



**[biblio.ugent.be](http://biblio.ugent.be)**

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

**Title: Asymptotic distribution of integers with certain prime factorizations**

**Authors: Hans Vernaevae, Jasson Vindas, and Andreas Weiermann**

**In: Journal of Number Theory, 136 (3), 87-99, 2014.**

**Link to the article:** <http://dx.doi.org/10.1016/j.jnt.2013.09.001>

**To refer to or to cite this work, please use the citation to the published version:**

**H. Vernaevae, J. Vindas, A. Weiermann (2014). Asymptotic distribution of integers with certain prime factorizations. *J. Number Theory* 136(3) 87-99. doi:10.1016/j.jnt.2013.09.001**

# ASYMPTOTIC DISTRIBUTION OF INTEGERS WITH CERTAIN PRIME FACTORIZATIONS

HANS VERNAEVE, JASSON VINDAS, AND ANDREAS WEIERMANN

ABSTRACT. Let  $p_1 < p_2 < \dots < p_\nu < \dots$  be the sequence of prime numbers and let  $m$  be a positive integer. We give a strong asymptotic formula for the distribution of the set of integers having prime factorizations of the form  $p_m^{k_1} p_m^{k_2} \dots p_m^{k_n}$  with  $k_1 \leq k_2 \leq \dots \leq k_n$ . Such integers originate in various combinatorial counting problems; when  $m = 2$ , they arise as Matula numbers of certain rooted trees.

## 1. INTRODUCTION

Let  $\{p_\nu\}_{\nu=1}^\infty$  be the sequence of all prime numbers arranged in increasing order and let  $m > 1$  be a fixed positive integer. We shall consider the class of integers only admitting prime factors from the subsequence  $\{p_m^k\}_{k=0}^\infty$ , that is, the set

$$(1.1) \quad A_m = \{p_m^{k_1} p_m^{k_2} \dots p_m^{k_n} \in \mathbb{N} : 0 \leq k_1 \leq k_2 \leq \dots \leq k_n\} .$$

The aim of this article is to provide an asymptotic formula for the distribution of  $A_m$ , namely, the following counting function

$$M_{2,m}(x) = \sum_{\substack{n \leq x \\ n \in A_m}} 1 .$$

The function  $M_{2,2}$  arises in various interesting combinatorial counting problems; particularly, in connection with rooted trees. In 1968 Matula gave an enumeration of (non-planar) rooted trees by prime factorization [20], the so-called Matula numbers. Number theoretic aspects of this rooted tree coding have been investigated in detail in [1, 13]. Such numbers may be used to deduce many intrinsic properties of rooted trees [4, 12, 15]. The set  $A_2$  in fact corresponds to a class of Matula numbers. In Section 2 we review Matula coding of rooted trees and give the interpretation of  $M_{2,2}$  as the counting function of rooted trees with height less or equal to 2, under Matula's enumeration. It is worth mentioning that the significance of Matula numbers comes from applications in organic chemistry, as they can be employed to develop efficient nomenclatures for representing molecules of a variety of organic compounds (cf. [5, 6, 7, 8, 14]). As explained in Section 2,  $M_{2,2}$  might also be regarded as a "transfinite counting function" for the ordinal  $\omega^\omega$  in a certain complexity norm [25].

In [25] Weiermann found the weak asymptotics of the counting function  $M_{2,2}$ . Using a Tauberian theorem by Kohlbecker for partitions [18], he showed that

---

2010 *Mathematics Subject Classification.* Primary 05A17, 11P82. Secondary 05A16, 05C30.

*Key words and phrases.* strong asymptotics for partition problems; Matula numbers; rooted trees; tree enumeration by prime factorization.

$$(1.2) \quad \log M_{2,2}(x) \sim \pi \sqrt{\frac{2 \log x}{3 \log 2}}.$$

The asymptotic relation (1.2) resembles the one obtained by Hardy and Ramanujan in 1917 for the celebrated (unrestricted) partition function,

$$(1.3) \quad \log p(n) \sim \pi \sqrt{\frac{2n}{3}},$$

which they [16], and independently Uspensky [24], greatly refined later to

$$(1.4) \quad p(n) \sim \frac{e^{\pi \sqrt{\frac{2n}{3}}}}{(4\sqrt{3})n}.$$

Naturally, the transition from (1.3) to (1.4) consists in finding missing asymptotic terms. The problem we address here is of similar nature. We shall fill the gap between (1.2) and the strong asymptotics by exhibiting hidden lower order terms in the approximation (1.2), as stated in the following theorem.

