



## MERTENS' PRIME PRODUCT FORMULA, DISSECTED

**Jared Duker Lichtman**

*Mathematical Institute, University of Oxford, Oxford, United Kingdom*  
 jared.d.lichtman@gmail.com

*Received: 1/13/21, Revised: 3/11/21, Accepted: 3/17/21, Published: 8/23/21*

### Abstract

In 1874, Mertens famously proved an asymptotic formula for the product of  $p/(p-1)$  over all primes  $p$  up to  $x$ . Observe that this product equals the reciprocal sum of all integers composed of prime factors up to  $x$ . It is natural to restrict such series to integers with a fixed number  $k$  of prime factors. In this article, we obtain formulae for these series for each  $k$ , which together dissect Mertens' original estimate. The proof is by elementary methods of a combinatorial flavor.

### 1. Introduction

We begin with the Euler-Mascheroni constant  $\gamma = 0.57721\dots$ , defined as the limit of the difference between the harmonic series up to  $x$  and  $\log x$ . The ubiquitous constant  $\gamma$  crops up in many contexts, notably, in the third of three results from a celebrated paper of Mertens [7] on the distribution of prime numbers.

As notation, throughout we write  $f(x) = O(g(x))$  and  $f(x) \ll g(x)$  to mean  $|f(x)/g(x)|$  is bounded, while  $f(x) \sim g(x)$  means  $\lim_{x \rightarrow \infty} f(x)/g(x) = 1$ . Also, let  $\log_2 x = \log \log x$ , and let  $p$  denote a prime number.

**Theorem 1.1** (Mertens, 1874). *There exists a constant  $\beta > 0$  for which*

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1), \quad \sum_{p \leq x} \frac{1}{p} = \log_2 x + \beta + O\left(\frac{1}{\log x}\right) \quad (1.1)$$

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right)^{-1} \sim e^\gamma \log x. \quad (1.2)$$

Here  $\beta = 0.26149\dots$  is Mertens' constant, which is known to satisfy

$$\beta - \gamma = \sum_p \left(\log\left(1 - \frac{1}{p}\right) + \frac{1}{p}\right) = - \sum_p \sum_{j \geq 2} \frac{p^{-j}}{j} = - \sum_{j \geq 2} \frac{Z(j)}{j}, \quad (1.3)$$

where  $Z(s) = \sum_p p^{-s}$  denotes the prime zeta function, for  $s > 1$ . See for instance Theorem 2.7 in [8, p.50].

Now by expanding Mertens' prime product in Equation (1.2), we have

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right)^{-1} = \prod_{p \leq x} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \dots\right) = \sum_{P^+(n) \leq x} \frac{1}{n} \tag{1.4}$$

where  $P^+(n)$  denotes the largest prime factor of  $n$ .

Consider "dissecting" the sum in Equation (1.4) according to the number of prime factors of  $n$  with multiplicity, denoted  $\Omega(n)$ . Our main result is an asymptotic formula for this dissected sum.

**Theorem 1.2.** *For each fixed  $k \geq 1$ , we have*

$$\sum_{\substack{\Omega(n)=k \\ P^+(n) \leq x}} \frac{1}{n} = \sum_{j=0}^k \frac{c_{k-j}}{j!} (\log_2 x + \beta)^j + O_k\left(\frac{(\log_2 x)^{k-1}}{\log x}\right) \tag{1.5}$$

where the sequence  $(c_k)_{k=0}^\infty$  is recursively defined by  $c_0 = 1$  and

$$c_k = \frac{1}{k} \sum_{j=2}^k c_{k-j} Z(j). \tag{1.6}$$

Theorem 1.2 may be viewed as a "dissection" of Mertens' prime product formula. Indeed, as shown later in Equation (2.7), the main term  $e^\gamma \log x$  in Equation (1.2) may be expressed as the series over all  $k \geq 1$  of the main terms in Equation (1.5) (i.e. the sum over  $j \leq k$ ). Here "dissection" is meant to highlight the formal compatibility of main terms. Whereas the estimate Equation (1.5) itself does not necessarily hold uniformly over all  $k \geq 1$ .

We note this terminology was introduced by Pollack [9], who dissected a classical mean value theorem of Hall and Tenenbaum.

### 1.1. Uniform Estimates via Complex Analysis

Classically, the analogous series to Equation (1.5) has been studied, replacing the condition  $P^+(n) \leq x$  with the more common  $n \leq x$ .

Mertens' 1<sup>st</sup> theorem implies, by induction on each fixed  $k \geq 1$ ,

$$\sum_{\substack{\Omega(n)=k \\ n \leq x}} \frac{1}{n} \sim \frac{(\log_2 x)^k}{k!} \tag{1.7}$$

as  $x \rightarrow \infty$ , see [8, p.228]. Note Equation (1.7) historically attributed to Landau [5]. This is another example of dissection, as the sum over all  $k$  of each side gives  $\sum_{n \leq x} \frac{1}{n}$  and  $\log x$ , respectively. We also note that the asymptotic (1.7) also holds with  $\Omega(n)$  replaced by  $\omega(n)$ , the number of distinct prime factors of  $n$ .

However, Equation (1.7) only holds for fixed  $k$ . The celebrated theorem of Sathe and Selberg implies the following uniform estimate for  $k$  less than  $2 \log_2 x$ .

