

# A NATURAL PROBABILISTIC MODEL ON THE INTEGERS AND ITS RELATION TO DICKMAN-TYPE DISTRIBUTIONS AND BUCHSTAB'S FUNCTION

ROSS G. PINSKY

ABSTRACT. Let  $\{p_j\}_{j=1}^\infty$  denote the set of prime numbers in increasing order, let  $\Omega_N \subset \mathbb{N}$  denote the set of positive integers with no prime factor larger than  $p_N$  and let  $P_N$  denote the probability measure on  $\Omega_N$  which gives to each  $n \in \Omega_N$  a probability proportional to  $\frac{1}{n}$ . This measure is in fact the distribution of the random integer  $I_N \in \Omega_N$  defined by  $I_N = \prod_{j=1}^N p_j^{X_{p_j}}$ , where  $\{X_{p_j}\}_{j=1}^\infty$  are independent random variables and  $X_{p_j}$  is distributed as  $\text{Geom}(1 - \frac{1}{p_j})$ . We show that  $\frac{\log n}{\log N}$  under  $P_N$  converges weakly to the Dickman distribution. As a corollary, we recover a classical result from multiplicative number theory—Mertens' formula. Let  $D_{\text{nat}}(A)$  denote the natural density of  $A \subset \mathbb{N}$ , if it exists, and let  $D_{\log\text{-indep}}(A) = \lim_{N \rightarrow \infty} P_N(A \cap \Omega_N)$  denote the density of  $A$  arising from  $\{P_N\}_{N=1}^\infty$ , if it exists. We show that the two densities coincide on a natural algebra of subsets of  $\mathbb{N}$ . We also show that they do not agree on the sets of  $n^{\frac{1}{s}}$ -smooth numbers  $\{n \in \mathbb{N} : p^+(n) \leq n^{\frac{1}{s}}\}$ ,  $s > 1$ , where  $p^+(n)$  denotes the largest prime divisor of  $n$ . This last consideration concerns distributions involving the Dickman function. We also consider the sets of  $n^{\frac{1}{s}}$ -rough numbers  $\{n \in \mathbb{N} : p^-(n) \geq n^{\frac{1}{s}}\}$ ,  $s > 1$ , where  $p^-(n)$  denotes the smallest prime divisor of  $n$ . We show that the probabilities of these sets, under the uniform distribution on  $[N] = \{1, \dots, N\}$  and under the  $P_N$ -distribution on  $\Omega_N$ , have the same asymptotic decay profile as functions of  $s$ , although their rates are necessarily different. This profile involves the Buchstab function. We also prove a new representation for the Buchstab function.

## 1. INTRODUCTION AND STATEMENT OF RESULTS

For a subset  $A \subset \mathbb{N}$ , the natural density  $D_{\text{nat}}(A)$  of  $A$  is defined by  $D_{\text{nat}}(A) = \lim_{N \rightarrow \infty} \frac{|A \cap [N]|}{N}$ , whenever this limit exists, where  $[N] = \{1, \dots, N\}$ .

---

2000 *Mathematics Subject Classification.* 60F05, 11N25, 11K65.

*Key words and phrases.* Dickman function, Dickman-type distribution, Buchstab function, prime number,  $k$ -free numbers .

The natural density is additive, but not  $\sigma$ -additive, and therefore not a measure. For each prime  $p$  and each  $n \in \mathbb{N}$ , define the nonnegative integer  $\beta_p(n)$ , the  $p$ -adic order of  $n$ , by  $\beta_p(n) = m$ , if  $p^m \mid n$  and  $p^{m+1} \nmid n$ . Let  $\delta_p(n) = \max(1, \beta_p(n))$  denote the indicator function of the set of positive integers divisible by  $p$ . It is clear that for each  $m \in \mathbb{N}$ , the natural density of the set  $\{n \in \mathbb{N} : \beta_p(n) \geq m\}$  of natural numbers divisible by  $p^m$  is  $(\frac{1}{p})^m$ . More generally, it is easy to see that for  $l \in \mathbb{N}$ ,  $\{m_j\}_{j=1}^l \subset \mathbb{N}$  and distinct primes  $\{p_j\}_{j=1}^l$ , the natural density of the set  $\{n \in \mathbb{N} : \beta_{p_j}(n) \geq m_j, j = 1, \dots, l\}$  is  $\prod_{j=1}^l (\frac{1}{p_j})^{m_j}$ . That is, the distribution of the random vector  $\{\delta_{p_j}\}_{j=1}^l$ , defined on the probability space  $[N]$  with the uniform distribution, converges weakly as  $N \rightarrow \infty$  to the random vector  $\{Y_{p_j}\}_{j=1}^l$  with independent components distributed according to the Bernoulli distributions  $\{\text{Ber}(\frac{1}{p_j})\}_{j=1}^l$ , and the distribution of the random vector  $\{\beta_{p_j}\}_{j=1}^l$  converges weakly as  $N \rightarrow \infty$  to the random vector  $\{X_{p_j}\}_{j=1}^l$  with independent components distributed according to the geometric distributions  $\text{Geom}(1 - \frac{1}{p_j})$  ( $P(X_{p_j} = m) = (\frac{1}{p_j})^m (1 - \frac{1}{p_j})$ ,  $m = 0, 1, \dots$ ). This fact is the starting point of probabilistic number theory.

Denote the primes in increasing order by  $\{p_j\}_{j=1}^\infty$ . In the sequel, we will assume that the random variables  $\{X_{p_j}\}_{j=1}^\infty, \{Y_{p_j}\}_{j=1}^\infty$  with distributions as above are defined as independent random variables on some probability space, and we will use the generic notation  $P$  to denote probabilities corresponding to these random variables.

A real-valued function  $f$  defined on  $\mathbb{N}$  is called a real arithmetic function. It is called *additive* if  $f(nm) = f(n) + f(m)$ , whenever  $(m, n) = 1$ . If in addition,  $f(p^m) = f(p)$ , for all primes  $p$  and all  $m \geq 2$ , then it is called *strongly additive*. Classical examples of additive arithmetic functions are, for example,  $\log \frac{\phi(n)}{n}$ , where  $\phi$  is the Euler totient function,  $\omega(n)$ , the number of distinct prime divisors of  $n$ ,  $\Omega(n)$ , the number of prime divisors of  $n$  counting multiplicities and  $\log \sigma(n)$ , where  $\sigma$  is the sum-of-divisors function. The first two of these functions are strongly additive while the last two are not.

If  $f$  is additive, then  $f(1) = 0$ . Writing  $n \in \mathbb{N}$  as  $n = \prod_{j=1}^\infty p_j^{\beta_{p_j}(n)}$ , we have for  $f$  additive,  $f(n) = \sum_{j=1}^\infty f(p_j^{\beta_{p_j}(n)})$ , and for  $f$  strongly additive,  $f(n) = \sum_{j=1}^\infty f(p_j^{\delta_{p_j}(n)}) = \sum_{j=1}^\infty f(p_j) \delta_{p_j}(n)$ . Equivalently, for each  $N \in \mathbb{N}$ ,

we have for  $f$  additive,

$$(1.1) \quad f(n) = \sum_{j=1}^N f(p_j^{\beta_{p_j}(n)}), \quad n \in [N],$$

and for  $f$  strongly additive,

$$(1.2) \quad f(n) = \sum_{j=1}^N f(p_j) \delta_{p_j}(n), \quad n \in [N].$$

In light of the above discussion, it is natural to compare (1.1) to

$$(1.3) \quad \mathcal{X}_N \equiv \sum_{j=1}^N f(p_j^{X_{p_j}}),$$

and to compare (1.2) to

$$(1.4) \quad \mathcal{Y}_N \equiv \sum_{j=1}^N f(p_j) Y_{p_j}.$$

Now  $\mathcal{Y}_N$  converges in distribution as  $N \rightarrow \infty$  if and only if it converges almost surely, and the almost sure convergence of  $\mathcal{Y}_N$  is characterized by the Kolmogorov three series theorem [8]. Since  $EY_{p_j} = EY_{p_j}^2 = \frac{1}{p_j}$ , it follows from that theorem that  $\mathcal{Y}_N$  converges almost surely if and only if the following three series converge: 1.  $\sum_{j:|f(p_j)| \leq 1} \frac{f(p_j)}{p_j}$ ; 2.  $\sum_{j:|f(p_j)| \leq 1} \frac{f^2(p_j)}{p_j}$ ; 3.  $\sum_{j:|f(p_j)| > 1} \frac{1}{p_j}$ . Since  $P(X_{p_j} \geq 2) = \frac{1}{p_j^2}$ , it follows from the Borel-Cantelli lemma that  $\sum_{j=1}^{\infty} 1_{\{X_{p_j} \geq 2\}}$  is almost surely finite; thus the very same criterion also determines whether  $\mathcal{X}_N$  converges almost surely. The Erdős-Wintner theorem [11] states that for additive  $f$ , the convergence of these three series is a necessary and sufficient condition for the convergence in distribution as  $N \rightarrow \infty$  of the random variable  $f(n)$  in (1.1) on the probability space  $[N]$  with the uniform distribution. In the same spirit, the Kac-Erdős theorem [12] states that if  $f$  is strongly additive and bounded, then a central limit theorem holds as  $N \rightarrow \infty$  for  $f(n)$  on the probability space  $[N]$  with the uniform distribution, if the conditions of the Feller-Lindeberg central limit theorem hold for  $\mathcal{Y}_N$ . An appropriate corresponding result can be stated for additive  $f$  and or unbounded  $f$ . There is also a weak law of large numbers result, which in the case of  $f = \omega$  goes by the name of the Hardy-Ramanujan theorem [14]. It should be noted that the original proof

of Hardy and Ramanujan was quite complicated and not at all probabilistic; however, the later and much simpler proof of Turan [24] has a strong probabilistic flavor. For a concise and very readable probabilistic approach to these results, see Billingsley [3]; for a more encyclopedic probabilistic approach, see Elliott [9, 10]; for a less probabilistic approach, see Tenenbaum [23].

Turan's paper with the proof of the Hardy-Ramanujan theorem, as well as the Erdős-Wintner theorem and several papers leading up to it, all appeared in the 1930's, and the Kac-Erdős theorem appeared in 1940. Now large deviations for independent and non-identically distributed random variables have been readily available since the 1970's, thus this author certainly finds it quite surprising that until very recently no one extended the parallel between (1.2) and (1.4), or (1.1) and (1.3), to study the large deviations of (1.2) or (1.1)! See [16, 17].

Another density that is sometimes used in number theory is the *logarithmic density*,  $D_{\log}$ , which is defined by

$$(1.5) \quad D_{\log}(A) = \lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{n \in A \cap [N]} \frac{1}{n},$$

for  $A \subset \mathbb{N}$ , whenever this limit exists. Using summation by parts, it is easy to show that if  $D_{\text{nat}}(A)$  exists, then  $D_{\log}(A)$  exists and coincides with  $D_{\text{nat}}(A)$  [23]. (On the other hand, there are sets without natural density for which the logarithmic density exists. The most prominent of these are the sets  $\{B_d\}_{d=1}^9$  associated with Benford's law, where  $B_d$  is the set of positive integers whose first digit is  $d$ . One has  $D_{\log}(B_d) = \log_{10}(1 + \frac{1}{d})$ .) Thus, also on the probability space  $[N]$  with the probability measure which gives to each integer  $n$  a measure proportional to  $\frac{1}{n}$ , the distribution of the random vector  $\{\beta_{p_j}\}_{j=1}^l$  converges weakly as  $N \rightarrow \infty$  to the random vector  $\{X_{p_j}\}_{j=1}^l$  with independent components distributed according to the geometric distributions  $\text{Geom}(1 - \frac{1}{p_j})$ .