**Theorem 1.** *The function  $M_{2,m}$  has asymptotic behavior*

$$(1.5) \quad M_{2,m}(x) \sim \frac{e^{K_m \sqrt{3} \log m}}{2\pi^2 \log 2} (\log x)^{\frac{\log\left(\frac{\pi}{\sqrt{6} \log m}\right)}{2 \log m}} \exp\left(\pi \sqrt{\frac{2 \log x}{3 \log m}} - \frac{(\log \log x)^2}{8 \log m}\right),$$

where

$$K_m = \frac{1}{2 \log m} \left( (\log \log m)^2 + \gamma^2 - 2\gamma \log \log m - \frac{\pi^2}{6} - \log^2 \left( \frac{\pi}{\sqrt{6} \log m} \right) \right) - C_{2,m},$$

$\gamma$  is the Euler-Mascheroni constant, and  $C_{2,m}$  is given by the convergent series

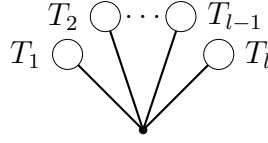
$$C_{2,m} = \sum_{k=1}^{\infty} \left( \log \log p_{m^k} - \log k - \log \log m - \frac{\log k}{k \log m} - \frac{\log \log m}{k \log m} \right).$$

We will provide a proof of Theorem 1 in Section 4. The proof is based on Ingham's method from [17]; however, it turns out that Ingham's original theorem for partitions [17, Thm. 2] is not directly applicable to our context. In Section 3, we shall slightly extend his result. It is likely that such an extension of Ingham's theorem might be useful for treating partition problems other than the one dealt with in this article.

## 2. TWO COUNTING PROBLEMS AND $M_{2,m}$

**2.1. Rooted trees.** Matula's coding of (non-planar) rooted trees in terms of prime factorizations provides a bijection between such trees and the positive integers. The same rooted tree enumeration was rediscovered by Göbel in [11]. It is defined as follows.

If we denote the trivial one-vertex tree by  $\bullet$ , then its Matula number is  $n(\bullet) := 1$ . Inductively, if  $T_1, T_2, \dots, T_l$  are trees and  $T$  is given as

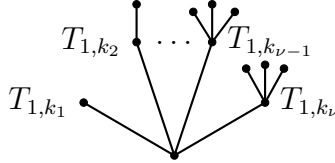


then its Matula number is defined as  $n(T) := p_{n(T_1)} \cdots p_{n(T_l)}$ .

If  $T_{1,k}$  is the tree of height one with  $k$  nodes above the root, then  $n(T_{1,k}) = p_1^k = 2^k$ . If  $T$  has height two, then

$$n(T) = p_{n(T_{1,k_1})} p_{n(T_{1,k_2})} \cdots p_{n(T_{1,k_{\nu-1}})} p_{n(T_{1,k_\nu})} = p_{2^{k_1}} p_{2^{k_2}} \cdots p_{2^{k_{\nu-1}}} p_{2^{k_\nu}},$$

where the  $j$ -th node connected to the root carries a tree  $T_{1,k_j}$ .



It is then clear that Matula coding gives a bijection between the set of rooted trees with height equal to 1 or 2 and the set  $A_2$  defined in (1.1). Consequently,  $M_{2,2}(x)$  counts the number of Matula numbers corresponding to trees with  $0 < \text{height}(T) \leq 2$ , that are below  $x$ , i.e.,

$$\sum_{\substack{n(T) \leq x \\ \text{height}(T) \leq 2}} 1 = M_{2,2}(x) + 1.$$

Thus, this rooted tree counting function has also asymptotics (1.5).

**2.2. Ordinal counting functions.** It might seem surprising at first sight that the counting function  $M_{2,2}$  is related to studying asymptotic properties of transfinite ordinals. Since transfinite ordinals rarely show up in a number-theoretic context we will explain some features of this connection in informal and general terms. The rest of the paper will not depend on the exposition given in this subsection, but it might be useful as a source of inspiration for further study.

In naive set theory ordinals generalize the ordering of the natural numbers  $0 < 1 < 2 < \dots$  by continuing beyond the first limit point  $\omega$  like  $0 < 1 < 2 < \dots < \omega + 1 < \omega + 2 < \dots$ . This process can be continued beyond the next limit  $\omega + \omega$  like  $0 < 1 < 2 < \dots < \omega + 1 < \omega + 2 < \dots < \omega \cdot 2 + 1 < \omega \cdot 2 + 2 < \dots$  and by iteration like  $0 < 1 < 2 < \dots < \omega + 1 < \omega + 2 < \dots < \omega \cdot 2 + 1 < \omega \cdot 2 + 2 < \dots < \omega \cdot 3 < \dots < \omega \cdot n < \dots$ . At a certain moment we reach the first limit of limits  $\omega \cdot \omega$  and, again by iteration, we reach limits of limits of limits and in the limit of this counting we reach  $\omega^\omega$  (an ordinal which – as will become clear soon – is of relevance to  $M_{2,m}$ ).

