Multiplicative arithmetic functions and the generalized Ewens measure

Dor Elboim and Ofir Gorodetsky

Abstract

Random integers, sampled uniformly from $[1, x]$, share similarities with random permutations, sampled uniformly from $S_n$. These similarities include the Erdös–Kac theorem on the distribution of the number of prime factors of a random integer, and Billingsley’s theorem on the largest prime factors of a random integer. In this paper we extend this analogy to non-uniform distributions.

Given a multiplicative function $\alpha : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$, one may associate with it a measure on the integers in $[1, x]$, where $n$ is sampled with probability proportional to the value $\alpha(n)$. Analogously, given a sequence $\{\theta_i\}_{i \geq 1}$ of non-negative reals, one may associate with it a measure on $S_n$ that assigns to a permutation a probability proportional to a product of weights over the cycles of the permutation. This measure is known as the generalized Ewens measure.

We study the case where the mean value of $\alpha$ over primes tends to some positive $\theta$, as well as the weights $\alpha(p) \approx (\log p)^\gamma$. In both cases, we obtain results in the integer setting which are in agreement with those in the permutation setting.

1 Introduction

The analogy between permutations and integers is a well-established one, see the surveys [6, Ch. 1] and [28]. The analogy leads to advancements both in permutations and in integers, see e.g. [26, 27, 17, 18]. This analogy always involves comparing a uniformly drawn integer in $[1, x]$ and a uniformly drawn permutation from $S_n$, where $n \approx \log x$. Our results suggest that the analogy persists even when the chosen measures are not uniform.

We begin with setup and notation. Let $S_n$ be the symmetric group on $\{1, 2, \ldots, n\}$. Given $\pi \in S_n$, we denote by $\ell_1(\pi) \geq \ell_2(\pi) \geq \ldots$ the lengths of the disjoint cycles of $\pi$, arranged in non-increasing order. They satisfy

$$\ell_1(\pi) + \ell_2(\pi) + \ldots = n.$$ 

We let $C_i(\pi)$ be the number of cycles of $\pi$ of length $i$ and denote by $C(\pi)$ the number of cycles in $\pi$.

In recent years, there has been significant activity in the study of permutations sampled according to cycle weights [58, 54, 37, 10, 41, 44, 43, 22, 13, 15, 51, 49]; this model is related to the study of the quantum Bose gas in statistical mechanics, see e.g. [8, 9, 19]. To state the model, let $\theta_1, \ldots, \theta_n$ be non-negative reals (not all zero). The probability of a permutation $\pi$ with respect to the weights $\theta_i$ is defined to be

$$P_{n,\theta_i}(\pi) = \frac{1}{h_n n!} \prod_{i=1}^n \theta_i^{C_i(\pi)} = \frac{1}{h_n n!} \prod_{c \in \pi} \theta_{|c|}$$

where $h_n$ is the normalization constant, known as the partition function, given by

$$h_n = \frac{1}{n!} \sum_{\pi \in S_n} \prod_{i=1}^n \theta_i^{C_i(\pi)}.$$ 

We let $\pi_{n,\theta_i}$ be the random permutation whose probability distribution is (1.1). The measure $P_{n,\theta_i}$ is called a generalized Ewens measure.
We now describe an analogous measure on the positive integers up to \( x \), which is the main object of study in this paper. Given a positive integer \( m \in \mathbb{N} \), denote by \( \rho_1(m) \geq \rho_2(m) \geq \ldots \) the prime factors of \( m \) (repeated according to their multiplicity), arranged in non-increasing order. We have
\[
\log \rho_1(m) + \log \rho_2(m) + \ldots = \log m.
\]
We denote by \( \Omega(m) \) the number of prime factors of \( m \), counted with multiplicity. If \( p^k \mid n \) and \( p^{k+1} \nmid n \), we write \( p^k \nmid n \). This \( k \) is known as the multiplicity of \( p \) in \( n \), and is denoted \( \nu_p(n) \).

A function \( \alpha : \mathbb{N} \to \mathbb{C} \) is called multiplicative if \( \alpha(1) = 1 \) and \( \alpha(nm) = \alpha(n)\alpha(m) \) for every coprime \( n, m \in \mathbb{N} \). Given a non-negative multiplicative function \( \alpha \), we define a measure on the positive integers up to \( x \) by
\[
P_{x,\alpha}(m) = \frac{1}{S(x)}\alpha(m) = \frac{1}{S(x)} \prod_{p^k \mid m} \alpha(p^k)
\]
where the product is over (maximal) prime powers dividing \( m \), and \( S(x) \) is the normalization constant
\[
S(x) = \sum_{m \leq x} \alpha(m).
\]
We let
\[
N_x = N_{x,\alpha}
\]
be the random integer whose probability distribution is \( \mathbb{P}_{x,\alpha} \). In this paper we consider two different families of multiplicative measures on the integers, and compare our results with corresponding generalized Ewens measures.

### 1.1 Constant mean value

We consider multiplicative functions \( \alpha : \mathbb{N} \to \mathbb{R}_{\geq 0} \) satisfying the following two conditions for some \( \theta > 0 \), \( d > -1 \), \( a \in (0,1) \), \( \eta \in (0,1/2] \) and \( r \in (0,2) \):

1. \( \sum_{p \leq x} \frac{\alpha(p) \log p}{p^d} = \theta x + O \left( \frac{x}{\log^a x} \right), \quad \ldots \quad (1.3) \)
2. \( \frac{\alpha(p)}{p^{d+1}} = O(p^{1/2-\eta}), \quad \frac{\alpha(p^k)}{p^{dk}} = O(r^k) \) for all \( k \geq 2 \). \quad \ldots \quad (1.4)

Here \( p \) denotes a prime number. Recall that the prime number theorem says that \( \sum_{p \leq x} \log p \sim x \). Thus, (1.3) should be interpreted as \( \alpha(p)/p^d \) being, on average, of size \( \theta \), and it is a common condition in multiplicative number theory. Condition (1.4) is of a more technical nature. We did not strive to find the most general conditions for our theorem to hold, but rather to find conditions which are easy to work with, lead to short proofs and are satisfied for natural examples. Our result for these weights is the following.

**Theorem 1.1.** Let \( \alpha : \mathbb{N} \to \mathbb{R}_{\geq 0} \) be a multiplicative function satisfying (1.3)–(1.4) with \( \theta > 0 \), \( d > -1 \), \( a \in (0,1) \), \( \eta \in (0,1/2] \) and \( r \in (0,2) \). As \( x \to \infty \) we have
\[
\frac{\Omega(N_x) - \theta \log \log x}{\sqrt{\theta \log \log x}} \sim N(0,1)
\]
and
\[
\left( \frac{\log \rho_1(N_x)}{\log x}, \frac{\log \rho_2(N_x)}{\log x}, \ldots \right) \overset{d}{\to} \text{PD}(\theta),
\]
where \( \text{PD}(\theta) \) is the Poisson–Dirichlet distribution with parameter \( \theta \) (defined in §).

Here the arrows indicate convergence in distribution. The proof of the first part of Theorem 1.1 applies to \( \omega(N_x) \) as well, where \( \omega(n) \) counts the number of prime factors of \( n \) without multiplicities, and the result is the same. The prototypical example of a function \( \alpha \) satisfying the conditions is \( \theta^\omega(n) \) (with \( d = 0 \), \( \eta = 1/2 \), \( r = 1 \) and any \( a > 0 \)).
The case $\alpha \equiv 1$ of (1.5) is the Erdős–Kac theorem [23]. Our proof of it is a generalization of a proof given by Billingsley [11] to the original Erdős–Kac theorem with $\alpha \equiv 1$. The case $\alpha \equiv 1$ of (1.6) is Billingsley’s theorem [12]. Our proof of it is a generalization of Donnelly and Grimmett’s proof [16], who elucidated Billingsley’s result.

The Ewens measure with parameter $\theta > 0$ is a measure on $S_n$, which may be defined by taking $\theta_i = \theta$ in the definition of the generalized Ewens measure. The partition function is $\left( \frac{n+\theta-1}{n} \right)$ in this case. This measure has first appeared in the study of population genetics [25]. The Ewens measure has found many practical applications, through its connection with Kingman’s coalescent process [36] and its occurrence in non-parametric Bayesian statistics [5].

Theorem 1.1 should be compared with two results on the Ewens measure, one on $C(\pi_{n,\theta})$ (the number of cycles in $\pi_{n,\theta}$) by Hansen [31], and another on $\left( \frac{\ell_1(\pi_{n,\theta})}{n}, \frac{\ell_2(\pi_{n,\theta})}{n}, \ldots \right)$ by Watterson [55].

**Theorem 1.2 (Hansen, Watterson).** Let $\theta > 0$. As $n \to \infty$ we have

$$\frac{C(\pi_{n,\theta}) - \theta \log n}{\sqrt{\theta \log n}} \xrightarrow{d} N(0,1)$$

and

$$\left( \frac{\ell_1(\pi_{n,\theta})}{n}, \frac{\ell_2(\pi_{n,\theta})}{n}, \ldots \right) \xrightarrow{d} PD(\theta).$$

The first part of Theorem 1.2 was proven under more general conditions, e.g. when $\sum_{i=1}^{n} \theta_i/n \to \theta$ sufficiently fast, see Lugo [37] and the works of Manstavičius [38, 39, 40].

The similarity of Theorem 1.1 and Theorem 1.2 is most apparent for functions $\alpha$ where $\alpha(p) \approx \theta$. It suggests an analogy between permutations chosen according to the Ewens measure and integers chosen according to multiplicative weights. We now discuss previous works.

1.1.1 Erdős–Kac

In a series of works, Alladi [1, 2, 3, 4] proved a generalization of Erdős–Kac involving weights $\alpha$ as well. His proof uses the combinatorial sieve and he requires $\alpha$ to satisfy a ‘level-of-distribution’ condition which is not always easily verified. A related (but simpler) sieve-theoretic approach to Erdős–Kac and its generalizations was introduced by Granville and Soundararajan [30]. This approach was used by Khan, Milinovich and Subedi to prove an Erdős–Kac theorem with weights being $d_k$, the $k$th divisor function [35].

See Elliott [20, 21] for a treatment of the Erdős–Kac theorem with weights being the standard divisor function $d_2$ and its real powers. Tenenbaum proved in [53, Cor. 2.5] an impressively general weighted Erdős–Kac theorem, but unlike Theorem 1.1 he requires $\alpha(p)/p^d$ to be uniformly bounded. Both Elliott and Tenenbaum use characteristic functions and complex analysis while we avoid these.

1.1.2 Billingsley

Arratia, Kochman and Miller proved an analogue of Billingsley’s theorem for normed arithmetic semigroups satisfying certain growth conditions [7, Thm. 2]. A commutative semigroup $S$ is called normed arithmetic semigroup if it contains an identity element and admits unique factorization into ‘prime’ elements. Furthermore, it should come equipped with a multiplicative norm function $s \mapsto |s| \in \mathbb{R}_{>0}$, such that $N(x) = \# \{ s \in S : |s| \leq x \}$ is a finite number for each $x > 0$. There is small overlap between [7, Thm. 2] and the second part of Theorem 1.1 as there are multiplicative functions $\alpha$ satisfying (1.3)–(1.4) and coinciding with $\alpha_S(n) := \# \{ s \in S : |s| = n \}$ for some normed arithmetic semigroup $S$.

1.2 Polynomialsly-growing weights

In the permutation setting, the measure $\mathbb{P}_{n,\theta_i}$ was studied extensively in the case of polynomially-growing cycle weights, that is

$$\theta_n \approx An^\gamma,$$

(1.7)

see [24, 41, 22, 15, 13]. Ercolani and Ueltschi proved the following in [22, Thm. 5.1].
**Theorem 1.3** (Ercolani and Ueltschi). Let $\gamma > 0$, and take

$$\theta_n = \frac{\Gamma(\gamma + n + 1)}{n!} = (1 + o(1))n^\gamma. \quad (1.8)$$

As $n \to \infty$ we have

$$EC(\pi_n, \theta_i) \sim n^{\gamma + 1} \left(\frac{\Gamma(\gamma)}{\gamma^\gamma}\right)^{\frac{1}{\gamma + 1}}.$$ 

The specific choice (1.8) simplifies the computations greatly. Maples, Nikeghbali and Zeindler \[41, Cor. 1.2\] were able to prove that $C(\pi_n, \theta_i)$ converges, after an explicit normalization, to a normal distribution (for $\theta_n$ as in (1.8) and also scalar multiples of it).

Next we describe a result of Ercolani and Ueltschi \[22, Thm. 5.1\] about a permutation statistic which we have yet to discuss, $L_1(\pi)$. This is the length of the cycle of a permutation $\pi$ which contains the element 1. In the Ewens case $\theta_1 = \theta$, it is known that $L_1(\pi_n, \theta_i)/n$ converges in distribution to a beta distribution \[22, \S6\]. For polynomially-growing weights, Ercolani and Ueltschi proved that $L_1(\pi_n, \theta_i)$ exhibits a very different behavior. First, the order of magnitude of $L_1(\pi_n, \theta_i)$ in this case is $n^{1/(\gamma+1)} = o(n)$ and not $n$. Second, the limiting distribution is a gamma distribution, whose definition is recalled in (1.8).

**Theorem 1.4** (Ercolani and Ueltschi). Let $\gamma > 0$, and take $\theta_n$ as in (1.8). Then, as $n \to \infty$,

$$\frac{L_1(\pi_n, \theta_i)}{n^{\gamma+1}} \overset{d}{\to} \text{gamma}(\gamma + 1, \Gamma(\gamma + 1)^{1/(\gamma+1)}).$$

See Dereich and Mörters \[15\] for finer results about $L_1(\pi_n, \theta_i)$ for similar weights.

We derive number-theoretic analogues of Theorems 1.3 and 1.4. Since there is no such thing as ‘a prime divisor of $n$ containing a fixed element’, we must turn to a different interpretation of $L_1(\pi)$.

**Definition 1.5.** Let $a = \{a_j\}_{j \geq 1}$ be a sequence of non-negative reals summing to $0 < S < \infty$. A size-biased sampling of an element from $a$ is a random variable $X$ whose distribution is given by

$$\mathbb{P}(X = a_j) = a_j \cdot \frac{\{i \geq 1 \mid a_i = a_j\}}{S}.$$ 

Suppose that $\mathbb{P}$ is some conjugation-invariant measure on $S_n$ (e.g. $\mathbb{P}_{n, \theta_i}$). If $\pi \in S_n$ is sampled according to $\mathbb{P}$, then the distribution of $L_1(\pi)$ coincides with the distribution of a typical cycle of $\pi$, that is: of a size-biased sampling of an element from $\{\ell_i(\pi)\}_{i \geq 1}$. See Lemma 2.3 below for the proof. It is now clear how to define an integer analogue of $L_1(\pi_n, \theta_i)$: given $N_x$, we define $P_1(N_x)$ by letting $\log P_1(N_x)$ be a size-biased sampling of an element from $\{\log p_i(N_x)\}_{i \geq 1}$. We think of $P_1(N_x)$ as a typical prime divisor of $N_x$.