Motivated by the background described above, in this paper we consider a sequence of probability measures on  $\mathbb{N}$  which may be thought of as a synthesis between the logarithmic density  $D_{\log}$  and the concept of approximating the natural density via a sequence of independent random variables.

Let us denote by

$$\Omega_N = \{n \in \mathbb{N} : p_j \nmid n, j > N\}$$

the set of positive integers with no prime divisor larger than  $p_N$ . By the Euler product formula,

$$(1.6) \quad C_N \equiv \sum_{n \in \Omega_N} \frac{1}{n} = \prod_{j=1}^N \left(1 - \frac{1}{p_j}\right)^{-1} < \infty.$$

Let  $P_N$  denote the probability measure on  $\Omega_N$  for which the probability of  $n$  is proportional to  $\frac{1}{n}$ ; namely,

$$(1.7) \quad P_N(\{n\}) = \frac{1}{C_N} \frac{1}{n}, \quad n \in \Omega_N.$$

The connection between  $P_N$  and the logarithmic density is clear; the connection between  $P_N$  and a sequence of independent random variables comes from the following proposition. Define a random positive integer  $I_N \in \Omega_N$  by

$$I_N = \prod_{j=1}^N p_j^{X_{p_j}}.$$

**Proposition 1.** *The distribution of  $I_N$  is  $P_N$ ; that is,*

$$P_N(\{n\}) = P(I_N = n), \quad n \in \Omega_N.$$

*Proof.* Let  $n = \prod_{j=1}^N p_j^{a_j} \in \Omega_N$ . We have

$$P(I_N = n) = \prod_{j=1}^N P(X_{p_j} = a_j) = \prod_{j=1}^N \left(\frac{1}{p_j}\right)^{a_j} \left(1 - \frac{1}{p_j}\right) = \frac{1}{C_N} \frac{1}{n} = P_N(\{n\}).$$

□

Let  $D_{\log\text{-indep}}$  denote the asymptotic density obtained from  $P_N$ :

$$D_{\log\text{-indep}}(A) = \lim_{N \rightarrow \infty} P_N(A \cap \Omega_N) = \lim_{N \rightarrow \infty} \frac{1}{C_N} \sum_{n \in A \cap \Omega_N} \frac{1}{n},$$

for  $A \subset \mathbb{N}$ , whenever the limit exists. Note that the weight functions used in calculating the asymptotic densities  $D_{\log\text{-indep}}$  and  $D_{\log}$  have the same profile, but the sequences of subsets of  $\mathbb{N}$  over which the limits are taken, namely  $\{\Omega_N\}_{N=1}^{\infty}$  and  $\{[N]\}_{N=1}^{\infty}$ , are different. As already noted, when  $D_{\text{nat}}(A)$  exists, so does  $D_{\log}(A)$  and they coincide. We will show below in Proposition 3 that the densities  $D_{\log\text{-indep}}$  and  $D_{\text{nat}}$  coincide on many

natural subsets of  $\mathbb{N}$ . However we will also show below in Theorem 2 that they disagree on certain important, fundamental subsets of  $\mathbb{N}$ .

For  $k \geq 2$ , a positive integer  $n$  is called *k-free* if  $p^k \nmid n$ , for all primes  $p$ . When  $k = 2$ , one uses the term *square-free*. Let  $S_k$  denote the set of all  $k$ -free positive integers. Let

$$\Omega_N^{(k)} = \Omega_N \cap S_k.$$

Note that  $\Omega_N^{(k)}$  is a finite set; it has  $k^N$  elements. The measure  $P_N$  behaves nicely under conditioning on  $S_k$ . For  $k \geq 2$ , define the measure  $P_N^{(k)}$  by

$$P_N^{(k)}(\cdot) = P_N(\cdot | S_k).$$

Let  $\{X_{p_j}^{(k)}\}_{j=1}^\infty$  be independent random variables with  $X_{p_j}^{(k)}$  distributed as  $X_{p_j}$  conditioned on  $\{X_{p_j} < k\}$ . (Assume that these new random variables are defined on the same space as the  $\{X_{p_j}\}_{j=1}^\infty$  so that we can still use  $P$  for probabilities.) Let

$$I_N^{(k)} = \prod_{j=1}^N p_j^{X_{p_j}^{(k)}}.$$

**Proposition 2.** *The distribution of  $I_N^{(k)}$  is  $P_N^{(k)}$ .*

*Proof.*

$$P_N^{(k)}(\{n\}) = P_N(\{n\} | S_k) = P(I_N = n | X_{p_j} < k, j \in [N]) = P(I_N^{(k)} = n),$$

where the second equality follows from Proposition 1.  $\square$

**Remark.** The measure  $P_N^{(2)}$  was considered by Cellarosi and Sinai in [6]. See also the remark after Theorem 1 below.

We will prove the following result, which identifies a certain natural algebra of subsets of  $\mathbb{N}$  on which  $D_{\log\text{-indep}}$  and  $D_{\text{nat}}$  coincide.

**Proposition 3.** *The densities  $D_{\log\text{-indep}}$  and  $D_{\text{nat}}$  coincide on the algebra of subsets of  $\mathbb{N}$  generated by the inverse images of  $\{\beta_{p_j}\}_{j=1}^\infty$  and the sets  $\{S_k\}_{k=2}^\infty$ .*

We will show that under the measure  $P_N$  as well as under the measure  $P_N^{(k)}$ , the random variable  $\frac{\log n}{\log N}$ , with  $n \in \Omega_N$  in the case of  $P_N$  and  $n \in \Omega_N^{(k)}$  in the case of  $P_N^{(k)}$ , converges in distribution as  $N \rightarrow \infty$  to the distribution whose density is  $e^{-\gamma}\rho(x)$ ,  $x \in [0, \infty)$ , where  $\gamma$  is Euler's constant, and  $\rho$  is

the Dickman function, which we now describe. The Dickman function is the unique continuous function satisfying

$$\rho(x) = 1, \quad x \in (0, 1],$$

and satisfying the differential-delay equation

$$x\rho'(x) + \rho(x-1) = 0, \quad x > 1.$$

By analyzing the Laplace transform of  $\rho$ , a rather short proof shows that  $\int_0^\infty \rho(x)dx = e^\gamma$ ; thus  $e^{-\gamma}\rho(x)$  is indeed a probability density on  $[0, \infty)$ . We will call this distribution the *Dickman distribution*. The distribution decays very rapidly; indeed, it is not hard to show that  $\rho(s) \leq \frac{1}{\Gamma(s+1)}$ . For an analysis of the Dickman function, see for example, [23] or [18].

**Theorem 1.** *Under both  $P_N$  and  $P_N^{(k)}$ ,  $k \geq 2$ , the random variable  $\frac{\log n}{\log N}$  converges weakly to the Dickman distribution.*

**Remark.** For  $P_N^{(2)}$ , Theorem 1 was first proved by Cellarosi and Sinai [6]. Their proof involved calculating characteristic functions and was quite tedious and long. Our short proof uses Laplace transforms and the asymptotic growth rate of the primes given by the Prime Number Theorem (henceforth PNT). After this paper was written, one of the authors of [13] pointed out to the present author that their paper also gives a simpler proof of the result in [6].

Using Theorem 1 we can recover a classical result from multiplicative number theory; namely,

**Mertens' formula.**

$$(1.8) \quad C_N = \sum_{n \in \Omega_N} \frac{1}{n} = \prod_{j=1}^N \left(1 - \frac{1}{p_j}\right)^{-1} \sim e^\gamma \log N, \quad \text{as } N \rightarrow \infty.$$

(Traditionally the formula is written as  $\prod_{p \leq N} \left(1 - \frac{1}{p}\right)^{-1} \sim e^\gamma \log N$ , where the product is over all primes less than or equal to  $N$ . To show that the two are equivalent only requires the fact that  $p_N = o(N^{1+\epsilon})$ , for any  $\epsilon > 0$ .) A nice, alternative form of the formula is

$$\frac{\sum_{n \in \Omega_N} \frac{1}{n}}{\sum_{n=1}^N \frac{1}{n}} \sim e^\gamma.$$

Here is the derivation of Mertens' formula from Theorem 1. From the definition of  $P_N$ , we have  $P_N(\frac{\log n}{\log N} \leq 1) = \frac{1}{C_N} \sum_{n=1}^N \frac{1}{n}$ . Thus, from Theorem 1, we have  $\lim_{N \rightarrow \infty} \frac{1}{C_N} \sum_{n=1}^N \frac{1}{n} = \int_0^1 e^{-\gamma} \rho(x) dx = e^{-\gamma}$ . Now (1.8) follows from this and the fact that  $\sum_{n=1}^N \frac{1}{n} \sim \log N$ .

A direct proof that  $C_N \sim c \log N$ , for some  $c$ , follows readily with the help of Mertens' second theorem (see (1.18)). The proof that the constant is  $e^{-\gamma}$  is quite nontrivial. Of course, our proof of Mertens' formula via Theorem 1 uses the fact that  $\int_0^\infty \rho(x) dx = e^{-\gamma}$ , but as noted, this result is obtained readily by analyzing the Laplace transform of  $\rho$ .

We now present a proof, independent of the proof we will give later for Theorem 1, that if the limiting distribution of  $\frac{\log n}{\log N}$  under  $P_N$  exists, then it must be the Dickman distribution. We believe that this is of independent interest. Let

$$J_N^+ = \max\{j \in [N] : X_{p_j} \neq 0\},$$

with  $\max \emptyset \equiv 0$ . By Proposition 1, the distribution of  $\frac{\log n}{\log N}$  under  $P_N$  is equal to the distribution of

$$(1.9) \quad D_N \equiv \frac{1}{\log N} \sum_{n=1}^N X_{p_j} \log p_j = \left(\frac{\log J_N^+}{\log N}\right) \frac{1}{\log J_N^+} \sum_{j=1}^{J_N^+-1} X_{p_j} \log p_j + X_{p_{J_N^+}} \frac{\log p_{J_N^+}}{\log N},$$

where, of course, the sum on the right hand side above is interpreted as equal to 0 if  $J_N^+ \leq 1$ , and where we define  $p_0 = 1$ . Our assumption is that  $\{D_N\}_{N=1}^\infty$  converges weakly to some distribution. Since  $P(J_N^+ \leq j) = \prod_{m=j+1}^N (1 - \frac{1}{p_m})$ , we have  $J_N^+ \rightarrow \infty$  a.s. as  $N \rightarrow \infty$ . Also, by the independence of  $\{X_{p_j}\}_{j=1}^\infty$ , we have  $\sum_{j=1}^{J_N^+-1} X_{p_j} \log p_j | \{J_N^+ = j_0\} \stackrel{\text{dist}}{=} \sum_{j=1}^{j_0-1} X_{p_j} \log p_j$ . Thus,  $\frac{1}{\log J_N^+} \sum_{j=1}^{J_N^+-1} X_{p_j} \log p_j$  converges weakly to the same distribution. Using no more than the weak form of Merten's formula (namely,  $\prod_{j=1}^N (1 - \frac{1}{p_j})^{-1} \sim c \log N$ , for some  $c$ ) for the asymptotic equivalence below, we have for  $0 < x < 1$ ,

$$(1.10) \quad P\left(\frac{\log J_N^+}{\log N} \leq x\right) = P(J_N^+ \leq N^x) = \prod_{j=[N^x+1]}^N \left(1 - \frac{1}{p_j}\right) \sim \frac{\log N^x}{\log N} = x.$$



Using only the fact that  $p_j = o(j^{(1+\epsilon)})$ , for any  $\epsilon > 0$ , it follows that (1.10) also holds with  $\frac{\log J_N^+}{\log N}$  replaced by  $\frac{\log p_{J_N^+}}{\log N}$ . Note that  $X_{p_{J_N^+}}$  conditioned on  $\{J_N^+ = j_0\}$  is distributed as  $X_{p_{j_0}}$  conditioned on  $\{X_{p_{j_0}} \geq 1\}$ . A trivial calculation shows that the conditional distribution of  $X_{p_j}$ , conditioned on  $X_{p_j} \geq 1$ , converges weakly to 1 as  $j \rightarrow \infty$ . From the above facts and (1.9) it follows that if  $D$  denotes a random variable distributed according to the limiting distribution of  $\{D_N\}_{N=1}^\infty$ , then

$$(1.11) \quad D \stackrel{\text{dist}}{=} DU + U, \quad U \stackrel{\text{dist}}{=} \text{Unif}([0, 1]), \quad U \text{ and } D \text{ independent.}$$

From this, it is a calculus exercise to show that  $D$  has a continuous density  $f$ , that  $f$  is equal to some constant  $c$  on  $(0, 1]$ , and that  $f$  satisfies the differential-delay equation satisfied by the Dickman function  $\rho$  on  $x > 1$ . (See, for example, [21].) Thus  $f = c\rho$ . Since  $f$  is a density and since  $\int_0^\infty \rho(x)dx = e^\gamma$ , it follows that the density of  $D$  is  $e^{-\gamma}\rho$ .