The ordinal  $\omega^\omega$  is not at all frightening since it appears as the order type of the polynomials in  $\mathbb{N}[x]$  under eventual domination or as the order type of the multisets of natural numbers. There is of course no bound in counting through the ordinals and by

further counting we reach  $\omega^{\omega^\omega}, \omega^{\omega^{\omega^\omega}}, \dots$ , but the higher we go the more complicated the description becomes. Slight extensions could still be dealt with by combinatorial means (which can still be formalized in Peano arithmetic) and stronger extensions will require from some moment onwards basic set theoretic machinery.

There is still some nice and accessible visualization of the ordinals less than  $\varepsilon_0$ , which is the limit of the finite powers of  $\omega$  showing up in the sequence  $\omega, \omega^\omega, \omega^{\omega^\omega}, \omega^{\omega^{\omega^\omega}}, \dots$ . For this we consider a subclass of Hardy's orders of infinity. Let  $\mathcal{E}$  be the class of unary functions  $f : \mathbb{N} \rightarrow \mathbb{N}$  such that

- (1) the function  $c_0$  is an element of  $\mathcal{E}$  where  $c_0(x) = x$  and
- (2) with two functions  $f, g$  in  $\mathcal{E}$  also the function  $h$  is in  $\mathcal{E}$ , where  $h(x) = x^{f(x)} + g(x)$ .

On  $\mathcal{E}$  we define the ordering of eventual domination as usual by  $f \prec g$  if and only if there exists a non-negative integer  $k$  such that  $f(x) < g(x)$  for all  $x \geq k$ . The structure  $\langle \mathcal{E}, \prec \rangle$  is isomorphic with  $\langle \{\alpha : \alpha < \varepsilon_0\}, < \rangle$  and so we can identify both structures. If we also write  $id = c_0$  for the identity function on  $\mathbb{N}$ , then the isomorphism maps  $\omega$  to  $id$ ,  $\omega^\omega$  to  $id^{id}$ ,  $\omega^{\omega^\omega}$  to  $id^{id^{id}}$ , etc. The ordinal  $\varepsilon_0$  is the proof-theoretic ordinal of first order Peano arithmetic  $PA$ .  $PA$  proves (after an appropriate formalization of the context) the scheme of transfinite induction for all strict initial segments of  $\varepsilon_0$  but not the scheme of transfinite induction for the full segment up to  $\varepsilon_0$ .

For treating  $\mathcal{E}$  in the context of arithmetic we need a specific (easily definable) coding of the elements of  $\mathcal{E}$  into the natural numbers. One of the standard devices for achieving this is provided by associating to the elements of  $\mathcal{E}$  their canonical counterparts in the finite non-planar rooted trees. Such a bijection  $t$  can be defined recursively as follows. First,  $t(c_0) := \bullet$ . Every  $f$  can be written as  $f = id^{g_1} + \dots + id^{g_n}$ , then let

$$t(f) := \begin{array}{c} \begin{array}{c} t(g_2) \circ \dots \circ t(g_{l-1}) \\ \diagdown \quad \quad \quad \diagup \\ t(g_1) \circ \quad \quad \quad \circ t(g_l) \end{array} \\ \diagdown \quad \quad \quad \diagup \\ \bullet \end{array} .$$

By this identification we can – using Matula's coding – canonically associate to  $f \in \mathcal{E}$  its Gödel number  $\lceil f \rceil := n(t(f))$ . This coding has been used explicitly by Troelstra and Schwichtenberg in [23, p. 320, Def. 10.1.5]. In this context we arrive at the following interpretation

$$1 + M_{2,2}(x) = \#\{f \in \mathcal{E} : f \prec id^{id} \wedge \lceil f \rceil \leq x\} .$$

For coding a larger segment of ordinals Schütte [22, Sec. V.8] used a related coding  $Nr$  which when restricted to  $\mathcal{E}$  has the property that  $1 + M_{2,4}(x)$  is the number of ordinals  $\alpha$  below  $\omega^\omega$  such that  $Nr(\alpha) \leq x$ .

