# A problem in number theory

Marko R. Riedel

May 20, 2001

## 1 Introduction

The purpose of this document is twofold. We discuss a number-theoretic limit and show how to evaluate this limit. In the process, we will demonstrate the use of a few straightforward techniques that are common in analytic number theory: Euler-Maclaurin summation, the Mellin-Perron formula, and the use of the Prime Number theorem. This first motivation for our work is pedagogical, because the problem is comparatively simple and our solution can be used as an introduction to the aforementioned techniques.

The second and more interesting function that this document can serve is that of illustrating how a relatively cumbersome, technical proof can be simplified to be entirely accessible to the layman, i.e. made into a "one-liner", or close to it. To this purpose we will incrementally simplify the proof until it becomes entirely elementary.

### 1.1 Problem definition

Let $n > 1$ be a positive integer, with prime factorization

$$n = \prod_{m=1}^{r} p_m^{v_m},$$

where the $p_m$ are distinct primes $2, 3 \ldots$ and the $v_m \geq 1$. We define the function

$$s(n) = \sum_{m=1}^{r} p_m v_m.$$

This function adds the prime factors of $n$, each counted according to its multiplicity. Denote by $\mathbb{P}$ the set of primes $\{2, 3, 5, \ldots\}$, by $\mathbb{P}(N)$ the set $\{p \mid p \in \mathbb{P} \wedge p \leq N\}$ and by $\mathbb{P}(M, N)$ the set $\{p \mid p \in \mathbb{P} \wedge M < p \leq N\}$.

Define the sums $S_1(N)$ and $S_2(N)$ by

$$S_1(N) = \sum_{1 < 2k+1 \leq N, 2k+1 \notin \mathbb{P}} s(2k+1)$$

and

$$S_2(N) = \sum_{1 < 2k \leq N, 2k \notin \mathbb{P}} s(2k),$$

i.e. the sum of the prime factors, counted according to their multiplicty, of the odd and even non-primes less than or equal to $N$. The definition of the problem is symmetric, but since all even numbers except 2 are non-primes, we may write

$$S_2(N) = \sum_{2 < 2k \leq N} s(2k).$$

We wish to evaluate the limit

$$\lim_{N \to \infty} \frac{S_1(N)}{S_2(N)}.$$

**Example.** Let $N = 10$. We have

$$S_1(10) = s(9) = 6$$

and

$$S_2(10) = s(4) + s(6) + s(8) + s(10) = 4 + 5 + 6 + 7 = 22$$

and hence

$$\frac{S_1(10)}{S_2(10)} = \frac{3}{11}.$$

We claim that

$$\lim_{N\to\infty} \frac{S_1(N)}{S_2(N)} = 1.$$

We give several proofs of this claim in the remainder of this document.

## 2 Starting point, or, the first proof

This proof is the one that a competently programmed computer algebra system is likely to discover, i.e. it uses a maximum of brute force calculation and a minimum of ingenuity; if we exhaustively searched the tree of possible representations of the limit, we would discover this proof.

Before we begin to discuss the details, we need to introduce some additional notation that will in fact simplify the presentation. Define

$$S(N) = \sum_{1<k\le N} s(k).$$

The key feature of this definition that the reader should bear in mind is that $S(N)$ does not exclude $s(k)$ for prime $k$, as do $S_1(N)$ and $S_2(N)$, but rather extends over all $k$. We clearly have

$$\begin{aligned} S_1(N) &= \sum_{1<2k+1\le N, 2k+1\notin\mathbb{P}} s(2k+1) = \sum_{1<k\le N, k\notin\mathbb{P}} s(k) - \sum_{1<2k\le N, 2k\notin\mathbb{P}} s(2k) \\ &= \sum_{1<k\le N} s(k) - \sum_{p\in\mathbb{P}(N)} s(p) - \sum_{2<2k\le N} s(2k) \\ &= S(N) - \sum_{p\in\mathbb{P}(N)} p - \sum_{1<k\le\lfloor N/2\rfloor} (2 + s(k)) = S(N) - \sum_{p\in\mathbb{P}(N)} p - S(\lfloor N/2\rfloor) - 2\lfloor N/2\rfloor + 2 \end{aligned}$$

and

$$S_2(N) = \sum_{2<2k\le N} s(2k) = S(\lfloor N/2\rfloor) + 2\lfloor N/2\rfloor - 2.$$

