Bachelor's project

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Robin Fissum

Digit sums and the number of prime factors of the factorial n!=1·2···n

Bachelor's project in BMAT
Supervisor: Prof. Kristian Seip
May 2020

NTNU
Kunnskap for en bedre verden

Robin Fissum

Digit sums and the number of prime factors of the factorial n!=1·2···n

Bachelor's project in BMAT
Supervisor: Prof. Kristian Seip
May 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Contents

1 Introduction **1**

2 Digit sums **2**
2.1 Basic properties 2
2.2 Some pointwise bounds 4
2.3 Sums of digit sums 10
2.4 Digit sums over primes 13

3 A result by Hardy & Ramanujan **18**
3.1 Introduction 18
3.2 The theorem 19

4 How we deduced the result **22**
4.1 Legendre's formula 22
4.2 The first summand 23
4.3 The second summand 24
4.4 A digit sum integral 26
4.5 Conclusion using digit sums 31

5 The full asymptotic expansion **33**

6 Summary **37**

List of symbols **38**

Appendix **40**

References **47**

1 Introduction

The purpose of this paper is to figure out, on average, the number of prime factors of $n!$ (factorial) as the integer n takes larger and larger values. The motivation for this arose from the fact that factorials appear, at least subtly, in almost every mathematical discipline. We will, as is common, let

$$\Omega(k)$$

denote the number of prime factors of the positive integer k, counted with multiplicity. Our problem of interest then, is to estimate the value of

$$\Omega(n!)$$

as n gets progressively larger.

There are at least two equivalent interpretations of this problem. The first is, as stated, estimating the number of prime factors of the factorial. The other (historically more present) interpretation is to estimate the «average order» of the function Ω. That is, to estimate

$$\frac{1}{n}\sum_{k=1}^{n}\Omega(k)$$

as n goes to infinity. The equivalence of these two problems is evident upon noticing that Ω is a completely additive function. I.e., satisfying

$$\Omega(mn) = \Omega(m) + \Omega(n),$$

for all positive integers m and n, so that

$$\Omega(n!) = \sum_{k=1}^{n}\Omega(k).$$

Unaware that the «exact solution» to this problem had been published 50 years prior, I started working on the problem using a certain formula due to Legendre. This eventually led to the discovery of an interesting asymptotic formula concerning digit sums. It is of this reason the paper is, in some sense, threefold. We begin by investigating certain properties of digit sums. Thereafter we go through a simple, but classic, theorem for the average order of Ω. We also show how one can, using digit sums, slightly improve this result. Finally, we consider a newer and much more precise result, being in some sense a resolution to the problem.

We have written the section on digit sums in a logically separate matter, but its relevance to our investigation of the factorial will become clear at the end[1] of section 4.1.

[1] Equation (11).

2 Digit sums

2.1 Basic properties

Let n be a non-negative integer and $k \geq 2$ an integer[2]. Then n can be represented in a unique way as

$$n = \sum_{t=0}^{m} d_t k^t, \tag{1}$$

where m is a non-negative integer, $d_t \in \{0, 1, \ldots, k-1\}$ for each t and $d_m \neq 0$. We call this the *base k representation of n*, and may write this compactly as $n = d_m d_{m-1} \ldots d_1 d_0$ when it is clear from context which base k we are working in. We call the integers d_t the *digits* of n in base k and $(m+1)$ the *number of digits* of n in base k.

Definition 2.1. *If* (1) *is the representation of n in base k, we define the digit sum of n in base k as*

$$S_k(n) \ = \ d_m + d_{m-1} + \ldots + d_0. \tag{2}$$

Notice that $S_k(n)$ is a function in two variables k and n taking non-negative integer values[3]. It is also clear from the definition that $S_k(n)$ takes the value zero if and only if n is equal to zero.

Proposition 2.2 (Basic properties). *If $n = d_m d_{m-1} \ldots d_0$ is the base k representation of n, then*

(i) $S_k(n) = 1$ *if and only if n is a power of k.*

(ii) $S_k(k^N n) = S_k(n)$ *for any integer $N \geq 0$, especially*

(iii) $S_k(kn) = S_k(n)$.

Proof. Since $S_k(n) = d_m + d_{m-1} + \ldots + d_0$ and each d_t is ≥ 0, this expression is equal to 1 if and only if exactly one of the d_t is equal to 1 and the others equal to 0. Since $d_m \neq 0$, this is equivalent with $d_m = 1$ and $d_{m-1} = \ldots = d_0 = 0$. But in this case $n = d_m k^m = 1k^m = k^m$. This proves (i).

To prove (ii), notice that if $n = \sum\limits_{t=0}^{m} d_t k^t$, then

$$k^N n = \sum_{t=0}^{m} d_t k^{t+N} = d_m k^{m+N} + \ldots + d_0 k^N + 0k^{N-1} + \ldots + 0k^0,$$

[2] $k = 1$ also works and is known as the *Unary numeral system*. This is however a special case we will not consider, as every digit would equal 1 and the number of digits would equal the number itself.

[3] I.e. $S_k(n) = f(k, n)$, where $f : \mathbb{N}_{\geq 2} \times \mathbb{N}_0 \longrightarrow \mathbb{N}_0$

and so

$$\begin{aligned} S_k(k^N n) &= d_m + d_{m-1} + \ldots + d_0 + 0 + \ldots + 0 \\ &= d_m + d_{m-1} + \ldots + d_0 \\ &= S_k(n). \end{aligned}$$

(iii) follows immediately from (ii). □

It would be interesting to know how the value of the digit sum $S_k(n)$ changes if n is replaced by, say $n+1$. Fortunately, we have the following result:

Proposition 2.3. *For all $n \in \mathbb{N}_0$ and $k \in \mathbb{N}_{\geq 2}$ we have*

$$S_k(n+1) = S_k(n) + 1 - \beta(k-1)$$

where β is the number of tailing digits equal to $k-1$ in the base k representation of n.

Proof. Suppose $n = d_m d_{m-1} \ldots d_1 d_0$ in base k. If d_0 is one of $0, 1, 2, \ldots, k-2$, then $n+1 = d_m d_{m-1} \ldots d_1 (d_0 + 1)$, and so $S_k(n+1) = S_k(n) + 1$ in this case. Otherwise, if $d_0 = k-1$, then adding one to n transforms the last digit d_0 into a zero $\rightarrow$ digit sum is reduced by $k-1$, but with a carry over to d_1. Then we repeat this process: The carry transforms d_1 into either $d_1 + 1$ or 0, etc. There are now only two possibilities: Either all of the d_t are equal to $k-1$, so that $100\ldots00$ is the base k representation of $n+1$, or $d_i \in \{0, 1, 2, \ldots, k-2\}$ for some $0 < i < m$ and $n+1 = d_m d_{m-1} \ldots d_{i+1}(d_i + 1)00\ldots00$ in base k. In any case, we see that $S_k(n+1)$ is equal to $S_k(n) + 1 - \beta(k-1)$, where β is the number of tailing $k-1$'s in the base k expansion of n. □

Some consequences of this are summarized in the following corollary.

Corollary 2.4. *For any n and k, it is true that*

(i) $\quad S_k(n+1) \leq S_k(n) + 1,\quad$ *especially*

(ii) $\quad S_k(n+N) \leq S_k(n) + N\quad$ *for every integer $N \geq 0$.*

(iii) $\quad S_k(n) \leq n.$

(iv) $\quad S_k(n+1) = S_k(n) + 1 \iff k \nmid n+1.$

Proof. (i) follows immediately from the Proposition 2.3, and (ii) follows by repeated application of (i). (iii) follows from (ii) by substituting $n = 0$. Finally, from Proposition 2.3 we have that $S_k(n+1) = S_k(n) + 1$ if and only if there are no tailing $k-1$'s in the base k representation of n. This is equivalent to the last digit of n being different from $k-1$, equivalently $k \nmid n+1$. □

2.2 Some pointwise bounds

From what we have seen, the value of $S_k(n)$ is bounded below by 1 and above by n, for nonzero n. The value 1 is assumed infinitely often, namely on powers of k. Regarding the upper bound n, we can do somewhat better by observing that if

$$n = d_m d_{m-1} \dots d_1 d_0 = \sum_{t=0}^{m} d_t k^t$$

in base k, then at most every digit of n is equal to $k-1$. Since there are $m+1$ digits, we have

$$S_k(n) \leq (k-1)(m+1).$$

Furthermore, since $k^m \leq n < k^{m+1}$, also $m \leq \log_k(n) < m+1$. Therefore $m = \lfloor \log_k(n) \rfloor$, and we get the following result:

Proposition 2.5. *For any n and k,*

$$S_k(n) \leq (k-1)\left(\lfloor \log_k(n) \rfloor + 1\right). \tag{3}$$

We have now looked at some properties of $S_k(n)$ as a function of n. For a precise result about the asymptotic behavior of $S_k(n)$ with k fixed, the reader might want to take a look at [2]. We will do the opposite; we fix the value of n and consider $S_k(n)$ as a function of the base k. One thing to point out is that whenever $k > n$, then $S_k(n)$ simply equals n since n has only one digit in base k, namely itself. Therefore, we only consider the values of k for which $2 \leq k \leq n$.

Figure 1 shows $S_k(n)$ as a function of k over $2 \leq k \leq n$ for certain fixed values of n. From these plots it seems as though that when (the fixed) value n gets larger, the graph of $S_k(n)$ becomes more and more like a perfect queue of triangles.

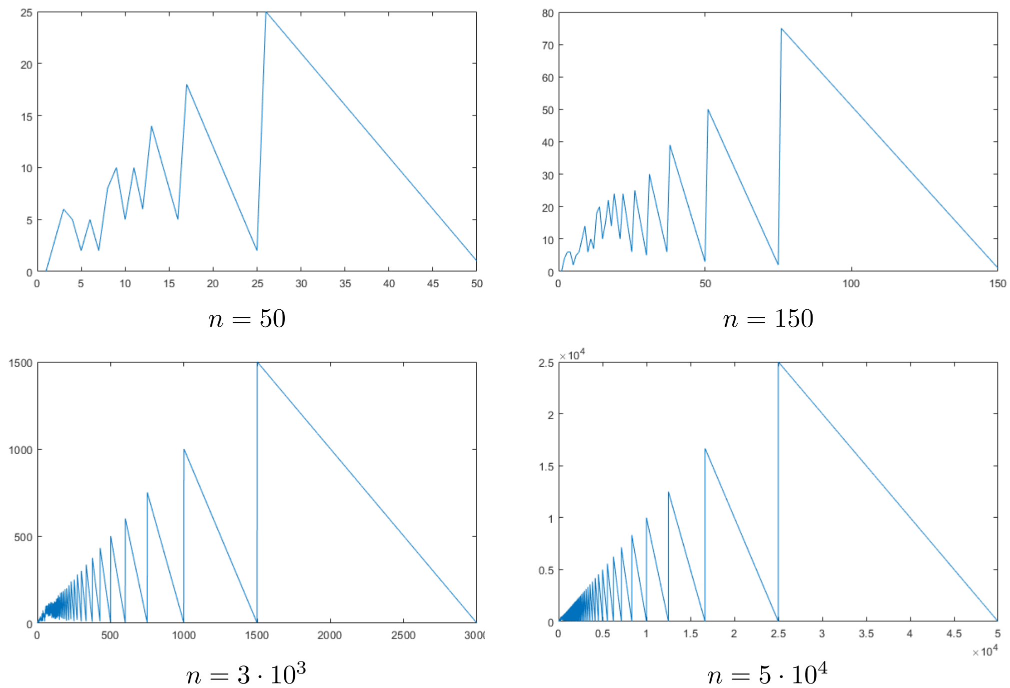

Figure 1

To explain why this is the case, consider first the following:

$$\begin{array}{c}n \text{ has } N \text{ digits}\\ \text{in base } k\end{array} \iff k^{N-1} \leq n < k^N \iff \sqrt[N]{n} < k \leq \sqrt[N-1]{n}.$$

The proof of this is straightforward. The point we wish to make is that when n is large, most of the integers k in $[2, n]$ are greater than $\sqrt{n}$. For such a value of k, n has exactly 2 digits in base k, say

$$n = d_1 k + d_0.$$

Furthermore, if M denotes the unique positive integer for which

$$\frac{n}{M+1} < k \leq \frac{n}{M},$$

then

$$Mk \leq n < (M+1)k,$$

and it is clear from this that we must have $d_1 = M$. Accordingly

$$n = Mk + d_0.$$

But in that case $d_0 = n - Mk$, and we deduce

$$S_k(n) = d_1 + d_0 = M + (n - Mk) = M + n - Mk.$$

Finally, since $\frac{n}{M+1} < k$ by assumption, we get

$$S_k(n) < M + n - M\frac{n}{M+1} = M + \frac{n}{M+1}.$$

We summarize our findings in the following:

Proposition 2.6. *Let n be a fixed positive integer. If $k \geq 2$ is an integer satisfying*

$$\sqrt{n} < k \leq n,$$

and M denotes the unique positive integer for which

$$\frac{n}{M+1} < k \leq \frac{n}{M},$$

then

$$\begin{aligned} (i) \qquad S_k(n) &= M + n - Mk \\ (ii) \qquad S_k(n) &< M + \frac{n}{M+1} \end{aligned}$$

both hold. It immediately follows that if k_1 and k_2 both satisfy the above hypothesis for the same value of M, but $k_1 < k_2$, then

$$(iii) \qquad S_{k_1}(n) > S_{k_2}(n).$$

Remark. Given that $S_k(n)$ is an integer, the strict inequality $S_k(n) < M + \frac{n}{M+1}$ from Proposition 2.6 can be strengthened to $S_k(n) \leq (M-1) + \frac{n}{M+1}$ if $M+1$ divides n, and $S_k(n) \leq M + \frac{n-t}{M+1}$ if $M+1$ does not divide n, where t is the remainder of n upon division by $M+1$. Especially, with $M = 1$ this tells us that $S_k(n)$ is less than or equal to $\frac{n+1}{2}$ on $(\frac{n}{2},\ n]$ if n is odd, and less than or equal to $\frac{n}{2}$ on the same interval if n is even. This is the best possible bound of this type in the sense that the value is attained. (Take $n = 6661$ and $k = 3331$, then $S_k(n) = \frac{n+1}{2}$).