Until now, the study of ordinal counting functions has found applications to logical limit laws for ordinals and to phase transitions for Gödel incompleteness results (it seems very interesting and intriguing to find additional applications). A further discussion of phase transitions will be beyond the scope of this exposition, but we want to include an intriguing example for a zero-one law (see [2] for an account on logical limit laws). As usual, we use  $\models$  for the satisfaction relation from model theory. Let  $\varphi$

be a sentence in the language of linear orders. Let

$$\delta_\varphi := \lim_{x \rightarrow \infty} \frac{\#\{f \in \mathcal{E} : f \prec id^{id} \wedge [f] \leq x \wedge \langle \{g \in \mathcal{E} : g \prec f\}, \prec \rangle \models \varphi\}}{\#\{f \in \mathcal{E} : f \prec id^{id} \wedge [f] \leq x\}}.$$

Then  $\delta_\varphi$  exists and either  $\delta_\varphi = 1$  or  $\delta_\varphi = 0$ . A proof of this and similar results has been obtained in [26] by an analysis of the asymptotic behavior of  $M_{2,2}$  and related counting functions.

At the beginning of this subsection it has been indicated that ordinals might provide a source of inspiration for further research and we will now indicate some possible options. For  $g \in \mathcal{E}$ , let  $c_g(x) := \#\{f \in \mathcal{E} : f \prec g \wedge [f] \leq x\}$ . For various choices of  $g$  some preliminary results on weak asymptotics for  $c_g$  have been obtained in [25]. Moreover, strong asymptotics for  $c_{id^k}$  can be obtained by elementary means.

We believe that the methods of this paper will allow one to provide strong asymptotics for  $c_{id^{id^k}}$  for any given fixed  $k$ . A strong asymptotic formula for  $c_{id^{id^{id}}}$  (which would resemble something like multiplicative double partitions) seems however to require new methods. A general challenge would be then to provide a general theorem on strong asymptotics for  $c_g$  for any fixed  $g$  and for analogous functions emerging from the Schütte coding.

### 3. AN EXTENSION OF INGHAM'S THEOREM FOR UNRESTRICTED PARTITIONS

As mentioned in the Introduction, we need an extension of Ingham's theorem for strong asymptotics of partition functions. The extension will follow from a complex Tauberian theorem for large asymptotic behavior of the Laplace transform, also due to Ingham [17, Thm. 1'].

Let  $0 < \lambda_0 < \lambda_1 < \dots < \lambda_k \rightarrow \infty$  be a sequence of real numbers and let

$$N(u) = \sum_{\lambda_k \leq u} 1$$

be its counting function. Consider the additive semigroup  $\Lambda$  generated by  $\{\lambda_k\}_{k=0}^\infty$ , i.e.,

$$\Lambda = \left\{ r \in \mathbb{R} : r = \sum_{k=0}^l n_k \lambda_k, n_k \in \mathbb{N} \right\}.$$

For  $r \in \Lambda$ , the partition function  $p(r)$  is defined as the number of ways of writing  $r$  as  $r = \sum_{k=0}^l n_k \lambda_k$ . We further set

$$P(u) = \sum_{\substack{r \leq u \\ r \in \Lambda}} p(r).$$

The following theorem obtains the asymptotic behavior of  $P(u)$  if one knows a certain average asymptotic behavior for  $N(u)$ . It slightly extends that of Ingham by allowing an extra term of the form  $B \log^2 u$  in the asymptotic expansion (3.1). As usual,  $\zeta$  stands for the Riemann zeta function and  $\Gamma$  for the Euler Gamma function. The constant  $\gamma_1$

denotes the Stieltjes constant, that is,

$$\gamma_1 = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{\log k}{k} - \frac{\log^2 n}{2}.$$

**Theorem 2.** *Suppose that*

$$(3.1) \quad \int_0^u \frac{N(t)}{t} dt = \frac{A}{\alpha} u^\alpha + B \log^2 u + C \log u + D + o(1),$$

with  $\alpha, A > 0$ . Then

$$(3.2) \quad P(u) \sim \left( \frac{1-\beta}{2\pi} \right)^{\frac{1}{2}} e^{D'} M^{-(C+\frac{1}{2})} u^{C-\beta C-\frac{\beta}{2}} \exp \left( \frac{Mu^\beta}{\beta} + B \log^2 \left( \frac{u^{1-\beta}}{M} \right) \right),$$

where

$$\beta = \frac{\alpha}{\alpha+1}, \quad M = (A\alpha\Gamma(\alpha+1)\zeta(\alpha+1))^{\frac{1}{\alpha+1}}, \quad D' = D + \left( \frac{\pi^2}{6} - 2\gamma_1 - \gamma^2 \right) B.$$

In order to deduce Theorem 2 from Ingham's Tauberian theorem, we proceed to find the asymptotic behavior of the Laplace-Stieltjes transform of  $P$ . Set

$$F(s) = \sum_{r \in \Lambda} e^{-sr} = \int_0^\infty e^{-su} dP(u)$$

and

$$f(s) = s \int_0^\infty \frac{N(u)}{e^{su} - 1} du.$$

The generating function identity  $F(s) = e^{f(s)}$  is well-known.