Therefore we may rewrite the limit as

$$\lim_{N\to\infty} \frac{S_1(N)}{S_2(N)} = \lim_{N\to\infty} \frac{S(N) - \sum_{p\in\mathbb{P}(N)} p - S(\lfloor N/2\rfloor) - 2\lfloor N/2\rfloor + 2}{S(\lfloor N/2\rfloor) + 2\lfloor N/2\rfloor - 2} = -1 + \lim_{N\to\infty} \frac{S(N) - \sum_{p\in\mathbb{P}(N)} p}{S(\lfloor N/2\rfloor) + 2\lfloor N/2\rfloor - 2}.$$

Hence the problem reduces to the study of $S(N)$.

The form of this quotient is such that we only require the first term in the asymptotic expansion of $S(N)$ in order to evaluate the limit.

### 2.1 The order of $S(N)$

The object of this section is find a formula for $S(N)$ and use it to obtain an approximation to the dominant term in the asymptotic expansion of $S(N)$.

Let $v_p(k)$ be the highest power of $p$ that divides $k$, i.e. $v_p(k) = \max\{r \mid (p^r, k) = p^r\}$. We rewrite $S(N)$, as follows.

$$
\begin{aligned}
S(N) &= \sum_{k\le N} s(k) = \sum_{k\le N}\sum_{p\in\mathbb{P}(N)} p v_p(k) = \sum_{p\in\mathbb{P}(N)}\sum_{r=1}^{\lfloor\log_2 N\rfloor}\sum_{k\le N, v_p(k)=r} pr \\
&= \sum_{r=1}^{\lfloor\log_2 N\rfloor}\sum_{p\in\mathbb{P}(N)}\sum_{k\le N, v_p(k)\ge r} p = \sum_{r=1}^{\lfloor\log_2 N\rfloor}\sum_{p\in\mathbb{P}(N)} p\lfloor N/p^r\rfloor = \sum_{r=1}^{\lfloor\log_2 N\rfloor}\sum_{s=1}^{\lfloor N/2^r\rfloor}\sum_{p\in\mathbb{P}(N),\lfloor N/p^r\rfloor=s} ps \\
&= \sum_{r=1}^{\lfloor\log_2 N\rfloor}\sum_{s=1}^{\lfloor N/2^r\rfloor} s \sum_{p\in\mathbb{P}((N/(s+1))^{1/r},(N/s)^{1/r})} p = \sum_{r=1}^{\lfloor\log_2 N\rfloor}\sum_{s=1}^{\lfloor N/2^r\rfloor}\sum_{p\in\mathbb{P}(1,(N/s)^{1/r})} p
\end{aligned}
$$

We require an approximation of $\sum_{p\in\mathbb{P}(x)} p$. If the primes were evenly distributed, the "average" $k$th prime in $\mathbb{P}(x)$ would be given by $kx/\pi(x)$, whence

$$
\sum_{p\in\mathbb{P}(x)} p \sim \sum_{k=1}^{\pi(x)} k\frac{x}{\pi(x)} = \frac{x}{\pi(x)}\sum_{k=1}^{\pi(x)} k = \frac{1}{2}x(\pi(x)+1).
$$

In the following we make use of the Prime Number theorem, i.e. that $\pi(x)\sim x/\ln x$. This is the first time that we use a theorem from advanced number theory, and we will see later that we can do without it. The theorem yields

$$
\sum_{s=1}^{\lfloor N/2^r\rfloor}\sum_{p\in\mathbb{P}(1,(N/s)^{1/r})} p \sim \frac{1}{2}\sum_{s=1}^{\lfloor N/2^r\rfloor}\frac{(N/s)^{2/r}}{1/r\ln N/s} = \frac{1}{2}rN^{2/r}\sum_{s=1}^{\lfloor N/2^r\rfloor}\frac{1}{s^{2/r}\ln N/s}.
$$