**Theorem 1.3** (Sathé–Selberg). Define  $\nu(z) = \frac{1}{\Gamma(z+1)} \prod_p (1 - \frac{z}{p})^{-1} (1 - \frac{1}{p})^z$ , and let  $r = k/\log_2 x$ . For any  $\varepsilon > 0$ , as  $x \rightarrow \infty$  we have uniformly for  $r \leq 2 - \varepsilon$ ,

$$\sum_{\substack{\Omega(n)=k \\ n \leq x}} \frac{1}{n} \sim \nu(r) \frac{(\log_2 x)^k}{k!}. \tag{1.8}$$

To see this, [8, Theorem 7.19] or [13, Theorem 6.5] gives an asymptotic in the stated range

$$\sum_{\substack{\Omega(n)=k \\ n \leq x}} 1 = \nu(r) \frac{x}{\log x} \frac{(\log_2 x)^{k-1}}{(k-1)!} \left(1 + O_\varepsilon\left(\frac{k}{\log_2 x}\right)\right).$$

Then Equation (1.8) follows by partial summation, combined with e.g. the Erdős–Sarközy upper bound  $O(k^4 2^{-k} x \log x)$  uniformly for all  $x, k \geq 1$ , see [2].

**Remark 1.** As  $\nu(r) = 1$  only when  $r = 0, 1$ , Landau’s estimate (1.7) holds if and only if  $k = o(\log_2 x)$  or  $k = (1 + o(1)) \log_2 x$ .

**Remark 2.** [13, Theorem 6.4] gives an analogous result with  $\Omega(n)$  replaced by  $\omega(n)$ , by substituting the function  $\nu(z)$  above with  $\lambda(z) = \frac{1}{\Gamma(z+1)} \prod_p (1 + \frac{z}{p-1})(1 - \frac{1}{p})^z$ .

**Remark 3.** The Sathé–Selberg theorem is proved through contour integration in the complex plane. Recently, Popa [10, 11] and Tenenbaum [12] have obtained results by similar analytic methods, for a generalized series that replaces the conditions  $\Omega(n) = k$  and  $n \leq x$  by the condition  $p_1 \cdots p_k \leq x$  over  $k$  independent prime variables. Or equivalently, they weight  $n$  by the number of its ordered prime factorizations.

The uniformity coming from sophisticated analytic tools exemplifies the larger tension within mathematics, between proving the strongest results and using the simplest arguments. Of particular interest historically is the case  $k = 1$ , i.e. the Prime Number Theorem. Hadamard and de la Vallée Poussin initially gave proofs in 1896 using complex analysis, and for decades many believed it impossible to prove by elementary means. It came as a great shock when Selberg and Erdős did so in 1948. For an intriguing historical account, see Spencer and Graham [4].

As such, we emphasize that in Theorem 1.2, our particular conditions  $\Omega(n) = k, P^+(n) \leq x$  in Equation (1.5) are directly amenable to elementary methods when  $k$  is fixed. Nevertheless, applying analytic tools to Equation (1.5) do lend the advantage of uniformity in  $k < (2 - \varepsilon) \log_2 x$ .

**Theorem 1.4.** Let  $r = k/\log_2 x$ . For any  $\varepsilon > 0$ , as  $x \rightarrow \infty$  we have uniformly for  $r \leq 2 - \varepsilon$ ,

$$\sum_{\substack{\Omega(n)=k \\ P^+(n) \leq x}} \frac{1}{n} \sim \nu(r) e^{r\gamma} \Gamma(r+1) \frac{(\log_2 x)^k}{k!}. \tag{1.9}$$

Hence by comparison with the Sathe–Selberg theorem, we obtain the following elegant relation between sums over  $P^+(n) \leq x$  with those over  $n \leq x$ .

**Corollary 1.1.** *Let  $r = k/\log_2 x$ . For any  $\varepsilon > 0$ , as  $x \rightarrow \infty$  we have uniformly for  $r \leq 2 - \varepsilon$ ,*

$$\sum_{\substack{\Omega(n)=k \\ P^+(n) \leq x}} \frac{1}{n} \sim e^{r\gamma} \Gamma(r+1) \sum_{\substack{\Omega(n)=k \\ n \leq x}} \frac{1}{n}. \tag{1.10}$$

**Remark 4.** One may prove an analogous result for  $\omega(n)$ , with the same factor  $e^{r\gamma} \Gamma(r+1)$ .

Note the factor  $e^{r\gamma} \Gamma(r+1) = 1$  if and only if  $r = 0$ . Hence Corollary 1.1 implies

$$\sum_{\substack{\Omega(n)=k \\ P^+(n) \leq x}} \frac{1}{n} \sim \sum_{\substack{\Omega(n)=k \\ n \leq x}} \frac{1}{n} \tag{1.11}$$

if and only if  $k$  is in the uniform range  $k = o(\log_2 x)$ . This is an example of friable regularity, in the following sense. Recall an integer  $n$  with  $P^+(n) \leq x$  is called  $x$ -smooth or  $x$ -friable.

**Definition 1.** A sequence  $(a_n)_{n \in \mathbb{N}}$  is *friably regular*<sup>1</sup> if  $\sum_{n \leq x} a_n \sim \sum_{P^+(n) \leq x} a_n$  as  $x \rightarrow \infty$ .

For example, the friable regularity of  $(\mu(n)/n)_{n \in \mathbb{N}}$  is equivalent to the prime number theorem. We also extend the definition to families of sequences.