In the integer setting, the polynomial weights \[1.7\] correspond to

$$\alpha(p) \approx K \log^\gamma p.$$ 

For our results, we require that for all primes $p$,

(I) $\alpha(p) = K \log^\gamma p + O(\log^{-2} p), \quad (1.9)$

(II) $\sum_{k \geq 2} \frac{ko(p^{k})}{p^{k}} = O\left(\frac{1}{p \log^2 p}\right). \quad (1.10)$

**Theorem 1.6.** Let $\alpha: \mathbb{N} \to \mathbb{R}_{\geq 0}$ be a multiplicative function satisfying (1.9)–(1.10) for some $K > 0$, $\gamma > 0$. As $x \to \infty$ we have

$$\frac{\log P_1(N_x)}{(\log x)^{\gamma+1}} \overset{d}{\to} \text{gamma}(\gamma + 1, (K \Gamma(\gamma + 1))^{1/(\gamma+1)}),$$

and

$$\mathbb{E}\Omega(N_x) \sim (\log x)^{\gamma+1} \left(\frac{K \Gamma(\gamma)}{\gamma^\gamma}\right)^{\frac{1}{\gamma+1}}.$$
The proof of the second part of Theorem 1.6 applies to \( \omega(N_x) \) as well. It is interesting that although the measure \( \mathbb{P}_{x, \alpha} \) assigns larger weights to larger primes, the typical prime factors of \( N_x \) are much smaller than in the case of uniformly drawn integers between 1 and \( x \). Indeed, it follows from the first part of Theorem 1.6 that \( \log P_1(N_x) = o(\log x) \), while from Theorem 1.1 it follows that \( \log P_1(N_x) = \Theta(\log x) \) for uniform integers.

The main ingredient in the proof of Theorem 1.6 is the asymptotics of \( \sum_{n \leq x} \alpha(n) \) for multiplicative \( \alpha \) obeying (1.9)–(1.10). This unusual sum was studied by Schwarz [50] and Marenich [42]. As they both appeal to the same Tauberian theorem [34, Thm. 1], no error term is obtained. We prove the following estimate.

**Theorem 1.7.** Let \( \alpha : \mathbb{N} \to \mathbb{R}_{\geq 0} \) be a multiplicative function satisfying (1.9)–(1.10) for some \( K > 0, \gamma > 0 \). There exists \( \varepsilon > 0 \) for which

\[
\frac{1}{x} \sum_{n \leq x} \alpha(n) = \left( 1 + O((\log x)^{-\varepsilon}) \right) A_\alpha \exp\left( B(\log x)^{\frac{\gamma}{\gamma + 1}} \right) \tag{1.11}
\]

as \( x \to \infty \), where \( A_\alpha \) is a positive constant and

\[
B = \left( 1 + \frac{1}{\gamma} \right) (K \Gamma(\gamma + 1))^{\frac{1}{\gamma + 1}}. \tag{1.12}
\]

The proof of Theorem 1.7 shows that

\[
A_\alpha = A \prod_p \frac{\sum_{k \geq 0} \alpha(p^k)}{p^{\log p} \log p} \tag{1.13}
\]

for a constant \( A \) depending only on \( K \) and \( \gamma \). Additionally, one may take \( \varepsilon = 1/(\gamma + 1) \) if \( \gamma > 2 \), and \( \varepsilon \) arbitrarily close to \( \gamma / (2(\gamma + 1)) \) otherwise.

**Conventions**

In the arguments below, we think of the function \( \alpha \) as fixed, and write \( N_x \) for \( N_{x, \alpha} \). We denote the set of prime numbers by \( \mathcal{P} \), and reserve the letter \( p \) for primes. The letters \( C \) and \( c \) always denote positive constants, which may vary from line to line. However, \( C \) and \( c \) depend only on the arithmetic function \( \alpha \) considered unless otherwise stated. The arguments in the proofs always hold for sufficiently large \( x \). The notation \( A \gg 1 \) indicates that \( A \) is sufficiently large.

**Acknowledgements**

The second author was supported by the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreements nos 786758 and 851318). We thank the anonymous referee for useful comments and suggestions.

**2 Preliminaries from probability theory**

We denote by beta\((\alpha, \beta)\) the beta distribution with shape parameters \( \alpha \) and \( \beta \) whose density with respect to Lebesgue measure on \([0, 1]\) is given by

\[
\frac{\Gamma(\alpha + \beta) x^{\alpha-1} (1-x)^{\beta-1}}{\Gamma(\alpha) \Gamma(\beta)}, \quad x \in [0, 1].
\]

We denote by gamma\((\alpha, \beta)\) the gamma distribution with shape parameters \( \alpha \) and \( \beta \) whose density with respect to Lebesgue measure on \([0, \infty)\) is given by

\[
\frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x \in [0, \infty).
\]
We define the Poisson–Dirichlet distribution with parameter $\theta$, denoted by $\text{PD}(\theta)$. Let $Y_1, Y_2, \ldots$ be an i.i.d. sequence of random variables with beta$(1, \theta)$ distribution. Define the sequence

$$Z_j = (1 - Y_1) \cdots (1 - Y_{j-1}) Y_j, \quad j \geq 1.$$ 

Intuitively $Z_1$ takes a beta$(1, \theta)$-distributed fraction of the unit interval. Conditioned on $Z_1$, $Z_2$ takes a beta$(1, \theta)$-distributed fraction of the remaining part of the interval, etc. Finally, the PD$(\theta)$ distribution is defined to be the distribution of the sequence $(Z_j)_{j \geq 1}$ arranged in non-increasing order.

In the proof of Theorem 1.1, we establish the convergence in distribution to PD$(\theta)$ by convergence of a certain sequence to a sequence of independent beta$(1, \theta)$ random variables. To this end we define the size-biased permutation of a sequence of random variables. Let $X_1, X_2, \ldots$ be a non-increasing sequence of random variables such that

$$\sum_{j=1}^{\infty} X_j = 1, \quad \text{almost surely.}$$

A sized-biased permutation $(\tilde{X}_i)_i$ of the sequence $(X_i)_i$ is a random reordering of the elements of the sequence such that for any $j \geq 1$

$$\mathbb{P}(\tilde{X}_1 = X_j \mid X_1, X_2, \ldots) = \frac{X_j \cdot \left(\{j' \geq 1 \mid X_{j'} = X_j\} - \{j' < k \mid \tilde{X}_{j'} = X_j\}\right)}{1 - X_1 - \ldots - X_{k-1}}.$$ 

and inductively for $k \geq 1$

$$\mathbb{P}(\tilde{X}_k = X_j \mid \tilde{X}_1, \ldots, \tilde{X}_{k-1}, X_1, X_2, \ldots) = \frac{X_j \cdot \left(\{j' \geq 1 \mid X_{j'} = X_j\} - \{j' < k \mid \tilde{X}_{j'} = X_j\}\right)}{1 - X_1 - \ldots - X_k}.$$ 

The sized-biased permutation can be used to reconstruct the beta$(1, \theta)$ random variables from the PD$(\theta)$ distribution in the following sense.

**Proposition 2.1.** Let $X_1, X_2, \ldots$ be a non-increasing sequence of random variables with

$$\sum_{j=1}^{\infty} X_j = 1, \quad \text{almost surely.}$$

Then, the sequence $(X_1, X_2, \ldots)$ has the PD$(\theta)$ distribution if and only if the sequence

$$\left(\frac{\tilde{X}_j}{1 - X_1 - \ldots - X_{j-1}}\right)^{\infty}_{j=1}$$ 

is an i.i.d. sequence of beta$(1, \theta)$ random variables.

For a discussion and references for Proposition 2.1 see the introduction of [48]. We use the following result in order to prove the convergence to PD$(\theta)$ in Theorem 1.1.

**Proposition 2.2.** For any $n \geq 1$, let $X_1^{(n)}, X_2^{(n)}, \ldots$ be a non-increasing sequence of random variables with

$$\sum_{j=1}^{\infty} X_j^{(n)} = 1, \quad \text{almost surely.}$$

Suppose that

$$\left(\frac{\tilde{X}_j^{(n)}}{1 - X_1^{(n)} - \ldots - X_{j-1}^{(n)}}\right)^{\infty}_{j=1} \xrightarrow{d} (Y_1, Y_2, \ldots), \quad n \to \infty,$$ 

(2.1)

where $Y_1, Y_2, \ldots$ is a sequence of i.i.d. beta$(1, \theta)$ random variables. Then, we have

$$(X_1, X_2, \ldots) \xrightarrow{d} \text{PD}(\theta), \quad n \to \infty.$$ 

(2.2)
Proof. Consider the function \( g: [0,1]^\mathbb{N} \to [0,1]^\mathbb{N} \) defined by
\[
g(y_1, y_2, \ldots) = ((1 - y_1) \cdots (1 - y_{k-1}) y_k)_{k=1}^{\infty},
\]
and the function
\[
r: \{ (z_1, z_2, \ldots) \in [0,1]^\mathbb{N} : \sum_{j=1}^{\infty} z_j \leq 1 \} \to [0,1]^\mathbb{N}
\]
that takes a sequence and returns the same sequence in non-increasing order. Since \( r \circ g \) is continuous in the product topology on \([0,1]^\mathbb{N}\) we have by (2.1) that
\[
(r \circ g)(\tilde{X}_1(n) - \tilde{X}_2(n) - \ldots - \tilde{X}_{j-1}(n))_{j=1}^{\infty} \overset{d}{\rightarrow} (r \circ g)(Y_1, Y_2, \ldots), \ n \to \infty. \tag{2.3}
\]
By the definition of the Poisson–Dirichlet distribution \((r \circ g)(Y_1, Y_2, \ldots) \sim PD(\theta)\) and moreover, since
\[
g^{-1}(z_1, z_2, \ldots) = \left( \frac{z_j}{1 - z_1 - \ldots - z_{j-1}} \right)_{j=1}^{\infty},
\]
we have that
\[
(r \circ g) \left( \left( \frac{\tilde{X}_j(n)}{1 - \tilde{X}_1(n) - \ldots - \tilde{X}_{j-1}(n)} \right)_{j=1}^{\infty} \right) = r(\tilde{X}_1, \tilde{X}_2, \ldots) = (X_1, X_2, \ldots).
\]
Thus, (2.3) simplifies to (2.2), as needed. \(\square\)

Lemma 2.3. Let \( \mathbb{P} \) be a conjugation-invariant measure on \( S_n \). Given \( \pi \in S_n \), let \( L_1(\pi) \) be the size of the cycle containing 1, and let \( Ty(\pi) \) be a random variable which is a size-biased sampling of a cycle of \( \pi \), according to Definition 1.5. Then \( L_1(\pi) \) and \( Ty(\pi) \) have the same distribution, where \( \pi \) is a permutation drawn according to \( \mathbb{P} \).

Proof. By using the conjugation-invariance, for any \( k \in \mathbb{N} \) we have
\[
\mathbb{P}(L_1(\pi) = k) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}(L_i(\pi) = k)
\]
where \( L_i(\pi) \) is the size of the cycle of \( \pi \) containing \( i \). Letting \( U_n \) be a uniformly drawn integer from \( \{1,2,\ldots,n\} \), we have just shown that
\[
\mathbb{P}(L_1(\pi) = k) = \mathbb{P}(L_{U_n}(\pi) = k).
\]
The size of the cycle of \( \pi \) which contains \( U_n \) is a size-biased sampling of a cycle of \( \pi \), which concludes the proof. \(\square\)

3 Asymptotically Ewens measure

3.1 Multiplicative number theory

Theorem 3.1 (La Bretèche and Tenenbaum [14]). Suppose a multiplicative function \( \alpha: \mathbb{N} \to \mathbb{R}_{\geq 0} \) satisfies
\[
\sum_{p \leq x} \alpha(p) \log p = \theta x + O \left( \frac{x}{\log x} \right), \quad x \to \infty,
\]
for some \( \theta > 0 \) and \( a \in (0,1) \). Suppose further that
\[
\sum_p \left( \frac{\alpha(p)^2}{p^{2a}} + \sum_{n \geq 2} \frac{\alpha(p^n)}{p^{n\sigma}} \right) < \infty
\]
for some \( \sigma \in (0,1) \). Then, for all sufficiently large \( x \),

\[
\sum_{n \leq x} \alpha(n) = (1 + O(\log^{-a} x)) \Gamma(\theta)^{-1} \prod_p \left( \sum_{i \geq 0} \frac{\alpha(p^i)}{p^i} \right)^{\theta} x \log^{\theta-1} x.
\]

**Corollary 3.2.** Suppose a multiplicative function \( \alpha: \mathbb{N} \to \mathbb{R}_{\geq 0} \) satisfies (1.3)–(1.4) for some \( \theta > 0 \), \( d > -1 \), \( a \in (0,1) \), \( \eta \in (0,1/2] \) and \( r \in (0,2) \). Then, for all sufficiently large \( x \),

\[
\sum_{n \leq x} \alpha(n) = (1 + O(\log^{-a} x)) A_\alpha x^{d+1} \log^{\theta-1} x
\]

where \( A_\alpha = (d + 1)^{-1} \Gamma(\theta)^{-1} \prod_p \left( \sum_{i \geq 0} \frac{\alpha(p^i)}{p^{i(d+1)}} \right)(1 - 1/p)^\theta > 0 \).

**Proof.** Let \( \alpha_d(n) = \alpha(n)/n^d \). This function is still multiplicative. If \( \alpha \) satisfies (1.3)–(1.4), then \( \alpha_d \) satisfies these conditions as well with the same parameters except that the parameter \( d \) is now 0.