The Dickman function arises in probabilistic number theory in the context of so-called *smooth* numbers; that is, numbers all of whose prime divisors are “small.” Let  $\Psi(x, y)$  denote the number of positive integers less than or equal to  $x$  with no prime divisors greater than  $y$ . Numbers with no prime divisors greater than  $y$  are called *y-smooth* numbers. Then for  $s \geq 1$ ,  $\Psi(N, N^{\frac{1}{s}}) \sim N\rho(s)$ , as  $N \rightarrow \infty$ . This result was first proved by Dickman in 1930 [7], whence the name of the function, with later refinements by de Bruijn [4]. (In particular, there are rather precise error terms.) See also [18] or [23]. Let  $p^+(n)$  denote the largest prime divisor of  $n$ . Then Dickman’s result states that the random variable  $\frac{\log N}{\log p^+(n)}$  on the probability space  $[N]$  with the uniform distribution converges weakly in distribution as  $N \rightarrow \infty$  to the distribution whose distribution function is  $1 - \rho(s)$ ,  $s \geq 1$ , and whose density is  $-\rho'(s) = \frac{\rho(s-1)}{s}$ ,  $s \geq 1$ . Since  $\frac{\log n}{\log N}$  on the probability space  $[N]$  with the uniform distribution converges weakly in distribution to 1 as  $N \rightarrow \infty$ , an equivalent statement of Dickman’s result is that the random variable  $\frac{\log n}{\log p^+(n)}$  on the probability space  $[N]$  with the uniform distribution converges weakly in distribution as  $N \rightarrow \infty$  to the distribution whose distribution function is  $1 - \rho(s)$ ,  $s \geq 1$ . For later use, we state this as

follows in terms of the natural density:

$$(1.12) \quad D_{\text{nat}}(\{n \in \mathbb{N} : p^+(n) \leq n^{\frac{1}{s}}\}) = D_{\text{nat}}(\{n \in \mathbb{N} : \frac{\log n}{\log p^+(n)} \geq s\}) = \rho(s), \quad s \geq 1.$$

We will call  $\{n \in \mathbb{N} : p^+(n) \leq n^{\frac{1}{s}}\}$  the set of  $n^{\frac{1}{s}}$ -smooth numbers.

The standard number-theoretic proof of Dickman's result is via induction. It can be checked that this inductive proof also works to obtain a corresponding result for  $k$ -free integers. Thus,

$$(1.13) \quad D_{\text{nat}}(\{n \in \mathbb{N} : p^+(n) \leq n^{\frac{1}{s}}\} | S_k) = D_{\text{nat}}(\{n \in \mathbb{N} : \frac{\log n}{\log p^+(n)} \geq s\} | S_k) = \rho(s),$$

for  $s \geq 1$  and  $k \geq 2$ .

**Remark.** Equivalent to (1.12) is the statement that  $\frac{\log p^+(n)}{\log n}$  on  $[N]$  with the uniform distribution converges weakly in distribution as  $N \rightarrow \infty$  to the distribution whose distribution function is  $\rho(\frac{1}{s})$ ,  $s \in [0, 1]$ . The corresponding density function is then  $\frac{-\rho'(\frac{1}{s})}{s^2} = \frac{1}{s}\rho(\frac{1}{s} - 1)$ . In the spirit of (1.11), it has been shown that if  $\hat{D}$  denotes a random variable with the above distribution, then

$$\hat{D} \stackrel{\text{dist}}{=} \max(1 - U, \hat{D}u), \quad U \stackrel{\text{dist}}{=} \text{Unif}([0, 1]), \quad U \text{ and } \hat{D} \text{ independent.}$$

In light of the comparison between (1.11) and the above equation, the distribution has been dubbed the *max-Dickman distribution* [20]. This distribution is the first coordinate of the Poisson-Dirichlet distribution on the infinite simplex  $\{x = (x_1, x_2, \dots) : x_i \geq 0, \sum_{i=1}^{\infty} x_i = 1\}$ . The Poisson-Dirichlet distribution can be defined as the decreasing order statistics of the GEM distribution, where the GEM distribution is the "stick-breaking" distribution: let  $\{U_n\}_{n=1}^{\infty}$  be IID uniform variables on  $[0, 1]$ ; let  $Y_1 = U_1$ , and let  $Y_n = U_n \prod_{r=1}^{n-1} (1 - U_r)$ ,  $n \geq 2$ ; then  $(Y_1, Y_2, \dots)$  has the GEM distribution. The  $n$ -dimensional density function for the distribution of the first  $n$  coordinates of the Poisson-Dirichlet distribution is given by

$$f^{(n)}(s_1, s_2, \dots, s_n) = \frac{1}{s_1 \cdots s_n} \rho\left(\frac{1 - s_1 - \cdots - s_n}{s_n}\right),$$

for  $0 < s_n < \cdots < s_1 < 1$  and  $\sum_{j=1}^n s_j < 1$ .

Let  $p_j^+(n)$  denote the  $j$ th largest distinct prime divisor of  $n$ , with  $p_j^+(n) = 1$  if  $n$  has fewer than  $j$  distinct prime divisors. In 1972 Billingsley [2] gave a probabilistic proof of the fact that  $\frac{1}{\log n}(\log p_1^+(n), \log p_2^+(n), \dots)$  on  $[N]$  with the uniform distribution converges weakly in distribution as  $N \rightarrow \infty$  to the Poisson-Dirichlet distribution. However, he did not identify it as such as the theory of the Poisson-Dirichlet distribution had not yet been developed. (We note that the random vector consisting of the lengths of the cycles of a uniformly random permutation of  $[N]$ , arranged in decreasing order, when normalized by dividing their lengths by  $N$ , also converges as  $N \rightarrow \infty$  to the Poisson-Dirichlet distribution [1].)

Proposition 3 shows that  $D_{\text{nat}}$  and  $D_{\text{log-indep}}$  coincide on a certain natural algebra of sets. We will prove that they disagree on the sets appearing in (1.12) or (1.13); namely on the sets of  $n^{\frac{1}{s}}$ -smooth numbers,  $s > 1$ , and on the intersection of such a set with the set of  $k$ -free numbers,  $S_k$ ,  $k \geq 2$ .

**Theorem 2.** *Under both  $P_N$  and  $P_N^{(k)}$  the random variable  $\frac{\log n}{\log p^+(n)}$  converges weakly as  $N \rightarrow \infty$  to  $D + 1$ , where  $D$  has the Dickman distribution; that is,*

$$(1.14) \quad D_{\text{log-indep}}(\{n \in \mathbb{N} : p^+(n) \leq n^{\frac{1}{s}}\}) = D_{\text{log-indep}}(\{n \in \mathbb{N} : \frac{\log n}{\log p^+(n)} \geq s\}) = e^{-\gamma} \int_{s-1}^{\infty} \rho(x) dx, \quad s \geq 1;$$

$$D_{\text{log-indep}}(\{n \in \mathbb{N} : p^+(n) \leq n^{\frac{1}{s}}\} | S_k) = D_{\text{log-indep}}(\{n \in \mathbb{N} : \frac{\log n}{\log p^+(n)} \geq s\} | S_k) = e^{-\gamma} \int_{s-1}^{\infty} \rho(x) dx, \quad s \geq 1, k \geq 2.$$

**Remark.** Recalling that whenever the natural density exists, the logarithmic one does too and they are equal, it follows from (1.12) that  $D_{\text{log}}(\{n \in \mathbb{N} : p^+(n) \leq n^{\frac{1}{s}}\}) = \rho(s)$ . Since, as we've noted, the weights used in calculating the densities  $D_{\text{log}}$  and  $D_{\text{log-indep}}$  have the same profile, but the sequences of subsets of  $\mathbb{N}$  over which the limits are taken, namely  $\{[N]\}_{N=1}^{\infty}$  and  $\{\Omega_N\}_{N=1}^{\infty}$ , are different, and since the integers in  $[N]$  and in  $\Omega_N$  are constructed from the same set  $\{p_j\}_{j=1}^N$  of primes, and  $[N] \subset \Omega_N$ , intuition

suggests that

$$(1.15) \quad \rho(s) \leq e^{-\gamma} \int_{s-1}^{\infty} \rho(x) dx, \quad s \geq 1;$$

that is, that under  $D_{\log\text{-indep}}$ ,  $n^{\frac{1}{s}}$ -smooth numbers are more likely than under  $D_{\text{nat}}$ . And indeed this is the case. Letting  $H(s) = e^{-\gamma} \int_{s-1}^{\infty} \rho(x) dx - \rho(s)$ , we have  $H(1) = H(\infty) = 0$ . Differentiating  $H$ , and using the differential-delay equation satisfied by  $\rho$ , one has  $H'(s) = -e^{-\gamma} \rho(s-1) - \rho'(s) = \rho(s-1)(\frac{1}{s} - e^{-\gamma})$ . Thus,  $H'(s)$  vanishes only at  $s = e^{\gamma}$ . Differentiating again and again using the differential-delay equation, one finds that  $H''(e^{\gamma}) < 0$ ; thus,  $H(s) \geq 0$ , for  $s \geq 1$ , proving (1.15).

We now consider integers all of whose prime divisors are “large.” Let  $\Phi(x, y)$  denote the number of positive integers less than or equal to  $x$  all of whose prime divisors are greater than or equal to  $y$ . Numbers with no prime divisors less than  $y$  are called  $y$ -rough numbers. The Buchstab function  $\omega(s)$ , defined for  $s \geq 1$ , is the unique continuous function satisfying

$$\omega(s) = \frac{1}{s}, \quad 1 \leq s \leq 2,$$

and satisfying the differential-delay equation

$$(s\omega(s))' = \omega(s-1), \quad s > 2.$$

In 1937, Buchstab proved [5] that for  $s > 1$ ,  $\Phi(N, N^{\frac{1}{s}}) \sim \frac{Ns\omega(s)}{\log N}$  as  $N \rightarrow \infty$ ; whence the name of the function. See also [18] or [23]. Let  $p^-(n)$  denote the smallest prime divisor of  $n$ . Then Buchstab’s result states that

$$(1.16) \quad \frac{|\{n \in [N] : p^-(n) \geq N^{\frac{1}{s}}\}|}{N} = \frac{|\{n \in [N] : \frac{\log N}{\log p^-(n)} \leq s\}|}{N} \sim \frac{s\omega(s)}{\log N},$$

for  $s > 1$ , as  $N \rightarrow \infty$ .