**Definition 2.** A one-parameter family  $(a_{n,x})_{n \in \mathbb{N}}$ , indexed by  $x \in \mathbb{R}$ , is *friably regular* if  $\sum_{n \leq x} a_{n,x} \sim \sum_{P^+(n) \leq x} a_{n,x}$  as  $x \rightarrow \infty$ .

In particular, Corollary 1.1 implies the family  $(\mathbf{1}_{\Omega(n)=k}/n)_{n \in \mathbb{N}}$ , indexed by  $k = k(x)$ , is friably regular if and only if  $k = o(\log_2 x)$ .

### 1.2. The Coefficients $c_k$

Finally, we emphasize an important feature of the combinatorial approach in Theorem 1.2. The recursion in Equation (1.6) enables rapid computation of the coefficients  $c_k$  to high precision, the first few displayed below.

$k$	$c_k$	$k$	$c_k$
0, 1	1, 0	6	0.0108213...
2	0.226123...	7	0.0054110...
3	0.058254...	8	0.0027375...
4	0.044814...	9	0.0013752...
5	0.020323...	10	0.0006903...

---

<sup>1</sup>This extends the notion of friable regularity as in [1],[3], from equality of limits of convergent series to asymptotic equality of (possibly non-convergent) partial sums.

At first glance, one might not expect the coefficients  $c_k$  arising from Equation (1.6) to exhibit any particular structure. However, the combinatorial approach shows  $c_k$  to satisfy exponentially precise asymptotics.

**Theorem 1.5.** *The coefficients satisfy  $c_k = \eta 2^{-k} + O(3^{-k})$ . Here the constant  $\eta$  is given by  $\eta = e^{-1} \prod_{p>2} (1 - \frac{2}{p})^{-1} e^{-2/p} = 0.71206\dots$ .*

**2. Elementary Combinatorial Proof for  $k$  Fixed**

In this section we prove Theorem 1.2. For  $x, s > 0$  define the (truncated) zeta functions

$$Z_k(s, x) = \sum_{\substack{\Omega(n)=k \\ P^+(n)\leq x}} n^{-s}, \quad Z(s, x) = Z_1(s, x) = \sum_{p\leq x} p^{-s}.$$

We first express  $Z_k(s, x)$  in terms of  $Z(s, x)$ .

**Proposition 2.1.** *For each  $k \geq 1$  and any  $x, s > 0$  we have the identity*

$$Z_k(s, x) = \sum_{n_1+2n_2+\dots=k} \prod_{j\geq 1} \frac{1}{n_j!} (Z(js, x)/j)^{n_j} \tag{2.1}$$

where the sum ranges over all partitions of  $k$ .

*Proof.* For any  $x, s > 0$  we have a formal power series identity in  $z$ ,

$$\sum_{k\geq 0} Z_k(s, x) z^k = \sum_{P^+(n)\leq x} \frac{z^{\Omega(n)}}{n^s} = \prod_{p\leq x} \left(1 + \frac{z}{p^s} + \frac{z^2}{p^{2s}} + \dots\right) = \prod_{p\leq x} \left(1 - \frac{z}{p^s}\right)^{-1}$$

since the function  $n \mapsto z^{\Omega(n)}/n^s$  is completely multiplicative. Thus expanding Taylor series,

$$\begin{aligned} \sum_{k\geq 0} Z_k(s, x) z^k &= \exp\left(-\sum_{p\leq x} \log(1 - zp^{-s})\right) = \exp\left(\sum_{p\leq x} \sum_{j\geq 1} \frac{(zp^{-s})^j}{j}\right) \\ &= \exp\left(\sum_{j\geq 1} \frac{Z(js, x)}{j} z^j\right) = \prod_{j\geq 1} \exp\left(\frac{Z(js, x)}{j} z^j\right) \\ &= \prod_{j\geq 1} \sum_{n_j\geq 0} \frac{1}{n_j!} \left(\frac{Z(js, x)}{j} z^j\right)^{n_j} \\ &= \sum_{k\geq 0} z^k \sum_{n_1+2n_2+\dots=k} \prod_{j\geq 1} \frac{1}{n_j!} (Z(js, x)/j)^{n_j}. \end{aligned} \tag{2.2}$$

Now Equation (2.1) follows by comparing the coefficients of  $z^k$ . □

**Remark 5.** This proposition generalizes [6, Proposition 3.1].

Next, the recursion for  $c_k$  in Equation (1.6) leads to the explicit formula,

$$c_k = \sum_{2n_2+3n_3+\dots=k} \prod_{j \geq 2} \frac{1}{n_j!} (Z(j)/j)^{n_j} \tag{2.3}$$

by the following lemma, for the choices  $A_1 = 0$  and  $A_j = Z(j)$  when  $j \geq 2$ .

**Lemma 2.1.** *Given any sequence  $(A_k)_{k=1}^\infty$ , the sequence  $(b_k)_{k=0}^\infty$  is given recursively by  $b_0 = 1$  and  $b_k = \frac{1}{k} \sum_{j=1}^k b_{k-j} A_j$ , if and only if  $(b_k)_{k=0}^\infty$  is given explicitly as*

$$b_k = \sum_{n_1+2n_2+\dots=k} \prod_{j \geq 1} \frac{(A_j/j)^{n_j}}{n_j!}.$$

Note the (unique) partition of  $k = 0$  has  $n_j = 0$  for all  $j \geq 1$ , so indeed  $b_0 = \prod_{j \geq 1} (A_j/j)^0/0! = 1$ .