Evidently, the conditions of Theorem 3.1 hold for \( \alpha_d \), with the same \( \theta \) and \( a \), and with any \( \sigma \in (0,1) \) that satisfies \( \sigma > \max\{\log r/\log 2, 1 - \eta\} \). Thus, we conclude from Theorem 3.1 that

\[
\sum_{n \leq x} \alpha_d(n) = (1 + O(\log^{-a} x)) A_\alpha x^{d+1} \log^{\theta-1} x
\]

as \( x \to \infty \). Using integration by parts and (3.2) we have

\[
\sum_{n \leq x} \alpha(n) = \sum_{n \leq x} n^d \alpha_d(n) = x^d \left( \sum_{n \leq x} \alpha_d(n) \right) - \int_1^x \left( \sum_{n \leq t} \alpha_d(n) \right) (t^d)' \, dt
\]

\[
= (1 + O(\log^{-a} x)) A_\alpha x^{d+1} \log^{\theta-1} x - \int_2^x A_\alpha(1 + O(\log^{-a} t))(t \log^{\theta-1} t)(dt^{d-1}) \, dt,
\]

which gives (3.1) with \( A_\alpha = A_{\alpha_d}/(d + 1) \). \( \square \)

**Lemma 3.3.** Suppose a multiplicative function \( \alpha: \mathbb{N} \to \mathbb{R}_{\geq 0} \) satisfies (1.3)–(1.4) for some \( \theta > 0 \), \( d > -1 \), \( a \in (0,1) \), \( \eta \in (0,1/2] \) and \( r \in (0,2) \). Let \( p \) be a prime. We have

\[
\sum_{n \leq x \atop p \mid n} \alpha(n) \leq \frac{C(1 + \alpha(p)p^{-d})}{p} x^{d+1} (\log x)^{\max\{\theta-1,0\}}.
\]

(3.3)

If furthermore \( p \leq \sqrt{x} \) we have

\[
\sum_{n \leq x \atop p \mid n} \alpha(n) \leq \frac{C(1 + \alpha(p)p^{-d})}{p} x^{d+1} (\log x)^{\theta-1}.
\]

(3.4)

**Proof.** By multiplicativity of \( \alpha \), we have

\[
\sum_{n \leq x \atop p \mid n} \alpha(n) = \sum_{k=1}^{[\log_p x]} \sum_{n \leq x \atop p^k \mid n} \alpha(n) \leq \sum_{k=1}^{[\log_p x]} \alpha(p^k) \sum_{l=1}^{[\log_p x]} \alpha(l) \leq \sum_{k=1}^{[\log_p x]} \alpha(p^k) \sum_{l=1}^{[\log_p x]} \alpha(l).
\]

(3.5)

By Corollary 3.2 there exists a constant \( C \) such that

\[
\sum_{n \leq y} \alpha(n) \leq Cy^d \log^{\theta-1} y \leq Cy^d \log(y+1)^{\max\{\theta-1,0\}}.
\]

(3.6)
for all \( y \geq 1 \) (one needs to consider \( y \in [1,2) \) separately from \( y \geq 2 \)). Note that the right-hand side of (3.6) is monotone increasing in \( y \) and so is \((\log(y+1))^{\max\{\theta-1,0\}}\). By (3.5) and (3.6) we have

\[
\sum_{n \leq x \atop p \mid n} \alpha(n) \leq C \sum_{k=1}^{\lfloor \log_p x \rfloor} \frac{\alpha(p^k)}{p^{k(d+1)}} x^{d+1}(\log(x+1))^{\max\{\theta-1,0\}}.
\]

(3.7)

From (1.4) and (3.7) we obtain

\[
\sum_{n \leq x \atop p \mid n} \alpha(n) \leq Cx^{d+1}(\log(x+1))^{\max\{\theta-1,0\}} \left( \frac{\alpha(p)}{p^{d+1}} + \sum_{k \geq 2} \left( \frac{r}{p} \right)^k \right) \leq C(1 + \alpha(p)p^{-d}) x^{d+1}(\log(x+1))^{\max\{\theta-1,0\}},
\]

and so (3.3) holds. To prove (3.4), we split the right-hand side of (3.3) into two sums, one for terms with \( p^k \leq \sqrt{x} \) and another for the rest. Set \( L := \lfloor \log_p \sqrt{x} \rfloor + 1 \geq 2 \). For \( 1 \leq k < L \), we have \( \log^{\theta-1}(x/p^k) \leq C \log^{\theta-1} x \), and so Corollary 3.2 gives

\[
\sum_{l=1}^{\lfloor \frac{x}{p^k} \rfloor} \alpha(l) \leq C \left( \frac{x}{p^k} \right)^{d+1} \log^{\theta-1} x.
\]

From (1.4) we obtain

\[
\sum_{k=1}^{L-1} \alpha(p^k) \sum_{l=1}^{\lfloor \frac{x}{p^k} \rfloor} \alpha(l) \leq Cx^{d+1} \log^{\theta-1} x \sum_{k=1}^{L-1} \frac{\alpha(p^k)}{p^{k(d+1)}} \leq C(1 + \alpha(p)p^{-d}) x^{d+1} \log^{\theta-1} x.
\]

(3.8)

When \( L \leq k \leq \lfloor \log_p x \rfloor \) we use the bound

\[
\sum_{l=1}^{\lfloor \frac{x}{p^k} \rfloor} \alpha(l) \leq C \left( \frac{x}{p^k} \right)^{d+1} \log^{\theta} x,
\]

which follows from (3.6), to obtain, from (1.4) again, that

\[
\sum_{k=L}^{\lfloor \log_p x \rfloor} \alpha(p^k) \sum_{l=1}^{\lfloor \frac{x}{p^k} \rfloor} \alpha(l) \leq Cx^{d+1} \log^{\theta} x \sum_{k=L}^{\infty} \frac{\alpha(p^k)}{p^{k(d+1)}} \leq Cx^{d+1} \log^{\theta} x \sum_{k=L}^{\infty} \left( \frac{x}{p} \right)^k \leq Cx^{d+1} \log^{\theta} x \left( \frac{x}{p} \right)^L
\]

(3.9)

From (3.8) and (3.9) we obtain (3.4). \( \square \)

Lemma 3.4. Let \( \alpha : \mathcal{P} \to \mathbb{R}_{\geq 0} \). Fix \( \theta \in \mathbb{R} \) and define a function \( E \) via

\[
\sum_{p \leq x} \alpha(p)p = \theta x + E(x).
\]

Let \( I \subseteq [2, \infty) \) be a finite closed interval and \( g : I \to \mathbb{R} \) be a differentiable function. Then

\[
\left| \sum_{p \in I} \alpha(p)g(p) - \theta \frac{g(t)}{\log t} dt \right| \leq \max_{t \in I} |g(t)\alpha(t)| + 2 \max_{t \in I} \left| \frac{g(t)}{\log t} \right| |E(t)| + \int_I \left| \left( \frac{g(t)}{\log t} \right)' \right| |E(t)| dt.
\]

\[
\sum_{p \leq x} \alpha(p) \log p = \theta x + E(x).
\]
Proof. We write $I = [x, y]$. We shall work with the half-open interval $(x, y]$; we may do so because the contribution of $[x, y] \setminus (x, y) = \{x\}$ to $\sum_{p \in I} \alpha(p)g(p)$ is at most $\max_{t \in I} |g(t)\alpha(t)|$. Using integration by parts,

$$\frac{\theta(y)}{\log y} \int_x^y \frac{g(t)}{t} \, dt = \frac{\theta(y)}{\log y} \int_x^y 1 \, dt - \int_x^y \theta(t-x) \left( \frac{g(t)}{\log t} \right)' \, dt = \frac{\theta(y)}{\log y} (y-x) - \int_x^y \theta(t-x) \left( \frac{g(t)}{\log t} \right)' \, dt$$  \hspace{1cm} (3.10)

and, by Abel’s summation formula,

$$\sum_{x < p \leq y} \alpha(p)g(p) = \sum_{x < n \leq y} 1 \{ n \text{ is prime} \} \alpha(n) \log n \cdot \frac{g(n)}{\log n}$$

$$= \frac{g(y)}{\log y} \left( \sum_{x < p \leq y} \alpha(p) \log p \right) - \int_x^y \left( \sum_{x < p \leq t} \alpha(p) \log p \right) \left( \frac{g(t)}{\log t} \right)' \, dt$$ \hspace{1cm} (3.11)

From (3.10), (3.11) and the definition of $E$ we obtain

$$\sum_{p \in (x, y]} \alpha(p)g(p) - \frac{\theta(y)}{\log y} \int_x^y \frac{g(t)}{t} \, dt = \frac{g(y)}{\log y} (E(y) - E(x)) - \int_x^y \left( \frac{g(t)}{\log t} \right)' (E(t) - E(x)) \, dt$$

$$= \frac{g(y)}{\log y} (E(y) - E(x)) - \int_x^y \left( \frac{g(t)}{\log t} \right)' E(t) \, dt + E(x) \left( \frac{g(y)}{\log y} - \frac{g(x)}{\log x} \right)$$

$$= \frac{g(y)}{\log y} E(y) - \frac{g(x)}{\log x} E(x) - \int_x^y \left( \frac{g(t)}{\log t} \right)' E(t) \, dt,$$

and now the required estimate follows by the triangle inequality. $\square$

3.2 Proof of second part of Theorem 1.1

3.2.1 Auxiliary lemma

Lemma 3.5. Fix $k \geq 1$. Let $\delta \in (0, 1)$ and suppose that $x$ is sufficiently large in terms of $\delta$. Let $p_1, \ldots, p_k$ be distinct primes such that $p_1 \cdots p_k \leq x^{1-\delta}$ and $p_i \geq x^\delta$ for all $1 \leq i \leq k$. Suppose a multiplicative function $\alpha : \mathbb{N} \to \mathbb{R}_{\geq 0}$ satisfies (1.3)–(1.4) for some $\theta > 0$, $d > -1$, $\alpha \in (0, 1)$, $\eta \in (0, 1/2]$ and $r \in (0, 2)$. We have

$$\sum_{n = \lfloor \delta x \rfloor \atop p_1, \ldots, p_k | n} \alpha(n) \geq (1 - 3\delta^{d+1})A_\alpha \left( \prod_{i=1}^k \frac{\alpha(p_i)}{p_i^{d+1}} \right) x^{d+1} \log^{\theta - 1} \left( \frac{x}{p_1 \cdots p_k} \right),$$

where $A_\alpha$ is the constant given in Corollary 3.2.

Proof. By the multiplicativity of $\alpha$,

$$\sum_{n = \lfloor \delta x \rfloor \atop p_1, \ldots, p_k | n} \alpha(n) = \sum_{l = \lfloor \delta x \rfloor \atop p_1, \ldots, p_k | l} \alpha(l \cdot p_1 \cdots p_k) = \alpha(p_1) \cdots \alpha(p_k) \sum_{l = \lfloor \delta x \rfloor \atop \forall i, p_i | l} \alpha(l).$$

For the sum in the right-hand side we have the naive lower bound

$$\sum_{l = \lfloor \delta x \rfloor \atop \forall i, p_i | l} \alpha(l) \geq \sum_{l = \lfloor \delta x \rfloor \atop p_i | l} \alpha(l) - \sum_{i=1}^k \sum_{l = \lfloor \delta x \rfloor \atop p_i | l} \alpha(l).$$  \hspace{1cm} (3.12)
The first sum in the right-hand side of (3.12) can be estimated by applying Corollary 3.2 twice, and the second sum can be bounded from above by (3.3). We thus obtain

\[
\sum_{t=\lceil \frac{x}{p_k} \rceil}^{\lfloor \frac{x}{p_k} \rfloor} \alpha(t) \geq A_\alpha \left(1 - \frac{5}{2}\delta^{d+1}\right) \left(\frac{x}{p_1 \cdot \cdots \cdot p_k}\right)^{d+1} \log^{\theta-1} \left(\frac{x}{p_1 \cdot \cdots \cdot p_k}\right)
\]

\[
- C_k \max_{1 \leq i \leq k} (p_i^{-1} + \alpha(p_i) p_i^{-d-1}) \left(\frac{x}{p_1 \cdot \cdots \cdot p_k}\right)^{d+1} \log^{\max\{\theta-1,0\}} \left(\frac{x}{p_1 \cdot \cdots \cdot p_k}\right)
\]

\[
\geq A_\alpha \left(1 - 3\delta^{d+1}\right) \left(\frac{x}{p_1 \cdot \cdots \cdot p_k}\right)^{d+1} \log^{\theta-1} \left(\frac{x}{p_1 \cdot \cdots \cdot p_k}\right)
\]

for sufficiently large \(x\), as needed. Here we made use of \(\alpha(p_i)/p_i^{d+1} \leq p_i^{-1/2}\) (by (1.4)) and \(p_i \geq x^\delta\).

\[\square\]

### 3.2.2 Conclusion of proof

From Corollary 3.2 it follows that for any given \(\varepsilon \in (0,1)\),

\[
\mathbb{P}\left(\frac{\log N_x}{\log x} \leq 1 - \varepsilon\right) = \mathbb{P}\left(N_x \leq x^{1-\varepsilon}\right) = \sum_{n \leq x} \frac{\alpha(n)}{\sum_{n \leq x} \alpha(n)} \sim x^{-\varepsilon(d+1)}(1 - \varepsilon)^{\theta-1} \to 0
\]
as \(x \to \infty\). Hence, \(\log N_x/\log x\) tends in distribution to 1. Thus, it suffices to prove that

\[
(X_j)_{j=1}^\infty := \left(\frac{\log p_j(N_x)}{\log N_x}\right)_{j=1}^\infty \overset{d}{\to} \text{PD}(\theta)
\]
as \(x \to \infty\). By Proposition 3.2, it suffices to prove the following proposition.

**Proposition 3.6.** We have

\[
(B_j)_{j=1}^\infty := \left(\frac{\tilde{X}_j}{1 - X_1 - \cdots - X_{j-1}}\right)_{j=1}^\infty \overset{d}{\to} (Y_1, Y_2, \ldots)
\]
as \(x \to \infty\), where \(Y_1, Y_2, \ldots\) is a sequence of i.i.d. beta\((1,\theta)\) random variables and where \((\tilde{X}_j)_{j=1}^\infty\) is a sized-biased permutation (defined in §4) of the sequence \((X_j)_{j=1}^\infty\).

**Remark 3.7.** To connect this proposition with number-theoretic terms, we introduce a sequence \(P_j\) of ‘typical prime divisors’ of \(N_x\), defined as \(P_j = N_x^{\tilde{X}_j}\). The asymptotic behavior of these typical primes might be of independent interest. It follows from the proposition that \(P_j\) for \(j \geq 1\) satisfy the following limit law

\[
\left(\frac{\log P_1(N_x)}{\log x}, \frac{\log P_2(N_x)}{\log x}, \frac{\log P_3(N_x)}{\log x}, \ldots\right) \overset{d}{\to} (Y_1, (1 - Y_1)Y_2, (1 - Y_1 - Y_2)Y_3, \ldots),
\]
as \(x \to \infty\), where once again, \(Y_1, Y_2, \ldots\) is a sequence of i.i.d. beta\((1,\theta)\).

