infty$. Murty and Murty [14] prove general theorems applicable to a wider class of functions arising as Fourier coefficients of modular forms.

What distinguishes these theorems from the classical theory is their "non-abelian" character. Indeed, latent in the Erdős-Kac theory is the intervention of the distribution of primes in cyclotomic fields, which are abelian extensions of the rational number field. In the case of Fourier coefficients arising from normalized Hecke eigenforms, one has the theory of l-adic representations (thanks to Deligne [1]) from which divisibility properties of the Fourier coefficients can be deduced via the Chebotarev density theorem. It is here that the generalized Riemann hypothesis intervenes, for in order to control the error terms in the expansions that arise, such a hypothesis is essential. In the classical case, the generalized Riemann hypothesis can be replaced by a direct application of the Bombieri-Vinogradov theorem (or as in the case of Erdős [3], the Siegel-Walfisz theorem combined with Brun's sieve).

Our purpose here is to indicate yet another new type of the Erdős-Kac theorem formulated in a conjecture of Erdős and Pomerance [5]. Let a be a natural number greater than 1. For any n coprime to a, define $f_a(n)$ to be the order of $a \pmod{n}$. Erdős and Pomerance [5] conjectured that for any $\alpha < \beta$,

$$\#\left\{ n \le x : (a,n) = 1, \alpha \le \frac{\omega\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2}{\frac{1}{\sqrt{3}}(\log\log n)^{3/2}} \le \beta \right\} \sim \Phi(\alpha,\beta)\frac{\phi(a)}{a}x,$$

as $x \to \infty$, where $\phi(a)$ denotes the number of positive integers less than a and coprime to a. In this paper we prove this conjecture assuming a hypothesis substantially weaker than the generalized Riemann hypothesis. Again, certain non-abelian extensions of $\mathbb{Q}$ intervene in a natural way.

In the course of our investigations, we prove a variety of interesting results related to the study of $f_a(n)$ and these we state in the next section. Most notable is Theorem 6, which is unconditional.

2 Statements of Theorems

Before stating the main results of the paper, we elucidate the precise nature of the generalized Riemann hypothesis, or quasi-Riemann hypothesis, invoked in Theorems 1 to 4.

In the study of $f_a(p)$, or more generally, $f_a(n)$, the non-abelian extensions

$$L_q = \mathbb{Q}(\omega_q, \sqrt[q]{a}),$$

where q is a prime and ω_q denotes a primitive q-th root of unity, intervene in a natural way. More precisely, if $\zeta_q(s)$ denotes the Dedekind zeta function of L_q, we assume that for some $\theta < 1$, $\zeta_q(s)$ has no zeros in the region $\mathrm{Re}(s) > \theta$, for every prime q. This we refer to as a quasi-Riemann hypothesis. Of course, $\theta = 1/2$ is the celebrated generalized Riemann hypothesis. In the last section of the paper, we discuss how we can weaken the assumption of a quasi-Riemann hypothesis.

Theorem 1 *Let $a \ge 2$ be squarefree. Assuming there is a $\theta < 1$ such that every*

$\zeta_q(s) \neq 0$ in $\mathrm{Re}(s) > \theta$, for every prime q, we have

$$\sum_{\substack{p\le x\\(a,p)=1}} \Big(\omega\big(f_a(p)\big) - \log\log p\Big)^2 \ll \pi(x)\log\log x. \tag{5}$$

Remark A quasi-GRH allows us to prove that the number of primes $p \leq x$ which split completely in L_q is

$$\frac{\mathrm{Li}\, x}{q\phi(q)} + O(x^{\theta}\log qax), \quad \text{where} \quad \mathrm{Li}\, x = \int_2^x \frac{dt}{\log t}.$$

If a were not squarefree, we would expect an analogous result as Theorem 1 to hold. However, the main term in the special case of the Chebotarev density theorem cited above will have to be slightly adjusted for those q that may divide the exponents of the prime powers in the unique factorization of a.

Theorem 2 *Under the same hypotheses as in Theorem 1, we have*

$$\#\left\{p \le x : (a,p)=1, \alpha \le \frac{\omega\big(f_a(p)\big) - \log\log p}{\sqrt{\log\log p}} \le \beta\right\} \sim \Phi(\alpha,\beta)\pi(x), \tag{6}$$

as $x \to \infty$.

We will deduce Theorem 2 from Theorem (4) of Halberstam [6]. If we let $i_a(p)$ denote the index of the subgroup generated by $a \pmod p$ in $\mathbb{F}_p^*$, then

$$\omega(p-1) = \omega\big(f_a(p)\big) + O\Big(\omega\big(i_a(p)\big)\Big).$$

A quasi-GRH is needed to ensure that for almost all p, $\omega\big(i_a(p)\big)$ is not as large as $\omega(p-1)$. The technical lemma we invoke to make the transition is of an independent interest in its own right.

Theorem 2 can be proved directly without invoking the theorem of Halberstam by considering all the higher moments, as was done in the thesis [20]. We relegate this proof to a future paper [21].