**Lemma 1.** *If (3.1) holds, then, as  $\sigma \rightarrow 0^+$ ,*

$$\sigma \int_0^\infty \frac{N(u)}{e^{\sigma u} - 1} du = \frac{A\Gamma(\alpha+1)\zeta(\alpha+1)}{\sigma^\alpha} + B \log^2 \sigma - C \log \sigma + D' + o(1).$$

Here

$$(3.3) \quad D' = D + \left( \frac{\pi^2}{6} - 2\gamma_1 - \gamma^2 \right) B.$$

*Proof.* We employ standard Schwartz distribution calculus in our manipulations. It might also be possible to give a classical proof along the lines of that of [19, Thm. IV.23.1]. For Schwartz distributions, we follow the notation exactly as in [9, Chap. 2]. Taking distributional derivative in (3.1), we obtain

$$\begin{aligned} \frac{N(\lambda u)}{u} &= A\lambda^\alpha u_+^{\alpha-1} + (B \log^2 \lambda + C \log \lambda + D)\delta(u) + 2B \text{Pf} \left( \frac{H(u) \log u}{u} \right) \\ &\quad + (2B \log \lambda + C) \text{Pf} \left( \frac{H(u)}{u} \right) + o(1), \quad \lambda \rightarrow \infty, \end{aligned}$$

distributionally in the space of tempered distributions  $\mathcal{S}'$  (cf. [9, Sec. 3.9], [21, Sec. 2.5]), where  $\delta$  stands for the Dirac delta distribution,  $H$  is the Heaviside function, and Pf denotes regularization via Hadamard finite part [9, Sec. 2.4]. Testing this

asymptotic expansion at the test function  $\psi(u) = u/(e^u - 1)$ , setting  $\sigma = 1/\lambda$ , and taking into account the well-known formula

$$(3.4) \quad \Gamma(s+1)\zeta(s+1) = \int_0^\infty \frac{u^s}{e^u - 1} du, \quad \Re s > 0,$$

we obtain

$$\sigma \int_0^\infty \frac{N(u)}{e^{\sigma u} - 1} du = \frac{A\Gamma(\alpha+1)\zeta(\alpha+1)}{\sigma^\alpha} + B \log^2 \sigma - (C + 2BK) \log \sigma + D' + o(1),$$

where  $D' = D + 2BK' + CK$ , and the constants  $K$  and  $K'$  are given by the Hadamard finite part at 0 of the integrals

$$K = \text{F. p.} \int_0^\infty \frac{du}{e^u - 1} \quad \text{and} \quad K' = \text{F. p.} \int_0^\infty \frac{\log u}{e^u - 1} du.$$

Hence, it remains to evaluate these two constants. We will do so by inspecting the Laurent expansion at  $s = 0$  of the analytic continuation of (3.4). In fact, the classical procedure of Marcel Riesz [9, 10] yields the analytic continuation of (3.4) to  $\mathbb{C} \setminus \{0, -1, -2, \dots\}$  as the finite part integral of the right hand side. By employing the Gelfand-Shilov Laurent expansion at  $s = 0$  for the distribution  $u_+^{s-1}$  [10, p. 87], we conclude

$$\text{F. p.} \int_0^\infty \frac{u^s}{e^u - 1} du = \frac{1}{s} + \sum_{n=0}^\infty \frac{s^n}{n!} \text{F. p.} \int_0^\infty \frac{\log^n u}{e^u - 1} du, \quad 0 < |s| < 1.$$

On the other hand, since

$$\zeta(s+1) = \frac{1}{s} + \gamma - \gamma_1 s + \dots \quad \text{and} \quad \Gamma(s+1) = 1 - \gamma s + \left(\frac{\gamma^2}{2} + \frac{\pi^2}{12}\right) s^2 + \dots,$$

we have

$$\Gamma(s+1)\zeta(s+1) = \frac{1}{s} + \left(\frac{\pi^2}{12} - \gamma_1 - \frac{\gamma^2}{2}\right) s + \dots, \quad 0 < |s| < 1,$$

and therefore  $K = 0$  and  $K' = \pi^2/12 - \gamma_1 - \gamma^2/2$ .  $\square$