**Proof.** Fix \(k \geq 1\) and \(0 < a_j < b_j < 1\) for any \(1 \leq j \leq k\). By the Portmanteau Lemma it suffices to show that

\[
\liminf_{x \to \infty} \mathbb{P}\left(\forall j \quad a_j \leq B_j \leq b_j\right) \geq \mathbb{P}\left(\forall j \quad a_j \leq Y_j \leq b_j\right).
\]

(3.13)

We have that

\[
B_j = \frac{\tilde{X}_j}{1 - X_1 - \cdots - X_{j-1}} = \frac{\log P_j}{\log \left(\frac{N_x}{p_1 \cdots p_{j-1}}\right)}
\]

and therefore

\[
\mathbb{P}\left(\forall j \quad a_j \leq B_j \leq b_j\right) = \sum_{n \leq x} \sum_{p_1, \ldots, p_k} \mathbb{P}(N_x = n, \forall j \quad P_j = p_j),
\]

where the sum is over all \(n \leq x\) and all \(k\)-tuples \((p_1, \ldots, p_k)\).
where the inner sum is over a sequence of $k$ primes $p_1, \ldots, p_k$ such that for any $1 \leq j \leq k$ we have $(n/(p_1 \cdots p_{j-1}))^{a_j} \leq p_j \leq (n/(p_1 \cdots p_{j-1}))^{b_j}$. Let $0 < \delta < \min\{a_1, \ldots, a_k\} \prod_{i=1}^{k} (1 - b_i)$ and put $x_j := x/(p_1 \cdots p_{j-1})$ for any $1 \leq j \leq k + 1$. We have the following lower bound:

\[
\mathbb{P}(\forall j \ a_j \leq B_j \leq b_j) \geq \sum_{[\delta x] \leq n \leq x} \sum_{p_1, \ldots, p_k: \text{distinct}} \mathbb{P}(N_x = n, \forall j \ P_j = p_j). \tag{3.14}
\]

By the definition of a sized-biased permutation, when $p_1, \ldots, p_k$ are distinct and $p_j \parallel n$ we have

\[
\mathbb{P}(N_x = n, \forall j \ P_j = p_j) = \frac{\alpha(n) \prod_{j=1}^{k} \log p_j}{\sum_{m \leq x} \alpha(m) \prod_{j=1}^{k} \log p_j / \log x} \geq \frac{\alpha(n) \prod_{j=1}^{k} \log p_j / \log x_j}{\sum_{m \leq x} \alpha(m) \prod_{j=1}^{k} \log p_j / \log x_j},
\]

for any $n \leq x$. Thus, changing the order of summation in (3.14) we obtain

\[
\mathbb{P}(\forall j \ a_j \leq B_j \leq b_j) \geq \sum_{x_k^k \leq p_k \leq (\delta x_k)^k} \sum_{p_k \notin \{p_1, \ldots, p_{k-1}\}} \mathbb{P}(N_x = n, \forall j \ P_j = p_j). \tag{3.15}
\]

Next, we would like to use Lemma 3.5 in order to lower bound the inner sum in (3.15). We have $x_j = p_j x_{j+1} \leq (\delta x_k)^k x_{j+1} \leq x_k^{\nu} x_{j+1}$ and therefore, for any $1 \leq j \leq k$ we have $x_{j+1} \geq x_k^{\nu} \geq \cdots \geq v^\nu$ where $\nu := \prod_{j=1}^{k} (1 - b_j)$. Thus $p_j \geq x^{a_j \nu} \geq x^\delta$ and moreover $p_1 \cdots p_k = x/x_{k+1} \leq x^{1-\delta} \leq x^{1-\delta}$. We get that the assumptions of Lemma 3.5 hold and therefore using also Corollary 3.2 we obtain

\[
\mathbb{P}(\forall j \ a_j \leq B_j \leq b_j) \geq \left(1 - 4 \delta^{d+1}\right) \sum_{x_k^a \leq p_k \leq (\delta x_k)^k} \frac{\alpha(p_k) \log p_k}{p_k^{d+1} \log x_k} \left(1 - \frac{\log p_k}{\log x_k}\right)^{\theta - 1} \sum_{\nu \prod_{j=1}^{k} (1 - b_j)} \frac{\alpha(p_k) \log p_k}{p_k^{d+1} \log x_k} \left(1 - \frac{\log p_k}{\log x_k}\right)^{\theta - 1} \tag{3.16}
\]

for sufficiently large $x$. Consider the innermost sum. We define the function $g = g_{nk}: [x_k^{a_k}, (\delta x_k)^k] \rightarrow \mathbb{R}$ by

\[
g(t) := \frac{\log t}{t} \left(1 - \frac{\log t}{\log x_k}\right)^{\theta - 1}.
\]

We apply Lemma 3.5 with $\alpha(n)/n^d$ and this $g$, obtaining

\[
\sum_{x_k^a \leq p_k \leq (\delta x_k)^k} \frac{\alpha(p_k) \log p_k}{p_k^{d+1} \log x_k} \left(1 - \frac{\log p_k}{\log x_k}\right)^{\theta - 1} = O_\delta \left(\frac{1}{\log x}\right) + \frac{1}{\log x_k} \sum_{x_k^a \leq p_k \leq (\delta x_k)^k} \frac{\alpha(p_k)}{p_k^{d+1}} g(pk)
\]

\[
= O_\delta \left(\frac{1}{\log x}\right) + \frac{1}{\log x_k} \int_{x_k^a}^{(\delta x_k)^k} \theta \left(1 - \frac{\log t}{\log x_k}\right)^{\theta - 1} dt = O_\delta \left(\frac{1}{\log x}\right) + \int_{a_k}^{b_k} \frac{b_k}{\log x_k} \theta(1 - y)^{\theta - 1} dy
\]

where in the last equality we performed the change of variables $y = \log t/\log x_k$. Substituting the last estimate into (3.16) we get the same expression with $k$ replaced by $k - 1$. Thus, for sufficiently large $x$
depending on \( \delta \),

\[
\mathbb{P} \left( \forall j \ a_j \leq B_j \leq b_j \right) \geq (1 - 5\delta^{d+1}) \prod_{j=1}^{k} \int_{a_j}^{b_j} \theta(1 - y)^{d-1} \, dy = (1 - 5\delta^{d+1}) \mathbb{P} \left( \forall j \ a_j \leq Y_j \leq b_j \right),
\]

and so

\[
\liminf_{x \to \infty} \mathbb{P} \left( \forall j \ a_j \leq B_j \leq b_j \right) \geq (1 - 5\delta^{d+1}) \mathbb{P} \left( \forall j \ a_j \leq Y_j \leq b_j \right).
\]

Since \( \delta \) is arbitrary it follows that [3.13] holds, as needed.

\section*{3.3 Proof of first part of Theorem [1.1]}

\subsection*{3.3.1 Preparatory results}

We need the following results from probability, which are given as Remarks 1, 2 and 3 in [11].

\begin{lemma}
Let \( D_x, E_x \) be random variables defined for any \( x \gg 1 \).

1. If \( D_x \xrightarrow{d} 1 \) and \( E_x \xrightarrow{d} 0 \), then \( U_x \xrightarrow{d} N(0,1) \) if and only if \( D_x U_x + E_x \xrightarrow{d} N(0,1) \).

2. Let \( X \sim N(0,1) \). If \( \mathbb{E} D_x^k \to \mathbb{E} X^k \) for each \( k \geq 1 \) then \( D_x \xrightarrow{d} N(0,1) \).

3. Let \( X \sim N(0,1) \). If \( D_x \xrightarrow{d} N(0,1) \) and if \( \sup_x \mathbb{E} |D_x|^{k+\varepsilon} < \infty \) for some \( \varepsilon > 0 \) then \( \mathbb{E} D_x^k \to \mathbb{E} X^k \).
\end{lemma}

Proof. Consider the multiplicative function \( \tilde{\alpha}(n) := \alpha(n) t^{\Omega(n) - \omega(n)} \), where we choose \( 1 < t < 2/r \), so that \( \tilde{\alpha} \) will still satisfy \( \tilde{\alpha} \) with the same parameters, except \( r \) replaced with \( rt \). Applying Corollary 3.2 with \( \alpha \) and \( \tilde{\alpha} \), we obtain that

\[
\mathbb{E}^{\Omega(N_x) - \omega(N_x)} = \frac{\sum_{n \leq x} \tilde{\alpha}(n)}{\sum_{n \leq x} \alpha(n)} = \frac{A_{\tilde{\alpha}}}{A_{\alpha}} (1 + o(1)) = O(1)
\]

which, by Jensen’s inequality for instance, implies that \( \mathbb{E} \left( \Omega(N_x) - \omega(N_x) \right) \) is bounded as \( x \to \infty \).

For \( x \gg 1 \), we define the subset

\[
\mathcal{P}_x := \left\{ p \in \mathcal{P} : \log^4 x \leq p \leq \exp \left( \exp \left( \log \log x - \log^\frac{1}{2} \log x \right) \right) \right\}.
\]

For each prime \( p \gg 1 \), define a Bernoulli random variable \( X_p \) such that \( \mathbb{P}(X_p = 1) = \alpha(p)/p^{d+1} \) (for \( p \gg 1 \), this is in \([0,1]\)) and such that the different \( X_p \)'s are independent. Define \( \sigma_x \) by

\[
\sigma_x^2 := \sum_{p \in \mathcal{P}_x} \frac{\alpha(p)}{p^{d+1}} \left( 1 - \frac{\alpha(p)}{p^{d+1}} \right).
\]

We define the random variables

\[
A_x := \frac{\omega(N_x) - \theta \log \log x}{\sqrt{\theta \log \log x}}, \quad B_x := \frac{\sum_{p \in \mathcal{P}_x} \left( \mathbb{1}_{\{p\mid N_x\}} - \frac{\alpha(p)}{p^{d+1}} \right)}{\sigma_x}, \quad C_x := \frac{\sum_{p \in \mathcal{P}_x} \left( X_p - \frac{\alpha(p)}{p^{d+1}} \right)}{\sigma_x}.
\]

\begin{lemma}
We have \( \sum_{p \in \mathcal{P}_x} \alpha(p)/p^{d+1} = \theta \log \log x + O(\log^\frac{1}{2} \log x) \) and \( \sigma_x^2 = \theta \log \log x + O(\log^\frac{1}{2} \log x) \).
\end{lemma}
Proof. Applying Lemma 3.10 with $\alpha(p)/p^d$ in place of $\alpha$, $g(t) = 1/t$ and the interval $[\log^d x, \exp(\exp(\log \log x - \log^{1/3} \log x))]$, we find that $\sum_{p \in \mathcal{P}_x} \alpha(p)/p^{d+1} = \theta \log \log x + O(\log^3 \log x)$. Since $\sum_{p \in \mathcal{P}} q(p)/p^{d+1} = O(1)$, the estimate for $\sigma_x^2$ follows from the first estimate.

Lemma 3.11. For each integer $k \geq 1$, we have $\sup_x |EC_x^k| < \infty$. In particular, $\sup_x \mathbb{E}[C_x^{2k}] < \infty$.

Proof. Let $Y_p := X_p - \alpha(p)/p^{d+1}$. We have the following expansion:

$$EC_x^k = \frac{1}{\sigma_x^m} \sum_{p \in \mathcal{P}_x} \mathbb{E}[Y_{p_1} \cdots Y_{p_k}] = \frac{1}{\sigma_x^m} \sum_{l_1, \ldots, l_m \geq 1} \left( \sum_{l_i = k}^k \right) S(l_1, \ldots, l_m), \quad (3.17)$$

where

$$S(l_1, \ldots, l_m) := \sum_{q_1 \cdots q_m \not\in \mathcal{P}_x} \mathbb{E}[Y_{q_1}] \mathbb{E}[Y_{q_2}] \cdots \mathbb{E}[Y_{q_m}].$$

As we have $\mathbb{E}Y_p = 0$, it follows that $S(l_1, \ldots, l_m)$ vanishes if $l_i = 1$ for some $i$, and so we may restrict the summation in (3.17) to $l_i \geq 2$. Since $|Y_p| \leq 1$ (for $p \geq 2$), it follows that if $l_i \geq 2$ then $|\mathbb{E}[Y_{q_i}]| \leq \mathbb{E}[Y_{q_i}]$, so that

$$S(l_1, \ldots, l_m) \leq \left( \sum_{l_i \in \mathcal{P}_x} \mathbb{E}[Y_{q_i}] \right)^m = \sigma_x^{2m}.\quad$$

As $\sum_{i=1}^m l_i = k$ and $l_i \geq 2$, it follows that $2m \leq k$. For $x \gg 1$ we have $\sigma_x \geq 1$ by Lemma 3.10 and so

$$|EC_x^k| \leq \sum_{m=1}^k \frac{\sigma_x^{2m}}{\sigma_x^m} \sum_{l_1, \ldots, l_m \geq 2} \left( \sum_{l_i = k}^k \right) \leq \sum_{m=1}^k \sum_{l_1, \ldots, l_m \geq 2} \left( \sum_{l_i = k}^k \right)$$

as needed.

Lemma 3.12. Fix $m \geq 1$. Let $q_1, \ldots, q_m \in \mathcal{P}_x$ be distinct primes. We have for $x \gg 1$

$$\mathbb{P}(\forall j, q_j \mid N_x) - \prod_{j=1}^m \frac{\alpha(q_j)}{q_j^{d+1}} \leq C_m \exp \left( -\log^t \log x \right) \prod_{j=1}^m \frac{1 + \alpha(q_j)q_j^{-d}}{q_j}.\quad$$

Proof. By multiplicativity of $\alpha$ we have

$$\sum_{n \leq x \atop q_1 \cdots q_m \mid n} \alpha(n) = \sum_{l_1=1}^\infty \cdots \sum_{l_m=1}^\infty \sum_{n \leq x \atop \forall j, q_j \mid n} \alpha(n) = \sum_{l_1=1}^\infty \cdots \sum_{l_m=1}^\infty \prod_{j=1}^m \alpha(q_j^{l_j}) \sum_{l=1}^\infty \prod_{l_j=1}^\infty \alpha(l). \quad (3.18)$$

The term corresponding to $(l_1, \ldots, l_m) = (1, \ldots, 1)$ in (3.18) may be estimated as

$$\prod_{j=1}^m \frac{\alpha(q_j)}{q_j^{d+1}} \leq C x^{d+1} \log^{\theta - 1} x \prod_{j=1}^m \frac{1 + \alpha(q_j)q_j^{-d}}{q_j^{d+1}} \leq C_m x^{d+1} \log^3 x \prod_{j=1}^m \frac{\alpha(q_j)}{q_j^{d+1}}, \quad (3.19)$$

where

$$\sum_{l=1}^\infty \prod_{l_j=1}^\infty \alpha(l) \leq \prod_{j=1}^m \alpha(q_j) \sum_{l_1=1}^\infty \cdots \sum_{l_m=1}^\infty \sum_{l_j=1}^\infty \alpha(l) \leq C_m x^{d+1} \log^3 x \prod_{j=1}^m \frac{\alpha(q_j)}{q_j^{d+1}}.$$
where in the second inequality we used (3.14) and (1.4). Using Corollary 3.2 we get

$$\prod_{j=1}^{m} \alpha(q_j) \sum_{l=1}^{\frac{m}{d+1}} \alpha(l) = (1 + O(\log^{-a} x)) A_\alpha x^{d+1} \log^{\theta-1} x \prod_{j=1}^{m} \alpha(q_j) \left( \frac{x}{q_1 \cdots q_m} \right) \prod_{j=1}^{m} \alpha(q_j)$$

$$= (1 + O \left( \frac{\log^a (q_1 \cdots q_m)}{\log^a x} \right)) A_\alpha x^{d+1} \log^{\theta-1} x \prod_{j=1}^{m} \alpha(q_j)$$

$$= (1 + O_m (\exp (-\log^4 \log x))) A_\alpha x^{d+1} \log^{\theta-1} x \prod_{j=1}^{m} \alpha(q_j).$$

(3.20)

The contribution of terms with \((l_1, \ldots, l_m) \neq (1, \ldots, 1)\) to \((3.18)\) may be bounded using Corollary 3.2 as follows:

$$\sum_{(l_1, \ldots, l_m) \neq (1, \ldots, 1)} \prod_{j=1}^{m} \alpha(q_j^{l_j}) \sum_{\forall j, q_l \mid l} \alpha(l) \leq C x^{d+1} \log^\theta x \sum_{(l_1, \ldots, l_m) \neq (1, \ldots, 1)} \prod_{j=1}^{m} \frac{\alpha(q_j^{l_j})}{q_j^{l_j}}$$

$$\leq C m x^{d+1} \log^\theta x \sum_{j=1}^{m} \frac{1}{q_j} \prod_{l_j=2}^{\infty} \alpha(q_j^{l_j}) \prod_{l_j=1}^{\infty} \frac{\alpha(q_j^{l_j})}{q_j^{l_j}} \leq C m x^{d+1} \log^{\theta-2} x \sum_{j=1}^{m} \frac{1}{q_j} \prod_{l_j=1}^{\infty} \frac{\alpha(q_j^{l_j})}{q_j^{l_j}}.$$

(3.21)

To conclude the proof, we observe that \(\mathbb{P} (\forall j, q_j \mid N_x) = \sum_{n \leq x} \alpha(n)/S(x)\), which combined with (3.18), (3.19), (3.20), (3.21) and Corollary 3.2 gives the desired bound.