Theorem 3 *Under the same hypotheses as in Theorem 1, we have*

$$\sum_{\substack{n\le x\\(a,n)=1}} \Big(\omega\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2\Big)^2 \ll x(\log\log x)^3. \tag{7}$$

Theorem 4 *Under the same hypotheses as in Theorem 1,*

$$\#\left\{n \le x : (a,n)=1, \alpha \le \frac{\omega\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2}{\frac{1}{\sqrt{3}}(\log\log n)^{3/2}} \le \beta\right\} \sim \Phi(\alpha,\beta)\frac{\phi(a)}{a}x, \tag{8}$$

as $x \to \infty$.

Remark This theorem establishes the conjecture of Erdős and Pomerance [5] assuming a quasi-GRH as defined in the remark following Theorem 1.

By invoking a method of Kubilius and Shapiro we deduce the following. Let

$$A(x) = \sum_{\substack{p \le x \\ (a,p)=1}} \frac{\omega\big(f_a(p)\big)}{p}, \quad B(x) = \sum_{\substack{p \le x \\ (a,p)=1}} \frac{\omega^2\big(f_a(p)\big)}{p}. \tag{9}$$

Theorem 5 *Without any hypothesis, we have*

$$\#\left\{ n \le x : (a,n) = 1, \alpha \le \frac{\omega\big(f_a(n)\big) - A(x)}{\sqrt{B(x)}} \le \beta \right\} \sim \Phi(\alpha,\beta)\frac{\phi(a)}{a}x, \tag{10}$$

as $x \longrightarrow \infty$.

Theorem 6 *For any* $\beta > 0$, *as* $x \longrightarrow \infty$,

$$\#\left\{ n \le x : (a,n) = 1, \omega\big(f_a(n)\big) < \frac{1}{2}(\log\log n)^2 + \frac{\beta}{\sqrt{3}}(\log\log n)^{3/2} \right\} \gtrsim \Phi(-\infty,\beta)\frac{\phi(a)}{a}x. \tag{11}$$

Even though Theorem 5 is "well-known" or can be deduced from general theory, we isolate it here to show where the quasi-GRH enters in the proof of Theorem 4. It is to determine the asymptotic behaviour of $A(x)$ and $B(x)$. Even a weak assertion such as

$$A(x) = \frac{1}{2}(\log\log x)^2 + o\big((\log\log x)^2\big)$$

would allow us to prove that $\omega\big(f_a(p)\big)$ has normal order $\log\log p$, without any hypotheses. But this seems to be beyond reach at the moment. The central assertion to be proved is

$$\sum_{p \le x} \frac{\omega\big(i_a(p)\big)}{p} = o\big((\log\log x)^2\big).$$

3 Technical Preparation

What we will need in the proofs of the theorems stated in the previous section are two technical lemmas which are of independent interest.

Lemma 1 *Let* f, g *and* h *be arithmetical functions, and suppose that for all* n *we have* $|f(n) - g(n)| \le h(n)$. *Let us define*

$$\mathfrak{A}(x) = \sum_{p \le x} \frac{f(p)}{p} \quad \textit{and} \quad \mathfrak{B}(x) = \sum_{p \le x} \frac{f^2(p)}{p}, \tag{12}$$

and assume that the following condition is satisfied:

$$\sum_{n\le x} h(n) = o\big(x\sqrt{\mathfrak{B}(x)}\big)\,. \tag{13}$$

Suppose that for all constants α and β (with $\alpha \le \beta$) we have:

$$\#\left\{ n \le x : \alpha \le \frac{f(n)-\mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \right\} \sim x\left(\frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-t^2/2}\,dt \right),$$

as $x \to \infty$, then also

$$\#\left\{ n \le x : \alpha \le \frac{g(n)-\mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \right\} \sim x\left(\frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-t^2/2}\,dt \right).$$

Proof First, let us define the functions:

$$S_f(x\,;\alpha,\beta) \overset{\text{def}}{=} \#\left\{ n \le x : \alpha \le \frac{f(n)-\mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \right\}, \quad \text{and} \tag{14}$$

$$S_g(x\,;\alpha,\beta) \overset{\text{def}}{=} \#\left\{ n \le x : \alpha \le \frac{g(n)-\mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \right\}.$$

What we will prove is that if $S_f(x\,;\alpha,\beta) \sim x\Phi(\alpha,\beta)$, then $S_g(x\,;\alpha,\beta) \sim x\Phi(\alpha,\beta)$. Fix $\epsilon > 0$. Then by the assumption, the number of $n \le x$, for which $|h(n)| > \epsilon\sqrt{\mathfrak{B}(x)}$ is $o(x)$. Hence for almost all $n \le x$, we may suppose that $|h(n)| \le \epsilon\sqrt{\mathfrak{B}(x)}$.