If we substitute this into the formula for $S(N)$, we obtain

$$
\frac{1}{2}\sum_{r=1}^{\lfloor\log_2 N\rfloor} rN^{2/r}\sum_{s=1}^{\lfloor N/2^r\rfloor}\frac{1}{s^{2/r}\ln N/s}.
$$

At this point we introduce a conjecture that is based on numerical experiments. We claim that

$$
S(N)\sim\frac{N^2}{\ln N}.
$$

This conjecture is verified if we show that

$$
\lim_{N\to\infty}\frac{S(N)\ln N}{N^2} = 1.
$$

Evidently

$$
\frac{S(N)\ln N}{N^2} = \frac{1}{2}\sum_{r=1}^{\lfloor\log_2 N\rfloor} rN^{2/r-2}\sum_{s=1}^{\lfloor N/2^r\rfloor}\frac{1}{s^{2/r}(1-\log_N s)}.
$$

Most of the terms in this sum can be treated with a straightforward upper bound, i.e.

$$
\begin{aligned}
&\frac{1}{2}\sum_{r>2}^{\lfloor\log_2 N\rfloor} rN^{2/r-2}\sum_{s=1}^{\lfloor N/2^r\rfloor}\frac{1}{s^{2/r}(1-\log_N s)} < \frac{1}{2}\sum_{r>2}^{\lfloor\log_2 N\rfloor} rN^{2/r-2}\sum_{s=1}^{\lfloor N/2^r\rfloor}\frac{1}{1-\log_N N/2^r} \\
=\ &\frac{1}{2}\sum_{r>2}^{\lfloor\log_2 N\rfloor} rN^{2/r-2}\sum_{s=1}^{\lfloor N/2^r\rfloor}\frac{1}{r\log_N 2} = \frac{1}{2}\sum_{r>2}^{\lfloor\log_2 N\rfloor}\log_2 N N^{2/r-2}\sum_{s=1}^{\lfloor N/2^r\rfloor} 1 \\
=\ &\frac{1}{2}\sum_{r>2}^{\lfloor\log_2 N\rfloor}\log_2 N N^{2/r-2}\lfloor N/2^r\rfloor < \frac{1}{2}\log_2 N\sum_{r>2}^{\lfloor\log_2 N\rfloor} N^{2/r-1}\frac{1}{2^r} \\
<\ &\frac{1}{2}\log_2 N\frac{1}{N^{1/3}}\sum_{r>2}^{\lfloor\log_2 N\rfloor}\frac{1}{8} < \frac{1}{16}(\log_2 N)^2\frac{1}{N^{1/3}}
\end{aligned}
$$

By repeated application of L'Hôpital's rule, it is easily verified that

$$\lim_{N\to\infty}(\log_2 N)^2\frac{1}{N^{1/3}} = 0.$$

It remains to check the terms that correspond to $r = 1$ and $r = 2$. In order to do this, we restrict our attention to $N = 2^k$. Note that

$$\log_{2^k} s = \frac{\ln s}{k\ln 2} = \frac{1}{k}\log_2 s.$$

We are now concerned with

$$\lim_{k\to\infty}\frac{1}{2}\left(\sum_{s=1}^{2^{k-1}}\frac{1}{s^2(1-1/k\log_2 s)} + \frac{2}{2^k}\sum_{s=1}^{2^{k-2}}\frac{1}{s(1-1/k\log_2 s)}\right).$$

### 2.1.1 Application of the Euler-Maclaurin summation formula

Let $\psi(u) = u - \lfloor u\rfloor - 1/2$, and let $a$ and $b$ be positive integers such that $a < b$. The simplest form of the Euler-Maclaurin summation formula states that

$$\sum_{a<s\le b} f(s) = \int_a^b f(u)du + \frac{1}{2}(f(b)-f(a)) + \int_a^b f'(u)\psi(u)du,$$

where $f(u)$ is continuous and differentiable in $[a,b]$. (We use $s$ as our discrete, and $u$ as our continuous variable.)