*Proof.* We prove the forward direction by induction on  $k$  (the reverse direction is similar). For  $k = 1$ , we have  $b_1 = b_0 A_1 = A_1$ .

Then assuming the claim for each  $r < k$ ,

$$\begin{aligned} kb_k &= \sum_{r=1}^k b_{k-r} A_r = \sum_{r=1}^k A_r \sum_{n_1+\dots=k-r} \prod_{j \geq 1} \frac{(A_j/j)^{n_j}}{n_j!} \\ &= \sum_{r=1}^k \sum_{n_1+\dots=k-r} \frac{A_r^{n_r+1}}{r^{n_r} n_r!} \prod_{j \neq r} \frac{(A_j/j)^{n_j}}{n_j!} = \sum_{r=1}^k \sum_{\substack{n_1+\dots=k \\ n_r \geq 1}} r n_r \prod_{j \geq 1} \frac{(A_j/j)^{n_j}}{n_j!} \\ &= \sum_{n_1+\dots=k} \prod_{j \geq 1} \frac{(A_j/j)^{n_j}}{n_j!} \sum_{\substack{1 \leq r \leq k \\ n_r \geq 1}} r n_r = k \sum_{n_1+\dots=k} \prod_{j \geq 1} \frac{(A_j/j)^{n_j}}{n_j!} \end{aligned}$$

In the last step, we dropped the condition  $n_r \geq 1$  (since  $r n_r = 0$  for  $n_r = 0$ ) which gives  $\sum_{r=1}^k r n_r = k$ . Dividing by  $k$  completes the induction.  $\square$

Now equipped with Proposition 2.1 and Equation (2.3) for  $c_k$ , we now prove Theorem 1.2.

*Proof of Theorem 1.2.* For  $Z(j, x)$  with  $j \geq 2$ , we trivially bound  $\sum_{p > x} p^{-j}$  by  $x^{2-j} \sum_{n > x} n^{-2} = O(x^{1-j})$ , which gives

$$Z(j, x) := \sum_{p \leq x} p^{-j} = Z(j) - \sum_{p > x} p^{-j} = Z(j) + O(x^{1-j}) \quad \text{for } j \geq 2.$$

Thus plugging into the identity for  $Z_k(1, x)$ , Proposition 2.1 with  $s = 1$  gives

$$Z_k(1, x) = \sum_{n_1+2n_2+\dots=k} \frac{Z(1, x)^{n_1}}{n_1!} \prod_{j \geq 2} \frac{1}{n_j!} \left( \frac{Z(j) + O(x^{1-j})}{j} \right)^{n_j}.$$

For any partition of  $k$ , the binomial theorem implies  $\prod_{j \geq 2} \frac{1}{n_j!} ([Z(j) + O(x^{1-j})]/j)^{n_j}$  equals  $\prod_{j \geq 2} \frac{1}{n_j!} (Z(j)/j)^{n_j}$  at negligible cost  $O_k(1/x)$ . Thus

$$Z_k(1, x) = \sum_{n_1+2n_2+\dots=k} \frac{Z(1, x)^{n_1}}{n_1!} \prod_{j \geq 2} \frac{1}{n_j!} (Z(j)/j)^{n_j} + O_k\left(\frac{Z(1, x)^k}{x}\right). \tag{2.4}$$

Then for  $Z(1, x)$ , we recall Mertens' 2<sup>nd</sup> theorem

$$Z(1, x) := \sum_{p \leq x} \frac{1}{p} = \log_2 x + \beta + E(x), \quad \text{with } E(x) = O\left(\frac{1}{\log x}\right) \tag{2.5}$$

so plugging in above gives

$$\begin{aligned} Z_k(1, x) & \tag{2.6} \\ &= \sum_{n_1+2n_2+\dots=k} \frac{1}{n_1!} \left( \log_2 x + \beta + E(x) \right)^{n_1} \prod_{j \geq 2} \frac{1}{n_j!} (Z(j)/j)^{n_j} + O_k\left(\frac{(\log_2 x)^k}{x}\right) \\ &= \sum_{n_1=0}^k \frac{1}{n_1!} (\log_2 x + \beta)^{n_1} \sum_{2n_2+\dots=k-n_1} \prod_{j \geq 2} \frac{1}{n_j!} (Z(j)/j)^{n_j} + O_k(E(x) (\log_2 x)^{k-1}) \end{aligned}$$

again by the binomial theorem. Here we used  $\prod_{j \geq 2} \frac{1}{n_j!} (Z(j)/j)^{n_j} = O_k(1)$ .