*Proof of Theorem 2.* We can now apply Ingham's Tauberian theorem [17, Thm. 1'] to the generating function  $F(s)$  with

$$\varphi(s) := \frac{A\Gamma(\alpha+1)\zeta(\alpha+1)}{s^\alpha} \quad \text{and} \quad \chi(s) := e^{D'} s^{-C} e^{B \log^2 s}.$$

Indeed, in view of Lemma 1,

$$F(\sigma) \sim \chi(\sigma) e^{\varphi(\sigma)}, \quad \sigma \rightarrow 0^+,$$

and the quoted theorem of Ingham immediately implies that

$$P(u) \sim \frac{\chi(\sigma(u)) \exp(\varphi(\sigma(u)) + u\sigma(u))}{\sqrt{2\pi\sigma^2(u)\varphi''(\sigma(u))}}, \quad u \rightarrow \infty,$$



where  $\sigma(u)$  is the inverse function of  $-\varphi'$ , i.e.,  $\sigma(u) = Mu^{-\frac{1}{\alpha+1}}$  with  $M = (A\alpha\Gamma(\alpha+1)\zeta(\alpha+1))^{\frac{1}{\alpha+1}}$ . Setting  $\beta = \alpha/(\alpha+1)$ , we have

$$e^{\varphi(\sigma(u))+u\sigma(u)} = e^{\frac{Mu^\beta}{\beta}},$$

$$\frac{\chi(\sigma(u))}{\sigma(u)\sqrt{2\pi\varphi''(\sigma(u))}} = \left(\frac{1-\beta}{2\pi}\right)^{\frac{1}{2}} e^{D'} M^{-(C+\frac{1}{2})} u^{C-\beta C-\frac{\beta}{2}} e^{B\log^2(u^{1-\beta}/M)}.$$

whence (3.2) follows.  $\square$

In the sequel, we will only use the case  $\alpha = 1$  of Theorem 2, which we state in the next corollary for the sake of convenience.

**Corollary 1.** *Suppose that*

$$\int_0^u \frac{N(t)}{t} dt = Au + B\log^2 u + C\log u + D + o(1).$$

Then

$$P(u) \sim \frac{e^{D'+B\log^2(\pi\sqrt{\frac{A}{6}})}}{2\sqrt{\pi}} \left(\pi\sqrt{\frac{A}{6}}\right)^{-(C+\frac{1}{2})} u^{\frac{C}{2}-\frac{1}{4}-B\log(\pi\sqrt{\frac{A}{6}})} \exp\left(\pi\sqrt{\frac{2Au}{3}} + \frac{B}{4}\log^2 u\right),$$

where  $D'$  is given by (3.3).

#### 4. THE ORIGINAL PROBLEM

We now proceed to give a proof of Theorem 1. We translate our original problem into an additive partition problem. Consider  $\lambda_k = \log p_{m^k}$ . Then, with the notation of the preceding section,

$$1 + M_{2,m}(e^u) = \sum_{\substack{r \leq u \\ r \in \Lambda}} p(r) = P(u).$$

Of course, here  $p(r) = 1$  for each  $r \in \Lambda = \left\{ \sum_{k=0}^l n_k \log p_{m^k} : n_k \in \mathbb{N} \right\}$ .

Thus, we are interested in the average asymptotics of the counting function

$$N(u) = \sum_{\log p_{m^k} \leq u} 1.$$

Observe that

$$(4.1) \quad \int_0^u \frac{N(t)}{t} dt = N(u) \log u - \sum_{\log p_{m^k} \leq u} \log \log p_{m^k}.$$

We first need to estimate  $\log \log p_{m^k}$ .

**Lemma 2.**

$$(4.2) \quad \log \log p_{m^k} = \log k + \log \log m + \frac{\log k}{k \log m} + \frac{\log \log m}{k \log m} + O\left(\frac{\log^2 k}{k^2}\right).$$

*Proof.* Using the prime number theorem, Cipolla [3] found in 1902 an asymptotic formula for  $p_n$ . Employing just two terms in the expansion, we have

$$p_n = n \log n + O(n \log \log n).$$

This leads to

$$\log p_n = \log n + \log \log n + O\left(\frac{\log \log n}{\log n}\right)$$

and

$$\log \log p_n = \log \log n + \frac{\log \log n}{\log n} + O\left(\frac{(\log \log n)^2}{\log^2 n}\right).$$

Thus, for  $n = m^k$ , we obtain the required formula.  $\square$

Lemma 2 immediately yields:

**Corollary 2.**

$$(4.3) \quad C_{2,m} = \sum_{k=1}^{\infty} \left( \log(\log p_{m^k}) - \log k - \log \log m - \frac{\log k}{k \log m} - \frac{\log \log m}{k \log m} \right)$$

converges.