Figure 2 illustrates the results from Proposition 2.6. Part (i) is illustrated by the fact that the graph is decreasing linearly from left to right on the intervals $(\frac{n}{M+1}, \frac{n}{M}]$, while the bounds in part (ii) correspond to the horizontal red lines. Notice from the figure how small $\sqrt{n}$ is compared to n, so n has two digits in base k for the wast majority of k less than or equal to n.

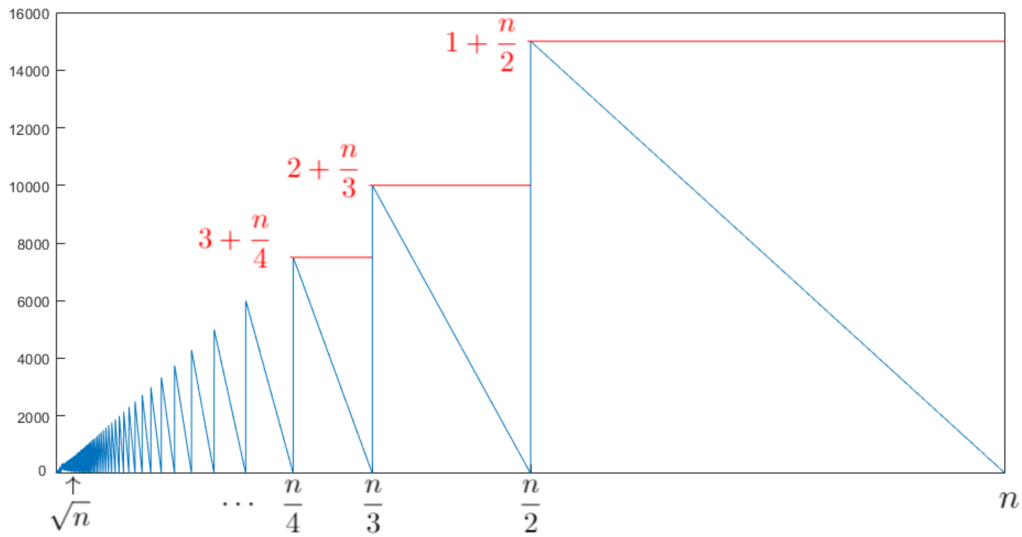

Figure 2: Schematic for $S_k(n)$ as a function of k. In this particular plot n is fixed equal to $3 \cdot 10^4$.

From an inspection of Figure 2 it seems reasonable that the leftmost red line is vertically below the middle red line, which in turn is vertically below the rightmost line. This is indeed the case, as made precise by the next proposition.

Proposition 2.7. *Let n be a fixed positive integer. If M and N are positive integers such that $M < N$,*

$$\sqrt{n} < \frac{n}{N+1} \qquad \text{and} \qquad \sqrt{n} < \frac{n}{M+1},$$

then

$$N + \frac{n}{N+1} \; < \; M + \frac{n}{M+1}.$$

Proof. The inequalities $\sqrt{n} < \frac{n}{N+1}$ and $\sqrt{n} < \frac{n}{M+1}$ together imply

$$n < \frac{n^2}{(N+1)(M+1)},$$

so that

$$(N+1)(M+1) < n.$$

Using $N - M > 0$, the sequence of implications given by

$$\begin{aligned}(N+1)(M+1) < n &\Longrightarrow \frac{(N-M)(N+1)(M+1)}{(N-M)} < n \\ &\Longrightarrow \frac{(N-M)(N+1)(M+1)}{(N+1)-(M+1)} < n \\ &\Longrightarrow \frac{N-M}{\frac{1}{M+1} - \frac{1}{N+1}} < n \\ &\Longrightarrow N - M < \frac{n}{M+1} - \frac{n}{N+1} \\ &\Longrightarrow N + \frac{n}{N+1} < M + \frac{n}{M+1}\end{aligned}$$

gives the desired conclusion. □

We have established horizontal bounds for the digit sum for all values of k except for those less than or equal to $\sqrt{n}$. We remedy this now.

Proposition 2.8. *If n and k are positive integers such that $2 \leq k \leq \sqrt{n}$, then*

$$S_k(n) \leq 2(\sqrt{n} - 1).$$

Proof. Let $m \geq 2$ be the unique integer such that $\sqrt[m+1]{n} < k \leq \sqrt[m]{n}$. By Proposition 2.5 we have $S_k(n) \leq (k-1)(m+1)$, and since $k \leq \sqrt[m]{n}$, we get

$$S_k(n) \leq (m+1)(\sqrt[m]{n} - 1).$$

Now, compare $X(n) = (m+1)(\sqrt[m]{n} - 1)$ with $Y(n) = 2(\sqrt{n} - 1)$.
When $m > 2$, $Y(n)$ will dominate $X(n)$ as $n \to +\infty$ because of the m-th root. Thus, there is a integer N_m such that

$$n \geq N_m \Longrightarrow X(n) < Y(n).$$

Furthermore, since $X'(n) = (1 + \frac{1}{m})n^{-\frac{m-1}{m}}$ is pointwise below $(1 + \frac{1}{m'})n^{-\frac{m'-1}{m'}}$ whenever $m \leq m'$, the sequence $(N_m)_m$ is decreasing. Computation gives

$$N_3 = 18,\ N_4 = 6,\ N_5 = 4,\ N_6 = 3,\ N_7 = 3,\ N_8 = 2.$$

Accordingly $N_m \leq 2$ for $m \geq 8$. This leaves verifying the statement for those k satisfying the hypothesis when $3 \leq m \leq 7$ and $1 \leq n \leq N_m$. But in that case $\sqrt[m]{n} \leq \sqrt[3]{17} \approx 2.57$, so the only k which can satisfy the hypothesis is $k = 2$, **if** it belongs to the interval $(\sqrt[m+1]{n}, \sqrt[m]{n}]$. With these restrictions on m and n, we find that $2 \in (\sqrt[m+1]{n}, \sqrt[m]{n}]$ only when $m = 3$ and $8 \leq n \leq 15$. Numerical calculations complete the proof for $m > 2$, as shown in the table on the next page.

n	8	9	10	11	12	13	14	15
$S_2(n)$	1	2	2	3	2	3	3	4
$2(\sqrt{n}-1)$	3.66	4	4.32	4.63	4.93	5.21	5.48	5.75

The case $m = 2$ still remains, but the argument above cannot be recycled in this case. Demanding that $m = 2$ means that n has 3 digits in base k, so we may write

$$n = ak^2 + bk + c,$$

where $a \neq 0$ and $0 \leq a, b, c \leq k-1$. Then $S_k(n) = a+b+c$ and the inequality we must prove is

$$a+b+c \leq 2\left(\sqrt{ak^2+bk+c}-1\right).$$

The left hand side is $a+b+c \leq 3(k-1) = 3k-3$, so whenever $a \geq 4$, we have

$$2\left(\sqrt{ak^2+bk+c}-1\right) \geq 2\left(\sqrt{ak^2}-1\right) \geq 2\left(\sqrt{4k^2}-1\right) = 4k-2 > 3k-3.$$

Therefore, we only have to check the cases $a = 1, 2, 3$.

<u>Case 1: $a = 1$</u>

Here $S_k(n) = 1+b+c \leq 1+2(k-1) = 2k-1$, but if at least one of b and c differs from $k-1$, this is improved to $S_k(n) \leq 2k-2$. Then

$$2\left(\sqrt{ak^2+bk+c}-1\right) \geq 2\left(\sqrt{a}\,k-1\right) = 2k-2,$$

and so it holds. Otherwise, if $b = c = k-1$, then $S_k(n) = 2k-1$ and

$$2\left(\sqrt{ak^2+bk+c}-1\right) = 2\left(\sqrt{k^2+(k-1)k+(k-1)}-1\right) = 2\sqrt{2k^2-1}-2,$$

an expression that is greater than $2k-1 = S_k(n)$ for all $k \geq 2$.

<u>Case 2: $a = 2$</u>

In this case, $S_k(n) = a+b+c = 2+b+c \leq 2+2(k-1) = 2k$. Now

$$2\left(\sqrt{ak^2+bk+c}-1\right) = 2\left(\sqrt{2k^2+bk+c}-1\right) \geq 2\left(\sqrt{2}\,k-1\right) \geq 2k,$$

for $k \geq 3$, so it suffices to check $k = 2$. However, as $a = 2$, we cannot have $k = 2$, since a base k-digit must be $\leq k-1$.

<u>Case 3: $a = 3$</u>

Here $S_k(n) = a+b+c = 3+b+c \leq 3+2(k-1) = 2k+1$, while

$$2\left(\sqrt{ak^2+bk+c}-1\right) = 2\left(\sqrt{3k^2+bk+c}-1\right) \geq 2\left(\sqrt{3}\,k-1\right) \geq 2k+1$$

for $k \geq 3$. Again, we don't have to consider $k = 2$, since the size of a prohibits this situation. This completes our proof of Proposition 2.8. □

2.3 Sums of digit sums

In this section we will consider the function defined on the positive integers by

$$D(n) := \sum_{2 \le k \le n} S_k(n).$$

Notice (!) that every term of the sum is dependent on n. Figure 3 shows a plot of $D(n)$ for $2 \le n \le 10^4$, being pointwise between $0.17n^2$ and $0.18n^2$. In other words, computational evidence seems to indicate that $D(n) \sim \delta n^2$ for some constant δ between 0.17 and 0.18. We will prove that this is true, but first we do some preparatory work.

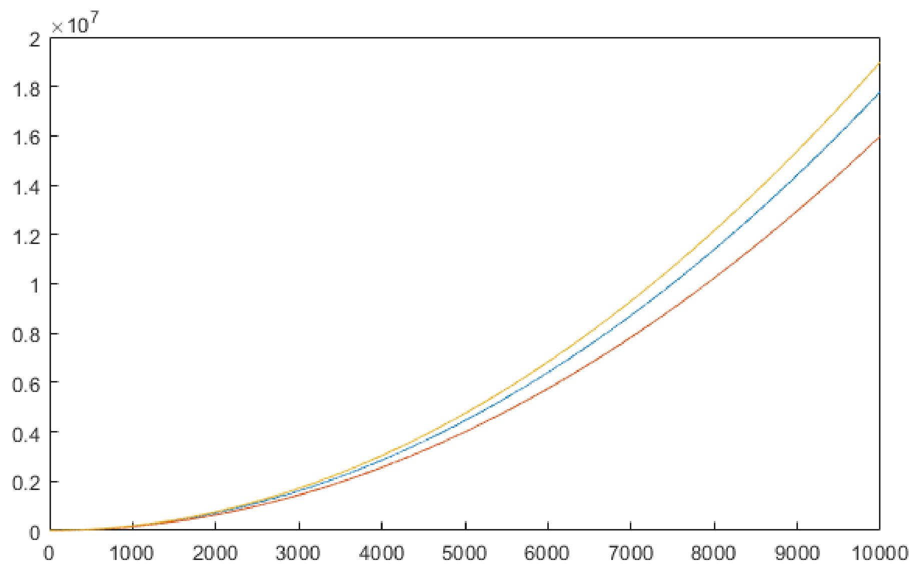

Figure 3: $D(n)$ [**blue**] vs. $0.17n^2$ [**orange**] and $0.18n^2$ [**yellow**].

We begin by deriving an expression for the number of integers in the half open interval $(\frac{n}{N+1}, \frac{n}{N}]$, if any. If both $\frac{n}{N+1}$ and $\frac{n}{N}$ are integers, the answer is simply $\frac{n}{N} - \frac{n}{N+1} = \frac{n}{N(N+1)}$. In the general case, we may apply the division algorithm to find integers q_1, q_2 and r_1, r_2 such that

$$n = q_1 N + r_1 \qquad \text{and} \qquad n = q_2(N+1) + r_2,$$

where $0 \leq r_1 < N$ and $0 \leq r_2 < N+1$. Then

$$\frac{n}{N+1} = \frac{q_2(N+1) + r_2}{N+1} \qquad \text{and} \qquad \frac{n}{N} = \frac{q_1 N + r_1}{N}.$$

This makes

$$\frac{q_2(N+1) + r_2 + [(N+1) - r_2]}{N+1} = \frac{n + (N+1) - (n \mod (N+1))}{N+1}$$

the smallest integer belonging to the interval, and

$$\frac{q_1 N + r_1 - r_1}{N} = \frac{n - (n \mod N)}{N}$$

the largest integer belonging to the interval. In total the number of integers belonging to the interval is

$$\begin{aligned}
\mathcal{A}(n, N) &:= \frac{n - (n \mod N)}{N} - \frac{n + (N+1) - (n \mod (N+1))}{N+1} + 1 \\
&= \frac{n}{N(N+1)} + \frac{n \mod (N+1)}{N+1} - \frac{n \mod N}{N}.
\end{aligned}$$

Especially, it holds true that

$$\frac{n}{N(N+1)} - 1 < \mathcal{A}(n, N) < \frac{n}{N(N+1)} + 1.$$

Now we consider the sum $\sum_{\frac{n}{N+1} < k \leq \frac{n}{N}} S_k(n)$ as $n \to +\infty$, where N is a fixed positive integer. We may suppose n is so large that $\sqrt{n} < \frac{n}{N+1}$. By Proposition 2.6 we have

$$\sum_{\frac{n}{N+1} < k \leq \frac{n}{N}} S_k(n) = \sum_{\frac{n}{N+1} < k \leq \frac{n}{N}} (N + n - Nk) = (N+n)\mathcal{A}(n, N) - N \sum_{\frac{n}{N+1} < k \leq \frac{n}{N}} k.$$

After substituting the expression for $\mathcal{A}(n, N)$, expanding the sum $\sum k$, and going through a tedious calculation (which we refer the reader to the Appendix for the full calculation) we eventually arrive at the result

$$\sum_{\frac{n}{N+1} < k \leq \frac{n}{N}} S_k(n) = \frac{1}{2N(N+1)^2} n^2 + O(n). \tag{4}$$

We will use (4) to derive the asymptotic formula for $D(n)$, but its validity is based on the following lemma:

Lemma 2.9. *The series*

$$\sum_{t=1}^{\infty} \frac{1}{2t(t+1)^2}$$

converges to $1 - \frac{\pi^2}{12}$.