Now recalling Equation (2.3) and  $E(x) = O(1/\log x)$  completes the proof of Theorem 1.2. □

From here, we may “dissect” Mertens' 3<sup>rd</sup> theorem. Indeed by Equation (1.4),

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right)^{-1} = \sum_{P^+(n) \leq x} \frac{1}{n} = \sum_{k \geq 0} Z_k(1, x)$$

and using the asymptotic formula for  $Z_k(1, x)$  from Theorem 1.2,

$$\begin{aligned} \sum_{k \geq 0} \sum_{j=0}^k \frac{c_{k-j}}{j!} (\log_2 x + \beta)^j &= \sum_{j \geq 0} \frac{1}{j!} (\log_2 x + \beta)^j \sum_{k \geq j} c_{k-j} = e^\beta \log x \sum_{m \geq 0} c_m \\ &= e^\gamma \log x, \end{aligned} \tag{2.7}$$

as desired, provided  $\sum_{m \geq 0} c_m = e^{\gamma-\beta}$ . This follows in turn by Equation (2.3),

$$\begin{aligned} \sum_{m \geq 0} c_m &= \sum_{m \geq 0} \sum_{2n_2+3n_3+\dots=m} \prod_{j \geq 2} \frac{1}{n_j!} (Z(j)/j)^{n_j} = \prod_{j \geq 2} \sum_{n_j \geq 0} \frac{1}{n_j!} (Z(j)/j)^{n_j} \\ &= \prod_{j \geq 2} \exp(Z(j)/j) = \exp\left(\sum_{j \geq 2} \frac{Z(j)}{j}\right) = e^{\gamma-\beta} \end{aligned}$$

recalling Equation (1.3). This shows the claim.

**3. Combinatorial Proof of Asymptotics for Coefficients  $c_k$**

In this section we prove a strengthening of Theorem 1.5. To this, we first rephrase the recursion for  $c_k$  in Equation (1.6).

Let  $A_1 = 0$  and  $A_k = \sum_p p^{-k}$  for  $k \geq 2$ . Then  $c_k$  is recursively defined by  $c_0 = 1$  and

$$kc_k = \sum_{j=1}^k c_{k-j}A_j. \tag{3.1}$$

Consider the following induced sequences  $A_{k,q}, c_{k,q}$  for each prime  $q$ : let  $A_{k,2} = A_k, c_{k,2} = c_k$ ; and if  $p$  is the prime preceding  $q > 2$ , let

$$A_{k,q} = A_{k,p} - p^{-k} \quad \text{for } k \geq 1, \tag{3.2}$$

$$c_{k,q} = c_{k,p} - p^{-1}c_{k-1,p} \quad \text{for } k \geq 1, \quad \text{and } c_{0,q} = c_0. \tag{3.3}$$

Explicitly we have

$$A_{k,q} = \sum_{r \geq q} r^{-k} \quad \text{for } k \geq 2, \quad \text{and } A_{1,q} = -\sum_{p < q} p^{-1}. \tag{3.4}$$

**Lemma 3.1.** *For each prime  $q$  and  $k \geq 0$ , we have the recursion*

$$kc_{k,q} = \sum_{j=1}^k c_{k-j,q}A_{j,q}. \tag{3.5}$$

*Proof.* We proceed by induction on the prime  $q$ . The base case  $q = 2$  holds by Equation (3.1).

Now assume Equation (3.5) for  $p < q$ . The difference of recursions in Equation



(3.5) for  $c_{k,p}$  and  $p^{-1} \cdot c_{p,k-1}$  is

$$\begin{aligned}
 kc_{k,p} - (k-1)p^{-1}c_{k-1,p} &= \sum_{j=1}^{k-1} (c_{k-j,p} - p^{-1}c_{k-j-1,p})A_{j,p} + c_{0,p}A_{k,p} \\
 &= \sum_{j=1}^{k-1} (c_{k-j,p} - p^{-1}c_{k-j-1,p})(p^{-j} + A_{j,q}) + c_{0,p}(p^{-k} + A_{k,q}) \\
 &= \sum_{j=1}^{k-1} (p^{-j}c_{k-j,p} - p^{-j-1}c_{k-j-1,p}) + c_{0,p}p^{-k} + \sum_{j=1}^{k-1} c_{k-j,q}A_{j,q} + c_{0,q}A_{k,q} \\
 &= p^{-1}c_{k-1,p} + \sum_{j=1}^k c_{k-j,q}A_{j,q}
 \end{aligned}$$

using Equations (3.2), (3.3) and telescoping series. Subtracting  $p^{-1}c_{k-1,p}$  gives

$$kc_{k,q} = k(c_{k,p} - p^{-1}c_{k-1,p}) = \sum_{j=1}^k c_{k-j,q}A_{j,q}.$$

□

Note Lemmas 2.1 and 3.1 together imply

$$c_{k,q} = \sum_{n_1+2n_2+\dots+kj \geq 1} \prod_{j \geq 1} \frac{(A_{j,q}/j)^{n_j}}{n_j!} \tag{3.6}$$

for each prime  $q$ ,  $k \geq 1$ .

Now with the recursion in hand, we bound the induced sequence  $c_{k,q}$ .

**Lemma 3.2.** *For each prime  $q$ , we have  $c_{k,q} \ll_q q^{-k}$  as  $k \rightarrow \infty$ .*

*Proof.* Fix  $q$  and let  $m_k = \max_{j \leq k} q^j |c_{j,q}|$ . We shall prove  $m_k \ll_q 1$ , and it suffices to show this along a subsequence, since  $m_k$  is itself a non-decreasing sequence. Namely, we consider the indices  $k$  for which  $m_k = q^k |c_{k,q}|$ .