Observe that

$$\sum_{s=1}^{2^{k-1}}\frac{1}{s^2(1-1/k\log_2 s)} = 1+\sum_{t=1}^{k-1}\sum_{2^{t-1}<s\le 2^t}\frac{1}{s^2(1-1/k\log_2 s)} = 1+k\sum_{t=1}^{k-1}\sum_{2^{t-1}<s\le 2^t}\frac{1}{s^2(k-\log_2 s)}$$

and similarly,

$$\sum_{s=1}^{2^{k-2}}\frac{1}{s(1-1/k\log_2 s)} = 1+\sum_{t=1}^{k-2}\sum_{2^{t-1}<s\le 2^t}\frac{1}{s(1-1/k\log_2 s)} = 1+k\sum_{t=1}^{k-2}\sum_{2^{t-1}<s\le 2^t}\frac{1}{s(k-\log_2 s)}.$$

We must therefore evaluate a sum of the form

$$\sum_{2^{t-1}<s\le 2^t}\frac{1}{s^m(k-\log_2 s)},$$

where $m = 2, 1$.

Given that $2^{t-1} < s \le 2^t$, we evidently have

$$\frac{1}{k+1-t} < \frac{1}{k-\log_2 s} \le \frac{1}{k-t}.$$

This gives a bound for our two sums.

$$\frac{1}{k+1-t}\sum_{2^{t-1}<s\le 2^t}\frac{1}{s^m} < \sum_{2^{t-1}<s\le 2^t}\frac{1}{s^m(k-\log_2 s)} \le \frac{1}{k-t}\sum_{2^{t-1}<s\le 2^t}\frac{1}{s^m}.$$

We need to know the first term in the expansion of

$$\sum_{2^{t-1}<s\le 2^t}\frac{1}{s^m},$$

where $m = 2, 1$; Euler-Maclaurin summation provides the answer.

**The case $m = 2$.** The sum is equal to

$$\int_{2^{t-1}}^{2^t} \frac{1}{u^2} du + \frac{1}{2}(2^{-2t} - 2^{2-2t}) - 2\int_{2^{t-1}}^{2^t} \frac{1}{u^3}\psi(u)du = \frac{1}{2^t} - \frac{3}{2}2^{-2t} - 2\int_{2^{t-1}}^{2^t} \frac{1}{u^3}\psi(u)du.$$

It is left as an exercise to the reader to verify that the integral is of lower order than the $2^{-t}$ term. (We also omit the corresponding step for $m = 1$.)

Using this result, we find that the upper bound is asympotically equivalent to

$$1 + k\sum_{t=1}^{k-1} \frac{1}{k-t}\frac{1}{2^t} = 1 + \frac{k}{2^k}\sum_{t=1}^{k-1}\frac{2^t}{t}$$

and the lower bound to

$$\begin{aligned} 1 + k\sum_{t=1}^{k-1} \frac{1}{k+1-t}\frac{1}{2^t} &= 1 + \frac{k}{2^k}\sum_{t=1}^{k-1}\frac{2^t}{t+1} = 1 + \frac{1}{2}\frac{k}{2^k}\sum_{t=1}^{k-1}\frac{2^{t+1}}{t+1} \\ &= 1 + \frac{1}{2}\frac{k}{2^k}\left(-1 + \sum_{t=0}^{k-2}\frac{2^{t+1}}{t+1} + \frac{2^k}{k}\right) = \frac{3}{2} - \frac{1}{2}\frac{k}{2^k} + \frac{1}{2}\frac{k}{2^k}\sum_{t=1}^{k-1}\frac{2^t}{t}. \end{aligned}$$

We define

$$s_k = \frac{k}{2^k}\sum_{t=1}^{k-1}\frac{2^t}{t}.$$

and ask what is

$$\rho = \lim_{k\to\infty} s_k.$$

Note that

$$\left(s_k\frac{2^k}{k} + \frac{2^k}{k}\right)\frac{k+1}{2^{k+1}} = \frac{1}{2}(s_k+1)\left(1+\frac{1}{k}\right) = s_{k+1}.$$