Now,

$$\alpha \le \frac{f(n)-\mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \Rightarrow \mathfrak{A}(x) + \alpha\sqrt{\mathfrak{B}(x)} \le f(n) \le \mathfrak{A}(x) + \beta\sqrt{\mathfrak{B}(x)},$$

and $g(n) - h(n) \le f(n) \le g(n) + h(n)$, together imply

$$g(n) - h(n) \le \mathfrak{A}(x) + \beta\sqrt{\mathfrak{B}(x)} \quad \text{and} \quad g(n) + h(n) \ge \mathfrak{A}(x) + \alpha\sqrt{\mathfrak{B}(x)}.$$

Hence for almost all integers $n \le x$, we have:

$$g(n) \le \mathfrak{A}(x) + \beta\sqrt{\mathfrak{B}(x)} + \epsilon\sqrt{\mathfrak{B}(x)} \quad \text{and} \quad g(n) \ge \mathfrak{A}(x) + \alpha\sqrt{\mathfrak{B}(x)} - \epsilon\sqrt{\mathfrak{B}(x)},$$

and hence also: $S_f(x\,;\alpha,\beta) \le S_g(x\,;\alpha-\epsilon,\beta+\epsilon) + o(x)$, or in other words:

$$\begin{aligned} S_g(x\,;\alpha,\beta) &\ge S_f(x\,;\alpha+\epsilon,\beta-\epsilon) + o(x) \\ &\sim x\Phi(\alpha+\epsilon,\beta-\epsilon) + o(x), \end{aligned} \tag{15}$$

by assumption (concerning the normality of the function $f(n)$), where we easily see that $\Phi(\alpha+\epsilon,\beta-\epsilon) = \Phi(\alpha,\beta) + O(\epsilon)$. Similarly, we can write:

$$\alpha \le \frac{g(n)-\mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \Rightarrow \mathfrak{A}(x) + \alpha\sqrt{\mathfrak{B}(x)} \le g(n) \le \mathfrak{A}(x) + \beta\sqrt{\mathfrak{B}(x)} \Rightarrow$$

$$f(n) - h(n) \le \mathfrak{A}(x) + \beta\sqrt{\mathfrak{B}(x)} \quad \text{and} \quad f(n) + h(n) \ge \mathfrak{A}(x) + \alpha\sqrt{\mathfrak{B}(x)}.$$

and so, for almost all integers $n \le x$ we have:

$$f(n) \le \mathfrak{A}(x) + \beta\sqrt{\mathfrak{B}(x)} + \epsilon\sqrt{\mathfrak{B}(x)} \quad \text{and} \quad f(n) \ge \mathfrak{A}(x) + \alpha\sqrt{\mathfrak{B}(x)} - \epsilon\sqrt{\mathfrak{B}(x)}.$$

This proves that $S_g(x\,;\alpha,\beta) \le S_f(x\,;\alpha-\epsilon,\beta+\epsilon) + o(x)$, or in other words:

$$\begin{aligned} (16) \qquad S_g(x\,;\alpha,\beta) &\le S_f(x\,;\alpha-\epsilon,\beta+\epsilon) + o(x) \\ &\sim x\Phi(\alpha-\epsilon,\beta+\epsilon) + o(x), \end{aligned}$$

and since again $\Phi(\alpha-\epsilon,\beta+\epsilon) = \Phi(\alpha,\beta) + O(\epsilon)$, the result follows. ■

Lemma 2 *Suppose that $\mathfrak{P} \subset \mathbb{N}$, i.e. $\mathfrak{P}$ is a subset of natural numbers, and let $\mathfrak{P}(x) = |\{n \in \mathfrak{P}, n \le x\}|$. Let f, g and h be arithmetical functions, and suppose that for all n we have $|f(n) - g(n)| \le ch(n)$, where $c > 0$ is a constant. Define $\mathfrak{A}(x)$ and $\mathfrak{B}(x)$ as in (12), and assume that the following condition is satisfied:*

$$(17) \qquad \sum_{n \le x} h(n) = o\big(\mathfrak{P}(x)\sqrt{\mathfrak{B}(x)}\big).$$

Suppose that for all constants α and β (with $\alpha \le \beta$) we have:

$$\#\left\{ n \le x, n \in \mathfrak{P} : \alpha \le \frac{f(n) - \mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \right\} \sim \mathfrak{P}(x)\left(\frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-t^2/2}\,dt \right),$$

as $x \to \infty$, then also

$$\#\left\{ n \le x, n \in \mathfrak{P} : \alpha \le \frac{g(n) - \mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \right\} \sim \mathfrak{P}(x)\left(\frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-t^2/2}\,dt \right).$$

Proof The proof is similar to Lemma 1, and we suppress it. ■

4 Proof of Theorem 1

Let $i_a(p)$ be the index in $\mathbb{F}_p^*$ of the subgroup generated by $a \pmod p$. It is easily seen that $q|i_a(p)$ if and only if p splits completely in L_q (see [12]). Clearly, we have

$$\omega(p-1) - \omega\big(i_a(p)\big) \le \omega\big(f_a(p)\big) \le \omega(p-1).$$

By [14], we know that (see (3))

$$\sum_{p \le x} \big(\omega(p-1) - \log\log p\big)^2 \ll \pi(x)\log\log x.$$

Recall that

$$\omega\big(f_a(p)\big) - \log\log p = \big(\omega(p-1) - \log\log p\big) + O\Big(\omega\big(i_a(p)\big)\Big).$$

Squaring both sides and summing over all $p \le x$, we have

$$\sum_{p\le x}\Big(\omega\big(f_a(p)\big)-\log\log p\Big)^2 \ll \sum_{p\le x}\big(\omega(p-1)-\log\log p\big)^2+\sum_{p\le x}\omega^2\big(i_a(p)\big).$$