Recalling (3.4), we have for all  $n \geq 1$

$$\sum_{1 \leq j \leq n} q^j A_{j,q} = qA_{1,q} + (n-1) + \sum_{2 \leq j \leq n} \sum_{r > q} (q/r)^j = n + O_q(1),$$

by summing the geometric series, and so the recursion in Equation (3.5) gives

$$\begin{aligned}
 kq^k |c_{k,q}| &= \left| \sum_{j=1}^k q^{k-j} c_{k-j,q} \cdot q^j A_{j,q} \right| \leq m_{k/2} \left| \sum_{k/2 < j \leq k} q^j A_{j,q} \right| + m_k \left| \sum_{1 \leq j \leq k/2} q^j A_{j,q} \right| \\
 &= m_{k/2}(k/2 + O(1)) + m_k(k/2 + O(1)).
 \end{aligned}$$

And by our choice of  $k$ , we have  $m_k = q^k |c_{k,q}|$  and so

$$m_k \leq m_{k/2} (1 + O(1/k)). \tag{3.7}$$

Hence by induction on  $k$ , we conclude

$$m_k \ll m_1 \prod_{2^i \leq k} (1 + O(2^{-i})) \ll \exp \sum_{2^i \leq k} O(2^{-i}) \ll 1.$$

□

Since Lemma 3.2 holds for every prime  $q$  and the sequences  $c_{k,q}$  are defined inductively on primes, Lemma 3.2 is self-improving. Indeed, for each pair of consecutive primes  $p < q$ ,

$$c_{k,p} - p^{-1} c_{k-1,p} = c_{k,q} = O(q^{-k}).$$

In other words, multiplying above by  $p^k$  the modified sequence  $c'_{k,p} := c_{k,p} p^k$  satisfies  $c'_{k,p} - c'_{k-1,p} = O((p/q)^k)$ , so  $(c'_{k,p})_{k \geq 1}$  is a Cauchy sequence for each prime  $p$ . Hence the limit

$$\eta_p := \lim_{k \rightarrow \infty} c'_{k,p} = \lim_{k \rightarrow \infty} c_{k,p} p^k$$

exists with  $c'_{k,p} = \eta_p + O((p/q)^k)$ . That is,

$$c_{k,p} = \eta_p p^{-k} + O_p(q^{-k}) \quad \text{for each prime } p. \tag{3.8}$$

To summarize, we expanded the definition of  $c_{k,q}$  and used a zeroth order expansion for each prime (Lemma 3.2) to prove a first order expansion for every prime simultaneously.

Continuing in this way, we obtain a  $h$ th order expansion for  $c_{k,q}$  by induction on the order  $h \geq 1$ , at each step proving the respective expansion for every prime simultaneously.

**Proposition 3.1.** *For any  $h \geq 1$ , we have*

$$c_{k,p_n} = \sum_{l=0}^{h-1} \eta_{p_{n+l}}^{(n)} p_{n+l}^{-k} + O_{n,h}(p_{n+h}^{-k}), \quad \text{where } \eta_{p_{n+l}}^{(n)} = \eta_{p_{n+l}} / \prod_{i=0}^{l-1} \left(1 - \frac{p_{n+i}}{p_{n+l}}\right) \tag{3.9}$$

for all  $n$  as  $k \rightarrow \infty$ . Here  $p_n$  denotes the  $n$ th prime, and  $\eta_p = \lim_{k \rightarrow \infty} c_{k,p} p^k$  as in Equation (3.8).

*Proof.* We proceed by induction on  $h$ . The base  $h = 1$  holds for all  $n$  by Equation (3.8), since  $\eta_{p_n}^{(n)} = \eta_{p_n}$ . Now assume Equation (3.9) holds with  $h$  for every  $n$ , and

write  $c_{k,p_n} = \sum_{l=0}^h \eta_{p_{n+l}}^{(n)} p_{n+l}^{-k} + E_{k,n,h}$ . By assumption  $E_{k,n,h} \ll p_{n+h}^{-k}$ , and we aim to show  $E_{k,n,h} \ll p_{n+1+h}^{-k}$ .

By Equation (3.3) and the induction hypothesis (3.9) for  $c_{k,p_{n+1}}$ ,

$$c_{k,p_{n+1}} = c_{k,p_n} - p_n^{-1} c_{k-1,p_n} \\ \sum_{l=0}^{h-1} \eta_{p_{n+1+l}}^{(n+1)} p_{n+1+l}^{-k} + O(p_{n+1+h}^{-k}) = \sum_{l=0}^h \eta_{p_{n+l}}^{(n)} \left(1 - \frac{p_{n+l}}{p_n}\right) p_{n+l}^{-k} + (E_{k,n,h} - p_n^{-1} E_{k-1,n,h})$$

Note by definition  $\eta_{p_{n+l}}^{(n+1)} = \eta_{p_{n+l}}^{(n)} \left(1 - \frac{p_{n+l}}{p_n}\right)$ , and so the above simplifies as

$$O(p_{n+1+h}^{-k}) = E_{k,n,h} - p_n^{-1} E_{k-1,n,h}.$$

Thus, similarly as with Equation (3.8), the modified sequence  $E'_{k,n,h} := E_{k,n,h} p_n^k$  converges as  $k \rightarrow \infty$  to some limit  $\ell_{n,h}$ , with  $E'_{k,n,h} = \ell_{n,h} + O((p_n/p_{n+1+h})^k)$ . That is,

$$E_{k,n,h} = \ell_{n,h} p_n^{-k} + O(p_{n+1+h}^{-k})$$

On the other hand,  $E_{k,n,h} \ll p_{n+h}^{-k}$  forces  $\ell_{n,h} = 0$ . Hence  $E_{k,n,h} = O(p_{n+1+h}^{-k})$  as desired.  $\square$

Next, we determine the expansion coefficients  $\eta_p$  from Equation (3.8).