Proof. A simple partial fraction decomposition will suffice. □

Proposition 2.10. *The function* $D(n)$ *is asymptotically equivalent to* $(1 - \frac{\pi^2}{12})n^2$.

Proof. Let N be the largest positive integer such that $\sqrt{n} < \frac{n}{N+1}$ (i.e. $N = \lfloor \sqrt{n} \rfloor - 1$). Then we may write

$$\sum_{2 \le k \le n} S_k(n) = \sum_{2 \le k \le \sqrt{n}} S_k(n) + \sum_{\sqrt{n} < k \le \frac{n}{N+1}} S_k(n) + \sum_{\frac{n}{N+1} < k \le n} S_k(n). \qquad (5)$$

By Proposition 2.8, the first summand of equation (5) is

$$\sum_{2 \le k \le \sqrt{n}} S_k(n) \le \sum_{2 \le k \le \sqrt{n}} 2(\sqrt{n} - 1) = O(n).$$

The second summand is by part (ii) of Proposition 2.6:

$$\sum_{\sqrt{n} < k \le \frac{n}{N+1}} S_k(n) = \sum_{\sqrt{n} < k \le \frac{n}{\lfloor \sqrt{n} \rfloor}} O\left(\frac{n}{\lfloor \sqrt{n} \rfloor}\right) = O\left(\frac{n}{\lfloor \sqrt{n} \rfloor}\right) O\left(\frac{n}{\lfloor \sqrt{n} \rfloor} - \sqrt{n}\right) = O(n)$$

The final, and most interesting summand of (5) is by (4) equal to:

$$\sum_{\frac{n}{N+1} < k \le n} S_k(n) = \sum_{t=1}^{N} \sum_{\frac{n}{t+1} < k \le \frac{n}{t}} S_k(n) = \sum_{t=1}^{N} \left(\frac{n^2}{2t(t+1)^2} + O(n) \right)$$

$$= n^2 \sum_{t=1}^{N} \frac{1}{2t(t+1)^2} \; + \; O(n^{\frac{3}{2}}).$$

Substituting these results back into equation (5) gives

$$\sum_{2 \le k \le n} S_k(n) \; = \; n^2 \sum_{t=1}^{N} \frac{1}{2t(t+1)^2} \; + \; O(n^{\frac{3}{2}}).$$

Finally, since $N = \lfloor \sqrt{n} \rfloor - 1$ goes to infinity with n, we get the desired conclusion from Lemma 2.9. □

Notice that $1 - \frac{\pi^2}{12} = 0.17753\ldots$, so the result matches the numerical data.

2.4 Digit sums over primes

In this section, we consider the related problem of estimating the growth of

$$h(n) := \sum_{p \le n} S_p(n),$$

the sum now taken over the primes less than or equal to n. Trivially,

$$\pi(n) \le h(n) \le n\pi(n)$$

for all n, so that $\varepsilon > 0$ implies[4]

$$h(n) \le (1+\varepsilon)\frac{n^2}{\log(n)}$$

for sufficiently large n. But we can do better:

Proposition 2.11. *For all n, we have*

$$h(n) \le \frac{n}{2}\pi(n) + \frac{1}{2}.$$

If $\frac{n+1}{2}$ is known not to be prime, the upper bound can be improved to $\frac{n}{2}\pi(n)$, for which it follows that $\varepsilon > 0$ implies

$$h(n) \le (1+\varepsilon)\frac{n^2}{2\log(n)},$$

for sufficiently large n.

Proof. In accordance with the remark following Proposition 2.6, as long as $\frac{n+1}{2}$ is not a prime number, we have

$$h(n) = \sum_{p \le n} S_p(n) \le \sum_{p \le n} \frac{n}{2} = \frac{1}{2} n \sum_{p \le n} 1 = \frac{1}{2} n\pi(n).$$

If $\frac{n+1}{2}$ happens to be a prime, we still have

$$\begin{aligned} h(n) = \sum_{p \le n} S_p(n) \;\le\; \frac{n+1}{2} + \sum_{\substack{p \le n \\ p \ne (n+1)/2}} \frac{n}{2} \;=\; \frac{n+1}{2} + \frac{n}{2} \sum_{\substack{p \le n \\ p \ne (n+1)/2}} 1 \\ = \frac{n+1}{2} + \frac{n}{2}(\pi(n) - 1) = \frac{n}{2}\pi(n) + \frac{1}{2}. \end{aligned}$$

□

It turns out that we can do quite a bit better than the above results. Numerical calculations for $n \le 10^9$ (see Figure 4) show that $h(n)$ is well approximated by $\delta \frac{n^2}{\log(n)}$ for some positive constant δ slightly below 0.2.

[4]See Appendix (P.N.T).

n	1	2	3	4	5	6	7	8	9
$\frac{h(10^n)\log(10^n)}{(10^n)^2}$	0.2303	0.2109	0.1946	0.1915	0.1888	0.1867	0.1854	0.1844	0.1836

Figure 4

Here is a heuristic argument to why this should be the case. Let ρ_n denote the «density» of primes in $[0, n]$, i.e. $\rho_n := \pi(n)/n$. If the primes in $[0, n]$ were evenly distributed (which isn't entirely true) and $S_k(n)$ is, on average, not too dependent on the primality of k, we would have

$$h(n) = \sum_{p \le n} S_p(n) \;\approx\; \rho_n \sum_{2 \le k \le n} S_k(n) \;\sim\; \rho_n \delta n^2 \;\sim\; \delta \frac{n^2}{\log(n)}$$

by Proposition 2.10 and the P.N.T., where $\delta = 1 - \frac{\pi^2}{12} = 0.1775\ldots$. We now show that our assertion is indeed true.

Proposition 2.12. *The function* $h(n) = \sum_{p \le n} S_p(n)$ *has asymptotic expansion*

$$h(n) = \delta \frac{n^2}{\log(n)} + C \frac{n^2}{\log^2(n)} + o\left(\frac{n^2}{\log^2(n)}\right),$$

where $\delta = 1 - \frac{\pi^2}{12} = 0.1775\ldots$ *and* C *is a constant approximately equal to* 0.1199.

Proof. We modify the proof of Proposition 2.10. Throughout, let $p(n, k)$ denote the number of primes in the half open interval $(\frac{n}{k+1}, \frac{n}{k}]$. Let $N = \lfloor \sqrt{n} \rfloor - 1$ be the largest positive integer such that $\sqrt{n} < \frac{n}{N+1}$. We may write

$$\sum_{p \le n} S_p(n) = \sum_{p \le \frac{n}{N+1}} S_p(n) + \sum_{t=1}^{N} \sum_{\frac{n}{t+1} < p \le \frac{n}{t}} S_p(n). \tag{6}$$

By proposition 2.8 we have

$$\sum_{p \le \frac{n}{N+1}} S_p(n) \le \sum_{p \le \frac{n}{N+1}} 2\left(\sqrt{n} - 1\right) < \sum_{p \le \frac{n}{\lfloor \sqrt{n} \rfloor}} 2\sqrt{n} \le \frac{3n^{3/2}}{\lfloor \sqrt{n} \rfloor} = O(n).$$

Therefore, the size of this summand is to be considered insignificant here. Also, using Proposition 2.6, we deduce

$$\begin{aligned}
&\sum_{t=1}^{N} \sum_{\frac{n}{t+1}<p\leq \frac{n}{t}} S_p(n) \\
&\quad = \sum_{t=1}^{N} \sum_{\frac{n}{t+1}<p\leq \frac{n}{t}} (t+n-tp) \\
&\quad = \sum_{t=1}^{N} \left(tp(n,t) + np(n,t) - t \sum_{\frac{n}{t+1}<p\leq \frac{n}{t}} p \right) \\
&\quad = \sum_{t=1}^{N} tp(n,t) \, + \, n \sum_{t=1}^{N} p(n,t) \, - \, \sum_{t=1}^{N} \left(t \sum_{\frac{n}{t+1}<p\leq \frac{n}{t}} p \right). \qquad (7)
\end{aligned}$$

Regarding the three summands of (7), the first is

$$\sum_{t=1}^{N} tp(n,t) < \sum_{t=1}^{N} \sqrt{n} p(n,t) = \sqrt{n} \sum_{t=1}^{N} p(n,t) \leq \sqrt{n}\pi(n) = O\left(\frac{n^{3/2}}{\log(n)}\right),$$

by the P.N.T.. This summand is therefore insignificant in this context. For the second summand of (7), the P.N.T. gives

$$n \sum_{t=1}^{N} p(n,t) = n\left(\pi(n) - \pi\left(\frac{n}{\lfloor \sqrt{n} \rfloor}\right)\right) = \frac{n^2}{\log(n)} + \widetilde{O}_n\left(\frac{n^2}{\log^2(n)}\right).$$

We now consider the third summand of equation (7). For this, let $S(x)$ denote the sum of the primes not exceeding x. It is known (see [1]) that

$$S(x) = \frac{x^2}{2\log(x)} + \frac{x^2}{4\log^2(x)} + \frac{x^2}{4\log^3(x)} + \frac{3x^2}{8\log^4(x)} + O\left(\frac{x^2}{\log^5(x)}\right), \qquad (8)$$

as x tends to infinity. Therefore, if t is any fixed positive integer, using the first two terms of equation (8), we can infer

$$\begin{aligned}
S\left(\frac{n}{t}\right) &= \frac{n^2}{2t^2\log(\frac{n}{t})} + \widetilde{O}_n\left(\frac{n^2}{4t^2\log^2(\frac{n}{t})}\right) \\
&= \frac{n^2}{2t^2\log(n)} + \frac{\log(t)}{2t^2}\frac{n^2}{\log(n)\log(\frac{n}{t})} + \widetilde{O}_n\left(\frac{n^2}{4t^2\log^2(n)}\right) \\
&= \frac{n^2}{2t^2\log(n)} + \widetilde{O}_n\left(\left[\frac{2\log(t)+1}{4t^2}\right]\frac{n^2}{\log^2(n)}\right),
\end{aligned}$$

and similarly

$$S\left(\frac{n}{t+1}\right) = \frac{n^2}{2(t+1)^2\log(\frac{n}{t+1})} + \widetilde{O}_n\left(\frac{n^2}{4(t+1)^2\log^2(\frac{n}{t+1})}\right)$$

$$= \frac{n^2}{2(t+1)^2\log(n)} + \widetilde{O}_n\left(\left[\frac{2\log(t+1)+1}{4(t+1)^2}\right]\frac{n^2}{\log^2(n)}\right).$$

Combined, they yield

$$\sum_{\frac{n}{t+1}<p\leq\frac{n}{t}} p = S\left(\frac{n}{t}\right) - S\left(\frac{n}{t+1}\right)$$

$$= \left[\frac{1}{2t^2} - \frac{1}{2(t+1)^2}\right]\frac{n^2}{\log(n)} + \widetilde{O}_n\left(c_t\frac{n^2}{\log^2(n)}\right)$$

$$= \left[\frac{2t+1}{2t^2(t+1)^2}\right]\frac{n^2}{\log(n)} + \widetilde{O}_n\left(c_t\frac{n^2}{\log^2(n)}\right),$$

where

$$c_t = \frac{2\log(t)+1}{4t^2} - \frac{2\log(t+1)+1}{4(t+1)^2}.$$

From this, (minus) the third summand of (7) is

$$\sum_{t=1}^{N}\left(t\sum_{\frac{n}{t+1}<p\leq\frac{n}{t}} p\right) = \sum_{t=1}^{N}\left[t\frac{(2t+1)}{2t^2(t+1)^2}\frac{n^2}{\log(n)} + t\,\widetilde{O}_n\left(c_t\frac{n^2}{\log^2(n)}\right)\right]$$

$$= \left[\sum_{t=1}^{N}\frac{2t+1}{2t(t+1)^2}\right]\frac{n^2}{\log(n)} + \widetilde{O}_n\left(\frac{n^2}{\log^2(n)}\sum_{t=1}^{N}tc_t\right)$$

$$= \left[\sum_{t=1}^{N}\frac{2t+1}{2t(t+1)^2}\right]\frac{n^2}{\log(n)} + \left[\sum_{t=1}^{N}tc_t\right]\widetilde{O}_n\left(\frac{n^2}{\log^2(n)}\right).$$

Substituting our obtained results back into (7), gives

$$\sum_{p\leq n} S_p(n)$$

$$= O(n) + O\left(\frac{n^{3/2}}{\log(n)}\right) + \left(\frac{n^2}{\log(n)} + \widetilde{O}_n\left(\frac{n^2}{\log^2(n)}\right)\right)$$

$$- \left[\sum_{t=1}^{N}\frac{2t+1}{2t(t+1)^2}\right]\frac{n^2}{\log(n)} - \left[\sum_{t=1}^{N}tc_t\right]\widetilde{O}_n\left(\frac{n^2}{\log^2(n)}\right)$$

$$= \left[1 - \sum_{t=1}^{N}\frac{2t+1}{2t(t+1)^2}\right]\frac{n^2}{\log(n)} + \left[1 - \sum_{t=1}^{N}tc_t\right]\widetilde{O}_n\left(\frac{n^2}{\log^2(n)}\right). \qquad (9)$$

Once again, since $N = \lfloor \sqrt{n} \rfloor - 1$ goes to infinity with n,

$$1 - \sum_{t=1}^{\infty} \frac{2t+1}{2t(t+1)^2} = 1 - \frac{\pi^2}{12} = 0.1775\ldots$$

and

$$C := 1 - \sum_{t=1}^{\infty} tc_t = 0.1199\ldots < \infty,$$

we get the desired conclusion from (9). $\square$

Remark. Some simplification (see Appendix) shows that

$$C = 1 - \frac{\pi^2}{24} - \frac{1}{2} \sum_{t=2}^{\infty} \frac{\log(t)}{t^2}.$$

3 A result by Hardy & Ramanujan

3.1 Introduction

In a famous paper from 1917, « *The normal number of prime factors of a number n*», G.H. Hardy and S. Ramanujan proved[5] that the so-called normal order of the functions $\omega(n)$ and $\Omega(n)$ is $\log(\log(n))$, where $\omega(n)$ is defined equal to the number of distinct prime factors of n. At the second page of that paper, where $f = \omega$ and $F = \Omega$, it says:

In fact it may be shewn[6]*, by purely elementary methods, that*

$$(1{\cdot}23) \qquad f(1) + f(2) + \ldots + f(n) \; = n \log\log n + An + O\left(\frac{n}{\log n}\right),$$

$$(1{\cdot}24) \qquad F(1) + F(2) + \ldots + F(n) \; = n \log\log n + Bn + O\left(\frac{n}{\log n}\right),$$

where A and B are certain constants.