Since  $\frac{|\{n \in [N] : \frac{\log N}{\log n} > 1 + \epsilon\}|}{N} = N^{-\epsilon}$ , it follows that (1.16) is equivalent to

$$(1.17) \quad \frac{|\{n \in [N] : p^-(n) \geq n^{\frac{1}{s}}\}|}{N} = \frac{|\{n \in [N] : \frac{\log n}{\log p^-(n)} \leq s\}|}{N} \sim \frac{s\omega(s)}{\log N},$$

for  $s > 1$ , as  $N \rightarrow \infty$ .

One has  $\lim_{s \rightarrow \infty} \omega(s) = e^{-\gamma}$ , and the rate of convergence is super-exponential [23]. We will call  $\{n \in [N] : p^-(n) \geq n^{\frac{1}{s}}\}$  the set of  $n^{\frac{1}{s}}$ -rough numbers. (We

note that the probability that the shortest cycle of a uniformly random permutation of  $[N]$  is larger or equal to  $\frac{N}{s}$  decays asymptotically as  $\frac{s\omega(s)}{N}$ .)

Note that (1.17) also holds for  $s = 1$ , since in this case (1.17) reduces to  $\frac{\Pi(N)}{N} \sim \frac{1}{\log N}$ ; that is, it reduces to the PNT. Buchstab assumed the PNT in proving (1.16).

What is the asymptotic probability of a prime number under the sequence of measures used to construct the logarithmic density  $D_{\log}$  and under the sequence  $\{P_N\}_{N=1}^{\infty}$  used to construct the density  $D_{\log\text{-indep}}$ ? Mertens' second theorem states that

$$(1.18) \quad \sum_{p \leq N} \frac{1}{p} = \log \log N + M_0 + O\left(\frac{1}{\log N}\right),$$

where the summation is over primes  $p$ , and where  $M_0$  is called the Meissel-Mertens constant [19]. By the PNT,  $p_N \sim N \log N$ , thus by Mertens' second theorem,

$$(1.19) \quad \sum_{j=1}^N \frac{1}{p_j} \sim \log \log(N \log N) \sim \log \log N.$$

From (1.18) we conclude that for the sequence of measures used to construct the logarithmic density  $D_{\log}$ , the probability of a prime is

$$(1.20) \quad \frac{1}{\log N} \sum_{p \leq N} \frac{1}{p} \sim \frac{\log \log N}{\log N}.$$

Since

$$P_N(\{n \in \Omega_N : n \text{ is prime}\}) = \frac{1}{C_N} \sum_{j=1}^N \frac{1}{p_j},$$

from (1.19) and Mertens formula given in (1.8), we conclude that for the sequence  $\{P_N\}_{N=1}^{\infty}$  use to construct the density  $D_{\log\text{-indep}}$ , the probability of a prime satisfies

$$(1.21) \quad P_N(\{n \in \Omega_N : n \text{ is prime}\}) \sim \frac{e^{-\gamma} \log \log N}{\log N}.$$

From (1.20) and (1.21) it is clear that (1.17) cannot hold when the sequence of uniform measures on  $[N]$ ,  $N = 1, 2, \dots$ , appearing on the left hand side there is replaced either by the sequence of measures used to calculate the logarithmic density  $D_{\log}$  or by the sequence  $\{P_N\}_{N=1}^{\infty}$  used to calculate

the density  $D_{\log\text{-indep}}$ . However, letting

$$a_s(n) = \begin{cases} 1, & p^-(n) \geq n^{\frac{1}{s}}, \\ 0, & \text{otherwise,} \end{cases},$$

and  $A_s(t) = \sum_{j=1}^{[t]} a_s(j)$ ,  $t \geq 1$ , a summation by parts gives

$$(1.22) \quad \sum_{n \leq N: p^-(n) \geq n^{\frac{1}{s}}} \frac{1}{n} = \sum_{n=1}^N \frac{a_s(n)}{n} = \frac{A_s(N)}{N} + \int_1^N \frac{A_s(t)}{t^2} dt.$$

By (1.17),  $\frac{A_s(t)}{t} \sim \frac{s\omega(s)}{\log t}$  as  $t \rightarrow \infty$ ; thus from (1.22) we have

$$\frac{1}{\log N} \sum_{n \leq N: p^-(n) \geq n^{\frac{1}{s}}} \frac{1}{n} \sim \log \log N \frac{s\omega(s)}{\log N}.$$

That is, modulo the change necessitated by comparing (1.20) to the PNT, Buchstab's result on  $n^{\frac{1}{s}}$ -rough numbers for the uniform measure in (1.17) carries over to the measures used in the construction of the logarithmic density.

Modulo the change necessitated by comparing (1.21) to the PNT, does Buchstab's result on  $n^{\frac{1}{s}}$ -rough numbers also carry over to the measures  $\{P_N\}_{N=1}^{\infty}$  used in the construction of the density  $D_{\log\text{-indep}}$ ? Since (1.12) and (1.14) show that the *positive* densities with respect  $D_{\text{nat}}$  and  $D_{\log\text{-indep}}$  of the  $n^{\frac{1}{s}}$ -smooth sets  $\{n \in \mathbb{N} : p^+(n) \leq n^{\frac{1}{s}}\}$  do not coincide, it is interesting to discover that the answer is indeed affirmative.

**Theorem 3.** For  $s \geq 1$ ,

$$(1.23) \quad P_N(\{n \in [N] : p^-(n) \geq n^{\frac{1}{s}}\}) = P_N\left(\frac{\log n}{\log p^-(n)} \leq s\right) \sim (e^{-\gamma} \log \log N) \frac{s\omega(s)}{\log N},$$

as  $N \rightarrow \infty$ .

Recalling the definition of the Buchstab function, note that  $V(s) \equiv s\omega(s)$  is the unique continuous function satisfying  $V(s) = 1$ ,  $1 \leq s \leq 2$ , and  $V'(s) = \frac{V(s-1)}{s-1}$ , for  $s > 2$ . In the proof of Theorem 3, we actually show that (1.23) holds with  $s\omega(s)$  on the right hand side replaced by

$$v(s) \equiv \sum_{L=1}^{[s]} \Lambda_L(s),$$

where

$$(1.24) \quad \begin{aligned} \Lambda_1(s) &= 1, \quad s \geq 1; \\ \Lambda_2(s) &= \int_1^{s-1} \frac{du_1}{u_1} = \log(s-1), \quad s \geq 2; \\ \Lambda_L(s) &= \int_{L-1}^{s-1} \int_{L-2}^{u_{L-1}-1} \cdots \int_1^{u_2-1} \prod_{j=1}^{L-1} \frac{du_j}{u_j}, \quad s \geq L \geq 3. \end{aligned}$$

Now  $\Lambda'_L(s) = \frac{1}{s-1} \Lambda_{L-1}(s-1)$ , for  $s \geq L \geq 2$ , while of course  $\Lambda'_1(s) = 0$ . Thus,  $v(s) = 1$ , for  $1 \leq s \leq 2$  and  $v'(s) = \frac{v(s-1)}{s-1}$ , for  $s > 2$ . This proves the following result.

**Proposition 4.**

$$(1.25) \quad s\omega(s) = 1 + \log(s-1) + \sum_{L=3}^{[s]} \int_{L-1}^{s-1} \int_{L-2}^{u_{L-1}-1} \cdots \int_1^{u_2-1} \prod_{j=1}^{L-1} \frac{du_j}{u_j}, \quad s \geq 3.$$

The representation of the Buchstab function  $\omega$  in (1.25) seems to be new. It is simpler than the following known representation [1, 15]:

$$s\omega(s) = 1 + \sum_{L=2}^{[s]} \frac{1}{L!} \int_{\substack{\frac{1}{s} \leq y_j \leq 1 \\ \frac{1}{s} \leq 1 - (y_1 + y_2 + \cdots + y_{L-1}) \leq 1}} \frac{1}{1 - (y_1 + y_2 + \cdots + y_{L-1})} \prod_{j=1}^{L-1} \frac{dy_j}{y_j}.$$

Since  $\lim_{s \rightarrow \infty} \omega(s) = e^{-\gamma}$ , we also obtain what seems to be yet another representation of Euler's constant:

$$e^{-\gamma} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{L=3}^N \int_{L-1}^N \int_{L-2}^{u_{L-1}-1} \cdots \int_1^{u_2-1} \prod_{j=1}^{L-1} \frac{du_j}{u_j}.$$

We prove Proposition 3 and Theorems 1-3 successively in sections 2-5 below.

## 2. PROOF OF PROPOSITION 3

For the proof of the proposition we need the following result which is obviously known; however, as we were unable to find it in a number theory text, we supply a proof in the appendix.

**Proposition 5.** For  $1 \leq l < k$ ,

$$(2.1) \quad D_{nat}(\beta_{p_j} \geq l | S_k) \equiv \frac{D_{nat}(\{\beta_{p_j} \geq l\} \cap S_k)}{D_{nat}(S_k)} = \frac{(\frac{1}{p_j})^l - (\frac{1}{p_j})^k}{1 - (\frac{1}{p_j})^k}.$$

**Remark.** When  $k = 2$  and  $l = 1$ , (2.1) becomes

$D_{\text{nat}}(\beta_{p_j} \geq 1 | S_2) = \frac{1}{1+p_j}$ . That is, among square-free numbers, the natural density of those divisible by the prime  $p_j$  is  $\frac{1}{p_j+1}$ .

*Proof of Proposition 3.* In light of Proposition 1, it follows immediately that for  $l \leq N$ , the random vector  $\{\beta_{p_j}\}_{j=1}^l$  under  $P_N$  has the distribution of  $\{X_{p_j}\}_{j=1}^l$  under  $P$ , this latter distribution being the weak limit as  $N \rightarrow \infty$  of the distribution of  $\{\beta_{p_j}\}_{j=1}^l$  on  $[N]$  with the uniform distribution. From this it follows that  $D_{\log\text{-indep}}$  and  $D_{\text{nat}}$  coincide on the algebra of sets generated by the inverse images of the  $\{\beta_{p_j}\}_{j=1}^\infty$ .

It is well-known that  $D_{\text{nat}}(S_k) = \frac{1}{\zeta(k)}$ , where  $\zeta(s) = \sum_{n=1}^\infty \frac{1}{n^s}$  is the Riemann zeta function [22]. On the other hand, by Proposition 1 we have

$$P_N(S_k) = P(X_{p_j} < k, j \in [N]) = \prod_{j=1}^N P(X_j < k) = \prod_{j=1}^N \left(1 - \frac{1}{p_j^k}\right),$$

and so by the Euler product formula we conclude that

$$D_{\log\text{-indep}}(S_k) = \lim_{N \rightarrow \infty} P_N(S_k) = \lim_{N \rightarrow \infty} \prod_{j=1}^N \left(1 - \frac{1}{p_j^k}\right) = \frac{1}{\zeta(k)}.$$

Thus, the two densities coincide on the algebra generated by  $\{S_k\}_{k=2}^\infty$ .