**Lemma 3.13.** For each integer \(k \geq 1\), we have \(\mathbb{E} B_x^k - \mathbb{E} C_x^k \to 0\).

**Proof.** As in (3.17), we have

$$\mathbb{E} \left( \sum_{p \in \mathcal{P}_x} X_p^k \right) = \sum_{m=1}^{k} \sum_{\sum l_i = k, l_i \geq 1} \binom{k}{l_1, \ldots, l_m} \sum_{q_1, \ldots, q_m \in \mathcal{P}_x, q_1 \leq \cdots \leq q_m} \mathbb{E} [X_{q_1}^{l_1} X_{q_2}^{l_2} \cdots X_{q_m}^{l_m}].$$

As \(X_p\) assumes the values 1 and 0 only, we have \(X_{q_i}^{l_i} = X_{q_i}\) and then \(\mathbb{E} [X_{q_1}^{l_1} X_{q_2}^{l_2} \cdots X_{q_m}^{l_m}] = \prod_{j=1}^{m} \frac{\alpha(q_j)}{q_j^d}\), so that

$$\mathbb{E} \left( \sum_{p \in \mathcal{P}_x} X_p^k \right) = \sum_{m=1}^{k} \sum_{\sum l_i = k, l_i \geq 1} \binom{k}{l_1, \ldots, l_m} \prod_{q_1, \ldots, q_m \in \mathcal{P}_x, q_1 \leq \cdots \leq q_m} \frac{\alpha(q_j)}{q_j^d}.$$

Similarly,

$$\mathbb{E} \left( \sum_{p \in \mathcal{P}_x} \mathbbm{1}_{\{p \mid N_x\}}^k \right) = \sum_{m=1}^{k} \sum_{\sum l_i = k, l_i \geq 1} \binom{k}{l_1, \ldots, l_m} \prod_{q_1, \ldots, q_m \in \mathcal{P}_x, q_1 \leq \cdots \leq q_m} \mathbb{P} (\forall j, q_j \mid N_x).$$

15
By Lemma 3.12 and Lemma 3.4 with \( g(t) = 1/t \),
\[
\left| \mathbb{E}\left( \sum_{p \in \mathcal{P}_x} \mathbb{1}_{\{p \mid N_x\}} \right)^k - \mathbb{E}\left( \sum_{p \in \mathcal{P}_x} X_p \right)^k \right|
\leq C_k \exp \left( -\log^{\frac{2}{3}} \log x \right) \sum_{m=1}^k \sum_{l_1, \ldots, l_m \geq 1} \binom{k}{l_1, \ldots, l_m} \sum_{q_1 < \cdots < q_m \in \mathcal{P}_p} \prod_{j=1}^m \frac{1 + \alpha(q_j)q_j^{-d}}{q_j}
\leq C_k \exp \left( -\log^{\frac{2}{3}} \log x \right) \sum_{m=1}^k \sum_{l_1, \ldots, l_m \geq 1} \binom{k}{l_1, \ldots, l_m} \left( \frac{\sum_{p \leq x} 1 + \alpha(p)p^{-d}}{p} \right)^m
\]
(3.22)
\[
\leq C_k \exp \left( -\log^{\frac{2}{3}} \log x \right) \left( \log \log x \right)^k.
\]

By the binomial theorem, (3.22) and Lemma 3.11, we have
\[
\left| \mathbb{E}B_x^k - \mathbb{E}C_x^k \right| \leq \sigma_x^{-k} \sum_{i=0}^k \binom{k}{i} \mathbb{E}\left( \sum_{p \in \mathcal{P}_x} \mathbb{1}_{\{p \mid N_x\}} \right)^i - \mathbb{E}\left( \sum_{p \in \mathcal{P}_x} X_p \right)^i \left( \sum_{p \in \mathcal{P}_x} \frac{\alpha(p)}{p^{d+1}} \right)^{k-i}
\leq \sigma_x^{-k} C_k \exp \left( -\log^{\frac{2}{3}} \log x \right) \sum_{i=0}^k \binom{k}{i} \left( \log \log x \right)^i \left( \log \log x \right)^{k-i}
\leq C_k \left( \log \log x \right)^{k/2} \exp \left( -\log^{\frac{2}{3}} \log x \right) \to 0
\]
as needed.

\[\square\]

### 3.3.2 Conclusion of proof

By Lemma 3.3 and the first part of Lemma 3.8 with \( D_x = 1 \) and \( E_x = (\Omega(N_x) - \omega(N_x))/\sqrt{\log \log x} \), it follows that (1.1) holds if and only if \( A_x \overset{d}{\rightarrow} N(0, 1) \). Now let \( D_x = \sqrt{\log \log x}/\sigma_x \) and
\[
E_x = \frac{\theta \log \log x - \sum_{p \in \mathcal{P}_x} \frac{\alpha(p)}{p^{d+1}} - \sum_{p \in \mathcal{P}_x \setminus \mathcal{P}_p} \mathbb{1}_{\{p \mid N_x\}}}{\sigma_x}.
\]
(3.23)

We have \( D_x \overset{d}{\rightarrow} 1 \). Moreover, in the sum in the second fraction in (3.23), there can be at most one non-zero term with \( p \) greater than \( \sqrt{x} \), and so we may use Lemma 3.11 and 3.4 to obtain that
\[
\mathbb{E}|E_x| \leq C(\log \log x)^{-\frac{3}{2}} + \frac{1}{\sigma_x} \sum_{p \in \mathcal{P}_x \setminus \mathcal{P}_p} \mathbb{P}(p \mid N_x) + \frac{1}{\sigma_x} \sum_{p \leq \sqrt{x}} \frac{1 + \alpha(p)p^{-d}}{p} \to 0,
\]
where the last expression tends to 0 by Lemma 3.4 with \( g(t) = 1/t \). Hence \( E_x \overset{d}{\rightarrow} 0 \). Since \( B_x = D_x A_x + E_x \), it follows by Lemma 3.11 and the first part of Lemma 3.8 that \( A_x \overset{d}{\rightarrow} N(0, 1) \) if and only if \( B_x \overset{d}{\rightarrow} N(0, 1) \).

By the second part of Lemma 3.8 to establish \( B_x \overset{d}{\rightarrow} N(0, 1) \) it suffices to show that \( \mathbb{E}B_x^k \to \mathbb{E}X^k \) for each \( k \), where \( X \sim N(0, 1) \). By Lemma 3.13 this is equivalent to \( \mathbb{E}C_x^k \to \mathbb{E}X^k \). Since the random variables \( X_p - \alpha(p)/p^{d+1} \) are uniformly bounded as \( p \) varies, and since the denominator of \( C_x \) tends to infinity by Lemma 3.11, we have \( C_x \overset{d}{\rightarrow} N(0, 1) \) by the Lindeberg–Feller theorem (also known as Central Limit Theorem for triangular arrays) [17, Thm. 4.7]. Thus, the moments of \( C_x \) converge to those of \( X \) by Lemma 3.11 and the last part of Lemma 3.8.

\[\square\]
4 Polynomials-grown weights

Lemma 4.1. Fix a function $f : \mathcal{P} \to \mathbb{R}_{>0}$ on primes such that $\log f(p) = o(\log p)$, and let

$$G_f(s) = \sum_p \frac{f(p)}{p^s}, \quad \Re s > 1.$$ 

Let $\alpha : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be a multiplicative function satisfying

(I) $\alpha(p) = f(p) + O(\log^{-2} p),$

(II) $\sum_{k \geq 2} k\alpha(p^k) p^k = O\left(\frac{1}{p \log^2 p}\right),$ \hspace{1cm} (4.1)

and define

$$F(s) = \sum_{n \geq 1} \frac{\alpha(n)}{n^s},$$

the Dirichlet series of $\alpha$. For $\Re s > 1$ we have

$$F(s) = \varphi(s) e^{G_f(s)}$$

where $\varphi$ is differentiable, bounded and has bounded derivative on $\Re s \geq 1$.

Proof. By multiplicativity of $\alpha$, we may write

$$F(s) = \prod_{p} \left(\sum_{k=0}^{\infty} \frac{\alpha(p^k)}{p^{ks}}\right) = e^{G_f(s)} \prod_{p} (1 + E_p(s)),$$

where

$$E_p(s) = -1 + \exp\left[-\frac{f(p)}{p^s}\right] \sum_{k=0}^{\infty} \frac{\alpha(p^k)}{p^{ks}}.$$ 

Using (4.1) and the triangle inequality, we obtain

$$\left| \exp\left[-\frac{f(p)}{p^s}\right] - \sum_{k=0}^{\infty} \frac{\alpha(p^k)}{p^{ks}} \right| \leq \left| \exp\left[-\frac{f(p)}{p^s}\right] - \left(1 + \frac{f(p)}{p^s}\right)\right| + \left| \frac{\alpha(p)}{p^s} - \frac{f(p)}{p^s}\right| + \sum_{k=2}^{\infty} \left| \frac{\alpha(p^k)}{p^{ks}} \right|$$

$$\leq C f^2(p) / p^2 + C / p \log^2 p + \sum_{k=2}^{\infty} \frac{\alpha(p^k)}{p^{ks}} \leq C / p \log^2 p$$

for $\Re s > 1$. Thus, $|E_p(s)| \leq C / (p \log^2 p)$ for $\Re s > 1$. We turn to bound the derivative of $E_p$. We have

$$E'_p(s) = \log p \cdot \exp\left[-\frac{f(p)}{p^s}\right] \left(\frac{f(p)}{p^s} \sum_{k=0}^{\infty} \frac{\alpha(p^k)}{p^{ks}} - \sum_{k=1}^{\infty} k\alpha(p^k) / p^{ks}\right)$$

and therefore

$$|E'_p(s)| \leq C \log p \left(\frac{f(p)}{p^s} - \frac{\alpha(p)}{p^s}\right) + \frac{f(p)}{p} \sum_{k=1}^{\infty} \frac{\alpha(p^k)}{p^{ks}} + \sum_{k=2}^{\infty} k\alpha(p^k) / p^{ks}\right) \leq C / p \log p.$$ 

Let $p_0 > 0$ so that for $p \geq p_0$ we have $|E_p(s)| \leq 1/2$ for any $s$ with $\Re s > 1$. We have that

$$F(s) = e^{G_f(s)} \cdot \psi_1(s) \cdot \psi_2(s),$$

where

$$\psi_1(s) := \prod_{p < p_0} (1 + E_p(s)), \quad \psi_2(s) := \prod_{p \geq p_0} (1 + E_p(s)) = \exp\left[\sum_{p \geq p_0} \log (1 + E_p(s))\right]$$

and

$$\prod_{p < p_0} (1 + E_p(s)) = \prod_{p < p_0} \left(1 + \frac{1}{p^{\Re s}}\right) = e^{\zeta'(1) / 2},$$

for $\Re s > 1$, where $\zeta(s)$ is the Riemann zeta function.
and \( \log z \) is the principal branch of the logarithm. The function \( \psi_1 \) is trivially differentiable, bounded and has bounded derivative on \( \Re s \geq 1 \). As for \( \psi_2 \), observe that

\[
|\log(\psi_2(s))| = \left| \sum_{p \geq p_0} \log (1 + E_p(s)) \right| \leq C \sum_{p \geq p_0} |E_p(s)| \leq \sum_{p \geq p_0} \frac{C}{p \log^2 p} < \infty
\]

and

\[
\left| \frac{d}{ds} \log(\psi_2(s)) \right| \leq \sum_{p \geq p_0} \left| \frac{E'_p(s)}{1 + E_p(s)} \right| \leq C \sum_{p \geq p_0} |E'_p(s)| \leq \sum_{p \geq p_0} \frac{C}{p \log p} < \infty.
\]

It follows that \( \psi_2 \) is also differentiable, bounded and has bounded derivative on \( \Re s \geq 1 \). Thus, the same is true for \( \varphi := \psi_1 \cdot \psi_2 \), as needed.

Lemma 4.2. Let \( \alpha : \mathbb{N} \to \mathbb{R}_{\geq 0} \) be a multiplicative function satisfying (1.10) and

\[
S(y/h)h \leq C \cdot S(y)
\]

for all \( y > 0, h \geq 2 \). Then \( |\mathbb{E} \Omega(N_x) - \mathbb{E} \omega(N_x)| \leq C \).

Proof. Writing \( \omega(N_x) \) as \( \sum_{p \leq x} 1_{p \mid N_x} \) and \( \Omega(N_x) \) as \( \sum_{p \leq x} \sum_{k \geq 1} 1_{p \mid N_x} \), we have

\[
0 \leq \mathbb{E} \Omega(N_x) - \omega(N_x) = \sum_{k \geq 2} \sum_{p \mid p^k} \mathbb{P}(p^k \mid N_x) \leq \sum_{k \geq 1} \sum_{p \mid p^k} \alpha(p^{k+1}) \frac{S(x/p^{k+1})}{S(x)} \leq C \sum_{k \geq 2} \sum_{p \mid p^k} \frac{\alpha(p^{k+1})}{p^{k+1}} \leq C \sum_{p} \frac{k \alpha(p)}{p^{k}} \leq C \sum_{p} \frac{1}{p \log p} \leq C,
\]

as needed.

Fix \( \gamma > 0 \) and define the Dirichlet series

\[
G(s) = \sum_p \frac{\log^\gamma p}{p^s}, \quad \Re s > 1.
\]

Lemma 4.3. Fix a non-negative integer \( k \). We have

\[
G^{(k)}(s) = (-1)^k \frac{\Gamma(\gamma + k)}{(s - 1)^{\gamma + k}} + B_k + O(|s - 1||s|)
\]

for \( \Re s > 1 \), where \( B_k \) is a real constant depending on \( \gamma \) and \( k \). Here \( (s - 1)^\gamma = \exp(\gamma \log(s - 1)) \) is defined using the principal branch of the logarithm.