However, they do not provide a full proof of these statements throughout the paper. Moreover, somewhat later they state:

This problem, however, we shall dismiss for the present, as results still more precise that (1·23) *and* (1·24) *can be found by transcendental methods.*

Here comes the interesting part: 53 years later, in 1970, a certain Bahman Saffari publishes a paper about asymptotic analysis, from which a complete asymptotic expansion for $\omega(n)$ and $\Omega(n)$ can be obtained. Saffari states in his paper, about Hardy and Ramanujan's claim of a result using «transcendental methods», that

To our knowledge, however, no such improvement has been published to date.

It would be interesting to know whether or not Hardy and Ramanujan actually had such a proof, but we may never know.

In this section, we will go through a simple proof of the above formulae. Before we jump into the proof, we state some results that will be relevant for our further work.

Theorem 3.1 (Mertens).

$$\sum_{p \le n} \frac{1}{p} = \log(\log(n)) + M + \varepsilon(n),$$

where M is a constant approximately equal to 0.2615 *and $\varepsilon(n)$ is a quantity that goes to zero as $n \to +\infty$. M is known as Mertens' constant.*

Remark. It is known[7] that the quantity $\varepsilon(n)$ is $O\left(\frac{1}{\log^k(n)}\right)$ for any $k > 0$, a fact we are going to use later.

[5] This is known as the Hardy-Ramanujan theorem.

[6] Old spelling of *shown*.

[7] In fact, it is even better. See [5].

Proposition 3.2. *The two series*

$$\sum_p \frac{1}{p(p-1)} \quad \textit{and} \quad \sum_{\substack{p^m \\ m\geq 2}} \frac{1}{p^m}$$

both converge to the same limit $\lambda = 0.7731\ldots$.

Proof. They are equal since

$$\sum_{\substack{p^m \\ m\geq 2}} \frac{1}{p^m} = \sum_p \left(\frac{1}{p^2} + \frac{1}{p^3} + \ldots\right) = \sum_p \frac{1}{p}\left(\frac{1}{p} + \frac{1}{p^2} + \ldots\right) = \sum_p \frac{1}{p(p-1)},$$

by the formula for a geometric series. They are convergent because

$$\sum_p \frac{1}{p(p-1)} < \sum_p \frac{1}{(p-1)^2} < \sum_k \frac{1}{k^2} = \frac{\pi^2}{6}.$$

□

Definition 3.3. *Let*

$$\theta := M + \lambda = 1.03465386\ldots$$

Per definition, we have the representation

$$\theta = \lim_{n\to\infty}\Big(-\log(\log(n)) + \sum_{p^m\leq n} \frac{1}{p^m}\Big).$$

We will also have occasion to bump into[8]

$$\gamma = \lim_{n\to\infty}\left(-\log(n) + \sum_{t=1}^{n} \frac{1}{t}\right) = 0.57721566\ldots.$$

The constant γ is known as the Euler–Mascheroni constant. We are not familiar with any names of θ and λ.

3.2 The theorem

We now go through a proof of the formulae from Hardy and Ramanujan's paper. We follow a proof that is a combination of that from [3] and [9], but with some comments and small modifications to make it easier to follow. Given that it appears in Hardy & Wright's *An Introduction to the Theory of Numbers* from 1938, a classical book in number theory, it is likely that this proof is the elementary proof mentioned above[9]. We are actually only interested in part (*ii*) of the theorem, but as we shall see, for the degree of precision under consideration, the results are in fact equivalent.

[8]We may deem a real, convergent series «simplified» if it is decomposed in terms of well understood constants such as M and γ.

[9]But not quite. We choose to present a (much more elegant) variant using the P.N.T., which only had a proof using complex analysis in 1917.

Theorem 3.4 (Hardy & Ramanujan). *The average order of both* $\omega(n)$ *and* $\Omega(n)$ *is* $\log(\log(n))$. *More precisely*

$$(i) \qquad \sum_{k\le n}\omega(k) = n\log(\log(n)) + Mn + O\left(\frac{n}{\log(n)}\right)$$

$$(ii) \qquad \sum_{k\le n}\Omega(k) = n\log(\log(n)) + \theta n + O\left(\frac{n}{\log(n)}\right).$$

Proof. Let

$$S_1 := \sum_{k\le n}\omega(k) = \sum_{k\le n}\sum_{p|k} 1 = \sum_{p\le n}\left\lfloor \frac{n}{p}\right\rfloor.$$

The last equality holds since there are exactly $\lfloor \frac{n}{p} \rfloor$ positive integers less than or equal to n that are multiples of a given prime p. Removing the floor bracket and then appealing to the prime number theorem gives

$$\begin{aligned} S_1 = \sum_{p\le n}\left(\frac{n}{p} - \left\{\frac{n}{p}\right\}\right) &= n\sum_{p\le n}\frac{1}{p} - \sum_{p\le n}\left\{\frac{n}{p}\right\} \\ &= n\sum_{p\le n}\frac{1}{p} + O(\pi(n)) \\ &= n\sum_{p\le n}\frac{1}{p} + O\left(\frac{n}{\log(n)}\right). \end{aligned}$$

An application of Mertens' theorem then gives

$$\begin{aligned} S_1 &= n\sum_{p\le n}\frac{1}{p} + O\left(\frac{n}{\log(n)}\right) \\ &= n\Big(\log(\log(n)) + M + \varepsilon(n)\Big) + O\left(\frac{n}{\log(n)}\right) \\ &= n\log(\log(n)) + Mn + O\left(\frac{n}{\log(n)}\right). \end{aligned}$$

In the last line we used that $\varepsilon(n)$ is $O\left(\frac{n}{\log(n)}\right)$ (See remark following Theorem 3.1). This proves part (i) of the theorem.

By similar reasoning to that as above, we have:

$$S_2 := \sum_{k\le n}\Omega(k) = \sum_{k\le n}\sum_{p^m|k} 1 = \sum_{p^m\le n}\left\lfloor \frac{n}{p^m}\right\rfloor.$$

Consider now the difference

$$A(n) := S_2 - S_1 = \sum_{k \le n} \Big(\Omega(k) - \omega(k) \Big) = \sum_{\substack{p^m \le n \\ m \ge 2}} \left\lfloor \frac{n}{p^m} \right\rfloor = \sum_{p} \sum_{m \ge 2} \left\lfloor \frac{n}{p^m} \right\rfloor,$$

where the second to last summation is extended over all primes p. On one hand, we have the upper bound

$$A(n) \le \sum_{p} \sum_{m \ge 2} \frac{n}{p^m} = n \sum_{\substack{p^m \\ m \ge 2}} \frac{1}{p^m} = \lambda n.$$

On the other hand, we notice that if $p^m \le n$ with $m \ge 2$, then $p \le \sqrt{n}$ and $m \le \log_p(n) = \frac{\log(n)}{\log(p)}$, so that

$$\begin{aligned} A(n) &\ge \sum_{p} \sum_{m \ge 2} \left(\frac{n}{p^m} - 1 \right) = \sum_{p \le \sqrt{n}} \sum_{2 \le m \le \log(n)/\log(p)} \left(\frac{n}{p^m} - 1 \right) \\ &= \sum_{p \le \sqrt{n}} \left(\frac{n}{p(p-1)} + O\left(\frac{\log(n)}{\log(p)} \right) \right) \\ &= n \sum_{p} \frac{1}{p(p-1)} + O(\sqrt{n}) \\ &= \lambda n + O(\sqrt{n}). \end{aligned}$$

Thus, we have showed

$$A(n) = \lambda n + O(\sqrt{n}),$$

and conclude that

$$S_2 = S_1 + A(n) = n \log(\log(n)) + \theta n + O\left(\frac{n}{\log(n)} \right).$$

This proves part (ii). □

4 How we deduced the result

We wish to show how one can arrive at the results from section 3 in a different way, and consider some of the interesting expressions it gives rise to. Our method is based on a version of a formula usually credited to Adrien-Marie Legendre, in which our previous work on digit sums will be rewarded. The formula is named after Legendre because it appears[10] in the introduction of his book «*Théorie des nombres*» from 1830.

4.1 Legendre's formula

For a positive integer n and prime number p, let $\mathcal{V}_p(n)$ denote the p-adic valuation of n, i.e. the largest integer k such that p^k divides n.

Theorem 4.1 (Legendre). *If n is a non-negative integer and p a prime, then*

$$\mathcal{V}_p(n!) = \sum_{t=1}^{\infty} \left\lfloor \frac{n}{p^t} \right\rfloor .$$

Proof. Among the numbers $p, 2p, 3p, \dots$ there are exactly $\left\lfloor \frac{n}{p} \right\rfloor$ of which are less than or equal to n. Among $p^2, 2p^2, 3p^2, \dots$ there are $\left\lfloor \frac{n}{p^2} \right\rfloor$ less than or equal to n. Etc. Among $p^t, 2p^t, 3p^t, \dots$ there are $\left\lfloor \frac{n}{p^t} \right\rfloor$ which are less than or equal to n. If we take the sum of all these for all values of the exponent t, we get the desired result. □

Remark (1)*.* It is possible that one of the numbers above appears in more than one list. For example, if $p = 3$ then surely $3p$ and p^2 is the same number, but from list 1 and 2, respectively. This does not pose a problem as the prime factor p is only counted twice anyway: Once in $\left\lfloor \frac{n}{p} \right\rfloor$ and once in $\left\lfloor \frac{n}{p^2} \right\rfloor$.

Remark (2)*.* Even though the upper index of summation is infinity, there are only finitely many nonzero terms. This is because the floor function evaluates to zero when $p^t > n$. Specifically,

$$\mathcal{V}_p(n!) = \sum_{t=1}^{\lfloor \log_p(n) \rfloor} \left\lfloor \frac{n}{p^t} \right\rfloor .$$

We are interested in the following version of Legendre's theorem, that might be more manageable in certain situations. Let, as usual, $S_k(n)$ denote the digit sum of n in base k.

Theorem 4.2. *If n is a nonnegative integer and p a prime, then*

$$\mathcal{V}_p(n!) = \frac{n - S_p(n)}{p - 1}. \tag{10}$$

[10]Legendre wrote $E(x)$ for $\lfloor x \rfloor$. The E stands for «Entier», meaning «whole» in french.

Proof. Let $n = d_k p^k + d_{k-1} p^{k-1} + \ldots + d_1 p + d_0$ be the representation of n in base p. Then $\lfloor \log_p(n) \rfloor = k$, and in accordance with Legendre's theorem

$$\begin{aligned}
\mathcal{V}_p(n!) &= \sum_{t=1}^{k} \left\lfloor \frac{n}{p^t} \right\rfloor = \sum_{t=1}^{k} \left\lfloor \frac{d_k p^k + d_{k-1} p^{k-1} + \ldots + d_1 p + d_0}{p^t} \right\rfloor \\
&= \sum_{t=1}^{k} \left\lfloor (d_k p^{k-t} + d_{k-1} p^{(k-1)-t} + \ldots + d_{t+1} p + d_t) + (d_{t-1} p^{-1} + \ldots + d_1 p^{1-t} + d_0 p^{-t}) \right\rfloor \\
&= \sum_{t=1}^{k} \Big(d_k p^{k-t} + d_{k-1} p^{(k-1)-t} + \ldots + d_{t+1} p + d_t \Big) = \sum_{t=1}^{k} \Big(d_t (1 + p + p^2 + \ldots + p^{t-1}) \Big) \\
&= \sum_{t=1}^{k} d_t \frac{p^t - 1}{p - 1} = \frac{1}{p-1} \sum_{t=1}^{k} (d_t p^t - d_t) = \frac{\sum_{t=0}^{k} d_t p^t - \sum_{t=0}^{k} d_t}{p-1} = \frac{n - S_p(n)}{p-1}.
\end{aligned}$$

□

The point of this rigmarole is that for integral n, the equalities

$$\sum_{k \le n} \Omega(k) \;=\; \Omega(n!) \;=\; \sum_{p \le n} \mathcal{V}_p(n!)$$

become, after an application of Theorem 4.2:

$$\Omega(n!) = n \sum_{p \le n} \frac{1}{p-1} \;-\; \sum_{p \le n} \frac{S_p(n)}{p-1}. \tag{11}$$

This equation reveals how the theory of digit sums is relevant for our investigation of the factorial. In the following sections we consider the summands of equation (11) one by one.