Also, for  $j \leq N$ ,  $k \geq 2$  and  $l < k$ , we have

$$\begin{aligned} P_N^{(k)}(\beta_{p_j} \geq l) &= P_N(\beta_{p_j} \geq l | S_k) = P(X_{p_j} \geq l | X_{p_i} < k, i = 1, \dots, N) = \\ P(X_{p_j} \geq l | X_{p_j} < k) &= \frac{\sum_{i=l}^{k-1} \left(\frac{1}{p_j}\right)^i \left(1 - \frac{1}{p_j}\right)}{\sum_{i=0}^{k-1} \left(\frac{1}{p_j}\right)^i \left(1 - \frac{1}{p_j}\right)} = \frac{\left(\frac{1}{p_j}\right)^l - \left(\frac{1}{p_j}\right)^k}{1 - \left(\frac{1}{p_j}\right)^k}. \end{aligned}$$

Thus,  $D_{\log\text{-indep}}(\beta_{p_j} \geq l | S_k) \equiv \frac{D_{\log\text{-indep}}(\{\beta_{p_j} \geq l\} \cap S_k)}{D_{\log\text{-indep}}(S_k)} = \frac{\left(\frac{1}{p_j}\right)^l - \left(\frac{1}{p_j}\right)^k}{1 - \left(\frac{1}{p_j}\right)^k}$ . Recalling

Proposition 5, we conclude that the two densities indeed coincide on the algebra generated by the inverse images of  $\{\beta_{p_j}\}_{j=1}^\infty$  and the sets  $\{S_k\}_{k=2}^\infty$ .

□

### 3. PROOF OF THEOREM 1

We first prove the theorem for  $P_N$ . Let  $E_N$  denote the expectation with respect to  $P_N$ . Using Proposition 1, we have

$$E_N \frac{\log n}{\log N} = \frac{1}{\log N} \sum_{j=1}^N E X_{p_j} \log p_j = \frac{1}{\log N} \sum_{j=1}^N \frac{\log p_j}{p_j - 1}.$$



Mertens' first theorem [19] states that  $\sum_{p \leq N} \frac{\log p}{p} \sim \log N$ , where the sum is over all primes less than or equal to  $N$ . Thus, using nothing more than the trivial bound  $p_N \leq N^k$ , for some  $k$ , it follows that  $\{E_N \frac{\log n}{\log N}\}_{N=1}^{\infty}$  is bounded, and therefore that the distributions of the nonnegative random variables  $\{\frac{\log n}{\log N}\}_{N=1}^{\infty}$  under  $\{P_N\}_{N=1}^{\infty}$  are tight. In the next paragraph we will prove that their Laplace transforms converge to  $\exp(-\int_0^1 \frac{1-e^{-tx}}{x} dx)$ . This proves that the distributions converge weakly. By the argument in the paragraph containing (1.9), it then follows that the limiting distribution is the Dickman distribution. Alternatively, the above function is known to be the Laplace transform of the Dickman distribution [18, 23].

By Proposition 1, we have for  $t \geq 0$ ,

$$(3.1) \quad E_N \exp(-t \frac{\log n}{\log N}) = E \exp(-\frac{t}{\log N} \sum_{j=1}^N X_{p_j} \log p_j) = \prod_{j=1}^N E \exp(-\frac{t \log p_j}{\log N} X_{p_j}).$$

For  $s \geq 0$ ,

$$(3.2) \quad E \exp(-s X_{p_j}) = \sum_{k=0}^{\infty} e^{-sk} \left(\frac{1}{p_j}\right)^k \left(1 - \frac{1}{p_j}\right) = \left(1 - \frac{1}{p_j}\right) \frac{1}{1 - \frac{e^{-s}}{p_j}} = \frac{1}{1 + \frac{1-e^{-s}}{p_j-1}}.$$

From (3.1) and (3.2) we have

$$(3.3) \quad \log E_N \exp(-t \frac{\log n}{\log N}) = - \sum_{j=1}^N \log \left(1 + \frac{1 - \exp(-t \frac{\log p_j}{\log N})}{p_j - 1}\right).$$

Now  $x - \frac{x^2}{2} \leq \log(1+x) \leq x$ , for  $x \geq 0$ , and by the bounded convergence theorem,  $\lim_{N \rightarrow \infty} \sum_{j=1}^N \left(\frac{1 - \exp(-t \frac{\log p_j}{\log N})}{p_j - 1}\right)^2 = 0$ ; thus,

$$(3.4) \quad \lim_{N \rightarrow \infty} \log E_N \exp(-t \frac{\log n}{\log N}) = - \lim_{N \rightarrow \infty} \sum_{j=1}^N \frac{1 - \exp(-t \frac{\log p_j}{\log N})}{p_j - 1}.$$

Let  $x_j^{(N)} = \frac{\log p_j}{\log N}$  and  $\Delta_j^{(N)} = x_{j+1}^{(N)} - x_j^{(N)}$ . By the PNT,  $p_j \sim j \log j$ , as  $j \rightarrow \infty$ ; thus

$$(3.5) \quad \log p_{j+1} - \log p_j \sim \log \frac{(j+1) \log(j+1)}{j \log j} = \log \left( \left(1 + \frac{1}{j}\right) \left(1 + \frac{\log(1 + \frac{1}{j})}{\log j}\right) \right) \sim \frac{1}{j} \sim \frac{\log p_j}{p_j}.$$

Consequently,

$$(3.6) \quad \Delta_j^{(N)} \sim \frac{\log p_j}{p_j \log N}, \text{ uniformly as } j, N \rightarrow \infty.$$

Note also that

$$(3.7) \quad \lim_{N \rightarrow \infty} x_1^{(N)} = 0, \quad \lim_{N \rightarrow \infty} x_N^{(N)} = 1.$$

We rewrite the summand on the right hand side of (3.4) as

$$(3.8) \quad \begin{aligned} \sum_{j=1}^N \frac{1 - \exp(-t \frac{\log p_j}{\log N})}{p_j - 1} &= \sum_{j=1}^N \frac{1 - \exp(-t \frac{\log p_j}{\log N})}{\frac{\log p_j}{\log N}} \frac{\log p_j}{(p_j - 1) \log N} = \\ &= \sum_{j=1}^N \frac{1 - \exp(-t x_j^{(N)})}{x_j^{(N)}} \frac{\log p_j}{(p_j - 1) \log N}. \end{aligned}$$

From (3.6)-(3.8) along with (3.4) we conclude that

$$(3.9) \quad \lim_{N \rightarrow \infty} E_N \exp(-t \frac{\log n}{\log N}) = \exp(- \int_0^1 \frac{1 - e^{-tx}}{x} dx).$$

This completes the proof of the theorem for  $P_N$ .

We now turn to  $P_N^{(k)}$ . Let  $E_N^{(k)}$  denote the expectation with respect to  $P_N^{(k)}$ . By Proposition 2,

$$(3.10) \quad E_N^{(k)} \exp(-t \frac{\log n}{\log N}) = E \exp(- \frac{t}{\log N} \sum_{j=1}^N X_{p_j}^{(k)} \log p_j) = \prod_{j=1}^N E \exp(- \frac{t \log p_j}{\log N} X_{p_j}^{(k)}).$$

For  $s \geq 0$ ,

$$(3.11) \quad E \exp(-s X_{p_j}^{(k)}) = \sum_{l=0}^{k-1} e^{-sl} \left(\frac{1}{p_j}\right)^l \frac{1 - \frac{1}{p_j}}{1 - (\frac{1}{p_j})^k} = \frac{1 - \frac{1}{p_j}}{1 - (\frac{1}{p_j})^k} \frac{1 - (\frac{e^{-s}}{p_j})^k}{1 - \frac{e^{-s}}{p_j}}.$$

Comparing the equality between the first and third expressions in (3.2) with (3.11), we have

$$(3.12) \quad E \exp(-s X_{p_j}^{(k)}) = \frac{1 - (\frac{e^{-s}}{p_j})^k}{1 - (\frac{1}{p_j})^k} E \exp(-s X_{p_j}) = \left(1 + \frac{(\frac{1}{p_j})^k (1 - e^{-sk})}{1 - (\frac{1}{p_j})^k}\right) E \exp(-s X_{p_j}).$$

Thus, from (3.1), (3.10) and (3.12) we have

$$(3.13) \quad E_N^{(k)} \exp(-t \frac{\log n}{\log N}) = E_N \exp(-t \frac{\log n}{\log N}) \prod_{j=1}^N \left(1 + \frac{(\frac{1}{p_j})^k (1 - \exp(-\frac{kt \log p_j}{\log N}))}{1 - (\frac{1}{p_j})^k}\right).$$

By the bounded convergence theorem,

$$(3.14) \quad \lim_{N \rightarrow \infty} \sum_{j=1}^N \frac{\left(\frac{1}{p_j}\right)^k (1 - \exp(-\frac{kt \log p_j}{\log N}))}{1 - \left(\frac{1}{p_j}\right)^k} = 0.$$

Thus, from (3.9), (3.13) and (3.14), we conclude that

$$\lim_{N \rightarrow \infty} E_N^{(k)} \exp(-t \frac{\log n}{\log N}) = \exp(-\int_0^1 \frac{1 - e^{-tx}}{x} dx).$$

□

#### 4. PROOF OF THEOREM 2

We prove the theorem for  $P_N$ ; the proof for  $P_N^{(k)}$  is done analogously. For definiteness and convenience, we define  $\frac{\log n}{\log p^+(n)}|_{n=1} = 0$ . Let

$$J_N^+ = \max\{j \in [N] : X_{p_j} \neq 0\},$$

with  $\max \emptyset$  defined to be 0. By Proposition 1,  $\frac{\log n}{\log p^+(n)}$  under  $P_N$  is equal in distribution to  $\frac{1_{\{J_N^+ \neq 0\}}}{\log p_{J_N^+}} \sum_{j=1}^N X_{p_j} \log p_j$ . On  $\{J_N^+ \neq 0\}$ , we write

$$\frac{1}{\log p_{J_N^+}} \sum_{j=1}^N X_{p_j} \log p_j = \frac{1}{\log p_{J_N^+}} \sum_{j=1}^{J_N^+-1} X_{p_j} \log p_j + X_{p_{J_N^+}}.$$

As noted in the paragraph containing (1.9),  $J_N^+ \rightarrow \infty$  a.s. as  $N \rightarrow \infty$ . Also, by the independence of  $\{X_{p_j}\}_{j=1}^\infty$ , we have  $\sum_{j=1}^{J_N^+-1} X_{p_j} \log p_j | \{J_N^+ = j_0\} \stackrel{\text{dist}}{=} \sum_{j=1}^{j_0-1} X_{p_j} \log p_j$ . Thus, it follows from Theorem 1 that  $\frac{1}{\log J_N^+} \sum_{j=1}^{J_N^+-1} X_{p_j} \log p_j$  converges weakly to the Dickman distribution. By the PNT,  $p_{J_N^+} \sim J_N^+ \log J_N^+$ ; thus also  $\frac{1}{\log p_{J_N^+}} \sum_{j=1}^{J_N^+-1} X_{p_j} \log p_j$  a.s. converges weakly to the Dickman distribution.

Note that  $X_{p_{J_N^+}}$  conditioned on  $\{J_N^+ = j_0\}$  is distributed as  $X_{p_{j_0}}$  conditioned on  $\{X_{p_{j_0}} \geq 1\}$ . A trivial calculation shows that  $X_{p_{j_0}}$  conditioned on  $\{X_{p_{j_0}} \geq 1\}$  converges weakly to 1 as  $j_0 \rightarrow \infty$ ; thus,  $X_{p_{J_N^+}}$  converges weakly to 1. Consequently,  $\frac{\log n}{\log p^+(n)}$  under  $P_N$  converges weakly to  $D+1$  as  $N \rightarrow \infty$ .