Next,

**Lemma 3.**

$$(4.4) \quad \int_0^u \frac{N(t)}{t} dt = \frac{u}{\log m} - \frac{\log^2 u}{2 \log m} + \frac{1}{2} \log u + D_m + o(1),$$

where

$$D_m = \frac{(\log \log m)^2}{2 \log m} + \log \left( \frac{1}{\log 2} \sqrt{\frac{\log m}{2\pi}} \right) - C_{2,m} - \frac{\gamma_1}{\log m} - \frac{\log \log m}{\log m} \gamma.$$

*Proof.* We first notice that

$$p_{m^k} \leq e^u \text{ if and only if } k \leq y := \frac{\log(\pi(e^u))}{\log m},$$

where  $\pi(x)$  is the distribution of the prime numbers. By the prime number theorem,

$$y = \frac{u}{\log m} - \frac{\log u}{\log m} + O\left(\frac{1}{u}\right).$$

Thus  $N(u) = \lfloor y \rfloor + 1$  and, combining equations (4.1), (4.2) and (4.3),

$$\begin{aligned} \int_0^u \frac{N(t)}{t} dt &= (\lfloor y \rfloor + 1) \log u - \log \log 2 - \sum_{k=1}^{\lfloor y \rfloor} \log(\log p_{m^k}) \\ &= (\lfloor y \rfloor + 1) \log u - \log \log 2 - C_{2,m} - \log(\lfloor y \rfloor!) - (\log \log m) \lfloor y \rfloor \\ &\quad - \frac{1}{\log m} \sum_{k=1}^{\lfloor y \rfloor} \frac{\log k}{k} - \frac{\log \log m}{\log m} \sum_{k=1}^{\lfloor y \rfloor} \frac{1}{k} + o(1). \end{aligned}$$

Using Stirling's formula and the defining formulas for  $\gamma$  and  $\gamma_1$

$$\log(n!) = n \log n - n + \frac{1}{2} \log n + \log(\sqrt{2\pi}) + o(1),$$

$$\gamma = \sum_{k=1}^n \frac{1}{k} - \log n + o(1) \quad \text{and} \quad \gamma_1 = \sum_{k=1}^n \frac{\log k}{k} - \frac{\log^2 n}{2} + o(1),$$

we have

$$\begin{aligned} \int_0^u \frac{N(t)}{t} dt &= \lfloor y \rfloor \log \left( \frac{u}{\lfloor y \rfloor \log m} \right) + \log u + \lfloor y \rfloor - \frac{\log^2 \lfloor y \rfloor}{2 \log m} - \left( \frac{1}{2} + \frac{\log \log m}{\log m} \right) \log \lfloor y \rfloor \\ &\quad - \log \log 2 - C_{2,m} - \log(\sqrt{2\pi}) - \frac{\gamma_1}{\log m} - \frac{\log \log m}{\log m} \gamma + o(1). \end{aligned}$$

Since

$$\lfloor y \rfloor \log \left( \frac{u}{\lfloor y \rfloor \log m} \right) = \lfloor y \rfloor \log \left( 1 + \frac{u - \lfloor y \rfloor \log m}{\lfloor y \rfloor \log m} \right) = \frac{u - \lfloor y \rfloor \log m}{\log m} + O \left( \frac{\log^2 u}{u} \right)$$

and

$$\log \lfloor y \rfloor = \log(y + O(1)) = \log y + O \left( \frac{1}{y} \right) = \log u - \log \log m + O \left( \frac{\log u}{u} \right),$$

we obtain (4.4), as required.  $\square$

The asymptotic formula (1.5) follows by combining Lemma 3 and Corollary 1 after a straightforward calculation. The proof of Theorem 1 is complete.