4.2 The first summand

Using $\frac{1}{p-1} = \frac{1}{p} + \frac{1}{p(p-1)}$ gives

$$\sum_{p \le n} \frac{1}{p-1} \;= \sum_{p \le n} \frac{1}{p} + \sum_{p \le n} \frac{1}{p(p-1)}.$$

By Mertens' theorem and Proposition 3.2, this equals

$$\begin{aligned}
&\Big(\log(\log(n)) + M + \varepsilon(n) \Big) + \left(\lambda - \sum_{p > n} \frac{1}{p(p-1)} \right) \\
&= \log(\log(n)) + \theta + \varepsilon(n) - \sum_{p > n} \frac{1}{p(p-1)},
\end{aligned}$$

where $\theta = M + \lambda$. Therefore, the first summand of equation (11) is

$$n \log(\log(n)) + \theta n + n\varepsilon(n) - n \sum_{p>n} \frac{1}{p(p-1)}.$$

At this point, we can infer from Theorem 3.4 that $n\varepsilon(n) - n \sum\limits_{p>n} \frac{1}{p(p-1)}$ plus the second term of equation (11) is $O(\frac{n}{\log(n)})$. However, we know that the size of Mertens' error $\varepsilon(n)$ is better than $O(\frac{1}{\log^k(n)})$ for any $k > 0$. Also, it can be shown that

$$n \sum_{p>n} \frac{1}{p(p-1)}$$

vanishes as n goes to infinity (see Appendix). Therefore, the major contribution to the error $O(\frac{n}{\log(n)})$ appearing in Theorem 3.4 must be coming from the second term of equation (11).

Before we go on to the next section, we list some numbers. The quantity $\kappa(n) = n \sum_{p>n} \frac{1}{p(p-1)}$ is positive and vanishes. Row 2 of Figure 5 gives the integer N such that $\kappa(n)$ is less than or equal to corresponding real number from row 1 for all $n \geq N$. Note that this is not a proof, only computational evidence for $n \leq 5 \cdot 10^6$.

$r > 0$	0.5	0.4	0.3	0.2	0.1
N	3	5	11	59	8689

Figure 5

4.3 The second summand

The second summand of equation (11) is

$$\sum_{p \leq n} \frac{S_p(n)}{p-1}.$$

This expression (in variable n) is unbounded and goes to infinity as $n \to +\infty$ (being pointwise above the series in Mertens' theorem). Note that this expression is not monotonically increasing, and that every term of the sum is dependent on n. Again, using $\frac{1}{p-1} = \frac{1}{p} + \frac{1}{p(p-1)}$, this expands into

$$\sum_{p \leq n} \frac{S_p(n)}{p} + \sum_{p \leq n} \frac{S_p(n)}{p(p-1)}.$$

The rightmost of these terms grows very slowly, so we ignore it for now and focus on the leftmost one. Suppose that n_0 is some fixed positive integer, and consider the similar looking expression

$$\sum_{p\leq n}\frac{S_p(n_0)}{p}.$$

Contrary to the previous expression, the summands here are not dependent on n. We rewrite it in terms of the indicator function $1_{\mathbb{P}}$ of the prime numbers as

$$\sum_{t=1}^{n}\frac{1_{\mathbb{P}}(t)S_t(n_0)}{t}.$$

We may let $(a_t)_{t\in\mathbb{N}}$ be the sequence defined by $a_t = 1_{\mathbb{P}}(t)S_t(n_0)$, and $\phi : (0,\infty) \to (0,\infty)$ be the function $\phi : t \mapsto \frac{1}{t}$. Then ϕ is continuously differentiable on $[1,n]$ and we can apply Abel's summation formula with

$$A(T) = \sum_{t=1}^{T} a_t = \sum_{t=2}^{T} 1_{\mathbb{P}}(t)S_t(n_0) \quad \text{and} \quad \frac{d\phi}{dt} = -\frac{1}{t^2}$$

to deduce

$$\begin{aligned}\sum_{t=1}^{n}\frac{1_{\mathbb{P}}(t)S_t(n_0)}{t} &= \sum_{t=1}^{n} a_t\phi(t) \;=\; a_n\phi(n) + \int_1^n \frac{A(\lfloor x\rfloor)}{x^2}dx \\ &= \frac{1}{n}\sum_{p\leq n} S_p(n_0) \;+\; \int_1^n \frac{\sum\limits_{p\leq x} S_p(n_0)}{x^2}dx.\end{aligned}$$

The point we wish to make is that in the deduction above, there are no restrictions on the positive integer n_0, so we are justified in setting n_0 equal to n. The effect is that we have showed that $\sum\limits_{p\leq n}\frac{S_p(n)}{p}$ is equal to the gruesome looking expression

$$\frac{1}{n}\sum_{p\leq n} S_p(n) \;+\; \int_1^n \frac{\sum\limits_{p\leq x} S_p(n)}{x^2}dx. \tag{12}$$

Notice that the summand to the left is just $\frac{1}{n}h(n)$, where h is the function we studied in section 2.4. The integral to the right is of a curious nature since n occurs in both the upper limit of integration and in the integrand. Is not a «trivial» integral. It is an instance of a function of the form $F(n) = \int_1^n f_n(x)dx$, where $(f_n)_{n\in\mathbb{N}}$ are different functions. Notice also that the numerator of the integrand is just a «partial sum» of $\sum_{p\leq n} S_p(n)$, i.e. the sum is just chopped of at the point x.

As a final effort toward a resolution of equation (11), we will study the integral from (12) in the next section. First however, we tie up one loose end. At the beginning of this section, we said that $\sum\limits_{p\le n} \frac{S_p(n)}{p(p-1)}$ grows very slowly. First of all, notice that this expression is bounded below by the converging series $\sum_{p\le n} \frac{1}{p(p-1)} \to \lambda = 0.773\ldots$. Also, by Proposition 2.5, an upper bound for the numerator gives

$$\begin{aligned}\sum_{p\le n} \frac{S_p(n)}{p(p-1)} &\le \sum_{p\le n} \frac{(p-1)(\lfloor \log_p(n) \rfloor + 1)}{p(p-1)} \\ &\le \sum_{p\le n} \frac{\log_p(n)+1}{p} \\ &= \log(n) \sum_{p\le n} \frac{1}{p\log(p)} + \sum_{p\le n} \frac{1}{p} \\ &\le (\log(n)+1) \sum_{p\le n} \frac{1}{p},\end{aligned}$$

so by Mertens' theorem this term grows at worst like $\log(n)\log(\log(n))$.

4.4 A digit sum integral

Before we can reach a conclusion similar to that of Theorem 3.4, we have to consider the integral from equation (12) given by

$$\mathcal{J}(n) := \int_1^n \frac{\sum\limits_{p\le x} S_p(n)}{x^2} dx.$$

The main goal of this section is to prove that $\mathcal{J}(n) \sim \mu \frac{n}{\log(n)}$, where μ is a constant approximately equal to 0.2453. During our proof, we will need the following lemma, whose proof can be found in the Appendix.

Lemma 4.3 (Mertens- type result). *Let t be a fixed positive integer. Then*

$$\sum_{\frac{n}{t+1} < p \le \frac{n}{t}} \frac{1}{p} = a_t \frac{1}{\log(n)} + b_t \frac{1}{\log^2(n)} + o\left(\frac{1}{\log^2(n)}\right)$$

as n tends to infinity, where

$$a_t = \log\left(1+\frac{1}{t}\right) \quad \textit{and} \quad b_t = \frac{1}{2}\log\left(1+\frac{1}{t}\right)\log\Big(t(t+1)\Big).$$

Proposition 4.4. *$\mathcal{J}(n)$ has the asymptotic expansion*

$$\mathcal{J}(n) = \mu \frac{n}{\log(n)} + O\left(\frac{n}{\log^2(n)}\right), \tag{13}$$

where $\mu = \frac{\pi^2}{12} - \gamma = 0.2452\ldots$, γ denoting the Euler-Macheroni constant.

Proof. Let $N = \lfloor \sqrt{n} \rfloor - 1$, so

$$\mathcal{J}(n) = \int_1^{\frac{n}{N+1}} \frac{\sum_{p \le x} S_p(n)}{x^2} dx + \sum_{t=1}^{N} \int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\sum_{p \le x} S_p(n)}{x^2} dx. \tag{14}$$

By propositions 2.6, 2.8 and 2.7, whenever $x \le \frac{n}{N+1}$ we have

$$\sum_{p \le x} S_p(n) \le \sum_{p \le x} \left(\lfloor \sqrt{n} \rfloor - 1 + \frac{n}{\lfloor \sqrt{n} \rfloor} \right) < \sum_{p \le x} 2\sqrt{n} = 2\sqrt{n}\, \pi(x),$$

so the first integral of equation (14) is

$$\int_1^{\frac{n}{N+1}} \frac{\sum_{p \le x} S_p(n)}{x^2} dx \le 2\sqrt{n} \int_1^{\frac{n}{N+1}} \frac{\pi(x)}{x^2} dx \le 2\sqrt{n} \int_1^{\frac{n}{\lfloor \sqrt{n} \rfloor}} \frac{dx}{x}$$

$$= 2\sqrt{n} \log\left(\frac{n}{\lfloor \sqrt{n} \rfloor}\right) = O\left(\sqrt{n} \log(n)\right). \tag{15}$$

To calculate the sum of integrals from (14), note that for $t = 1, 2, \ldots, N$, we can split up the integrals as

$$\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\sum_{p \le x} S_p(n)}{x^2} dx$$

$$= \int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\sum_{p \le \frac{n}{t+1}} S_p(n) + \sum_{\frac{n}{t+1} < p \le x} S_p(n)}{x^2} dx$$

$$= \left(\sum_{p \le \frac{n}{t+1}} S_p(n) \right) \int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{dx}{x^2} + \int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\sum_{\frac{n}{t+1} < p \le x} S_p(n)}{x^2} dx$$

$$= \left(\sum_{p \le \frac{n}{t+1}} S_p(n) \right) \frac{1}{n} + \int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\sum_{\frac{n}{t+1} < p \le x} S_p(n)}{x^2} dx. \tag{16}$$

By recycling some of the proof of Proposition 2.12, with appropriate modifications[11], it is not too difficult to show that

$$\sum_{p \leq \frac{n}{t+1}} S_p(n) = \left[\delta - \sum_{k=1}^{t} \frac{1}{2k(k+1)^2}\right] \frac{n^2}{\log(n)} + \left[C - \sum_{k=1}^{t} kc_k\right] \widetilde{O}_n\left(\frac{n^2}{\log^2(n)}\right),$$

where δ, C and the c_k are the constants from Proposition 2.12. From this it follows that

$$\begin{aligned}\frac{1}{n}\sum_{p \leq \frac{n}{t+1}} S_p(n) &= \left[\delta - \sum_{k=1}^{t} \frac{1}{2k(k+1)^2}\right] \frac{n}{\log(n)} + \left[C - \sum_{k=1}^{t} kc_k\right] \widetilde{O}_n\left(\frac{n}{\log^2(n)}\right) \\ &= \left[\sum_{k=t+1}^{\infty} \frac{1}{2k(k+1)^2}\right] \frac{n}{\log(n)} + \left[\sum_{k=t+1}^{\infty} kc_k\right] \widetilde{O}_n\left(\frac{n}{\log^2(n)}\right).\end{aligned} \tag{17}$$

Next, we show

$$\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\sum_{\frac{n}{t+1} < p \leq x} S_p(n)}{x^2} dx = \left[\log\left(1 + \frac{1}{t}\right) - \frac{2t+3}{2(t+1)^2}\right] \frac{n}{\log(n)} + O\left(\frac{n}{\log^2(n)}\right). \tag{18}$$

To that end, let $S(x) = \sum_{p \leq x} p$ be the function from section 2.4. By an application of Proposition 2.6, the numerator of the integrand can be expressed

$$\begin{aligned}\sum_{\frac{n}{t+1} < p \leq x} S_p(n) &= \sum_{\frac{n}{t+1} < p \leq x} (n + t - tp) \\ &= (n+t) \sum_{\frac{n}{t+1} < p \leq x} 1 \;-\; t \sum_{\frac{n}{t+1} < p \leq x} p \\ &= (n+t)\left(\pi(x) - \pi\left(\frac{n}{t+1}\right)\right) - t\left(S(x) - S\left(\frac{n}{t+1}\right)\right) \\ &= (n+t)\pi(x) - tS(x) + \left(tS\left(\frac{n}{t+1}\right) - (n+t)\pi\left(\frac{n}{t+1}\right)\right).\end{aligned}$$

This gives us

$$\begin{aligned}\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\sum_{\frac{n}{t+1} < p \leq x} S_p(n)}{x^2} dx &= (n+t) \int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\pi(x)}{x^2} dx - t \int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{S(x)}{x^2} dx \\ &\quad + \frac{1}{n}\left(tS\left(\frac{n}{t+1}\right) - (n+t)\pi\left(\frac{n}{t+1}\right)\right).\end{aligned} \tag{19}$$

[11] I.e. modify the lower index of summation of equation (6).