□

## 5. PROOF OF THEOREM 3

As noted after the statement of the theorem, we will prove (1.23) with  $s\omega(s)$  replaced by  $\sum_{L=1}^{\lfloor s \rfloor} \Lambda_L(s)$ , where  $\Lambda_L$  is as in (1.24). That is, we will prove that

$$(5.1) \quad P_N\left(\frac{\log n}{\log p^+(n)} \leq s\right) \sim (e^{-\gamma} \log \log N) \frac{\sum_{L=1}^{\lfloor s \rfloor} \Lambda_L(s)}{\log N}, \quad s \geq 1.$$

We will first prove (5.1) for  $s \in [1, 2]$ , then for  $s \in [2, 3]$ , and then for  $s \in [3, 4]$ . After treating these three particular cases, an inductive argument for the general case of  $s \in [L, L+1]$  will be explained succinctly.

For definiteness and convenience, we define  $\frac{\log n}{\log p^-(n)}|_{n=1} = 0$ . Of course,  $\frac{\log n}{\log p^-(n)} \geq 1$ , for  $n \geq 2$ . Let

$$J_N^- = \min\{j \in [N] : X_{p_j} \neq 0\},$$

with  $\min \emptyset$  defined to be 0. Note that by (1.8),

$$(5.2) \quad P\left(\frac{\log n}{\log p^-(n)} < 1\right) = P(J_N^- = 0) = C_N^{-1} \sim \frac{e^{-\gamma}}{\log N}.$$

By Proposition 1,  $\frac{\log n}{\log p^-(n)}$  under  $P_N$  is equal in distribution to  $\frac{1_{\{J_N^- \neq 0\}}}{\log p_{J_N^-}} \sum_{j=1}^N X_{p_j} \log p_j$ .

Thus, we have

$$(5.3) \quad P_N\left(L \leq \frac{\log n}{\log p^-(n)} \leq s\right) = \sum_{a=1}^N P\left(L \log p_a \leq \sum_{j=a}^N X_{p_j} \log p_j \leq s \log p_a \mid J_N^- = a\right) P(J_N^- = a), \quad \text{for } L \in \mathbb{N},$$

and

$$(5.4) \quad P(J_N^- = a) = \frac{1}{p_a} \prod_{j=1}^{a-1} \left(1 - \frac{1}{p_j}\right).$$

Under the conditioning  $\{J_N^- = a\}$ , the random variables  $\{X_{p_j}\}_{j=a}^N$  are still independent, and for  $j > a$ ,  $X_{p_j}$  is distributed as before, namely according to  $\text{Geom}(1 - \frac{1}{p_j})$ ; however  $X_{p_a}$  is now distributed as a  $\text{Geom}(1 - \frac{1}{p_a})$  random variable conditioned to be positive.

Consider first  $L = 1$  and  $s \in [1, 2]$ . For  $s \neq 2$ , the inequality  $\log p_a \leq \sum_{j=a}^N X_{p_j} \log p_j \leq s \log p_a$  in (5.3) under the conditional probability  $P(\cdot \mid J_N^- = a)$  will hold if and only if  $X_{p_a} = 1$  and  $X_{p_j} = 0$ , for  $a+1 \leq j \leq N$ . For

$s = 2$  it will hold if and only if  $X_{p_a}$  is equal to either 1 or 2 and  $X_{p_j} = 0$ , for  $a + 1 \leq j \leq N$ . Thus, we have

$$(5.5) \quad P\left(\log p_a \leq \sum_{j=a}^N X_{p_j} \log p_j \leq s \log p_a \mid J_N^- = a\right) = \begin{cases} \prod_{j=a}^N \left(1 - \frac{1}{p_j}\right), & s \in [1, 2); \\ \prod_{j=a}^N \left(1 - \frac{1}{p_j}\right) + \frac{1}{p_a} \prod_{j=a}^N \left(1 - \frac{1}{p_j}\right), & s = 2. \end{cases}$$

From (5.2)-(5.5), along with (1.8) and (1.19) and the fact that  $\Lambda_1(s) \equiv 1$  for  $s \in [1, 2]$ , we obtain

$$(5.6) \quad P_N\left(\frac{\log n}{\log p^-(n)} \leq s\right) \sim C_N^{-1} \sum_{a=1}^N \frac{1}{p_a} \sim (e^{-\gamma} \log \log N) \frac{\Lambda_1(s)}{\log N}, \quad s \in [1, 2].$$

Now consider  $L = 2$  and  $s \in [2, 3]$ . Let

$$J_{a,1}(s) = \max\{j : p_j \leq p_a^{s-1}\}.$$

(Note that  $J_{a,1}(s) \geq a$ , for  $s \geq 2$ .) Then for  $s \in [2, 3)$ , the inequality  $2 \log p_a \leq \sum_{j=a}^N X_{p_j} \log p_j \leq s \log p_a$  in (5.3) under the conditional probability  $P(\cdot \mid J_N^- = a)$  will hold if and only if either  $X_{p_a} = 2$  and  $X_{p_j} = 0$  for  $a + 1 \leq j \leq N$ , or  $X_{p_a} = 1$ ,  $X_{p_j} = 1$  for exactly one  $j$  satisfying  $a + 1 \leq j \leq J_{a,1}(s) \wedge N$ , and  $X_{p_j} = 0$  for all other  $j$  satisfying  $a + 1 \leq j \leq N$ . Thus, we have

$$(5.7) \quad P\left(2 \log p_a \leq \sum_{j=a}^N X_{p_j} \log p_j \leq s \log p_a \mid J_N^- = a\right) = \frac{1}{p_a} \prod_{j=a}^N \left(1 - \frac{1}{p_a}\right) + \sum_{l=a+1}^{J_{a,1}(s) \wedge N} \frac{1}{p_l} \prod_{j=a}^N \left(1 - \frac{1}{p_j}\right), \quad s \in [2, 3),$$

where, of course, the sum on the right hand side above is interpreted as 0 if  $J_{a,1}(s) = a$ . For the case  $s = 3$ , there is also the possibility of  $X_{p_a} = 3$  and  $X_{p_j} = 0$  for  $a + 1 \leq j \leq N$ . The  $P(\cdot \mid J_N^- = a)$ -probability of this is  $\frac{1}{p_a^2} \prod_{j=a}^N \left(1 - \frac{1}{p_a}\right)$ . Thus, with  $s = 3$ , (5.7) has the additional term  $\frac{1}{p_a^2} \prod_{j=a}^N \left(1 - \frac{1}{p_a}\right)$  on the right hand side. However, this term does not contribute to the leading order asymptotics. From (5.3), (5.4) and (5.7), we obtain

$$(5.8) \quad P_N\left(2 \leq \frac{\log n}{\log p^-(n)} \leq s\right) = C_N^{-1} \left( \sum_{a=1}^N \frac{1}{p_a^2} + \sum_{a=1}^N \frac{1}{p_a} \sum_{l=a+1}^{J_{a,1}(s) \wedge N} \frac{1}{p_l} \right), \quad s \in [2, 3).$$

Since  $p_a \sim a \log a$  as  $a \rightarrow \infty$ , it follows that

$$(5.9) \quad J_{a,1}(s) \log J_{a,1}(s) \sim (a \log a)^{s-1}, \text{ as } a \rightarrow \infty.$$

Taking the logarithm of each side in (5.9), we obtain

$$(5.10) \quad \lim_{a \rightarrow \infty} \frac{\log J_{a,1}(s)}{\log a} = s - 1.$$

Using Mertens' second theorem in the form (1.18) along with the fact that  $p_j \sim j \log j$ , we have

$$(5.11) \quad \sum_{l=a+1}^{J_{a,1}(s)} \frac{1}{p_l} \sim \log \log (J_{a,1}(s) \log J_{a,1}(s)) - \log \log (a \log a) \sim \log \frac{\log J_{a,1}(s)}{\log a}, \text{ as } a \rightarrow \infty,$$

and thus, by (5.10),

$$(5.12) \quad \lim_{a \rightarrow \infty} \sum_{l=a+1}^{J_{a,1}(s)} \frac{1}{p_l} = \log(s - 1).$$

Now choose any  $b \in (0, \frac{1}{s})$ . Then  $(N^b \log N^b)^s < N$  for all large  $N$ . By (5.9),

$$(5.13) \quad J_{a,1}(s) \leq N, \text{ for } a \leq N^b \text{ and sufficiently large } N.$$

By Mertens' second theorem in the form (1.19), we have

$$(5.14) \quad \sum_{a=1}^N \frac{1}{p_a} = \sum_{a=1}^{N^b} \frac{1}{p_a} + O(1) \sim \log \log N.$$

From (5.12)-(5.14), we obtain

$$(5.15) \quad \sum_{a=1}^N \frac{1}{p_a} \sum_{l=a+1}^{J_{a,1}(s) \wedge N} \frac{1}{p_l} \sim \sum_{a=1}^{N^b} \frac{1}{p_a} \sum_{l=a+1}^{J_{a,1}(s)} \frac{1}{p_l} \sim (\log \log N) \log(s - 1).$$

Recalling the asymptotic behavior of  $C_N$ , recalling from (1.24) that  $\Lambda_2(s) = \log(s - 1)$  for  $s \geq 2$ , and using (5.8) and (5.15), we conclude that

$$(5.16) \quad P_N(2 \leq \frac{\log n}{\log p^-(n)} \leq s) \sim (e^{-\gamma} \log \log N) \frac{\Lambda_2(s)}{\log N}, \quad s \in [2, 3],$$

where the inclusion of the right endpoint  $s = 3$  follows from the remarks made after (5.7). From (5.6) with  $s = 2$  and (5.16), along with the fact that  $\Lambda_1(s) \equiv 1$ , we obtain

$$(5.17) \quad P_N(\frac{\log n}{\log p^-(n)} \leq s) \sim (e^{-\gamma} \log \log N) \frac{\Lambda_1(s) + \Lambda_2(s)}{\log N}, \quad s \in [2, 3].$$

Now consider  $L = 3$  and  $s \in [3, 4]$ . In fact we will work with  $s \in [3, 4)$  since the case  $s = 4$  is slightly different but leads to the same asymptotics, similar to the remarks after (5.7). Then the inequality  $3 \log p_a \leq \sum_{j=a}^N X_{p_j} \log p_j \leq s \log p_a$  in (5.3) under the conditional probability  $P(\cdot | J_N^- = a)$  will hold if and only if one of the following four situations obtains:

(5.18)

(1)  $X_{p_a} = 3; X_{p_j} = 0$ , for  $a + 1 \leq j \leq N$ .

(2)  $X_{p_a} = 2; X_{p_j} = 1$  for exactly one  $j$  satisfying  $a + 1 \leq j \leq J_{a,1}(s - 1) \wedge N$ ;  
 $X_{p_j} = 0$  for all other  $j$  satisfying  $a + 1 \leq j \leq N$ .

(3)  $X_{p_a} = 1; X_{p_j} = 1$  for exactly one  $j$  satisfying  $J_{a,1}(3) < j \leq J_{a,1}(s) \wedge N$ ;  
 $X_{p_j} = 0$  for all other  $j$  satisfying  $a + 1 \leq j \leq N$ .

(4)  $X_{p_a} = 1$ ; there exist  $j_1, j_2$ , satisfying  $a + 1 \leq j_1 \leq j_2 \leq N$  and  $p_{j_1} p_{j_2} \leq p_a^{s-1}$ ,  
such that  $X_{j_1} = X_{j_2} = 1$ , if  $j_1 \neq j_2$  and  $X_{j_1} = 2$  if  $j_1 = j_2$ ;  
 $X_{p_j} = 0$  for all other  $j$  satisfying  $a + 1 \leq j \leq N$ .