Proof. We start with the case \( k = 0 \). Let \( E(t) = (\sum_{p \leq t} \log p) - t \) be the error term in the prime number theorem. Using integration by parts we obtain, for \( \Re s > 1 \), that

\[
\sum_{p} \frac{\log^\gamma p}{p^s} = \sum_{n=2}^\infty 1_{(n \text{ is prime})} \log n \cdot \frac{\log^\gamma - 1}{n^s} = - \int_2^\infty \left( \sum_{p \leq t} \log p \right) \left( \frac{\log^\gamma - 1}{t^s} \right)' dt = \int_2^\infty (t + E(t)) \frac{s \log^\gamma - 1 - (\gamma - 1) \log^\gamma - 2}{t^{s+1}} dt = \psi_1(s) + \psi_2(s),
\]

where

\[
\psi_1(s) := \int_2^\infty E(t) \frac{s \log^\gamma - 1 - (\gamma - 1) \log^\gamma - 2}{t^{s+1}} dt, \quad \psi_2(s) := \int_2^\infty \frac{s \log^\gamma - 1 - (\gamma - 1) \log^\gamma - 2}{t^s} dt.
\]
When $\Re s \geq 1$ we may bound $\psi_1$ as follows, using the known result that $E(t) \ll te^{-c\sqrt{\log t}}$:

$$|\psi_1(s)| \leq C|s| \int_2^\infty \left| e^{-c\sqrt{\log t}} \frac{\log^{\gamma-1} t}{t^s} \right| dt \leq C|s| \int_2^\infty e^{-c\sqrt{\log t}} \frac{\log^{\gamma-1} t}{t^s} dt \leq C|s| \int_0^\infty e^{-c\sqrt{\gamma} t} dt \leq C|s|,$$

and similarly $\psi'_1$ may be bounded by

$$|\psi'_1(s)| \leq C|s| \int_2^\infty \left| e^{-c\sqrt{\log t}} \frac{\log^{\gamma-1} t}{t^s} \right| dt \leq C|s| \int_0^\infty e^{-c\sqrt{\gamma} t} dt \leq C|s|,$$

and so

$$\psi_1(s) = \psi_1(1) + O(|s-1|)$$  \hspace{1cm} (4.4)

in $\Re s \geq 1$. We turn to estimate $\psi_2$. Using integration by parts we obtain

$$\int_2^\infty \frac{\log^{\gamma-1} t}{t^s} dt = \frac{\log^{\gamma-1} t}{(1-s)t^{s-1}} \bigg|_{t=2}^{t=\infty} - \int_2^\infty \frac{(\gamma-1)\log^{\gamma-2} t}{(1-s)t^s} dt = \frac{\log^{\gamma-1} 2}{(s-1)2^{s-1}}$$

$$+ \frac{1}{s-1} \left( \int_2^\infty \frac{s\log^{\gamma-1} t}{t^s} dt - \psi_2(s) \right)$$

for $\Re s > 1$, so that

$$\psi_2(s) = \int_2^\infty \frac{\log^{\gamma-1} t}{t^s} dt + \frac{\log^{\gamma-1} 2}{2^{s-1}}.$$  

Setting

$$\psi_3(s) := \int_1^2 \frac{\log^{\gamma-1} t}{t^s} dt$$

we obtain

$$\psi_2(s) = \int_1^\infty \frac{\log^{\gamma-1} t}{t^s} dt - \psi_3(s) + \log^{\gamma-1} 2 + O(|s-1|)$$

$$= \int_1^\infty \frac{\log^{\gamma-1} t}{t^s} dt - \psi_3(1) + \log^{\gamma-1} 2 + O(|s-1|),$$  \hspace{1cm} (4.5)

where in the last equality we used that $|\psi_3(s)|, |\psi'_3(s)| \leq C$ for $\Re s \geq 1$. In order to compute the integral in the right-hand side of (4.5) we perform the change of variables $w = (s-1)\log t$, obtaining

$$\int_1^\infty \frac{\log^{\gamma-1} t}{t^s} dt = \frac{1}{(s-1)\gamma} \int_0^{(s-1)e} w^{\gamma-1} e^{-w} dw.$$  \hspace{1cm} (4.6)

Letting $C_R$ be the circular contour from $R$ to $R \frac{s-1}{|s-1|}$, we compute that $\lim_{R \to \infty} \int_{C_R} w^{\gamma-1} e^{-w} dw = 0$ when $\Re s > 1$. Thus, we may deform the contour $\{(s-1)\log t : t \geq 0\}$ in (4.6) to the positive real line, obtaining

$$\int_1^\infty \frac{\log^{\gamma-1} t}{t^s} dt = \frac{1}{(s-1)\gamma} \int_0^{(s-1)e} w^{\gamma-1} e^{-w} dw = \frac{\Gamma(\gamma)}{(s-1)\gamma}. \hspace{1cm} (4.7)$$

Substituting (4.7) into (4.5) and then substituting (4.3) and (4.5) into (3.4), we obtain (4.2) with $k = 0$ and $B_0 = \log^{\gamma-1} 2 + \psi_1(1) - \psi_3(1)$.
Next we consider the case \( k \geq 1 \). We have
\[
G^{(k)}(s) = (-1)^k \sum_p \frac{\log^{\gamma+k} p}{p^s}.
\]
Hence, repeating the above arguments with \( \gamma \) replaced by \( \gamma + k \) gives the desired result for any \( k \geq 1 \).

The following lemma bounds \( \Re(G(\sigma + it)) \) when \( t \) is not too large.

**Lemma 4.4.** There exists \( c > 0 \) with the following property. For \( \sigma > 1 \) sufficiently close to 1, and for any \( t \in \mathbb{R} \) with \( 1 \leq |t| \leq e^{1/(\sigma-1)} \) we have
\[
\Re(G(\sigma + it)) \leq (1 - c) \cdot G(\sigma).
\]

To prove Lemma 4.4 we use bounds on primes in short intervals. Information of the form we need is already found in a work of Hoheisel [33]. For ease of presentation, we use a stronger result.

**Theorem 4.5.** [32] For any sufficiently large \( m \), and any \( n \geq m \) with \( n - m \geq n^{3/4} \), we have
\[
|\{p : p \in [m, n] \text{ prime}\}| \geq \frac{n - m}{2\log n}. \tag{4.8}
\]

**Proof of Lemma 4.4.** Let \( t \in \mathbb{R} \) with \( 1 \leq |t| \leq e^{1/(\sigma-1)} \). As \( \Re(G(\sigma + it)) \) is an even function of \( t \), we may assume that \( t > 0 \). Consider the set of integers
\[
M := \left\{ n \geq 0 : \frac{10}{\sigma - 1} \leq \frac{2\pi}{t} \left(n + \frac{3}{4}\right) \leq \frac{20}{\sigma - 1} \right\}.
\]
Clearly, for \( \sigma \) close enough to 1 we have \( |M| \geq t/(\sigma - 1) \). For any \( n \in M \) we define
\[
A_n := \left\{ p : t \log p \in \left[2\pi n + \frac{\pi}{2}, 2\pi n + \frac{3\pi}{2}\right] \right\} = \{ p : p \in [e^{-\frac{\pi}{t}} x_n, x_n] \},
\]
where
\[
x_n := \exp \left(\frac{2\pi}{t} \left(n + \frac{3}{4}\right)\right).
\]
For any \( n \in M \) and \( p \in A_n \) we have that \( \cos(t \log p) \leq 0 \) and therefore
\[
\Re(G(\sigma + it)) = \sum_p \frac{\log^\gamma p}{p^\sigma} \cos(t \log p) \leq G(\sigma) - \sum_{n \in M} \sum_{p \in A_n} \frac{\log^\gamma p}{p^\sigma}. \tag{4.9}
\]

Since \( \log^\gamma p/p^\sigma \) is decreasing for sufficiently large \( p \) (independently of \( \sigma \geq 1 \)) and as \( \min_{n \in M} \min_{A_n} \rightarrow \infty \) as \( \sigma \to 1^+ \) (uniformly in \( t \geq 1 \)) we get that, when \( \sigma \) is close enough to 1,
\[
\sum_{p \in A_n} \frac{\log^\gamma p}{p^\sigma} \geq |A_n| \log^\gamma x_n x_n^\gamma \geq \frac{1 - e^{-\frac{\pi}{t}}}{2\log x_n} \geq c \frac{\log^\gamma - 1 x_n}{t \cdot x_n^\gamma - 1} \geq c \left(\frac{1}{\sigma - 1}\right)^{\gamma - 1} \tag{4.10}
\]
where in the second inequality we used (4.3). Indeed the conditions of Theorem 4.4 hold for \([e^{-\pi/t} x_n, x_n]\) as for any \( n \in M \) we have
\[
x_n \geq \exp \left(\frac{10}{\sigma - 1}\right) \geq t^{10}
\]
and so
\[
x_n - e^{-\frac{\pi}{t}} x_n = x_n \left(1 - e^{-\frac{\pi}{t}}\right) \geq c t x_n \geq c x_n^{0.9}.
\]

Summing (4.10) over \( n \in M \) we get
\[
\sum_{n \in M} \sum_{p \in A_n} \frac{\log^\gamma p}{p^\sigma} \geq c |M| \left(\frac{1}{\sigma - 1}\right)^{\gamma - 1} \geq c \left(\frac{1}{\sigma - 1}\right)^{\gamma} \geq c \cdot G(\sigma), \tag{4.11}
\]
where in the last inequality we used Lemma 4.3 with \( k = 0 \). From (4.9) and (4.11) we obtain the desired bound. \( \square \)
It turns out that when $|t| \geq e^{1/(\sigma-1)}$, the result of Lemma 4.4 does not hold and $\Re(G(\sigma + it))$ might be as large as $G(\sigma)$. We shall show that the reals $t \in \mathbb{R}$ for which $\Re(G(\sigma + it))$ is as large as $G(\sigma)$ are quite rare. More precisely, we will show in the following lemma that if $t_1$, $t_2$ are such that $\Re(G(\sigma + it))$ is large then the same holds for $t_1 - t_2$, and therefore by Lemma 4.4 $t_1$ and $t_2$ must be far away from each other.

**Lemma 4.6.** Let $\{a_n\}_{n \geq 1}$ be a sequence of non-negative reals with $\sum_{n=1}^{\infty} a_n < \infty$. Consider the function

$$f(t) = \sum_{n=1}^{\infty} a_n \cos(t \log n), \quad t \in \mathbb{R}.$$  

For any $0 < \varepsilon < 1$ and any $t_1, t_2 \in \mathbb{R}$ with

$$f(t_1) \geq (1 - \varepsilon)f(0), \quad f(t_2) \geq (1 - \varepsilon)f(0)$$

we have

$$f(t_1 - t_2) \geq (1 - 8\sqrt{\varepsilon})f(0).$$

**Proof.** Let $0 < \varepsilon < 1$ and define

$$A_i := \{n \geq 1 : \cos(t \log n) > 1 - \sqrt{\varepsilon}\}$$

for $t \in \{t_1, t_2\}$. By the assumption on $t_1$ we have

$$(1 - \varepsilon)f(0) \leq f(t_1) = \sum_{n=1}^{\infty} a_n \cos(t_1 \log n) \leq \sum_{n \in A_{t_1}} a_n + (1 - \sqrt{\varepsilon}) \sum_{n \notin A_{t_1}} a_n = f(0) - \sqrt{\varepsilon} \sum_{n \notin A_{t_1}} a_n,$$

so that

$$\sum_{n \notin A_{t_1}} a_n \leq \sqrt{\varepsilon}f(0). \tag{4.12}$$

The same argument shows that (4.12) holds for $t_2$ in place of $t_1$ as well. Therefore, by the union bound,

$$\sum_{n \notin A_{t_1} \cap A_{t_2}} a_n \leq 2\sqrt{\varepsilon}f(0) \quad \text{and} \quad \sum_{n \in A_{t_1} \cap A_{t_2}} a_n \geq (1 - 2\sqrt{\varepsilon})f(0). \tag{4.13}$$

Now, for any $n \in A_{t_1} \cap A_{t_2}$ and $i = 1, 2$ we have

$$|\sin(t_i \log n)| = \sqrt{1 - \cos^2(t_i \log n)} \leq \sqrt{1 - (1 - \sqrt{\varepsilon})^2} \leq \sqrt{1 - (1 - 2\sqrt{\varepsilon})} = \sqrt{2\varepsilon + \varepsilon^2}. $$

Thus, for any $n \in A_{t_1} \cap A_{t_2}$,

$$\cos((t_1 - t_2) \log n) = \cos(t_1 \log n) \cos(t_2 \log n) + \sin(t_1 \log n) \sin(t_2 \log n) \geq (1 - \sqrt{\varepsilon})^2 - 2\sqrt{\varepsilon} \geq 1 - 4\sqrt{\varepsilon}. \tag{4.14}$$

From (4.13) and (4.14) we obtain that

$$f(t_1 - t_2) = \sum_{n=1}^{\infty} a_n \cos((t_1 - t_2) \log n) \geq (1 - 4\sqrt{\varepsilon}) \sum_{n \in A_{t_1} \cap A_{t_2}} a_n - \sum_{n \notin A_{t_1} \cap A_{t_2}} a_n \geq (1 - 4\sqrt{\varepsilon})(1 - 2\sqrt{\varepsilon})f(0) - 2\sqrt{\varepsilon}f(0) \geq (1 - 8\sqrt{\varepsilon})f(0),$$

as needed. \hfill \square

Fix $K > 0$. The function $G'(s)$ is monotone-increasing for real $s > 1$, with $\lim_{s \to \infty} G'(s) = 0$ and $\lim_{s \to 1^+} G'(s) = -\infty$. For $x > 1$, we let $\sigma = \sigma_x$ be the unique real solution to

$$K \cdot G'(\sigma) = -\log x. \tag{4.15}$$

The point $\sigma$ plays the role of the saddle point in the proof of Theorem 4.7. The following is a corollary of Lemma 4.3.
Corollary 4.7. As \( x \to \infty \) we have
\[
\sigma - 1 = \left( 1 + O \left( \frac{1}{\log x} \right) \right) \left( K\Gamma(\gamma + 1) \right)^{\frac{1}{\gamma + 1}} \left( \log x \right)^{-\frac{1}{\gamma + 1}},
\]
(4.16)
and moreover
\[
G(\sigma) = \frac{\Gamma(\gamma)}{(K\Gamma(\gamma + 1))^{\frac{1}{\gamma + 1}}} \left( \log x \right)^{\frac{1}{\gamma + 1}} + B_0 + O \left( \left( \log x \right)^{-\frac{1}{\gamma + 1}} \right),
\]
\[
G''(\sigma) = \frac{\Gamma(\gamma + 2)}{(K\Gamma(\gamma + 1))^{\frac{2}{\gamma + 1}}} \left( \log x \right)^{\frac{2}{\gamma + 1}} + O \left( \left( \log x \right)^{-\frac{1}{\gamma + 1}} \right),
\]
\[
G'''(\sigma) \sim -\frac{\Gamma(\gamma + 3)}{(K\Gamma(\gamma + 1))^{\frac{3}{\gamma + 1}}} \left( \log x \right)^{\frac{3}{\gamma + 1}}.
\]

Proof. Since \( \lim_{s \to 1^+} G'(s) = -\infty \), it follows that \( \sigma_x = O(1) \) for \( x \geq 2 \). Using Lemma 4.3 (4.15) becomes
\[
K \frac{\Gamma(\gamma + 1)}{(\sigma - 1)\gamma + 1} = \log x + O(1),
\]
from which we deduce (4.16). Applying the estimates for \( G^{(k)}(s) \) in Lemma 4.3 for \( s = \sigma \) and \( k = 0, 2, 3 \), and using (4.16), we obtain the stated estimates for \( G(\sigma), G''(\sigma), G'''(\sigma) \).