An application of (3) shows that it suffices to prove:

$$\sum_{p\le x}\omega^2\big(i_a(p)\big) \ll \pi(x)\log\log x,$$

in order to establish Theorem 1. In fact, we will show that a quasi-GRH implies the stronger result:

$$\sum_{p\le x}\omega^2\big(i_a(p)\big) \ll \pi(x). \tag{18}$$

Indeed, defining

$$\omega_y(n)=\sum_{\substack{p|n\\ p<y}}1,$$

we have for $y=x^\delta$, $\delta<1/4$,

$$\sum_{p\le x}\omega^2\big(i_a(p)\big) \ll \sum_{p\le x}\omega_y^2\big(i_a(p)\big)+O\big(\pi(x)\big)$$

But

$$\sum_{p\le x}\omega_y^2\big(i_a(p)\big) = \sum_{\substack{q_1,q_2\le y^2\\ q_1\ne q_2}}\pi(x,L_{q_1q_2})+\sum_{q\le y}\pi(x,L_q),$$

where $\pi(x,L_k)$ denotes the number of primes $p\le x$ which split completely in L_k, because as was pointed out earlier, $q|i_a(p)$ if and only if p splits completely in L_q.

A quasi-GRH (with no zeros of $\zeta_q(s)$ for $\mathrm{Re}(s)>\theta$) gives

$$\pi(x,L_k)=\frac{\mathrm{Li}\,x}{k\phi(k)}+O(x^\theta\log kax). \tag{19}$$

Hence, if we choose $2\delta<\theta$, we obtain

$$\sum_{p\le x}\omega^2\big(i_a(p)\big) \ll \pi(x),$$

which completes the proof of Theorem 1. ■

Corollary 1 *With the same hypotheses as Theorem 1,*

$$\sum_{p\le x}\omega\big(f_a(p)\big)=\pi(x)\log\log x+O\big(\pi(x)\big), \quad \textit{and} \tag{20}$$

$$\sum_{p\le x}\omega^2\big(f_a(p)\big)=\pi(x)(\log\log x)^2+O\big(\pi(x)\log\log x\big). \tag{21}$$

Corollary 2 *With the same hypotheses as Theorem 1,*

$$\sum_{p\leq x}\Omega\big(f_a(p)\big) = \pi(x)\log\log x + O\big(\pi(x)\big), \quad \textit{and} \tag{22}$$

$$\sum_{p\leq x}\Omega^2\big(f_a(p)\big) = \pi(x)(\log\log x)^2 + O\big(\pi(x)\log\log x\big). \tag{23}$$

By a partial summation, we deduce:

Corollary 3 *Under the same hypotheses as Theorem 1,*

$$\sum_{p\leq x}\frac{\Omega\big(f_a(p)\big)}{p} = \frac{1}{2}(\log\log x)^2 + O(\log\log x), \quad \textit{and} \tag{24}$$

$$\sum_{p\leq x}\frac{\Omega^2\big(f_a(p)\big)}{p} = \frac{1}{3}(\log\log x)^3 + O\big((\log\log x)^2\big). \tag{25}$$

5 Proof of Theorem 2

By the theorem (4) of Halberstam [6] we know that

$$\frac{\omega(p-1)-\log\log p}{\sqrt{\log\log p}}$$

obeys a normal distribution. Since $\omega(p-1) = \omega\big(f_a(p)\big) + O\big(\omega\big(i_a(p)\big)\big)$, we may apply Lemma 2 of Section 3 to deduce that

$$\frac{\omega\big(f_a(p)\big)-\log\log p}{\sqrt{\log\log p}}$$

has a normal distribution. This is because

$$\sum_{p\leq x}\omega\big(i_a(p)\big) \ll \pi(x) = o\big(\pi(x)\sqrt{\log\log x}\big),$$

so that the hypotheses of Lemma 2 are satisfied. This completes the proof of Theorem 2. ■

6 Transition From $\Omega\big(f_a(n)\big)$ to $\omega\big(f_a(n)\big)$

It will be convenient to prove the analogs of Theorems 3 and 4 for $\Omega\big(f_a(n)\big)$, and then deduce the corresponding result for $\omega\big(f_a(n)\big)$. In this section we will indicate how this can be done, and then, in the next section, we establish Theorem 3 for $\Omega\big(f_a(n)\big)$. The same strategy will be applied in our proof of Theorem 4. It will be efficacious to extend the definition of $f_a(n)$ when $(a,n) \neq 1$, by setting $f_a(1) = 1$, and $f_a(n) = f_a(n_2)$, where $n = n_1 n_2$, with $(n_1, n_2) = 1$ and $p|n_1 \Rightarrow p|a$.