It follows that

$$\frac{1}{2}(\rho + 1) = \rho,$$

i.e. $\rho = 1$. We conclude that the upper and lower bounds of the $m = 2$ sum are asymptotically equivalent to 2, i.e. that

$$\lim_{k\to\infty}\sum_{s=1}^{2^{k-1}} \frac{1}{s^2(1 - 1/k\log_2 s)} = 2.$$

**The case $m = 1$.** The Euler-MacLaurin integral yields

$$\int_{2^{t-1}}^{2^t} \frac{1}{u} du + \frac{1}{2}(2^{-t} - 2^{1-t}) - \int_{2^{t-1}}^{2^t} \frac{1}{u^2}\psi(u)du = \ln 2 - \frac{1}{2}2^{-t} - \int_{2^{t-1}}^{2^t} \frac{1}{u^2}\psi(u)du.$$

We substitute the dominant term, i.e. $\ln 2$, into the bounds for the sum. The upper bound is

$$\frac{2}{2^k}\left(1 + k\sum_{t=1}^{k-2}\frac{1}{k-t}\ln 2\right) = \frac{2}{2^k}\left(1 + k\ln 2\sum_{t=2}^{k-1}\frac{1}{t}\right)$$

and the lower bound

$$\frac{2}{2^k}\left(1 + k\sum_{t=1}^{k-2}\frac{1}{k+1-t}\ln 2\right) = \frac{2}{2^k}\left(1 + k\ln 2\sum_{t=3}^{k}\frac{1}{t}\right) = \frac{2}{2^k}\left(1 + \ln 2 - \frac{1}{2}k\ln 2 + k\ln 2\sum_{t=2}^{k-1}\frac{1}{t}\right).$$

There are at least two ways to proceed. The simplest is to note that

$$k\sum_{t=2}^{k-1}\frac{1}{t} < k^2$$

and hence both the upper and lower bound vanish in the limit as $k$ goes to infinity. This shows that

$$\lim_{k\to\infty} \frac{2}{2^k} \sum_{s=1}^{2^{k-2}} \frac{1}{s(1-1/k\log_2 s)} = 0.$$

An alternate means of verifying this last result is to use harmonic numbers. They are not strictly necessary at this point, but we introduce them here because we will need them later. The $n$th harmonic number is given by

$$H_n = \sum_{k=1}^{n} \frac{1}{k}$$

and has the asymptotic expansion

$$H_n \sim \ln n + \gamma + \frac{1}{2n} + O\left(\frac{1}{n^2}\right).$$

The reader is strongly encouraged to verify this expansion, either by repeated application of Euler-Maclaurin summation or by the Mellin summation formula.

This asymptotic expansion shows that

$$k \sum_{t=2}^{k-1} \frac{1}{t} \sim k \ln k$$

(this is a more accurate estimate than the first one), and we may once more conclude that both the upper and the lower bound of the $m = 1$ sum vanish in the limit.

### 2.1.2 Conclusion of the first proof

At this point we recall that our purpose was to show that

$$\lim_{N\to\infty} \frac{S(N)\ln N}{N^2} = 1.$$

As we have seen, we can represent the quotient as half of an iterated sum in the variable $r$, which ranges from 1 to $\lfloor \log_2 N \rfloor$. We began by showing that the sum of the terms with $r > 2$ vanishes, as does the $r = 2$ term (our $m = 1$). We showed that the $r = 1$ term (our $m = 2$) goes to 2 in the limit. Collecting terms, we see that the limit of the half the sum is $1/2(2 + 0 + 0) = 1$, and this verifies the claim.

It follows that

$$\begin{aligned}\lim_{N\to\infty} \frac{S_1(N)}{S_2(N)} &= -1 + \lim_{N\to\infty} \frac{S(N) - \sum_{p\in\mathbb{P}(N)} p}{S(\lfloor N/2 \rfloor) + 2\lfloor N/2 \rfloor - 2} \\ &= -1 + \lim_{N\to\infty} \frac{\frac{1}{2}\frac{N^2}{\ln N}}{\frac{1}{4}\frac{N^2}{\ln N - \ln 4} + 2\lfloor N/2 \rfloor - 2} = -1 + \lim_{N\to\infty} \frac{1/2}{1/4\frac{1}{1-1/\log_4 N}} = 1,\end{aligned}$$

which was to be shown.