We do as usual and consider the summands of equation (19) one by one, starting with the first. Let $\phi : t \mapsto -\frac{1}{t}$. Then $\frac{d\phi}{dt} : t \mapsto \frac{1}{t^2}$ and Abel-summation gives

$$\begin{aligned}\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\pi(x)}{x^2}dx &= \pi\left(\frac{n}{t}\right)\phi\left(\frac{n}{t}\right) - \pi\left(\frac{n}{t+1}\right)\phi\left(\frac{n}{t+1}\right) - \sum_{\frac{n}{t+1}<k\leq\frac{n}{t}} 1_{\mathbb{P}}(k)\phi(k) \\ &= -\frac{\pi(\frac{n}{t})}{\frac{n}{t}} + \frac{\pi(\frac{n}{t+1})}{\frac{n}{t+1}} + \sum_{\frac{n}{t+1}<p\leq\frac{n}{t}} \frac{1}{p}.\end{aligned}$$

Into the above equation we now substitute

$$\frac{\pi(\frac{n}{k})}{\frac{n}{k}} = \frac{1}{\log(n)} + \widetilde{O}_n\left(\left[1+\log(k)\right]\frac{1}{\log^2(n)}\right)$$

for $k = t$ and $k = t+1$, as well as the expression from Lemma 4.3. It then turns into

$$\log\left(1+\frac{1}{t}\right)\frac{1}{\log(n)} + \widetilde{O}_n\left(r_t\frac{1}{\log^2(n)}\right),$$

where

$$r_t = \log\left(1+\frac{1}{t}\right) + \frac{1}{2}\log^2(t+1) - \frac{1}{2}\log^2(t).$$

From this it follows that

$$(n+t)\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\pi(x)}{x^2}dx = \log\left(1+\frac{1}{t}\right)\frac{n}{\log(n)} + \widetilde{O}_n\left(r_t\frac{n}{\log^2(n)}\right). \tag{20}$$

In a similar vein, an application of (8) gives

$$\begin{aligned} & t\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{S(x)}{x^2}dx \\ &= t\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\frac{x^2}{2\log(x)} + O(\frac{x^2}{\log^2(x)})}{x^2}dx \\ &= \frac{t}{2}\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{dx}{\log(x)} + O\left(\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{dx}{\log^2(x)}\right) \\ &= \frac{t}{2}\left(\mathrm{li}\left(\frac{n}{t}\right) - \mathrm{li}\left(\frac{n}{t+1}\right)\right) + O\left(\frac{n}{\log^2(n)}\right) \\ &= \frac{t}{2}\left(\frac{\frac{n}{t}}{\log(\frac{n}{t})} - \frac{\frac{n}{t+1}}{\log(\frac{n}{t+1})} + O\left(\frac{n}{\log^2(n)}\right)\right) + O\left(\frac{n}{\log^2(n)}\right)\end{aligned}$$

$$= \frac{t}{2}\left(\frac{1}{t(t+1)}\frac{n}{\log(n)} + O\left(\frac{n}{\log^2(n)}\right)\right) + O\left(\frac{n}{\log^2(n)}\right)$$
$$= \frac{1}{2(t+1)}\frac{n}{\log(n)} + O\left(\frac{n}{\log^2(n)}\right), \tag{21}$$

where we used that the *Logarithmic integral* li satisfies $\mathrm{li}(u) = \frac{u}{\log(u)} + O(\frac{u}{\log^2(u)})$ and that $\mathrm{li}(x) - \frac{x}{\log(x)}$ is an antiderivative of $\frac{1}{\log^2(x)}$. For the last summand of equation (19) we have

$$\frac{1}{n}\left(tS\left(\frac{n}{t+1}\right) - (n+t)\pi\left(\frac{n}{t+1}\right)\right)$$
$$= \frac{t}{n}\left(\frac{(\frac{n}{t+1})^2}{2\log(\frac{n}{t+1})} + \frac{(\frac{n}{t+1})^2}{4\log^2(\frac{n}{t+1})} + \widetilde{O}_n\left(\frac{(\frac{n}{t+1})^2}{4\log^3(\frac{n}{t+1})}\right)\right)$$
$$- \frac{n+t}{n}\left(\frac{\frac{n}{t+1}}{\log(\frac{n}{t+1})} + \widetilde{O}_n\left(\frac{\frac{n}{t+1}}{\log^2(\frac{n}{t+1})}\right)\right)$$
$$= -\frac{t+2}{2(t+1)^2}\frac{n}{\log(n)} + \widetilde{O}_n\left(s_t\frac{n}{\log^2(n)}\right), \tag{22}$$

where

$$s_t = -\frac{(t+2)\log(t+1) + \frac{3}{2}t + 2}{2(t+1)^2}.$$

We begin putting our obtained results together. When we substitute the results from (20) - (22) into equation (19), the result is

$$\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\sum\limits_{\frac{n}{t+1}<p\leq x} S_p(n)}{x^2}dx = \left[\log\left(1+\frac{1}{t}\right) - \frac{2t+3}{2(t+1)^2}\right]\frac{n}{\log(n)} + O\left(\frac{n}{\log^2(n)}\right). \tag{23}$$

Then, substituting (17) and (23) into equation (16), we get

$$\int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\sum\limits_{p\leq x} S_p(n)}{x^2}dx$$
$$= \left[\log\left(1+\frac{1}{t}\right) - \frac{2t+3}{2(t+1)^2} + \sum_{k=t+1}^{\infty}\frac{1}{2k(k+1)^2}\right]\frac{n}{\log(n)} + O\left(\frac{n}{\log^2(n)}\right). \tag{24}$$

This allows us to calculate the following sum from equation (14):

$$\sum_{t=1}^{N} \int_{\frac{n}{t+1}}^{\frac{n}{t}} \frac{\sum_{p \le x} S_p(n)}{x^2} dx$$
$$= \sum_{t=1}^{N} \left[\log\left(1+\frac{1}{t}\right) - \frac{2t+3}{2(t+1)^2} + \sum_{k=t+1}^{\infty} \frac{1}{2k(k+1)^2} \right] \frac{n}{\log(n)} + O\left(\frac{n}{\log^2(n)}\right). \tag{25}$$

At last, substituting (15) and (25) into equation (14), we have

$$\mathcal{J}(n) = \sum_{t=1}^{N} \left[\log\left(1+\frac{1}{t}\right) - \frac{2t+3}{2(t+1)^2} + \sum_{k=t+1}^{\infty} \frac{1}{2k(k+1)^2} \right] \frac{n}{\log(n)} + O\left(\frac{n}{\log^2(n)}\right). \tag{26}$$

Since the limit

$$\mu := \lim_{N\to\infty} \sum_{t=1}^{N} \left[\log\left(1+\frac{1}{t}\right) - \frac{2t+3}{2(t+1)^2} + \sum_{k=t+1}^{\infty} \frac{1}{2k(k+1)^2} \right]$$

simplifies (see Appendix) to

$$\mu \ = \ \frac{\pi^2}{12} - \gamma,$$

where γ denotes the famous Euler-Mascheroni constant, we get the desired result from (26). □

4.5 Conclusion using digit sums

We have showed $\Omega(n!)$ is exactly equal to

$$n\log(\log(n)) \ + \ \theta n + \ n\varepsilon(n) - n\sum_{p>n} \frac{1}{p(p-1)} - \sum_{p\le n} \frac{S_p(n)}{p(p-1)}$$
$$- \frac{1}{n}\sum_{p\le n} S_p(n) - \int_1^n \frac{\sum_{p\le x} S_p(n)}{x^2} dx.$$

From our work with these expressions we know know that

$$\left(n\varepsilon(n) - n\sum_{p>n} \frac{1}{p(p-1)} - \sum_{p\le n} \frac{S_p(n)}{p(p-1)} \right) = O\left(\frac{n}{\log^k(n)}\right)$$

for any $k > 0$. From our work in section 2.4, we have

$$\frac{1}{n} \sum_{p \leq n} S_p(n) = \delta \frac{n}{\log(n)} + C \frac{n}{\log^2(n)} + o\left(\frac{n}{\log^2(n)}\right),$$

and from the calculations of the previous section

$$\int_1^n \frac{\sum_{p \leq x} S_p(n)}{x^2} dx = \mu \frac{n}{\log(n)} + O\left(\frac{n}{\log^2(n)}\right),$$

where $\delta = 1 - \frac{\pi^2}{12}$, C and $\mu = \frac{\pi^2}{12} - \gamma$ are constants. Setting

$$\Delta := -(\delta + \mu) = \gamma - 1 = -0.4227\ldots,$$

the pinnacle of our effort is summarized in the equation

$$\sum_{k \leq n} \Omega(k) = \Omega(n!) = n \log(\log(n)) + \theta n + \Delta \frac{n}{\log(n)} + O\left(\frac{n}{\log^2(n)}\right).$$

Even though our approach using digit sums was long and tedious, it did indeed allow us to calculate the next term of the asymptotic expansion.

5 The full asymptotic expansion

We consider in this section, a complete asymptotic expansion for the average order of ω and Ω. First we make very explicit the fact we observed in the proof of Theorem 3.4, and mentioned at the beginning of section 2.3:

Lemma 5.1. *If*

$$A(n) = \sum_{k\leq n} \Omega(k) - \sum_{k\leq n} \omega(k),$$

then

$$A(n) = \lambda n + O(\sqrt{n}),$$

where λ is the constant from Proposition (3.2).

The $O(\sqrt{n})$ error is in fact much smaller than the error we will obtain from the asymptotic expansions under consideration. Therefore, as $n \to +\infty$, we consider these problems equivalent as an asymptotic expansion for one will also give an asymptotic expansion for the other.

Even though the result we are interested in is not explicitly stated in [8], it is an easy consequence of the following theorem (Theorem 1 is Saffari's paper):

Theorem 5.2 (B. Saffari). *For two given, relatively prime integers $k, l, (k \geq 1)$, let $\omega_{k,l}(n)$ be the number of distinct prime divisors p of n such that $p \equiv l \pmod{k}$. Then, as $x \to +\infty$, we have for every whole integer $m \geq 1$:*

$$\sum_{1\leq n\leq x} \omega_{k,l}(n) = \frac{x\log(\log(x))}{\varphi(k)} + B_{k,l}x + \sum_{r=1}^{m} \frac{C_r}{\varphi(k)} \frac{x}{(\log(x))^r} + O\left[\frac{x}{(\log(x))^{m+1}}\right],$$

$B_{k,l}$ being a constant depending on k and l, $\varphi(k)$ Euler's totient function, and the constants C_r being defined in the following way:

$$C_r = -\int_1^\infty \frac{\{\, t \,\}}{t^2}(\log(t))^{r-1}dt = \frac{(-1)^{r-1}}{r} \cdot \frac{d^r}{ds^r}\left(\frac{(s-1)\zeta(s)}{s}\right)_{s=1}.$$

In particular, $C_1 = \gamma - 1$.

Here, $\zeta(s)$ denotes the Riemann-zeta function.

We will not go through the proof of Theorem 5.2, but point out that it makes clever use of a method known as the «Dirichlet hyperbola method». It is based on the observation that $\omega_{k,l}$ is the product of $1_{k,l}$ and 1 under Dirichlet-convolution, where 1 denotes the identity function and $1_{k,l}$ denotes the indicator function of the primes p for which $p \equiv l \pmod{k}$. However, we easily obtain:

Theorem 5.3. *As $x \to +\infty$, we have for every integer $m \geq 1$:*

$$\sum_{1\leq n\leq x} \omega(n) = x\log(\log(x)) + Mx + \sum_{r=1}^{m} C_r \frac{x}{(\log(x))^r} + O\left[\frac{x}{(\log(x))^{m+1}}\right],$$

where M denotes Mertens' constant and C_r are the constants defined above.

Proof.

$$\sum_{1\le n\le x}\omega(n)=\frac{x}{2}+O(1)+\sum_{1\le n\le x}\omega_{1,4}(n)+\sum_{1\le n\le x}\omega_{3,4}(n),$$

where $\frac{x}{2}+O(1)$ is the number of even positive integers less than or equal to x, and $\omega_{1,4}(n)$ and $\omega_{3,4}(n)$ count the (distinct) odd prime factors over the same interval, that are $\equiv 1 \pmod 4$ and $\equiv 3 \pmod 4$, respectively[12]. Substituting the result from Theorem 5.2 together with $\varphi(4)=2$ yields

$$\sum_{1\le n\le x}\omega(n)=x\log(\log(x))+\left(\frac{1}{2}+B_{1,4}+B_{3,4}\right)x$$
$$+\sum_{r=1}^{m}C_r\frac{x}{(\log(x))^r}+O\left[\frac{x}{(\log(x))^{m+1}}\right].$$

Since we know from Theorem 3.4 that the coefficient of the x-term is M, the proof is complete. □

We have at once:

Theorem 5.4. *As $x\to+\infty$, we have for every integer $m\ge 1$:*

$$\sum_{1\le n\le x}\Omega(n)=x\log(\log(x))+\theta x+\sum_{r=1}^{m}C_r\frac{x}{(\log(x))^r}+O\left[\frac{x}{(\log(x))^{m+1}}\right],$$

where θ is the constant from Definition (3.3).

Proof. Follows by substituting the results from Lemma 5.1 and Theorem 5.3 into the equation

$$\sum_{1\le n\le x}\Omega(n)=A(n)+\sum_{1\le n\le x}\omega(n),$$

and using the fact that $\theta=M+\lambda$. □

Notice that the C_r are all negative and increasing in absolute value, so the $x\log(\log(x))+\theta x$ is ultimately an overestimate of the actual value. We have the following bound[13] for the constants:

$$|C_r|\le(r-1)!$$

Indeed,

$$|C_r|\le\int_1^\infty\frac{\log(t)^{r-1}}{t^2}dt=\int_0^\infty\frac{x^{r-1}}{e^x}dx=(r-1)!,$$

for $r=1,2,3\ldots$, where we utilized the substitution $t=e^x$.