For all $3 < y \leq x$, let us define the truncated functions

$$\omega_y\big(f_a(n)\big) = \sum_{\substack{p|f_a(n)\\ p<y}} 1, \quad \text{and} \quad \Omega_y\big(f_a(n)\big) = \sum_{\substack{p^\alpha \| f_a(n)\\ p^\alpha<y}} \alpha, \tag{26}$$

and observe that $\omega\big(f_a(n)\big) \leq \Omega\big(f_a(n)\big)$. We will be using:

Lemma 3 *For all $x \geq 2$, and $2 \leq k \leq x$,*

$$\sum_{\substack{p\leq x\\ p\equiv 1 (\text{mod } k)}} \frac{1}{p} = \frac{\log\log x}{\phi(k)} + O\Big(\frac{\log k}{k}\Big). \tag{27}$$

Proof This follows easily by partial summation and the Brun-Titchmarsh theorem (see Norton [16] or Pomerance [17] for a complete proof). ■

The following lemma will also be applied several times in the discussion below:

Lemma 4 *If, for all x, we have*

$$\sum_{n\leq x}\Big(\Omega\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2\Big)^2 \ll x(\log\log x)^3, \quad \textit{then}$$

$$\sum_{n\leq x}\Big(\omega\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2\Big)^2 \ll x(\log\log x)^3. \tag{28}$$

Proof Let $\omega_y^+(n)$ be the number of primes $> y$, dividing n. Similarly we define $\Omega_y^+(n)$. Then $\omega(n) = \omega_y(n) + \omega_y^+(n)$, so that if $y = (\log x)^k$, for some constant k, then

$$\sum_{n\leq x}\Big(\omega_y^+\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2 + \omega_y(f_a(n))\Big)^2 \ll \sum_{n\leq x}\Big(\omega_y^+\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2\Big)^2 + \sum_{n\leq x}\omega_y\big(f_a(n)\big)^2. \tag{29}$$

Similarly, the same thing can be said about $\Omega\big(f_a(n)\big)$. We have

$$\sum_{n\leq x}\Big(\Omega_y^+\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2 + \Omega_y\big(f_a(n)\big)\Big)^2 \ll \sum_{n\leq x}\Big(\Omega_y^+\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2\Big)^2 + \sum_{n\leq x}\Omega_y\big(f_a(n)\big)^2. \tag{30}$$

Notice that

$$\begin{aligned}
\sum_{n\le x}\omega_y\big(f_a(n)\big)^2 &\le \sum_{n\le x}\Omega_y\big(f_a(n)\big)^2 \le \sum_{n\le x}\Big(\sum_{\substack{p\mid n\\ p\le y}}\Omega\big(f_a(p)\big) + O\big(\Omega_y(n)\big)\Big)^2\\
&\ll \sum_{n\le x}\Big\{\Big(\sum_{\substack{p\mid n\\ p\le y}}\Omega\big(f_a(p)\big)\Big)^2 + \Omega_y(n)^2\Big\}\\
&\ll \sum_{p_1,p_2\le y}\Omega\big(f_a(p_1)\big)\,\Omega\big(f_a(p_2)\big)\,\frac{x}{p_1p_2} + O\big(x(\log\log y)^2\big)\\
&\ll x\left(\sum_{p\le y}\frac{\Omega\big(f_a(p)\big)}{p}\right)^2 \ll x(\log\log y)^4,
\end{aligned}$$

by Corollary 3. Thus, with our choice of y it suffices to prove that

$$\sum_{n\le x}\Big(\omega_y^+\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2\Big)^2 \ll x(\log\log x)^3. \tag{31}$$

Let us recall that $f_a(n) = \operatorname{lcm}\{f_a(p^\alpha) : p^\alpha \parallel n\}$, and $f_a(p^\alpha)$ divides $p^{\alpha-1}f_a(p)$, for all α. So if $q^2|f_a(n)$, then either (A) $q|n$, or (B) $q^2|f_a(p)$, for some $p|n$, or (C) there exist two primes p_1, p_2, such that $q|f_a(p_1)$, $q|f_a(p_2)$, and $p_1p_2|n$. Also,

$$\omega_y^+\big(f_a(n)\big) \le \Omega_y^+\big(f_a(n)\big) \le \omega_y^+\big(f_a(n)\big) + \Omega\big(f_a(n)\big)\,\delta(n),$$

where $\delta(n) = 1$, if $\exists q^2|f_a(n)$, $q > y$, and $\delta(n) = 0$ otherwise. We immediately have

$$\Omega_y^+\big(f_a(n)\big) = \omega_y^+\big(f_a(n)\big) + O\Big(\delta(n)\Omega\big(f_a(n)\big)\Big),$$

so that

$$\begin{aligned}
\sum_{n\le x}\Big(\omega_y^+\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2\Big)^2 &\ll \sum_{n\le x}\Big(\Omega_y^+\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2\Big)^2\\
&\quad + \sum_{n\le x}\delta(n)\Omega\big(f_a(n)\big)^2\\
&\ll \sum_{n\le x}\Big(\Omega_y^+\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2\Big)^2\\
&\quad + \sum_{\mathrm{A}} + \sum_{\mathrm{B}} + \sum_{\mathrm{C}},
\end{aligned}$$

where the sums $\sum_{\mathrm{A}}$, $\sum_{\mathrm{B}}$, and $\sum_{\mathrm{C}}$ correspond to the cases (A), (B), (C), respectively.