Because  $\sum_{a=1}^{\infty} \frac{1}{p_a^2} < \infty$ , the probabilities from situations (1) and (2) in (5.18) do not contribute to the leading order asymptotics of  $P_N(3 \leq \frac{\log n}{\log p^-(n)} \leq s)$ , just as in the case  $L = 2$  and  $s \in [2, 3)$ , the probability from the case  $X_{p_a} = 2$  did not contribute to the leading order asymptotics there. (The contribution there from the case  $X_{p_a} = 2$  is the term  $C_N^{-1} \sum_{a=1}^N \frac{1}{p_a^2}$  in (5.8).)

The analysis of the contribution from situation (3) in (5.18) follows the same line of analysis as above when  $L = 2$  and  $s \in [2, 3)$  for the case  $X_{p_a} = 1$ ,  $X_{p_j} = 1$  for exactly one  $j$  satisfying  $a + 1 \leq j \leq J_{a,1}(s) \wedge N$ , and  $X_{p_j} = 0$  for all other  $j$  satisfying  $a + 1 \leq j \leq N$ . The difference is that there one had  $X_{p_j} = 1$  for exactly one  $j$  satisfying  $a + 1 \leq j \leq J_{a,1}(s) \wedge N$ , while here one has  $X_{p_j} = 1$  for exactly one  $j$  satisfying  $J_{a,1}(3) < j \leq J_{a,1}(s) \wedge N$ . Thus, whereas the corresponding contribution there was the term  $\sum_{a=1}^N \frac{1}{p_a} \sum_{l=a+1}^{J_{a,1}(s) \wedge N} \frac{1}{p_l}$  in (5.8), the contribution here will be  $\sum_{a=1}^N \frac{1}{p_a} \sum_{l=J_{a,1}(3)+1}^{J_{a,1}(s) \wedge N} \frac{1}{p_l}$ . Similar to (5.11), we have  $\sum_{l=J_{a,1}(3)+1}^{J_{a,1}(s) \wedge N} \frac{1}{p_l} \sim \log \frac{\log J_{a,1}(s)}{\log J_{a,1}(3)}$ , and from (5.10) we have  $\lim_{a \rightarrow \infty} \frac{\log J_{a,1}(s)}{\log J_{a,1}(3)} = \frac{s-1}{3-1} = \frac{s-1}{2}$ . Thus, similar to (5.15), we obtain

(5.19)

$$\sum_{a=1}^N \frac{1}{p_a} \sum_{l=J_{a,1}(3)+1}^{J_{a,1}(s) \wedge N} \frac{1}{p_l} \sim \sum_{a=1}^{N^b} \frac{1}{p_a} \sum_{l=J_{a,1}(3)+1}^{J_{a,1}(s)} \frac{1}{p_l} \sim (\log \log N) (\log(s-1) - \log 2).$$

And finally, similar to (5.16), the contribution to  $P_N(3 \leq \frac{\log n}{\log p^{-(n)}} \leq s)$  from situation (3), which we denote by  $\rho_3(s)$ , satisfies

$$(5.20) \quad \rho_3(s) \sim (e^{-\gamma} \log \log N) \frac{\Lambda_2(s) - \Lambda_2(3)}{\log N}, \quad s \in [3, 4],$$

where the inclusion of the right endpoint  $s = 4$  follows from the remarks made at the beginning of the treatment of the case  $s \in [3, 4]$ .

We now analyze the contribution from situation (4) in (5.18). From (5.3) and (5.4), the contribution to  $P_N(3 \leq \frac{\log n}{\log p^{-(n)}} \leq s)$  from situation (4), which we will denote by  $\rho_4(s)$ , is

$$(5.21) \quad \rho_4(s) = C_N^{-1} \sum_{a=1}^N \frac{1}{p_a} \sum_{\substack{a+1 \leq j_1 \leq j_2 \leq N \\ p_{j_1} p_{j_2} \leq p_a^{s-1}}} \frac{1}{p_{j_1} p_{j_2}}.$$

Define

$$J_a(s, j_1) = \max\{j : p_j \leq \frac{p_a^{s-1}}{p_{j_1}}\}, \quad J_{a,2}(s) = \max\{j : p_j^2 \leq p_a^{s-1}\}.$$

Then

$$(5.22) \quad \rho_4(s) = C_N^{-1} \sum_{a=1}^N \frac{1}{p_a} \sum_{j_1=a+1}^{J_{a,2}(s) \wedge N} \frac{1}{p_{j_1}} \sum_{j_2=j_1}^{J_a(s, j_1) \wedge N} \frac{1}{p_{j_2}}.$$

Since  $p_j \sim j \log j$ , it follows that  $J_{a,2}(s) \log J_{a,2}(s) \sim (a \log a)^{\frac{s-1}{2}}$ , as  $a \rightarrow \infty$ . Taking the logarithm of both sides above, it follows that  $\log J_{a,2}(s) \sim \frac{s-1}{2} \log a$  as  $a \rightarrow \infty$ . Thus

$$(5.23) \quad J_{a,2}(s) \sim \frac{2}{s-1} a^{\frac{s-1}{2}} (\log a)^{\frac{s-3}{2}}, \quad \text{as } a \rightarrow \infty.$$

Consider now  $J_a(s, j_1)$ , for  $a+1 \leq j_1 \leq J_{a,2}(s)$ . Similarly as in the above paragraph, it follows that  $J_a(s, j_1) \log J_a(s, j_1) \sim \frac{(a \log a)^{s-1}}{j_1 \log j_1}$ , as  $j_1, a \rightarrow \infty$ . Since  $j_1 \leq J_{a,2}(s)$ , it follows from (5.23) that  $j_1 = o(a^{s-1})$ . Thus, taking the logarithm of both sides above, we have

$$(5.24) \quad \log J_a(s, j_1) \sim (s-1) \log a - \log j_1, \quad \text{as } j_1, a \rightarrow \infty.$$

Therefore,

$$(5.25) \quad J_a(s, j_1) \sim \frac{a^{s-1} (\log a)^{s-1}}{j_1 \log j_1 ((s-1) \log a - \log j_1)}, \quad \text{as } j_1, a \rightarrow \infty.$$

In light of (5.23) and (5.25), we can choose  $b \in (0, 1)$ , depending on  $s$ , such that  $J_a(s, j_1) \leq N$  and  $J_{a,2}(s) \leq N$ , for all  $a \leq N^b$  and all sufficiently large



$N$ . Thus, from (5.14) and (5.22), we have, similar to the first asymptotic equivalence in (5.15),

$$(5.26) \quad \rho_4(s) \sim C_N^{-1} \sum_{a=1}^{N^b} \frac{1}{p_a} \sum_{j_1=a+1}^{J_{a,2}(s)} \frac{1}{p_{j_1}} \sum_{j_2=j_1}^{J_a(s,j_1)} \frac{1}{p_{j_2}}.$$

By (1.19) and (5.24), we have

$$(5.27) \quad \sum_{j_2=j_1}^{J_a(s,j_1)} \frac{1}{p_{j_2}} \sim \log \frac{\log J_a(s,j_1)}{\log j_1} \sim \log \frac{(s-1) \log a - \log j_1}{\log j_1} = \log \left( (s-1) \frac{\log a}{\log j_1} - 1 \right),$$

as  $j_1, a \rightarrow \infty$ .

Using (5.27), (5.23) and the fact that  $p_j \sim j \log j$  as  $j \rightarrow \infty$ , we have

$$(5.28) \quad \begin{aligned} & \sum_{j_1=a+1}^{J_{a,2}(s)} \frac{1}{p_{j_1}} \sum_{j_2=j_1}^{J_a(s,j_1)} \frac{1}{p_{j_2}} \sim \sum_{j_1=a+1}^{J_{a,2}(s)} \frac{1}{j_1 \log j_1} \log \left( (s-1) \frac{\log a}{\log j_1} - 1 \right) \sim \\ & \int_a^{J_{a,2}(s)} \frac{1}{x \log x} \log \left( (s-1) \frac{\log a}{\log x} - 1 \right) dx \sim \\ & \int_a^{a^{\frac{s-1}{2}}} \frac{1}{x \log x} \log \left( (s-1) \frac{\log a}{\log x} - 1 \right) dx, \text{ as } a \rightarrow \infty, \end{aligned}$$

where the final asymptotic equivalence follows from the iterated logarithmic growth rate of the indefinite integral of the integrand appearing in the equation. Making the substitution

$$x = a^{\frac{s-1}{u_2}}$$

reveals that the second integral in (5.28) is in fact independent of  $a$ . We obtain

$$(5.29) \quad \begin{aligned} & \int_a^{a^{\frac{s-1}{2}}} \frac{1}{x \log x} \log \left( (s-1) \frac{\log a}{\log x} - 1 \right) dx = \int_2^{s-1} \frac{du_2}{u_2} \log(u_2 - 1) = \\ & \int_2^{s-1} \int_1^{u_2-1} \frac{du_1}{u_1} \frac{du_2}{u_2} = \Lambda_3(s). \end{aligned}$$

From (5.28) and (5.29), we conclude that

$$(5.30) \quad \lim_{a \rightarrow \infty} \sum_{j_1=a+1}^{J_{a,2}(s)} \frac{1}{p_{j_1}} \sum_{j_2=j_1}^{J_a(s,j_1)} \frac{1}{p_{j_2}} = \Lambda_3(s).$$

Thus, recalling the asymptotic behavior of  $C_N$ , from (5.30), (5.26) and (5.14) we conclude that

$$(5.31) \quad \rho_4(s) \sim (e^{-\gamma} \log \log N) \frac{\Lambda_3(s)}{\log N}, \quad s \in [3, 4],$$

where the inclusion of the right endpoint  $s = 4$  follows from the remarks made at the beginning of the treatment of the case  $s \in [3, 4]$ . From (5.20) and (5.31), we conclude that

$$(5.32) \quad P_N\left(3 \leq \frac{\log n}{\log p^-(n)} \leq s\right) \sim (e^{-\gamma} \log \log N) \frac{(\Lambda_2(s) - \Lambda_2(3)) + \Lambda_3(s)}{\log N}, \quad s \in [3, 4].$$

From (5.17) with  $s = 3$  and (5.32), and recalling that  $\Lambda_1(s) \equiv 1$ , we have

$$P_N\left(\frac{\log n}{\log p^-(n)} \leq s\right) \sim (e^{-\gamma} \log \log N) \frac{\Lambda_1(s) + \Lambda_2(s) + \Lambda_3(s)}{\log N}, \quad s \in [3, 4].$$

We now consider the general case that  $s \in [L, L + 1]$ . By induction, we have

$$(5.33) \quad P_N\left(\frac{\log n}{\log p^-(n)} \leq s\right) \sim (e^{-\gamma} \log \log N) \frac{\sum_{l=1}^{[s]} \Lambda_l(s)}{\log N}, \quad s \leq L.$$

Making a list similar to (5.18), and analyzing the situations as was done there, one concludes that the situations with  $X_{p_a} \geq 2$  do not contribute to the leading order asymptotics of  $P_N(L \leq \frac{\log n}{\log p^-(n)} \leq s)$ , while the situations with  $X_{p_a} = 1$  do contribute. When  $X_{p_a} = 1$ , we obtain  $L - 1$  situations, with all but one of them of the form already treated in the case of  $L - 1$ . (In (5.18), where  $L = 3$ , there were 2 such situations—labeled there (3) and (4), and one of them, namely (3), was of the form already treated for  $L = 2$ .) Thus, by induction and by the argument used to show that the contribution from situation (3) in (5.18) is as it appears in (5.20), these terms will give asymptotic contributions

$$(5.34) \quad (e^{-\gamma} \log \log N) \frac{\Lambda_1(s) - \Lambda_1(L)}{\log N}, \dots, (e^{-\gamma} \log \log N) \frac{\Lambda_{L-1}(s) - \Lambda_{L-1}(L)}{\log N}.$$