[12] This includes all the primes.

[13] In actuality, $|C_r|$ is closer to $\frac{1}{2}(r-1)!$; The $\{t\}$ in the integrand almost correspond to multiplying the integral by $\frac{1}{2}$.

Let us calculate some of the constants C_r, using:

Lemma 5.5. *For $r \in \mathbb{N}$ and $n \in \mathbb{N}_0$, the following are identities:*

$$(i) \qquad \int \frac{\log(t)^{r-1}}{t} dt = \frac{1}{r}\log(t)^r + const.$$

$$(ii) \qquad \int \frac{\log(t)^n}{t^2} dt = -\frac{\log(t)^n}{t} + n \int \frac{\log(t)^{n-1}}{t^2} dt + const.$$

By substituting $n = r-1$ in equation (ii) and repeatedly expanding the integral, we infer:

$$\int \frac{\log(t)^{r-1}}{t^2} dt = -\sum_{l=0}^{r-1} \frac{(r-1)!}{l!} \frac{\log(t)^l}{t} + const. \tag{27}$$

Therefore

$$\begin{aligned}
C_r &= -\int_1^\infty \frac{\{t\}}{t^2} (\log(t))^{r-1} dt \\
&= \sum_{k=1}^\infty \int_k^{k+1} \frac{\lfloor t \rfloor - t}{t^2} (\log(t))^{r-1} dt \\
&= \sum_{k=1}^\infty \int_k^{k+1} \frac{k-t}{t^2} (\log(t))^{r-1} dt \\
&= \sum_{k=1}^\infty \int_k^{k+1} k\frac{\log(t)^{r-1}}{t^2} - \frac{\log(t)^{r-1}}{t} dt \\
&= \sum_{k=1}^\infty \left[-k \sum_{l=0}^{r-1} \frac{(r-1)!}{l!} \frac{\log(t)^l}{t} - \frac{1}{r}\log(t)^r \right]_{t=k}^{t=k+1} \\
&= \sum_{k=1}^\infty \left[\frac{\log(k)^r - \log(k+1)^r}{r} + \sum_{l=0}^{r-1} \frac{(r-1)!}{l!} \left(\log(k)^l - \frac{k}{k+1}\log(k+1)^l \right) \right] \\
&= \sum_{k=1}^\infty \left[\frac{\alpha_k^r - \alpha_{k+1}^r}{r} + \sum_{l=0}^{r-1} \frac{(r-1)!}{l!} \left(\alpha_k^l - \frac{k}{k+1}\alpha_{k+1}^l \right) \right],
\end{aligned}$$

where we utilized part (i) of the lemma and equation (27) in the fourth line, and set $\alpha_m = \log(m)$ for better readability. This is very computable. We obtain the first constants, as shown in Figure 6.

r	1	2	3	4	5
C_r	-0.42278	-0.49559	-1.00078	-2.99897	-11.97534

r	6	7	8	9	10
C_r	-59.62423	-354.26889	-2431.84993	-18791.20146	-159952.86263

Figure 6: First 10 constants C_r, rounded to 5 decimal places.

6 Summary

In this paper we have seen that the factorial $n! = 1 \cdot 2 \cdots n$ has roughly $n \log(\log(n))$ prime factors in total, as n goes to infinity. Or equivalently, that the average order of Ω and ω is $\log(\log(n))$. We have also examined interesting properties of digit sums, and seen a link between these matters via Legendre's formula. Historically, we have found that the «problem» of the compositeness of $n!$ can be traced back, at the very least, to the beginning of the nineteenth century.

This project has thought us a lot about scientific work, and there is a lot more work behind it than it may seem. Much time was spent typesetting, reading ancient articles in number theory and writing numerical programs to verify results. We owe a special thanks to the community of Mathematics Stack Exchange for being helpful throughout.

Given the limited time at disposal, our project naturally doesn't cover every piece of the puzzle. If we were to continue, we could probably make a complete exposition of the history of this problem. At least two questions remain unsolved: 1. What is this result using «transcendental methods» that Hardy and Ramanujan speak of? And if there was such a result, why was it not published? 2. Is it a coincidence that, nowhere in the literature the problem is mentioned with the point of view, of calculating the total number of prime factors of $n!$? (pun intended).

List of symbols

General

$\lfloor x \rfloor$	Integer part of x
$\{x\}$	Fractional part of x
$S_k(n)$	Digit sum of n in base k
$S_p(n)$	Digit sum of n in base p, p a prime number
$\pi(n)$	Prime counting function
$\omega(n)$	Number of distinct prime factors of n
$\Omega(n)$	Total number of prime factors of n
$n!$	Factorial of n
$S(x)$	Sum of all primes less than or equal to x
$\mathcal{A}(n, N)$	Number of integers in $(\frac{n}{N+1}, \frac{n}{N}]$
$p(n, k)$	Number of primes in $(\frac{n}{k+1}, \frac{n}{k}]$
$D(n)$	$\sum_{2 \leq k \leq n} S_k(n)$
$h(n)$	$\sum_{p \leq n} S_p(n)$
$\varepsilon(n)$	Error term in Mertens' theorem
$A(n)$	The difference $\sum_{k \leq n} \Omega(n) - \sum_{k \leq n} \omega(n)$
$\log(n)$	Natural logarithm
$\log_m(n)$	Base m logarithm ($= \frac{\log(n)}{\log(m)}$)
$\mathcal{V}_p(n)$	p-adic valuation of n with respect to the prime p
$\kappa(n)$	The vanishing quantity $n \sum_{p>n} \frac{1}{p(p-1)}$
$1_{\mathbb{P}}(n)$	Indicator function of the prime numbers
$\mathbb{N}, \mathbb{N}_0, \mathbb{N}_{\geq 2}$	The positive integers, including zero or excluding 1
$\mathcal{J}(n)$	$\int_1^n x^{-2} \sum_{p \leq x} S_p(n) dx$
$\mathrm{li}(n)$	«Offset» Logarithmic integral $\int_2^n \frac{1}{\log(x)} dx$
$\omega_{k,l}(n)$	Number of prime factors p of n with $p \equiv l \pmod{k}$
$a \equiv b \pmod{n}$	Modular congruence
$\varphi(n)$	Euler's totient function
$\zeta(s)$	Riemann-zeta function

Symbols of asymptotic analysis

We use the following symbols for functions $f : S \subseteq (-\infty, \infty) \to (-\infty, \infty)$, where S contains arbitrary large, positive numbers. In general, the «o-notations» mean both a collection of functions and *some* element of that collection.

$f(n) \sim g(n)$	Asymptotic equivalence; $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$
$f(n,t) = \widetilde{O}_n(g(n,t))$	Means $f(n) \sim g(n)$, when t is considered fixed
$f(n) = o(g(n))$	Little o-notation; $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$
$f(n) = O(g(n))$	Big O-notation; There exists real numbers N and $C \geq 0$, such that $n \geq N$ implies $\lvert f(n) \rvert \leq C \lvert g(n) \rvert$

Constants

δ	$= 1 - \frac{\pi^2}{12} = 0.1775\ldots$
C	$= 0.1199\ldots$
M	$= 0.2614\ldots$ (Mertens' constant)
λ	$= \sum_p \frac{1}{p(p-1)} = 0.7731\ldots$
θ	$= M + \lambda = 1.0346\ldots$
γ	$= 0.5772\ldots$ (Euler-Mascheroni constant)
μ	$= \frac{\pi^2}{12} - \gamma = 0.2452\ldots$
Δ	$= -(\delta + \mu) = \gamma - 1 = -0.4227\ldots$
$B_{k,l}$	Constant(s) depending only on k and l
C_r	Constant(s) defined by $-\int_1^\infty t^{-2}\{t\}\log(t)^{r-1}dt$

Appendix

Equation 4, chapter 2.3

$$\sum_{\frac{n}{N+1}<k\leq \frac{n}{N}} S_k(n) = \frac{1}{2N(N+1)^2}n^2 + O(n)$$

Proof. In section 2.3 we showed that

$$\sum_{\frac{n}{N+1}<k\leq \frac{n}{N}} S_k(n) = (N+n)\mathcal{A}(n,N) - N \sum_{\frac{n}{N+1}<k\leq \frac{n}{N}} k, \tag{28}$$

where $\mathcal{A}(n,N)$ is the number of integers in $(\frac{n}{N+1}, \frac{n}{N}]$. Throughout the rest of this proof, let n_N and n_{N+1} denote $(n \mod N)$ and $(n \mod (N+1))$, respectively. The red expression we derived for $\mathcal{A}(n,N)$ in section 2.3 then becomes

$$\mathcal{A}(n,N) = \frac{n}{N(N+1)} + \frac{n_{N+1}}{N+1} - \frac{n_N}{N},$$

and so

$$\begin{aligned}(N+n)\mathcal{A}(n,N) &= (N+n)\left(\frac{n}{N(N+1)} + \frac{n_{N+1}}{N+1} - \frac{n_N}{N}\right) \\ &= \frac{1}{N(N+1)}n^2 + \left(\frac{1+n_{N+1}}{N+1} - \frac{n_N}{N}\right)n + \left(\frac{Nn_{N+1}}{N+1} - n_N\right).\end{aligned}$$

To calculate $N \sum_{\frac{n}{N+1}<k\leq \frac{n}{N}} k$ we use that the smallest and largest integer of this half open interval of summation is

$$L = \frac{n+(N+1)-n_{N+1}}{N+1} \qquad \text{and} \qquad U = \frac{n-n_N}{N},$$

respectively (as calculated in section 2.3). Using that the sum of the first m positive integers is $\frac{1}{2}m(m+1)$ gives

$$\begin{aligned}N \sum_{\frac{n}{N+1}<k\leq \frac{n}{N}} k &= N\left[\frac{1}{2}U(U+1) - \frac{1}{2}L(L+1)\right] \\ &= \frac{2N+1}{2N(N+1)^2}n^2 - \frac{f_1(n,N)}{2N(N+1)^2}n + \frac{f_2(n,N)}{2N(N+1)^2},\end{aligned}$$

where

$$\begin{aligned}f_1(n,N) &= 2N^3 + (1+2n_N-2n_{N+1})N^2 + (4n_N-1)N + 2n_N \qquad \text{and} \\ f_2(n,N) &= -2N^4 - (n_N+4-3n_{N+1})N^3 + (n_N^2 - 2n_N - 2 - n_{N+1}^2 + 3n_{N+1})N^2 \\ &\quad + (2n_N^2 - n_N)N + n_N^2.\end{aligned}$$

Substituting these formulae into (28) yields

$$\sum_{\frac{n}{N+1}<k\le\frac{n}{N}} S_k(n) = \frac{1}{2N(N+1)^2}n^2 + \frac{f_3(n,N)}{2(N+1)^2}n + \frac{f_4(n,N)}{2N(N+1)^2},$$

where

$$\begin{aligned} f_3(n,N) =& 2N^3+3N+(1+2n_{N+1}) \\ f_4(n,N) =& 2N^4+(5n_{N+1}-3n_N-4)N^3+(n_{N+1}^2-n_N^2-n_{N+1} \\ & -2n_N+2)N^2-(n_N+2n_N)N-n_N^2. \end{aligned}$$

Since this is just $\frac{1}{2N(N+1)^2}n^2+O(n)$, the proof is complete. □

Simplification of C
If

$$c_t = \frac{2\log(t)+1}{4t^2} - \frac{2\log(t+1)+1}{4(t+1)^2},$$

then

$$\begin{aligned} tc_t &= \frac{2\log(t)+1}{4t} - \frac{2t\log(t+1)+t}{4(t+1)^2} \\ &= \frac{\log(t)}{2t} + \frac{1}{4t} - \frac{t\log(t+1)}{2(t+1)^2} - \frac{t}{4(t+1)^2} - \frac{\log(t+1)}{2(t+1)^2} + \frac{\log(t+1)}{2(t+1)^2} \\ &= \frac{\log(t)}{2t} - \frac{\log(t+1)}{2(t+1)} + \frac{1}{4t} - \frac{t}{4(t+1)^2} + \frac{\log(t+1)}{2(t+1)} \\ &= \frac{\log(t)}{2t} - \frac{\log(t+1)}{2(t+1)} + \frac{1}{4t} - \frac{t}{4(t+1)^2} + \frac{\log(t+1)}{2(t+1)} - \frac{1}{4(t+1)^2} + \frac{1}{4(t+1)^2} \\ &= \frac{\log(t)}{2t} - \frac{\log(t+1)}{2(t+1)} + \frac{1}{4t} - \frac{1}{4(t+1)} + \frac{\log(t+1)}{2(t+1)^2} + \frac{1}{4(t+1)^2}. \end{aligned}$$

This gives

$$\begin{aligned} \sum_{t=1}^{\infty} tc_t &= \sum_{t=1}^{\infty}\left(\frac{\log(t)}{2t} - \frac{\log(t+1)}{2(t+1)}\right) + \sum_{t=1}^{\infty}\left(\frac{1}{4t} - \frac{1}{4(t+1)}\right) + \sum_{t=1}^{\infty}\left(\frac{1+2\log(t+1)}{4(t+1)^2}\right) \\ &= 0 + \frac{1}{4} + \frac{1}{4}\sum_{t=1}^{\infty}\frac{1}{(t+1)^2} + \frac{1}{2}\sum_{t=1}^{\infty}\frac{\log(t+1)}{(t+1)^2} \\ &= \frac{\pi^2}{24} + \frac{1}{2}\sum_{t=2}^{\infty}\frac{\log(t)}{t^2}. \end{aligned}$$

Thus:

$$C := 1 - \sum_{t=1}^{\infty} tc_t = 1 - \frac{\pi^2}{24} - \frac{1}{2}\sum_{t=2}^{\infty}\frac{\log(t)}{t^2}.$$

Theorem (Prime number theorem, P.N.T.).

$$\pi(n) \sim \frac{n}{\log(n)},$$

as $n \to +\infty$*. Equivalently, given any* $\varepsilon > 0$*, there is a positive integer* N_ε *such that*

$$(1-\varepsilon)\frac{n}{\log(n)} \leq \pi(n) \leq (1+\varepsilon)\frac{n}{\log(n)},$$

for all $n \geq N_\varepsilon$*.*

Furthermore, it can be show that

$$\pi(n) = \frac{n}{\log(n)} + \widetilde{O}_n\left(\frac{n}{\log^2(n)}\right)$$

holds, and that the inequality

$$\frac{n}{\log(n)} < \pi(n)$$

holds for every $n \geq 17$.