But since $\Omega^2\big(f_a(n)\big) = O\big((\log n)^2\big)$, we see that

$$\sum_A \ll \sum_{n\le x} \Omega(n) \ll x \log\log x,$$

$$\sum_B = (\log x)^2 \sum_{\substack{n\le x\\ n\in B}} \delta(n) \ll (\log x)^2 \sum_{q>y} \sum_{p\equiv 1(\operatorname{mod} q^2)} \frac{x}{p} \ll x\log\log x,$$

$$\sum_C = (\log x)^2 \sum_{\substack{n\le x\\ n\in C}} \delta(n) \ll (\log x)^2 \sum_{q>y} \sum_{\substack{p_1\equiv 1(\operatorname{mod} q)\\ p_2\equiv 1(\operatorname{mod} q)}} \frac{x}{p_1p_2} \ll x(\log\log x)^2,$$

if we choose $y = (\log x)^2$, for example. Here we have used Lemma 3 in estimation of the sums $\sum_B$ and $\sum_C$. This finishes the proof of (31). ■

7 Proof of Theorem 3

By the previous section, it suffices to prove Theorem 3 with $\omega\big(f_a(n)\big)$ replaced by $\Omega\big(f_a(n)\big)$. Evidently

$$\sum_{p|n} \Omega\big(f_a(p)\big) \le \Omega\big(f_a(n)\big) \le \sum_{p|n} \Omega\big(f_a(p)\big) + \Omega(n). \tag{32}$$

Therefore

$$\sum_{\substack{n\le x\\ (a,n)=1}} \Omega\big(f_a(n)\big) = \sum_{\substack{n\le x\\ (a,n)=1}} \Big(\sum_{p|n} \Omega\big(f_a(p)\big)\Big) + O\Big(\sum_{n\le x} \Omega(n)\Big). \tag{33}$$

Now, the condition $(a,n) = 1$ can be removed, if we insert the expression

$$\sum_{d|n,a} \mu(d),$$

where μ is the Möbius function, into our expression above. Thus

$$\begin{aligned}\sum_{d|a} \mu(d) \sum_{p\le x} \Omega\big(f_a(p)\big) \sum_{\substack{n\le x\\ d|n, p|n}} 1 &= \sum_{d|a} \mu(d) \sum_{\substack{p\le x\\ (a,p)=1}} \Omega\big(f_a(p)\big)\left[\frac{x}{pd}\right]\\ &= \frac{\phi(a)}{2a} x(\log\log x)^2 + O(x\log\log x),\end{aligned}$$

by Corollary 3. Hence

$$\sum_{\substack{n\le x\\ (a,n)=1}} \Omega\big(f_a(n)\big) = \frac{\phi(a)}{2a} x(\log\log x)^2 + O(x\log\log x). \tag{34}$$

Similarly, it is possible to prove the second moment estimate along the same lines,

$$\sum_{\substack{n\leq x\\(a,n)=1}} \Omega^2\big(f_a(n)\big) = \frac{\phi(a)}{4a}x(\log\log x)^4 + O\big(x(\log\log x)^3\big)\,. \tag{35}$$

Putting together the estimates (34) and (35) gives us our Theorem 3.

8 Proof of Theorem 4

The function

$$F(n) = \sum_{p|n} \Omega\big(f_a(p)\big)$$

is evidently strongly additive (that is, F satisfies the conditions: $F(mn) = F(m)+F(n)$ for $(m,n)=1$, and $F(p^{\alpha}) = F(p)$ for all α). Recall the extended definition of $f_a(n)$, when $\gcd(a,n)\neq 1$. For all $n\in\mathbb{N}$, we write $n = n_1n_2$, where $p|n_1 \Rightarrow p|a$, and $(n_1,n_2)=1$, and we set $f_a(n) = f_a(n_2)$, and $f_a(1)=1$. We will work with this extended function first

We recall the theorem of Kubilius and Shapiro for strongly additive functions: Let f be strongly additive, and $\mathfrak{A}(x)$ and $\mathfrak{B}(x)$ defined as in (12). If we suppose that for every fixed $\epsilon>0$, as $x\to\infty$, we have

$$\sum_{\substack{p\leq x\\|f(p)|>\epsilon\mathfrak{B}(x)^{1/2}}} \frac{f^2(p)}{p} = o\big(\mathfrak{B}(x)\big)\,, \tag{36}$$

then for any real constant γ,

$$\lim_{x\to\infty}\frac{\#\{n : n\leq x, \frac{f(n)-\mathfrak{A}(x)}{\mathfrak{B}(x)^{1/2}}\leq\gamma\}}{x} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\gamma} e^{-t^2/2}\,dt,$$

or equivalently, for any real constants α and β, with $\alpha<\beta$:

$$\lim_{x\to\infty}\frac{\#\{n : n\leq x, \alpha<\frac{f(n)-\mathfrak{A}(x)}{\mathfrak{B}(x)^{1/2}}\leq\beta\}}{x} = \frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\beta} e^{-t^2/2}\,dt. \tag{37}$$