4.1 Proof of Theorem 1.7

If we replace \( x \) with \( |x| + \frac{1}{2} \) in (4.11), then the left-hand side does not change, while the function in the right-hand side is changed by a factor of \( 1 + O(1/x) \), which can be absorbed in the relative error term. Thus, in proving Theorem 1.7 we may assume without loss of generality that \( x \) is of the form \( m + \frac{1}{2} \) for some positive integer \( m \) (i.e. half-integer). We denote by \( F(s) \) the Dirichlet series of \( \alpha \), which by Lemma 4.1 is of the form \( F(s) = \varphi(s) e^{K\cdot G(s)} \) where \( \varphi \) is differentiable, bounded and has bounded derivative on \( \Re s \geq 1 \). By an effective version of Perron’s formula [52 Thm. II.2.3] we have
\[
\sum_{n \leq x} \alpha(n) = \frac{1}{2\pi i} \int_{\sigma - iT}^{\sigma + iT} F(s)x^s \frac{ds}{s} + O \left( x^\sigma \sum_{n \geq 1} \frac{\alpha(n)}{n^{\sigma}(1 + T|\log(x/n)|)} \right)
\]
for any \( T \geq 1 \), where \( \sigma = \sigma_x \) is defined in (4.15). We choose \( T = x^2 \), obtaining
\[
\sum_{n \leq x} \alpha(n) = \frac{1}{2\pi} \int_{-x^2}^{x^2} \varphi(x + it)e^{K\cdot G(x + it)} \frac{x^{\sigma + it}}{\sigma + it} \frac{dt}{\sigma + it} + O(E(x)),
\]
(4.17)
where
\[
E(x) = x^\sigma \sum_{n \geq 1} \frac{\alpha(n)}{n^{\sigma}(1 + x^2|\log(x/n)|)} \leq C \frac{x^\sigma}{x} \sum_{n \geq 1} \frac{\alpha(n)}{n^{\sigma}} = C \frac{x^\sigma F(x)}{x} \leq C \frac{x^{\sigma} e^{K\cdot G(x)}}{x}.
\]
(4.18)
(We have used the fact that \( x \) is an half-integer and so \( |\log(x/n)| \geq C/x \).) Set
\[
t_x := (\log x)^{\frac{1}{\gamma(\gamma + 1)}} \left( \log x \right)^{-\frac{1}{\gamma + 1}}.
\]
for some sufficiently small \( \delta > 0 \). We decompose the integral in the right-hand side of (4.17) into three parts, to be estimated in the following ways:
\[
|t| \leq t_x : \text{ estimated using Corollary 4.7}
\]
\[
t_x \leq |t| \leq 1 : \text{ bounded using Lemma 4.3}
\]
\[
1 \leq |t| \leq x^2 : \text{ bounded using Lemmas 4.4 and 4.6}
\]
We denote by $I_1, I_2, I_3$ the integrals over these respective domains. We begin by computing the asymptotics for $I_1$, which gives the main term. When $|t| \leq t_x$ we have, by Corollary 4.7

$$\frac{\varphi(\sigma + it)}{\sigma + it} = \varphi(1) (1 + O(t_x + \sigma - 1)) = \varphi(1) \left(1 + O\left(\log^{-\frac{\alpha}{1+\epsilon}} x\right)\right)$$

(4.19)

since $\varphi$ has bounded derivative. A second-order Taylor approximation of $G(\sigma + it)$ around $t = 0$ gives

$$K \cdot G(\sigma + it) = K \cdot G(\sigma) + itK \cdot G'(\sigma) - \frac{t^2}{2} K \cdot G''(\sigma) + O\left(|t|^3 G'''(\sigma)\right)$$

(4.20)

for $|t| \leq t_x$, where in the first equality we used the fact that $|G'''(\sigma + it)| \leq |G'''(\sigma)|$ and in the second equality we used Corollary 4.7 and the definition of $\sigma$ in (4.13). From (4.19) and (4.20), we get

$$\varphi(\sigma + it)e^{K \cdot G(\sigma + it)} \frac{x^{\sigma + it}}{\sigma + it} = \varphi(1)x^\sigma e^{K \cdot G(\sigma) - \frac{t^2}{2} K \cdot G''(\sigma)} (1 + O\left(\log^{-x} z\right))$$

where $z := \min\{\gamma - \delta, 2\}/(2(\gamma + 1))$. We thus have

$$I_1 = \left(1 + O\left(\log^{-x} x\right)\right) A_0 x(\log x)^{-\frac{x^2}{4(\gamma + 1)}} \exp\left[1 + O\left(\log^{-x} x\right)\right] \frac{\varphi(1)x^\sigma e^{K \cdot G(\sigma)}}{\sqrt{2\pi K \cdot G''(\sigma)}},$$

which by Corollary 4.7 can be simplified to

$$I_1 = \left(1 + O\left(\log^{-x} x\right)\right) A_0 x(\log x)^{-\frac{x^2}{4(\gamma + 1)}} \exp\left[1 + O\left(\log^{-x} x\right)\right],$$

(4.21)

where $B$ is defined in (4.12), and $A_0$ is defined in (4.13). Next we bound $I_2$. Using Lemma 4.3 with $s = \sigma + it$ where $t_x \leq |t| \leq 1$, we get

$$\Re(G(\sigma + it)) \leq |G(\sigma + it)| \leq \frac{\Gamma(\gamma)}{|\sigma - 1 + it|^\gamma} + C \leq \frac{\Gamma(\gamma)}{|\sigma - 1 + it|^\gamma} + C \leq |G(\sigma + it_x)| + C,$$

and so a second-order Taylor approximation shows that

$$\Re(G(\sigma + it)) \leq \left|G(\sigma) + it_x G'(\sigma) - \frac{t_x^2}{2} G''(\sigma)\right| + C$$

$$= \sqrt{G'(\sigma)^2 - t_x^2 (G'(\sigma) G''(\sigma) - G''(\sigma)^2)} + C \leq \sqrt{G'(\sigma)^2 - c(\log x)^{\frac{\alpha}{3(1+\epsilon)}}} + C \leq G'(\sigma) - c(\log x)^{\frac{\alpha}{3(1+\epsilon)}} + C \leq G(\sigma) - c(\log x)^{\frac{\alpha}{3(1+\epsilon)}}$$

(4.22)

where the second inequality holds for sufficiently small $\delta$ and follows from Corollary 4.7. Thus

$$|I_2| \leq C x^\sigma e^{K \cdot G(\sigma) - c(\log x)^{\frac{\alpha}{3(1+\epsilon)}}} \leq \frac{C}{\log x} |I_1|.$$
Lemma 4.8. Let \( x \) be sufficiently large. Let \( 2 \leq h \leq x \). When \( \log h \leq (\log x)^{(\gamma + 4)/(4\gamma + 4)} \), we have that

\[
S(x/h) \leq C \cdot S(x) \tag{4.24}
\]

for \( x \geq 1, h \geq 1 \) by Theorem 1.6.

Lemma 4.8. Let \( \alpha : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0} \) be a multiplicative function satisfying (1.9)–(1.10) and suppose that \( x \) is sufficiently large. Let \( 2 \leq h \leq x \). When \( \log h \leq (\log x)^{(\gamma + 4)/(4\gamma + 4)} \), we have that

\[
S(x/h) = \frac{1}{h} \exp \left[ -\frac{B\gamma}{\gamma + 1} \frac{\log h}{(\log x)^{\frac{\gamma-1}{\gamma+1}}} + O\left(\frac{1}{(\log x)^{\gamma}}\right) \right].
\]

When \( \log h \geq (\log x)^{(\gamma + 4)/(4\gamma + 4)} \), we have that

\[
S(x/h) \leq \frac{1}{h} e^{-(\log x)^\gamma}.
\]

Proof. Let \( h_x := \exp((\log x)^{(\gamma + 4)/(4\gamma + 4)}) \). Suppose that \( h \leq h_x \). By a first-order Taylor approximation, we get

\[
(\log(x/h))^{\frac{\gamma}{\gamma+1}} = (\log x)^{\frac{\gamma}{\gamma+1}} \left(1 - \frac{\log h}{\log x}\right)^{\frac{\gamma}{\gamma+1}} = (\log x)^{\frac{\gamma}{\gamma+1}} - \frac{\gamma}{\gamma+1} \frac{\log h}{(\log x)^{\frac{\gamma}{\gamma+1}}} + O\left(\frac{1}{(\log x)^{\gamma}}\right).
\]

We now study the integral over \( S \). Applying Lemma 4.6, we find that for any \( t_1, t_2 \in S \) we have that \( \Re(G(\sigma + it(t_1 - t_2))) \geq (1 - 8\sqrt{2})G(\sigma) \) and therefore, by Lemma 4.4 and the choice of \( \epsilon \), either \( |t_1 - t_2| \leq 1 \) or \( |t_1 - t_2| \geq e^{1/(\sigma - 1)} \). It follows that

\[
S \subseteq \bigcup_{j=0}^{\infty} [a_j, b_j]
\]

for some \( a_{j+1} > b_j > a_j \geq 0 \) with

\[
b_j - a_j \leq 1, \quad a_j \geq j \cdot e^{\frac{1}{\sigma}} \quad \text{and} \quad a_0 = 0.
\]

Thus, for sufficiently large \( x \),

\[
\int_{[1, x^2] \cap S} \left| \varphi(\sigma + it)e^{K-G(\sigma) + it} \right| \leq C x^\sigma e^{K-G(\sigma)} \sum_{1 \leq j \leq x^2} \int_{a_j}^{b_j} \frac{dt}{t} \leq C x^\sigma e^{K-G(\sigma)} \sum_{1 \leq j \leq x^2} \frac{b_j - a_j}{a_j}
\]

\[
\leq C x^\sigma e^{K-G(\sigma)} e^{-\frac{1}{4\gamma+1}} \sum_{1 \leq j \leq x^2} \frac{1}{j} \leq C x^\sigma e^{K-G(\sigma)} e^{-\frac{1}{4\gamma+1}} \log x
\]

\[
\leq \frac{C}{\log x} |I_1|.
\]

Combining the estimates for the integrals over \([1, x^2] \setminus S\) and \([1, x^2] \cap S\), we obtain

\[
|I_3| \leq \frac{C}{\log x} |I_1|.
\]

We conclude the proof by plugging the estimates (4.18), (4.21), (4.22) and (4.23) in (4.17). \(\square\)

### 4.2 Proof of Theorem 1.6

#### 4.2.1 Auxiliary results

An important step in the proof is understanding the asymptotic behavior of \( \mathbb{P}(p \mid N_x) \). We shall see that \( \mathbb{P}(p \mid N_x) \approx \alpha(p)S(x/p)/S(x) \), and so begin by studying the ratio \( S(x/h)/S(x) \). Observe that

\[
S(x/h)h \leq C \cdot S(x)
\]

for \( x \geq 1, h \geq 1 \) by Theorem 1.6.

Let \( \alpha : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0} \) be a multiplicative function satisfying (1.9)–(1.10) and suppose that \( x \) is sufficiently large. Let \( 2 \leq h \leq x \). When \( \log h \leq (\log x)^{(\gamma + 4)/(4\gamma + 4)} \), we have that

\[
S(x/h) = \frac{1}{h} \exp \left[ -\frac{B\gamma}{\gamma + 1} \frac{\log h}{(\log x)^{\frac{\gamma-1}{\gamma+1}}} + O\left(\frac{1}{(\log x)^{\gamma}}\right) \right].
\]

When \( \log h \geq (\log x)^{(\gamma + 4)/(4\gamma + 4)} \), we have that

\[
S(x/h) \leq \frac{1}{h} e^{-(\log x)^\gamma}.
\]
Thus, by Theorem 1.4 applied with \( x \) and \( x/h \), we obtain the first part of the lemma. We turn to prove the second part of the lemma. Using the first part of the lemma and (4.24) we get that when \( h \geq h_x \)

\[
\frac{S(x/h)}{S(x)} = \frac{S(x/h)}{S(x/h_x)} \leq C h_x \frac{1}{h} \exp \left[ -\epsilon \frac{\log h_x}{(\log x)^{\gamma+1}} \right] \leq \frac{1}{h} e^{-\epsilon (\log x)^c},
\]

as needed.

**Lemma 4.9.** Let \( \alpha : \mathbb{N} \to \mathbb{R}_{\geq 0} \) be a multiplicative function satisfying \( \log \alpha(p) = o(\log p) \), (1.10) and (4.24). For sufficiently large \( x \) and any prime \( p \leq x \) we have that

\[
\mathbb{P}(p \mid N_x) = \frac{\alpha(p)S(x/p)}{S(x)} + O\left( \frac{1}{p \log^2 p} \right).
\]

**Proof.** By (4.24) we have for any \( y > 1 \)

\[
\sum_{n \leq y} \sum_{p \mid n} \alpha(n) \leq \sum_{k \geq 1} \sum_{p \mid \gamma| n} \alpha(n) \leq \sum_{k \geq 1} \alpha(p^k)S\left( \frac{y}{p^k} \right) \leq C \cdot S(y) \sum_{k \geq 1} \frac{\alpha(p^k)}{p^k} \leq C \cdot S(y) \sqrt{p}.
\]

Hence

\[
\mathbb{P}(p \mid N_x) = \frac{1}{S(x)} \sum_{n \leq x} \alpha(n) = \frac{\alpha(p)}{S(x)} \sum_{m \leq x/p} \alpha(m) = \frac{\alpha(p)S(x/p)}{S(x)} + O\left( \frac{\alpha(p)S(x/p)}{\sqrt{p}S(x)} \right)
\]

\[
= \frac{\alpha(p)S(x/p)}{S(x)} + O\left( \frac{1}{p \log^2 p} \right),
\]

where in the last inequality we use (4.24) again. Using the same arguments as in (4.25), we have that

\[
\mathbb{P}(p^2 \mid N_x) \leq C \sum_{k=2}^\infty \frac{\alpha(p^k)}{p^k} \leq \frac{C}{p \log^2 p}.
\]

As \( \mathbb{P}(p \mid N_x) = \mathbb{P}(p \mid N_x) + \mathbb{P}(p^2 \mid N_x) \), the proof is concluded.