We now look at the new situation that arises; namely the one in which  $X_{p_a} = 1$  and there exist  $j_1, \dots, j_{L-1}$  satisfying  $a + 1 \leq j_1 \leq \dots \leq j_{L-1} \leq N$  and  $\prod_{i=1}^{L-1} p_{j_i} \leq p_a^{s-1}$ , such that for  $j \in \{a + 1, \dots, N\}$ ,  $X_j$  is equal to the number of times  $j$  appears among the  $\{j_i\}_{i=1}^{L-1}$ . From (5.3) and (5.4), the

contribution to  $P_N(L \leq \frac{\log n}{\log p^-(n)} \leq s)$  from this situation, similar to (5.21) in the case  $L = 3$ , is

$$(5.35) \quad C_N^{-1} \sum_{a=1}^N \frac{1}{p_a} \sum_{\substack{a+1 \leq j_1 \leq \dots \leq j_{L-1} \leq N \\ \prod_{i=1}^{L-1} p_{j_i} \leq p_a^{s-1}}} \frac{1}{\prod_{i=1}^{L-1} p_{j_i}}.$$

An analysis analogous to that implemented between (5.21) and (5.30) gives

$$(5.36) \quad \lim_{a \rightarrow \infty} \sum_{\substack{a+1 \leq j_1 \leq \dots \leq j_{L-1} \leq N \\ \prod_{i=1}^{L-1} p_{j_i} \leq p_a^{s-1}}} \frac{1}{\prod_{i=1}^{L-1} p_{j_i}} = \int_{L-1}^{s-1} \int_{L-2}^{u_{L-1}-1} \dots \int_1^{u_2-1} \prod_{j=1}^{L-1} \frac{du_j}{u_j} = \Lambda_L(s).$$

From (5.35) and (5.36) it follows that the contribution to the leading order asymptotics of  $P_N(L \leq \frac{\log n}{\log p^-(n)} \leq s)$  from this situation is  $(e^{-\gamma} \log \log N) \frac{\Lambda_L(s)}{\log N}$ . We conclude from this and (5.34) that

$$(5.37) \quad P_N(L \leq \frac{\log n}{\log p^-(n)} \leq s) \sim (e^{-\gamma} \log \log N) \frac{\Lambda_L(s) + \sum_{l=1}^{L-1} (\Lambda_l(s) - \Lambda_l(L))}{\log N}, \quad s \in [L, L+1].$$

From (5.33) with  $s = L$  and from (5.37), we conclude that

$$P_N(L \leq \frac{\log n}{\log p^-(n)} \leq s) \sim (e^{-\gamma} \log \log N) \frac{\sum_{l=1}^L \Lambda_l(s)}{\log N}, \quad s \in [L, L+1].$$

This completes the proof of (5.1).  $\square$

## 6. APPENDIX: PROOF OF PROPOSITION 5

For notational convenience, we will work with  $p$  instead of  $p_j$ . The proof is via the inclusion-exclusion principle along with the fact that  $D_{\text{nat}}(S_k) = \frac{1}{\zeta(k)}$ , where  $S_k$  denotes the  $k$ -free integers, as was noted with a reference in the proof of Proposition 3. Recall that  $1 \leq l < k$ . We have

$$\begin{aligned} I_N &\equiv |\{n : p^l | n, n \leq N, n \in S_k\}| = |\{n_1 : n_1 \leq [\frac{N}{p^l}], n_1 \in S_k, p^{k-l} \nmid n_1\}| = \\ &|\{n_1 : n_1 \leq [\frac{N}{p^l}], n_1 \in S_k\}| - |\{n_1 : n_1 \leq [\frac{N}{p^l}], n_1 \in S_k, p^{k-l} | n_1\}| \equiv I_{N,1} - I_{N,2}. \end{aligned}$$

Similarly,

$$\begin{aligned} I_{N,2} &= |\{n_2 : n_2 \leq [\frac{N}{p^k}], n_2 \in S_k, p^l \nmid n_2\}| = \\ &|\{n_2 : n_2 \leq [\frac{N}{p^k}], n_2 \in S_k\}| - |\{n_2 : n_2 \leq [\frac{N}{p^k}], n_2 \in S_k, p^l | n_2\}| \equiv I_{N,3} - I_{N,4}, \end{aligned}$$

and

$$I_{N,4} = |\{n_3 : n_3 \leq [\frac{N}{p^{k+l}}], n_3 \in S_k, p^{k-l} \nmid n_3\}| =$$

$$|\{n_3 : n_3 \leq [\frac{N}{p^{k+l}}], n_3 \in S_k\}| - |\{n_3 : n_3 \leq [\frac{N}{p^{k+l}}], n_3 \in S_k, p^{k-l} | n_3\}| \equiv I_{N,5} - I_{N,6}.$$

So up to this point, we have

$$I_N = I_{N,1} - I_{N,3} + I_{N,5} - I_{N,6}.$$

Now  $\lim_{N \rightarrow \infty} \frac{I_{N,1}}{N} = \frac{1}{p^l} D_{\text{nat}}(S_k) = \frac{1}{p^l \zeta(k)}$ ,  $\lim_{N \rightarrow \infty} \frac{I_{N,3}}{N} = \frac{1}{p^k} D_{\text{nat}}(S_k) = \frac{1}{p^k \zeta(k)}$  and  $\lim_{N \rightarrow \infty} \frac{I_{N,5}}{N} = \frac{1}{p^{k+l}} D_{\text{nat}}(S_k) = \frac{1}{p^{k+l} \zeta(k)}$ . Continuing this process of inclusion-exclusion, we have

$$I_N = \sum_{m=0}^{\infty} I_{N,4m+1} - \sum_{m=0}^{\infty} I_{N,4m+3},$$

where for each  $N$  only a finite number of the summands above are non-zero.

Now

$$\lim_{N \rightarrow \infty} \frac{I_{N,4m+1}}{N} = \frac{1}{p^{mk+l} \zeta(k)}, \quad m = 0, 1, \dots,$$

and

$$\lim_{N \rightarrow \infty} \frac{I_{N,4m+3}}{N} = \frac{1}{p^{(m+1)k} \zeta(k)}, \quad m = 0, 1, \dots$$

From this we conclude that

$$D_{\text{nat}}(\beta_p \geq l, S_k) = \frac{1}{\zeta(k)} \sum_{m=0}^{\infty} \frac{1}{p^{mk+l}} - \frac{1}{\zeta(k)} \sum_{m=0}^{\infty} \frac{1}{p^{(m+1)k}} = \frac{1}{\zeta(k)} \frac{\frac{1}{p^l} - \frac{1}{p^k}}{1 - \frac{1}{p^k}}.$$

Thus,  $D_{\text{nat}}(\beta_p \geq l | S_k) = \frac{\frac{1}{p^l} - \frac{1}{p^k}}{1 - \frac{1}{p^k}}$ . □

#### REFERENCES

- [1] Arratia, R., Barbour, A. and Tavaré, S., *Logarithmic Combinatorial Structures: A Probabilistic Approach*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zurich, (2003).
- [2] Billingsley, P., *On the distribution of large prime divisors*, Collection of articles dedicated to the memory of Alfréd Rényi, I. Period. Math. Hungar. 2 (1972), 283–289.
- [3] Billingsley, P., *The probability theory of additive arithmetic functions*, Ann. Probability 2 (1974), 749–791.
- [4] de Bruijn, N., *On the number of positive integers  $\leq x$  and free of prime factors  $> y$* , Nederl. Acad. Wetensch. Proc. Ser. A. 54, (1951), 50–60.
- [5] Buchstab, A., *An asymptotic estimation of a general number-theoretic function*, Mat. Sbornik, 44, (1937), 1239–1246.

- [6] Cellarosi, F. and Sinai, Y., *Non-standard limit theorems in number theory*, Prokhorov and Contemporary Probability Theory, 197–213, Springer Proc. Math. Stat., 33, Springer, Heidelberg, (2013).
- [7] Dickman, K., *On the frequency of numbers containing prime factors of a certain relative magnitude*, Ark. Math. Astr. Fys. 22, 1-14 (1930).
- [8] Durrett, R., *Probability Theory and Examples*, third edition, Brooks/Cole, Belmont, CA (2005).
- [9] Elliott, P.D.T.A., *Probabilistic Number Theory. I. Mean-Value Theorems*, Grundlehren der Mathematischen Wissenschaften, 239, Springer-Verlag, New York-Berlin, (1979).
- [10] Elliott, P.D.T.A., *Probabilistic Number Theory. II. Central Limit Theorems*, Grundlehren der Mathematischen Wissenschaften, 240, Springer-Verlag, New York-Berlin, (1980).
- [11] Erdős, P and Wintner, A., *Additive arithmetical functions and statistical independence*, Amer. J. Math. 61, (1939), 713–721.
- [12] Erdős, P. and Kac, M., *The Gaussian law of errors in the theory of additive number theoretic functions*, Amer. J. Math. 62, (1940), 738–742.
- [13] Giuliano, R. and Macci, C., *Asymptotic results for weighted means of random variables which converge to a Dickman distribution, and some number theoretical applications*, ESAIM Probab. Stat. 19 (2015), 395–413.
- [14] Hardy, G. H. and Ramanujan, S., *Proof that almost all numbers  $n$  are composed of about  $\log \log n$  prime factors*, Proc. London Math. Soc. 16 (1917), *Collected papers of Srinivasa Ramanujan*, 242–243, AMS Chelsea Publ., Providence, RI, (2000).
- [15] Lagarias, J., *Euler’s constant: Euler’s work and modern developments*, Bull. Amer. Math. Soc. (N.S.) 50 (2013), 527–628.
- [16] Mehrdad, B. and Zhu, L., *Moderate and large deviations for the Erdős-Kac theorem*, Q. J. Math. 67 (2016), 147–160.
- [17] Mehrdad, B. and Zhu, L., *Limit theorems for empirical density of greatest common divisors*, Math. Proc. Cambridge Philos. Soc. 161 (2016), 517-533.
- [18] Montgomery, H. and Vaughan, R., *Multiplicative Number Theory. I. Classical Theory*, Cambridge Studies in Advanced Mathematics, 97, Cambridge University Press, Cambridge, (2007).
- [19] Nathanson, M., *Elementary Methods in Number Theory*, Graduate Texts in Mathematics, 195, Springer-Verlag, New York, (2000).
- [20] Penrose, M. and Wade, A., *Random minimal directed spanning trees and Dickman-type distributions*, Adv. in Appl. Probab. 36 (2004), 691–714.
- [21] Pinsky, R., *On the strange domain of attraction to generalized Dickman distributions for sums of independent random variables*, preprint (2017).
- [22] Shapiro, H., *Introduction to the Theory of Numbers*, John Wiley & Sons, Inc., New York, (1983).

- [23] Tenenbaum, G., *Introduction to Analytic and Probabilistic Number Theory*, Cambridge Studies in Advanced Mathematics, 46, Cambridge University Press, Cambridge, (1995).
- [24] Turán, P., *On a theorem of Hardy and Ramanujan*, J. London Math. Soc. 9, (1934), 274–276.

DEPARTMENT OF MATHEMATICS, TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY,  
HAIFA, 32000, ISRAEL

*E-mail address:* `pinsky@math.technion.ac.il`

*URL:* <http://www.math.technion.ac.il/~pinsky/>