Now we can apply the Kubilius-Shapiro theorem to $F(n)$, as long as the condition (36) is satisfied. But note that for a prime q, one has $F(q) = \Omega\big(f_a(q)\big)$, and hence we can write:

$$\sum_{\substack{q\le x\\ |F(q)|>\epsilon\sqrt{\mathfrak{B}(x)}}} \frac{F^2(q)}{q} = \sum_{\substack{q\le x\\ |F(q)|>\epsilon\sqrt{\mathfrak{B}(x)}}} \frac{\Omega^2\big(f_a(q)\big)}{q} < \sum_{\substack{q\le x\\ \Omega(q-1)>\epsilon\sqrt{\mathfrak{B}(x)}}} \frac{\Omega^2\big(f_a(q)\big)}{q} < \sum_{\substack{q\le x\\ \Omega(q-1)>\epsilon\sqrt{\mathfrak{B}(x)}}} \frac{\Omega^2(q-1)}{q} = o\big(\mathfrak{B}(x)\big)\,, \tag{38}$$

the last inequality coming from a computation done by Erdős & Pomerance (see [5, p. 348]). But $\Omega\big(f_a(n)\big) = \sum_{p|n}\Omega\big(f_a(p)\big) + O\big(\Omega(n)\big)$, so that we can apply Lemma 1 to deduce the following result:

Theorem 4′ *With the extended definition of $f_a(n)$, under the assumption of a quasi-*GRH, *for all $x > 0$ we have*

$$\lim_{x\to\infty} \frac{H_a(x,\alpha,\beta)}{x} = \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-t^2/2}\,dt, \tag{39}$$

where $H_a(x,\alpha,\beta)$ is defined as

$$H_a(x,\alpha,\beta) \overset{def}{=} \#\left\{ n\le x : \alpha \le \frac{\Omega\big(f_a(n)\big) - \frac{1}{2}(\log\log n)^2}{\frac{1}{\sqrt{3}}(\log\log n)^{3/2}} \le \beta \right\}.$$

Now it is easy to prove Theorem 4, after introducing the condition $(a,n) = 1$ in our enumeration. Let

$$S = \left\{ n\le x : \alpha \le \frac{\Omega\big(f_a(n)\big) - \mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \right\}, \tag{40}$$

with $\mathfrak{A}(x)$ and $\mathfrak{B}(x)$ defined as in (12), and $S(x)$ the cardinality of S. Evidently

$$S(x) \sim x\cdot\Phi(\alpha,\beta),$$

and since

$$\sum_{\substack{n\in S\\ (a,n)=1}} 1 = \sum_{n\in S}\sum_{\substack{d|a\\ d|n}} 1 = \sum_{d|a}\mu(d)\sum_{\substack{n\in S\\ d|n}} 1,$$

we want to count the number of elements $S_d(x)$ in S_d, where

$$S_d = \left\{ n\le x : d|n, \alpha \le \frac{\Omega\big(f_a(n)\big) - \mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \right\}, \tag{41}$$

because then the quantity in Theorem 4 is simply

$$\sum_{d|a} \mu(d) S_d(x). \tag{42}$$

Writing $n = dm$, we see that

$$\Omega\big(f_a(m)\big) \le \Omega\big(f_a(n)\big) \le \Omega\big(f_a(m)\big) + \Omega\big(f_a(d)\big),$$

and since $d|a$, and a is fixed, d itself is less than a constant, giving us

$$\Omega\big(f_a(n)\big) = \Omega\big(f_a(m)\big) + O(1).$$

Therefore

$$S_d(x) = \#\left\{ m \le \frac{x}{d} : \alpha \le \frac{\Omega\big(f_a(n)\big) - \mathfrak{A}(x)}{\sqrt{\mathfrak{B}(x)}} \le \beta \right\} \sim \frac{x}{d} \cdot \Phi(\alpha, \beta). \tag{43}$$

This completes the proof of Theorem 4. ■

9 Proof of Theorems 5 and 6

The analysis of Section 8 clearly shows that a couple of important unconditional theorems could be deduced. As stated earlier, we want to isolate for later clinical study the precise role of the GRH. By invoking Lemma 1, we can deduce Theorems 5 and 6 from the following. We have:

Theorem 5′ *For all constants α and β (with $\alpha \le \beta$), as $x \to \infty$,*

$$\#\left\{ n \le x : \gcd(a,n) = 1, \alpha \le \frac{\Omega\big(f_a(n)\big) - A_1(x)}{\sqrt{B_1(x)}} \le \beta \right\} \sim x \cdot \frac{\phi(a)}{a} \cdot \Phi(\alpha, \beta),$$

where

$$A_1(x) = \sum_{p \le x} \frac{\Omega\big(f_a(p)\big)}{p} \quad \text{and} \quad B_1(x) = \sum_{p \le x} \frac{\Omega^2\big(f_a(p)\big)}{p}.$$

But in order to accurately estimate the behaviour of $A_1(x)$ and $B_1(x)$ one needs to assume a certain GRH.

However, utilizing trivial upper bounds (*i.e.* replacing $f_a(p)$ by $p-1$, and using unconditional results known for $p-1$) on the size of $A_1(x)$ and $B_1(x)$, we have the following useful unconditional theorem:

Theorem 6′ *For any constant $\alpha > 0$, as $x \to \infty$,*

$$\#\left\{ n \le x : \gcd(a,n) = 1, \Omega\big(f_a(n)\big) < \frac{1}{2}(\log\log n)^2 + \frac{\alpha}{\sqrt{3}}(\log\log x)^{3/2} \right\}$$
$$\gtrsim x \cdot \frac{\phi(a)}{a} \cdot \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-t^2/2}\, dt \right).$$

10 Concluding Remarks

It would be desirable if one could remove the assumption of a GRH from our results. However, at this point in time, that doesn't seem to be very realistic, mainly because we are lacking any kind of "mean value" theorems that could serve as replacements of the Bombieri-Vinogradov and the Brun-Titchmarsh theorems in the non-abelian situations. Until one can establish some versions of these theorems (at least in some special cases, or smaller ranges), all we can hope for is to weaken the hypotheses we are using.