## REFERENCES

- [1] R. de la Bretèche, G. Tenenbaum, *Sur certaines équations fonctionnelles arithmétiques*, Ann. Inst. Fourier (Grenoble) **50** (2000), 1445–1505.
- [2] S. N. Burris, *Number theoretic density and logical limit laws*, Mathematical Surveys and Monographs, 86, American Mathematical Society, Providence, RI, 2001.
- [3] M. Cipolla, *La determinazione assintotica dell'  $n^{\text{imo}}$  numero primo*, Napoli Rend. **3** (1902), 132–166.
- [4] E. Deutsch, *Rooted tree statistics from Matula numbers*, Discrete Appl. Math. **160** (2012), 2314–2322.
- [5] S. B. Elk, *A problem with the application of Matula's method of prime numbers and rooted trees for canonical nomenclatures of alkanes*, Graph theory notes (New York) **18** (1989), 40–43.
- [6] S. B. Elk, *A canonical ordering of polybenzenes and polymantanes using a prime number factorization technique*, J. Math. Chem. **4** (1990), 55–68.
- [7] S. B. Elk, *Expansion of Matula numbers to heteroatoms and to ring compounds*, J. Chem. Inf. Comput. Sci. **35** 1995, 233–236.
- [8] S. B. Elk, *The Structure-Nomenclature Cycle of Chemistry*, in: Mathematical Chemistry Monographs, vol. 11, Univ. Kragujevac, Kragujevac, 2011.
- [9] R. Estrada, R. P. Kanwal, *A distributional approach to asymptotics. Theory and applications*, Birkhäuser Boston, Inc., Boston, MA, 2002.
- [10] I. M. Gel'fand, G. E. Shilov, *Generalized functions. Vol. I: Properties and operations*, Academic Press, New York-London, 1964.
- [11] F. Göbel, *On a 1-1-correspondence between rooted trees and natural numbers*, J. Combin. Theory B **29** (1980), 141–143.

- [12] I. Gutman, A. Ivić, *Graphs with maximal and minimal Matula numbers*, Bull. Acad. Serbe des Sciences et des Arts Sér. Math. **18** (1994), 65–74.
- [13] I. Gutman, A. Ivić, *On Matula numbers*, Discrete Mathematics **150** (1996), 131–142.
- [14] I. Gutman, A. Ivić, S. B. Elk, *Matula numbers for coding chemical structures and some of their properties*, J. Serb. Chem. Soc. **58** (1993), 193–201.
- [15] I. Gutman, Y. N. Yeh, *Deducing properties of trees from their Matula numbers*, Publ. Inst. Math. (Beograd) (N.S.) **53**(67) (1993), 17–22.
- [16] G. H. Hardy, S. Ramanujan, *Asymptotic formulae in combinatory analysis*, Proc. London Math. Soc. **17** (1918), 75–115.
- [17] A. E. Ingham, *A Tauberian theorem for partitions*, Ann. of Math. (2) **42** (1941), 1075–1090.
- [18] E. E. Kohlbecker, *Weak asymptotic properties of partitions*, Trans. Amer. Math. Soc. **88** (1958), 346–365.
- [19] J. Korevaar, *Tauberian theory. A Century of developments*, Grundlehren der Mathematischen Wissenschaften, 329, Springer-Verlag, Berlin, 2004.
- [20] D. W. Matula, *A natural rooted tree enumeration by prime factorization*, Siam Rev. **10** (1968), 273.
- [21] S. Pilipović, B. Stanković, J. Vindas, *Asymptotic behavior of generalized functions*, Series on Analysis, Applications and Computation, 5, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2012.
- [22] K. Schütte, *Proof Theory*, Grundlehren der Mathematischen Wissenschaften, 225, Springer-Verlag, Berlin-New York, 1977.
- [23] A. S. Troelstra, H. Schwichtenberg, *Basic proof theory*, Cambridge Tracts in Theoretical Computer Science, 43, Cambridge University Press, Cambridge, 2000.
- [24] J. V. Uspensky, *Asymptotic expressions of numerical functions occurring in problems of partitions of numbers*, Bull. Acad. Sci. de Russie **14** (1920), 199–218.
- [25] A. Weiermann, *Analytic combinatorics of the transfinite: a unifying Tauberian perspective*, in: Logic Colloquium 2007, pp. 238–267, Lect. Notes Log., 35, Assoc. Symbol. Logic, La Jolla, CA, 2010.
- [26] A. Weiermann, A. R. Woods, *Some natural zero one laws for ordinals below  $\varepsilon_0$* , in: Cooper, S. Barry (ed.) et al., How the world computes. Turing centenary conference and 8th conference on computability in Europe, CiE 2012, Cambridge, UK, June 18–23, 2012. Proceedings. Berlin. Springer Lecture Notes in Computer Science 7318, pp. 723–732.

DEPARTMENT OF MATHEMATICS, GHENT UNIVERSITY, KRIJGSLAAN 281 GEBOUW S22, B 9000 GENT, BELGIUM

*E-mail address:* `hvernaev@cage.UGent.be`

*E-mail address:* `jvindas@cage.Ugent.be`

*E-mail address:* `weierman@cage.UGent.be`