Theorem (Rosser). *If* p_n *denotes the* n*-th prime, then*

$$p_n > n\log(n)$$

for every integer $n \geq 1$*.*

Proposition.

$$\lim_{n\to\infty} n\sum_{p>n}\frac{1}{p(p-1)} = 0$$

Proof.

$$n\sum_{p>n}\frac{1}{p(p-1)} < n\sum_{p>n}\frac{1}{(p-1)^2} = n\sum_{t=\pi(n)+1}^{\infty}\frac{1}{(p_t-1)^2},$$

since $p_{\pi(n)+1}$ is the smallest prime larger than n. By Rosser's theorem, this is strictly less than

$$n\sum_{t=\pi(n)+1}^{\infty}\frac{1}{(t\log(t)-1)^2} < n\sum_{t=\pi(n)+1}^{\infty}\frac{1}{t^2(\log(t)-1)^2}.$$

But when $n \geq 17$, we have $t = \pi(n)+1 > \frac{n}{\log(n)}$, so the expression above is less than

$$
\begin{aligned}
n\sum_{t=\pi(n)}^{\infty}\frac{1}{t^2(\log(\frac{n}{\log(n)})-1)^2} &= \frac{n}{(\log(\frac{n}{\log(n)})-1)^2}\sum_{t=\pi(n)}^{\infty}\frac{1}{t^2}\\
&\le \frac{n}{(\log(\frac{n}{\log(n)})-1)^2}\left(\frac{1}{\pi(n)^2}+\int_{n/\log(n)}^{\infty}\frac{1}{t^2}dt\right)\\
&\le \frac{n}{(\log(\frac{n}{\log(n)})-1)^2}\left(\frac{\log^2(n)}{n^2}+\frac{\log(n)}{n}\right)\\
&= \frac{\log^2(n)}{n\log^2(\frac{n}{e\log(n)})}+\frac{\log(n)}{\log^2(\frac{n}{e\log(n)})},
\end{aligned}
$$

a quantity that clearly vanishes. □

Lemma 4.3 (Mertens- type result). *Let t be a fixed positive integer. Then*

$$
\sum_{\frac{n}{t+1}<p\le\frac{n}{t}}\frac{1}{p}=a_t\frac{1}{\log(n)}+b_t\frac{1}{\log^2(n)}+o\left(\frac{1}{\log^2(n)}\right)
$$

as n tends to infinity, where

$$
a_t=\log\left(1+\frac{1}{t}\right)\qquad\text{and}\qquad b_t=\frac{1}{2}\log\left(1+\frac{1}{t}\right)\log\Big(t(t+1)\Big).
$$

Proof. By Mertens' theorem

$$
\sum_{p\le\frac{n}{j}}\frac{1}{p}=\log\left(\log\left(\frac{n}{j}\right)\right)+M+\varepsilon\left(\frac{n}{j}\right),
$$

for $j=t$ and $j=t+1$, where the $\varepsilon(x)$ is $O\left(\frac{1}{\log^k(x)}\right)$ for any $k>0$. Therefore, for any $k>1$:

$$
\sum_{\frac{n}{t+1}<p\le\frac{n}{t}}\frac{1}{p} = \log\left(\log\left(\frac{n}{t}\right)\right)-\log\left(\log\left(\frac{n}{t+1}\right)\right)+O\left(\frac{1}{\log^k(n)}\right),
$$

where

$$
\begin{aligned}
\log\left(\log\left(\frac{n}{t}\right)\right)-\log\left(\log\left(\frac{n}{t+1}\right)\right) &= \log\left(\frac{\log(n)-\log(t)}{\log(n)-\log(t+1)}\right)\\
= \log\left(1+\frac{\log(t+1)-\log(t)}{\log(n)-\log(t+1)}\right) &= \log\left(1+\frac{\log(1+\frac{1}{t})}{\log\left(\frac{n}{t+1}\right)}\right).
\end{aligned}
$$

Then

$$
\begin{aligned}
&\log(n) \sum_{\frac{n}{t+1} < p \leq \frac{n}{t}} \frac{1}{p} \\
&= \log(n) \log\left(1 + \frac{\log(1+\frac{1}{t})}{\log\left(\frac{n}{t+1}\right)}\right) + O\left(\frac{1}{\log^{k-1}(n)}\right) \\
&= \log\left[\left(1 + \frac{\log(1+\frac{1}{t})}{\log\left(\frac{n}{t+1}\right)}\right)^{\log(n)}\right] + O\left(\frac{1}{\log^{k-1}(n)}\right) \\
&= \log\left[\left(1 + \frac{\log(1+\frac{1}{t})}{\log\left(\frac{n}{t+1}\right)}\right)^{\log\left(\frac{n}{t+1}\right)} \left(1 + \frac{\log(1+\frac{1}{t})}{\log\left(\frac{n}{t+1}\right)}\right)^{\log(t+1)}\right] + O\left(\frac{1}{\log^{k-1}(n)}\right) \\
&\to \log\left[e^{\log(1+\frac{1}{t})}\right] \\
&= \log\left(1 + \frac{1}{t}\right),
\end{aligned}
$$

where we in the second to last line have let $n \to +\infty$ and used the identity $e = \lim_{u\to\infty}(1 + \frac{1}{u})^u$. Thus

$$
\sum_{\frac{n}{t+1} < p \leq \frac{n}{t}} \frac{1}{p} \sim \log\left(1 + \frac{1}{t}\right) \frac{1}{\log(n)}.
$$

To get the second term, consider

$$
\begin{aligned}
&\log^2(n) \left[\sum_{\frac{n}{t+1} < p \leq \frac{n}{t}} \frac{1}{p} - \log\left(1 + \frac{1}{t}\right) \frac{1}{\log(n)}\right] \\
&= \log^2(n) \left[\log\left(1 + \frac{\log\left(1+\frac{1}{t}\right)}{\log(\frac{n}{t+1})}\right) - \log\left(1 + \frac{1}{t}\right) \frac{1}{\log(n)} + O\left(\frac{1}{\log^k(n)}\right)\right] \\
&= \log^2(n) \log\left(1 + \frac{\log\left(1+\frac{1}{t}\right)}{\log(\frac{n}{t+1})}\right) - \log\left(1 + \frac{1}{t}\right) \log(n) + O\left(\frac{1}{\log^{k-1}(n)}\right).
\end{aligned}
$$

To calculate the limit of this expression as n goes to infinity, set $a = \log(t)$, $b = \log(t+1)$ and $x = \frac{1}{\log(n)}$. Then $x \to 0$ as $n \to +\infty$ and

$$
\log^2(n) \log\left(1 + \frac{\log\left(1+\frac{1}{t}\right)}{\log(\frac{n}{t+1})}\right) - \log\left(1 + \frac{1}{t}\right) \log(n)
$$

becomes

$$
\frac{\log(1 - ax) - \log(1 - bx) - x(b - a)}{x^2}.
$$

Using the Maclaurin series

$$\log(1+y) = y - \frac{1}{2}y^2 + O(y^3)$$

on the above expression gives

$$\frac{1}{2}(b^2 - a^2) + O\left(x\right),$$

and we find that the desired limit is equal to

$$\frac{1}{2}(b^2 - a^2) = \frac{1}{2}\Big(\log^2(t+1) - \log^2(t)\Big) = \frac{1}{2}\log\left(1+\frac{1}{t}\right)\log\Big(t(t+1)\Big).$$

□

Simplification of μ

$$\begin{aligned}
\mu &:= \lim_{N\to\infty}\sum_{t=1}^{N}\left[\left(\log\left(1+\frac{1}{t}\right) - \frac{2t+3}{2(t+1)^2}\right) + \left(1 - \frac{\pi^2}{12} - \sum_{k=1}^{t}\frac{1}{2k(k+1)^2}\right)\right] \\
&= \lim_{N\to\infty}\sum_{t=1}^{N}\left[\left(\log\left(1+\frac{1}{t}\right) - \frac{2t+3}{2(t+1)^2}\right) + \sum_{k=t+1}^{\infty}\frac{1}{2k(k+1)^2}\right] \\
&= \lim_{N\to\infty}\sum_{t=1}^{N}\left[\log\left(1+\frac{1}{t}\right) - \frac{t}{(t+1)^2}\right] - \frac{3}{2}\sum_{t=1}^{\infty}\frac{1}{(t+1)^2} + \sum_{t=1}^{\infty}\sum_{k=t+1}^{\infty}\frac{1}{2k(k+1)^2}.
\end{aligned}$$

The middle summand is easily seen to equal $-\frac{3}{2}(\frac{\pi^2}{6} - 1)$. We calculate the other two summands, starting with the rightmost one.

$$\begin{aligned}
\sum_{t=1}^{\infty}\sum_{k=t+1}^{\infty}\frac{1}{2k(k+1)^2} &= \sum_{k=2}^{\infty}\frac{k-1}{2k(k+1)^2} = \frac{1}{2}\sum_{k=2}^{\infty}\frac{1}{(k+1)^2} - \sum_{k=2}^{\infty}\frac{1}{2k(k+1)^2} \\
&= \frac{1}{2}\left(\frac{\pi^2}{6} - 1 - \frac{1}{4}\right) - \left(1 - \frac{\pi^2}{12} - \frac{1}{8}\right) = \frac{\pi^2}{6} = \frac{\pi^2}{6} - \frac{3}{2}.
\end{aligned}$$

For the leftmost summand, exchange first N with $N-1$:

$$\lim_{N\to\infty}\sum_{t=1}^{N}\left[\log\left(1+\frac{1}{t}\right) - \frac{t}{(t+1)^2}\right] = \lim_{N\to\infty}\sum_{t=1}^{N-1}\left[\log\left(1+\frac{1}{t}\right) - \frac{t}{(t+1)^2}\right].$$

To simplify this, notice how

$$\sum_{t=1}^{N-1}\log\left(1+\frac{1}{t}\right) = \log\left(\prod_{t=1}^{N-1}\left(1+\frac{1}{t}\right)\right) = \log(N),$$

and in addition that

$$\sum_{t=1}^{N-1} \frac{t}{(t+1)^2} = \sum_{t=2}^{N} \frac{t-1}{t^2} = \sum_{t=1}^{N} \frac{1}{t} - \sum_{t=1}^{N} \frac{1}{t^2}.$$

In all,

$$\lim_{N\to\infty} \sum_{t=1}^{N-1} \left[\log\left(1+\frac{1}{t}\right) - \frac{t}{(t+1)^2}\right] = \lim_{N\to\infty} \left[\log(N) - \sum_{t=1}^{N} \frac{1}{t}\right] + \sum_{t=1}^{\infty} \frac{1}{t^2} = -\gamma + \frac{\pi^2}{6}.$$

Putting the three pieces together, we get:

$$\mu = -\frac{3}{2}\left(\frac{\pi^2}{6} - 1\right) + \left(\frac{\pi^2}{6} - \frac{3}{2}\right) + \left(-\gamma + \frac{\pi^2}{6}\right) = \frac{\pi^2}{12} - \gamma.$$

References

[1] Axler, C. *On the sum of the first n prime numbers* (2019). arXiv:1409.1777 [math.NT]

[2] Bush, L.E. *An asymptotic Formula for the Average Sum of the Digits of Integers* (1940), The American Mathematical Monthly, 47:3, 154-156, DOI: 10.1080/00029890.1940.11990954

[3] Hardy, G.H.; Wright, E.M *An Introduction to the Theory of Numbers* (1960)(fourt edition), Oxford University Press (p.355 - 356). First edition published (1938).

[4] Hardy, G.H.; Ramanujan, S.A. *The normal number of prime factors of a number n* (1917), Quarterly Journal of Mathematics, XLVIII, (p.76 – 92).

[5] Landau, E. *Handbuch der Lehre von der Verteilung der Primzahlen* (1909), Vol. 1, Teubner, Leipzig , § 55, (p.200 - 201).

[6] Legendre, A.M. *Théorie des nombres*, 2. Vol (1830), In the Introduction part XVI, (p.10 - 12). Can be accessed at `https://gallica.bnf.fr/ark:/12148/bpt6k426107.texteImage`

[7] Rosser, B. *The n-th Prime is Greater than nlogn.* (1939), Proceedings of the London Mathematical Society, s2-45, (p.21 - 44). DOI:10.1112/plms/s2-45.1.21

[8] Saffari, B. *Sur quelques applications de la "méthode de l'hyperbole" de Dirichlet à la théorie des nombres premiers.* (1970) Enseign. Math. (2) 14, (p.205 – 224). Available at `https://www.unige.ch/math/EnsMath/EM_en/welcome.html`

[9] Tenenbaum,G. *Introduction to analytic and probabilistic number Theory* (1995), Cambridge University Press, (p.40 - 42).

NTNU

Kunnskap for en bedre verden