As was already noted, the full strength of the Generalized Riemann Hypothesis is not necessary. In fact, a quasi-GRH is always sufficient. Namely, it is enough to assume that there exists a constant $\delta > 0$, such that:

$$\pi\big(x, L_q(a)\big) = \frac{\pi(x)}{q(q-1)} + O(x^{1-\delta}). \tag{44}$$

More generally, for any square-free $k \geq 2$, all we need to do is to assume the Chebotarev Density Theorem with the following error term:

$$\pi\big(x, L_k(a)\big) = \frac{\pi(x)}{k\phi(k)} + O(x^{1-\delta}), \tag{45}$$

for some $\delta > 0$. Currently, the best unconditional error version of (42) is still due to Lagarias & Odlyzko [9] from 1977. They proved that there is a constant $A > 0$ such that

$$\pi\big(x, L_k(a)\big) = \frac{\pi(x)}{k\phi(k)} + O\left(x \exp\left(-A\sqrt{\frac{\log x}{k\phi(k)}}\right)\right). \tag{46}$$

An unconditional result of this kind is good enough to deduce asymptotic behaviour of the sum

$$\sum_{p \leq x} \omega_y\big(f_a(p)\big)$$

for $y < (\log x)^{1/2-\epsilon}$. One can extend the range to $y < (\log x)^{1-\epsilon}$ with some work. But it seems to be beyond reach at the present moment to push this to $y = \exp(\frac{c\log x}{\log\log x})$, for some constant $c > 0$. Such a result would enable us to deduce all of the theorems in this paper unconditionally.

Acknowledgments We thank the referee for several useful remarks.

References

[1] P. Deligne, *Formes modulaires et représentations l-adiques*. Sem. Bourbaki 355, Lecture Notes in Math. **179**, 139–172, Springer Verlag, 1971.

[2] P. D. T. A. Elliott, *Probabilistic Number Theory*. Volume I. & II, Springer Verlag, 1979.

[3] P. Erdős, *On the normal order of prime factors of* $p-1$ *and some related problems concerning Euler's* ϕ*-function*. Quart. J. Math. Oxford Ser. **6**(1935), 205–213.

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

[5] P. Erdős and C. Pomerance, *On the normal number of prime factors of $\phi(n)$*. Rocky Mountain J. Math. **15**(1985), 343–352.

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

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

[8] J. Kubilius, *Probabilistic methods in number theory*. Transl. Math. Monogr. **11**, Rhode Island, 1964.

[9] J. Lagarias and A. Odlyzko, *Effective versions of the Tchebotarev density theorem*. In: Algebraic Number Fields, (ed. A. Fröhlich), Proceedings of the 1975 Durham Symposium, Academic Press, 1975.

[10] D. H. Lehmer, *Ramanujan's function $\tau(n)$*. Duke Math. J. **10**(1943), 483–492.

[11] ———, *The vanishing of Ramanujan's function $\tau(n)$*. Duke Math. J. **14**(1947), 429–433.

[12] R. Murty, *On Artin's conjecture*. J. Number Theory **16**(1983), 147–168.

[13] ———, *Problems in Analytic Number Theory*. Springer Verlag **206**, New York, 2001.

[14] V. K. Murty and R. M. Murty, *Prime divisors of Fourier coefficients of modular forms*. Duke Math. J. **51**(1984), 57–76.

[15] K. Murty and R. Murty, *An analogue of the Erdős-Kac theorem for Fourier coefficients of modular forms*. Indian J. Pure Appl. Math. **15**(1984), 1090–1101.

[16] K. K. Norton, *On the number of restricted prime factors of an integer (I)*. Illinois J. Math. **20**(1976), 681–705.

[17] C. Pomerance, *On the distribution of amicable numbers*. J. Reine Angew. Math. **293/294**(1977), 217–222.

[18] S. Ramanujan, *Highly composite numbers*. Proc. London Math. Soc. (2) **14**(1915), 347–409, see also Collected Works, Oxford, 1927.

[19] F. Saidak, *An elementary proof of a theorem of Délange*. Mathematical Reports of the Royal Society of Canada, **24**(2002), 144–151.

[20] ———, *Non-Abelian Generalizations of the Erdős-Kac Theorem*. Ph.D. Thesis, Queen's University, Kingston, 2001.

[21] ———, *Erdős-Kac type theorems for $\omega\big(f_a(p)\big)$ and $\Omega\big(f_a(p)\big)$ via higher moments*. Acta Math. Univ. Com., submitted.

[22] H. Shapiro, *Distribution functions of additive arithmetic functions*. Proc. Nat. Acad. Sci. USA **42**(1956), 426–430.

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

Department of Mathematics
Queen's University
Kingston, Ontario
K7L 3N6
e-mail: murty@mast.queensu.ca
e-mail: filips@math.ucalgary.ca