**4.2.2 Conclusion of proof**

We begin with the first part of the theorem. We abbreviate \( P_1(N_x) \) as \( P_1 \). Fix \( 0 < a < b < \infty \). It suffices to show that

\[
\liminf_{x \to \infty} \mathbb{P}\left( a \leq \frac{\log P_1}{(\log x)^{\gamma+1}} \leq b \right) \geq \mathbb{P}(a \leq Y \leq b)
\]

(4.26)

where \( Y \) has gamma\((\gamma + 1, (K\Gamma(\gamma + 1))^{1/(\gamma+1)})\) distribution.

For any prime \( p \) such that \( a(\log x)^{\gamma+1} \leq \log p \leq b(\log x)^{\gamma+1} \), we have, by Lemmas 4.8 and 4.9

\[
\mathbb{P}(P_1 = p) = \sum_{n \leq x \atop p \mid n} \alpha(n) \frac{\nu_p(n) \log p}{\log n} \geq \frac{\log p}{\log x} \mathbb{P}(p \mid N_x) \geq \frac{\log p}{\log x} \left( K \log \gamma+1 p S(x/p) \right) + O\left( \frac{1}{p \log^2 p} \right)
\]

\[
\geq \left( 1 + O\left( \frac{1}{(\log x)^c} \right) \right) \frac{K \log \gamma+1 p}{\log x} \exp \left[ -\frac{B\gamma}{\gamma+1} \frac{\log p}{(\log x)^{\gamma+1}} \right],
\]

where the error term \( 1/(p \log^2 p) \) is absorbed in the last error term. Thus,

\[
\mathbb{P}\left( a \leq \frac{\log P_1}{(\log x)^{\gamma+1}} \leq b \right) \geq \left( 1 + O\left( \frac{1}{(\log x)^c} \right) \right) \sum_{n_x^y \leq p \leq n_x^b} \frac{K \log \gamma+1 p}{p \log x} \exp \left[ -\frac{B\gamma}{\gamma+1} \frac{\log p}{(\log x)^{\gamma+1}} \right]
\]

25
where $n_x := \exp((\log x)^{1/(\gamma+1)})$. By Lemma 3.4 with $\alpha \equiv 1$, the interval $[n_x^a, n_x^b]$ and
\[
g(t) = K \frac{\log^{\gamma+1} t}{t \log x} \exp \left[ - \frac{B \gamma}{\gamma+1} \frac{\log t}{(\log x)^{1+\gamma}} \right]
\]
we have
\[
\mathbb{P} \left( a \leq \frac{\log P_1}{(\log x)^{1/(\gamma+1)}} \leq b \right) \geq \left( 1 + O \left( \frac{1}{(\log x)^c} \right) \right) \left( \int_{n_x^a}^{n_x^b} K \log^\gamma t \frac{\log t}{t \log x} \exp \left[ - \frac{B \gamma}{\gamma+1} \frac{\log t}{(\log x)^{1+\gamma}} \right] dt + O(\log \frac{1}{\gamma+1} x) \right). \tag{4.27}
\]
The change of variables $t = n_x^z$ in the last integral shows that it equals $\mathbb{P}(a \leq Y \leq b)$. Taking $x$ to infinity in (4.27) we obtain (4.26), as needed.

We turn to the second part of the theorem. By Lemma 4.9 and (1.9) we have
\[
\mathbb{E}\omega(N_x) = \sum_{p \leq x} \mathbb{P}(p \mid N_x) = \sum_{p \leq x} \frac{K \log^\gamma p S(x/p)}{S(x)} + O \left( \sum_p \frac{1}{p \log^2 p} \right).
\]
The error term is bounded by a constant. In order to estimate the sum, we split it into three sums $S_1$, $S_2$ and $S_3$, over the respective ranges $p < \exp((\log x)^{\delta_1})$, $\exp((\log x)^{\delta_1}) \leq p \leq \exp((\log x)^{\delta_2})$ and $\exp((\log x)^{\delta_2}) < p \leq x$, where $\delta_1 = 1/(2(\gamma+1))$ and $\delta_2 = (\gamma+4)/(4\gamma+4)$. We bound $S_1$ using (4.24):
\[
S_1 \leq C \sum_{p \leq \exp((\log x)^{\delta_1})} \frac{\log^\gamma p}{p} \leq C \log^{\delta_1} x,
\]
where in the last inequality we used Lemma 8.4 with $\alpha \equiv 1$ and $g(t) = \log^\gamma t/t$. We bound $S_3$ using the second part of Lemma 4.8 which gives
\[
S_3 \leq e^{-(\log x)^c} \sum_{\exp((\log x)^{\delta_2}) < p \leq x} \frac{K \log^\gamma p}{p} \leq e^{-(\log x)^c}.
\]
We now estimate $S_2$. By Lemma 4.8
\[
S_2 = \left( 1 + O \left( \frac{1}{(\log x)^c} \right) \right) \sum_{(\log x)^{\delta_1} \leq p \leq (\log x)^{\delta_2}} \frac{K \log^\gamma p}{p} \exp \left[ - \frac{B \gamma}{\gamma+1} \frac{\log p}{(\log x)^{1+\gamma}} \right]. \tag{4.28}
\]
From (4.28) and Lemma 3.4 with $\alpha \equiv 1$ and
\[
g(t) = \frac{K \log^\gamma t}{t} \exp \left[ - \frac{B \gamma}{\gamma+1} \frac{\log t}{(\log x)^{1+\gamma}} \right]
\]
we obtain
\[
S_2 = \left( 1 + O \left( \frac{1}{(\log x)^c} \right) \right) \left( \int_{\exp((\log x)^{\delta_2})}^{\exp((\log x)^{\delta_1})} g(t) \frac{dt}{\log t} + O \left( \log \max\{0, \frac{1}{\gamma+1}\} x \right) \right).
\]
The change of variables $t = n_x^z$ in the last integral shows that it equals
\[
\int_{(\log x)^{\delta_1}}^{(\log x)^{\delta_2}} K z^{\gamma-1} \exp \left( - \frac{B \gamma z}{\gamma+1} \right) dz
\]
\[
= (\log x)^{\frac{\gamma+1}{\gamma+1}} K \Gamma(\gamma) \left( \frac{2 \gamma + 1}{B \gamma} \right) \left( 1 + O \left( (\log x)^{(\delta_1-\delta_2)} \right) \right).
\]
Since all the accumulated error terms are of order smaller than $(\log x)^{\gamma/(\gamma+1)}$, the expectation of $\omega$ is estimated. The expectation of $\Omega$ behaves the same by Lemma 4.2.\qed
References

[1] K. Alladi. Additive functions and special sets of integers. In Number theory (Mysore, 1981), volume 938 of Lecture Notes in Math., pages 1–49. Springer, Berlin-New York, 1982.

[2] K. Alladi. Moments of additive functions and sieve methods. In Number theory (New York, 1982), volume 1052 of Lecture Notes in Math., pages 1–25. Springer, Berlin, 1984.

[3] K. Alladi. Moments of additive functions and the sequence of shifted primes. Pacific J. Math., 118(2):261–275, 1985.

[4] K. Alladi. A study of the moments of additive functions using Laplace transforms and sieve methods. In Number theory (Ootacamund, 1984), volume 1122 of Lecture Notes in Math., pages 1–37. Springer, Berlin, 1985.

[5] C. E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Ann. Statist., 2:1152–1174, 1974.

[6] R. Arratia, A. D. Barbour, and S. Tavaré. Logarithmic combinatorial structures: a probabilistic approach. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich, 2003.

[7] R. Arratia, F. Kochman, and V. S. Miller. Extensions of Billingsley’s theorem via multi-intensities. arXiv preprint arXiv:1401.1555, 2014.

[8] V. Betz and D. Ueltschi. Spatial random permutations and infinite cycles. Comm. Math. Phys., 285(2):469–501, 2009.

[9] V. Betz and D. Ueltschi. Spatial random permutations with small cycle weights. Probab. Theory Related Fields, 149(1-2):191–222, 2011.

[10] V. Betz, D. Ueltschi, and Y. Velenik. Random permutations with cycle weights. Ann. Appl. Probab., 21(1):312–331, 2011.

[11] P. Billingsley. On the central limit theorem for the prime divisor functions. Amer. Math. Monthly, 76:132–139, 1969.

[12] P. Billingsley. On the distribution of large prime divisors. Period. Math. Hungar., 2:283–289, 1972. Collection of articles dedicated to the memory of Alfréd Rényi, I.

[13] A. Cipriani and D. Zeindler. The limit shape of random permutations with polynomially growing cycle weights. ALEA Lat. Am. J. Probab. Math. Stat., 12(2):971–999, 2015.

[14] R. de la Bretèche and G. Tenenbaum. Remarks on the Selberg-Delange method. Acta Arith., 200(4):349–369, 2021.

[15] S. Dereich and P. Mörters. Cycle length distributions in random permutations with diverging cycle weights. Random Structures Algorithms, 46(4):635–650, 2015.

[16] P. Donnelly and G. Grimmett. On the asymptotic distribution of large prime factors. J. London Math. Soc. (2), 47(3):395–404, 1993.

[17] S. Eberhard, K. Ford, and B. Green. Permutations fixing a k-set. Int. Math. Res. Not. IMRN, (21):6713–6731, 2016.

[18] S. Eberhard, K. Ford, and B. Green. Invariable generation of the symmetric group. Duke Math. J., 166(8):1573–1590, 2017.

[19] D. Elboim and R. Peled. Limit distributions for Euclidean random permutations. Comm. Math. Phys., 369(2):457–522, 2019.

[20] P. D. T. A. Elliott. Central limit theorems for classical cusp forms. Ramanujan J., 36(1-2):81–98, 2015.
[21] P. D. T. A. Elliott. Corrigendum to: “Central limit theorems for classical cusp forms”. Ramanujan J., 36(1-2):99–102, 2015.

[22] N. M. Ercolani and D. Ueltschi. Cycle structure of random permutations with cycle weights. Random Structures Algorithms, 44(1):109–133, 2014.

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

[24] M. M. Erlihson and B. L. Granovsky. Limit shapes of Gibbs distributions on the set of integer partitions: the expansive case. Ann. Inst. Henri Poincaré Probab. Stat., 44(5):915–945, 2008.

[25] W. J. Ewens. The sampling theory of selectively neutral alleles. Theoret. Population Biology, 3:87–112; erratum, ibid. 3 (1972), 240; erratum, ibid. 3 (1972), 376, 1972.

[26] A. Granville. Cycle lengths in a permutation are typically Poisson. Electron. J. Combin., 13(1):Research Paper 107, 23, 2006.

[27] A. Granville. Prime divisors are Poisson distributed. Int. J. Number Theory, 3(1):1–18, 2007.

[28] A. Granville. The anatomy of integers and permutations. preprint, 2008.

[29] A. Granville and D. Koukoulopoulos. Beyond the LSD method for the partial sums of multiplicative functions. Ramanujan J., 49(2):287–319, 2019.

[30] A. Granville and K. Soundararajan. Sieving and the Erdős-Kac theorem. In Equidistribution in number theory, an introduction, volume 237 of NATO Sci. Ser. II Math. Phys. Chem., pages 15–27. Springer, Dordrecht, 2007.

[31] J. C. Hansen. A functional central limit theorem for the Ewens sampling formula. J. Appl. Probab., 27(1):28–43, 1990.

[32] D. R. Heath-Brown. The number of primes in a short interval. J. Reine Angew. Math., 389:22–63, 1988.

[33] G. Hoheisel. Primzahlprobleme in der Analysis. Sitzungsber. Preuß. Akad. Wiss., Phys.-Math. Kl., 1930:580–588, 1930.

[34] A. E. Ingham. A Tauberian theorem for partitions. Ann. of Math. (2), 42:1075–1090, 1941.

[35] R. Khan, M. B. Milinovich, and U. Subedi. A Weighted Version of the Erdős-Kac Theorem. Journal of Number Theory, 2021.

[36] J. F. C. Kingman. The coalescent. Stochastic Process. Appl., 13(3):235–248, 1982.

[37] M. Lugo. Profiles of permutations. Electron. J. Combin., 16(1):Research Paper 99, 20, 2009.

[38] E. Manstavičius. Mappings on decomposable combinatorial structures: analytic approach. Combin. Probab. Comput., 11(1):61–78, 2002.

[39] E. Manstavičius. An analytic method in probabilistic combinatorics. Osaka J. Math., 46(1):273–290, 2009.

[40] E. Manstavičius. On mean values of multiplicative functions on the symmetric group. Monatsh. Math., 182(2):359–376, 2017.

[41] K. Maples, A. Nikeghbali, and D. Zeindler. On the number of cycles in a random permutation. Electron. Commun. Probab., 17:no. 20, 13, 2012.

[42] E. E. Marenich. Summation of the values of arithmetic functions. Math. USSR, Sb., 46:51–73, 1983.

[43] A. Nikeghbali, J. Storm, and D. Zeindler. Large cycles and a functional central limit theorem for generalized weighted random permutations. arXiv preprint arXiv:1302.5938, 2013.
[44] A. Nikeghbali and D. Zeindler. The generalized weighted probability measure on the symmetric group and the asymptotic behavior of the cycles. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(4):961–981, 2013.

[45] R. W. K. Odoni. A problem of Rankin on sums of powers of cusp-form coefficients. *J. London Math. Soc. (2)*, 44(2):203–217, 1991.

[46] R. W. K. Odoni. Solution of a generalised version of a problem of Rankin on sums of powers of cusp-form coefficients. *Acta Arith.*, 104(3):201–223, 2002.

[47] V. V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1995. Sequences of independent random variables, Oxford Science Publications.

[48] J. Pitman and M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25(2):855–900, 1997.

[49] N. Robles and D. Zeindler. Random permutations with logarithmic cycle weights. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(3):1991–2016, 2020.

[50] W. Schwarz. Einige Anwendungen Tauberscher Sätze in der Zahlentheorie. *A. J. Reine Angew. Math.*, 219:67–96, 1965.

[51] J. Storm and D. Zeindler. The order of large random permutations with cycle weights. *Electron. J. Probab.*, 20:no. 126, 34, 2015.

[52] G. Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 163 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, third edition, 2015. Translated from the 2008 French edition by Patrick D. F. Ion.

[53] G. Tenenbaum. Moyennes effectives de fonctions multiplicatives complexes. *Ramanujan J.*, 44(3):641–701, 2017.

[54] A. N. Timashev. Random permutations with cycle lengths in a given finite set. *Diskret. Mat.*, 20(1):25–37, 2008.

[55] G. A. Watterson. The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Probability*, 13(4):639–651, 1976.

[56] E. Wirsing. Das asymptotische Verhalten von Summen über multiplikative Funktionen. *Math. Ann.*, 143:75–102, 1961.

[57] E. Wirsing. Das asymptotische Verhalten von Summen über multiplikative Funktionen. II. *Acta Math. Acad. Sci. Hungar.*, 18:411–467, 1967.

[58] A. L. Yakymiv. Random A-permutations: convergence to a Poisson process. *Mat. Zametki*, 81(6):939–947, 2007.

Department of Mathematics, Princeton University, Princeton, NJ 08544, USA
E-mail address: delboim@math.princeton.edu

Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK
E-mail address: ofir.goro@gmail.com