**Proposition 3.2.** *For any prime  $p$ , the coefficient  $\eta_p = \lim_{k \rightarrow \infty} c_{k,p} p^k$  equals*

$$\eta_p = e^{-\sum_{q \leq p} p/q} \prod_{q > p} \left(1 - \frac{p}{q}\right)^{-1} e^{-p/q}. \tag{3.10}$$

*Proof.* Consider the generating function  $C_p(z) = \sum_{k \geq 0} c_{k,p} z^k$ . On one hand, the explicit formula for  $c_{k,p}$  in Equation (3.6) implies

$$C_p(z) = \sum_{k \geq 0} c_{k,p} z^k = \sum_{k \geq 0} z^k \sum_{n_1+2n_2 \dots = k} \prod_{j \geq 1} \frac{(A_{j,p}/j)^{n_j}}{n_j!} = \prod_{j \geq 1} \sum_{n_j \geq 0} \frac{1}{n_j!} (A_{j,p} z^j / j)^{n_j} \\ = \prod_{j \geq 1} \exp(A_{j,p} z^j / j) = \exp\left(\sum_{j \geq 1} A_{j,p} z^j / j\right).$$

Then recalling  $A_{j,p} = \sum_{q \geq p} q^{-j}$  for  $j \geq 2$ ,

$$C_p(z) = \exp\left(z A_{1,p} + \sum_{q \geq p} \sum_{j \geq 2} (z/q)^j / j\right) = e^{z A_{1,p}} \exp\left(-\sum_{q \geq p} [\log(1 - z/q) + z/q]\right) \\ = e^{z A_{1,p}} \prod_{q \geq p} (1 - z/q)^{-1} e^{-z/q}. \tag{3.11}$$

On the other hand, by the expansion for  $c_k$  in Equation (3.8) we have

$$C_p(z) = \sum_{k \geq 0} c_k z^k = \eta_p \sum_{k \geq 0} (z/p)^k + O_p\left(\sum_{k \geq 0} (z/q)^k\right) = \frac{\eta_p}{1 - z/p} + \frac{O_p(1)}{1 - z/q}. \tag{3.12}$$

since  $A_{1,p} = -\sum_{q < p} q^{-1}$ . So comparing  $C_p(z)$  from Equations (3.11) and (3.12) at the pole  $z = p$ ,

$$\eta_p = \lim_{z \rightarrow p} C_p(z)(1 - z/p) = e^{pA_{1,p-1}} \prod_{q > p} (1 - p/q)^{-1} e^{-p/q}.$$

Hence the result follows since  $A_{1,p} = -\sum_{q < p} q^{-1}$ . □

Finally, we obtain an expansion for the original sequence  $c_{k,p_1} = c_k$  to arbitrary order, which gives a considerable refinement of Theorem 1.5.

**Theorem 3.1.** *For each prime  $q$ ,*

$$c_k = \sum_{p < q} \alpha_p p^{-k} + O_q(q^{-k})$$

where  $\alpha_p := e^{-1} \prod_{q \neq p} (1 - \frac{p}{q})^{-1} e^{-p/q}$ . In particular  $c_k = \alpha_2 2^{-k} + O(3^{-k})$ .

*Proof.* Setting  $n = 1$  in Proposition 3.1, the sequence  $c_{k,p_1} = c_k$  satisfies

$$c_k = \sum_{p < q} \eta_p^{(1)} p^{-k} + O_q(q^{-k})$$

where Proposition 3.2 gives, by definition of  $\eta_p^{(1)}$  in Equation (3.9),

$$\eta_p^{(1)} := \eta_p / \prod_{q < p} (1 - \frac{p}{q}) = e^{-1} \prod_{q \neq p} (1 - \frac{p}{q})^{-1} e^{-p/q} = \alpha_p.$$

□

#### 4. Analytic Proof for $k$ in Uniform Range

We prove Theorem 1.4 which quantitative error, which we state below.

**Theorem 4.1.** *Let  $r = k/\log_2 x$  and define  $\eta(z) = e^{\gamma z} \prod_p (1 - \frac{1}{p})^z (1 - \frac{z}{p})^{-1}$ . For any  $\varepsilon > 0$ , as  $x \rightarrow \infty$  we have uniformly for  $r \leq 2 - \varepsilon$ ,*

$$\sum_{\substack{\Omega(n)=k \\ P^+(n) \leq x}} \frac{1}{n} = \eta(r) \frac{(\log_2 x)^k}{k!} \left(1 + O_\varepsilon\left(\frac{k}{(\log_2 x)^2}\right)\right).$$

*Proof.* By Cauchy’s residue formula, we have for any  $r < 2$ ,

$$Z_k(1, x) = \frac{1}{2\pi i} \int_{|z|=r} f_x(z) \frac{dz}{z^{k+1}}, \tag{4.1}$$

where  $f_x$  is given by the power series

$$\begin{aligned} f_x(z) &= \sum_{k \geq 0} Z_k(1, x) z^k = \sum_{P^+(n) \leq x} \frac{z^{\Omega(n)}}{n} \\ &= \prod_{p \leq x} \left( 1 + \frac{z}{p} + \frac{z^2}{p^2} + \dots \right) = \prod_{p \leq x} \left( 1 - \frac{z}{p} \right)^{-1} \\ &= (1 + O(E(x))) e^{z\gamma} (\log x)^z \prod_{p \leq x} \left( 1 - \frac{z}{p} \right)^{-1} \left( 1 - \frac{1}{p} \right)^z \\ &= (1 + O(E(x))) \eta(z) (\log x)^z, \end{aligned}$$

as  $\prod_{p \leq x} (1 - \frac{1}{p})^{-1} = (1 + E(x)) e^\gamma \log x$  by Merten’s 2<sup>nd</sup> theorem in quantitative form (this also follows from the prime number theorem.)

Hence Equation (4.1) becomes

$$Z_k(1, x) = \frac{1 + O(E(x))}{2\pi i} \int_{|z|=r} \eta(z) (\log x)^z \frac{dz}{z^{k+1}}. \tag{4.2}$$

The desired main term in Theorem 1.4 is given by evaluating  $\eta(z)$  at  $z = r$ , namely

$$\frac{\eta(r)}{2\pi i} \int_{|z|=r} (\log x)^z \frac{dz}{z^{k+1}} = \eta(r) \frac{(\log_2 x)^k}{k!}. \tag{4.3}$$

For the error we follow the argument in [8, p. 233], which we provide for completeness. Recall  $E(x) \ll 1/\log x$ . For  $|z| = r = k/\log_2 x$ , integration by parts gives

$$\frac{1}{2\pi i} \int_{|z|=r} (z - r) (\log x)^z \frac{dz}{z^{k+1}} = \frac{(\log_2 x)^{k-1}}{(k-1)!} - \frac{r(\log_2 x)^k}{k!} = 0, \tag{4.4}$$

and 
$$\eta(z) - \eta(r) - \eta'(r)(z - r) = \int_r^z (z - w) \eta''(w) dw \ll |z - r|^2. \tag{4.5}$$

Thus subtracting Equations (4.3) from (4.2), the error is

$$\begin{aligned} \eta(r) \frac{(\log_2 x)^k}{k!} - Z_k(1, x) &\ll \int_{|z|=r} [\eta(r) - \eta(z)] (\log x)^z \frac{dz}{z^{k+1}} \\ &\stackrel{(4.4)}{=} \int_{|z|=r} [\eta(r) - \eta(z) - \eta'(r)(z - r)] (\log x)^z \frac{dz}{z^{k+1}} \\ &\stackrel{(4.5)}{\ll} \int_{|z|=r} |z - r|^2 (\log x)^z \frac{dz}{z^{k+1}} \\ &\ll r^{2-k} \int_{-1/2}^{1/2} (\sin \pi \theta)^2 e^{k \cos(2\pi \theta)} d\theta \\ &\ll r^{2-k} e^k \int_0^\infty \theta^2 e^{-8k\theta^2} d\theta \ll r^{2-k} e^k k^{-3/2} \\ &= (\log_2 x)^{k-2} (e/k)^k k^{1/2} \ll k (\log_2 x)^{k-2} / k! \end{aligned}$$

Here we used  $|\sin x| \leq x$ ,  $\cos(2\pi\theta) \leq 1 - 8\theta^2$  for  $|\theta| \leq 1/2$ , and Stirling's formula.  $\square$

**Acknowledgments.** The author is grateful to Paul Kinlaw and James Maynard for stimulating conversations, as well as to Paul Pollack and Carl Pomerance for valuable feedback. In addition, the author thanks the anonymous referee for comments to clarify the paper. The author is supported by a Clarendon Scholarship at the University of Oxford.

**References**

[1] R. J. Duffin, Representation of Fourier integrals as sums, III, *Proc. Amer. Math. Soc.* **8** (1957), 272–277.  
 [2] P. Erdős and A. Sárközy, On the number of prime factors of integers, *Acta Sci. Math.* **42** (1980), 237–246.  
 [3] E. Fouvry and G. Tenenbaum, Entiers sans grand facteur première en progressions arithmétiques, *Proc. London Math. Soc.* **63** (1991), 449–494.  
 [4] J. Spencer and R. Graham, The elementary proof of the prime number theorem, *Math. Intelligencer* **31** (2009), 18–23.  
 [5] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*, Erster Band. Leipzig, Berlin: B. G. Teubner, 1909.  
 [6] J. D. Lichtman, Almost primes and the Banks–Martin conjecture, *J. Number Theory* **211** (2020), 513–529.  
 [7] F. Mertens, Ein Beitrag zur analytischen Zahlentheorie, *J. reine angew. Math.* **78** (1874), 46–62.  
 [8] H. L. Montgomery and R. C. Vaughan, *Multiplicative Number Theory I: Classical Theory*, Cambridge University Press, Cambridge, 2006.

- [9] P. Pollack, A generalization of the Hardy-Ramanujan inequality and applications, *J. Number Theory* **210** (2020), 171–182.
- [10] D. Popa, A double Mertens type evaluation, *J. Math. Anal. Appl.* **409** (2014), 1159–1163.
- [11] D. Popa, A triple Mertens evaluation, *J. Math. Anal. Appl.* **444** (2016), 464–474.
- [12] G. Tenenbaum, Generalized Mertens sums, in *Analytic Number Theory, Modular Forms and  $q$ -hypergeometric Series* – in honor of Krishna Alladi’s 60th birthday, G. Andrews and F. Garvan, eds., Springer Proc. Math. Stat. **221**, Springer, Cham, 2017, 477–495.
- [13] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, Graduate Studies in Mathematics **163**, Amer. Math. Soc., 2015.