# 3 Second proof, or, the first simplification

We were able to evaluate the limit because we showed that

$$S(N) \sim \frac{N^2}{\ln N}.$$

We now retrace our steps almost to the beginning of our evaluation of $S(N)$. We recall that

$$S(N) = \sum_{k\le N} s(k) = \sum_{k\le N} \sum_{p\in\mathbb{P}(N)} p v_p(k).$$

Instead of splitting the inner sum over fixed values of $v_p(k)$, as we did earlier, we rewrite it as follows.

$$\sum_{k\le N} \sum_{p\in\mathbb{P}(N)} p v_p(k) = \sum_{p\in\mathbb{P}(N)} p \sum_{k\le N} v_p(k).$$

We can compute the order of $S(N)$ if we have an asymptotic expansion of the inner sum.

## 3.1 Application of the Mellin-Perron formula

The Mellin-Perron formula is actually a class of formulas associated to a non-negative exponent $m$. The one that interests us here is the case $m = 0$, where we have

$$\sum_{1\le k\le N} \lambda_k - \frac{1}{2}\lambda_N = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \left(\sum_{k\ge 1} \frac{\lambda_k}{k^s}\right) N^s \frac{ds}{s}.$$

This holds for $c$ in the half-plane of convergence of the Dirichlet series

$$\sum_{k\ge 1} \frac{\lambda_k}{k^s}.$$

We set $\lambda_k = v_p(k)$ and

$$V_p(s) = \sum_{k\ge 1} \frac{v_p(k)}{k^s}$$

With $v_p(k) \le \log_2 k$ we know that we may choose $c = 3/2 > 1$. We will evaluate the integral in the formula by the Cauchy residue theorem. Hence we need to locate the poles of the integrand, and $V_p(s)$ in particular. We have

$$V_p(s) = \sum_{k\ge 1} \frac{v_p(k)}{k^s} = \sum_{kp\ge 1} \frac{v_p(kp)}{(kp)^s} = \frac{1}{p^s}\zeta(s) + \frac{1}{p^s}V_p(s).$$

We obtain

$$V_p(s) = \frac{\zeta(s)}{p^s - 1},$$

hence the poles are at $s = 1$ and $m\frac{2\pi i}{\ln p}$, $m$ an integer, and all are simple. Combining these observations, we find

$$\sum_{1\le k\le N} v_p(k) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{\zeta(s)}{p^s - 1} N^s \frac{ds}{s} + \frac{1}{2}\lambda_N.$$

We point out that the integrand has a simple pole at $s = 1$, a double pole at $s = 0$ and an infinite number of simple poles at $s = m\frac{2\pi i}{\ln p}$, $m$ a non-zero integer. We are only interested in the dominant term of the asymptotic expansion, however, and hence we will only use the pole at $s = 1$. The interested reader may of course compute the entire expansion. The pole at $s = 1$ yields

$$\lim_{s\to 1} \frac{\zeta(s)(s-1)}{p^s - 1} N^s \frac{1}{s} = \frac{N}{p-1} \lim_{s\to 1} \zeta(s)(s-1) = \frac{N}{p-1}$$

### 3.1.1 Remark

We digress here for a moment in order to point out a feature of the problem that is remarkable. The above result shows that the "expected value" of $v_p(k)$ in the range $1 \le k \le N$ is $1/(p-1)$, in fact independent of $N$. Recall our definition of $s(k)$ as

$$\sum_{m=1}^{r} p_m v_m$$

if the prime factorization of $k$ is

$$\prod_{m=1}^{r} p_m^{v_m}.$$

Note that

$$\sum_{p\in\mathbb{P}(k)} p_m \frac{1}{p_m - 1} \sim \pi(k).$$

*The "expected value" of* $s(k)$ *is* $\pi(k)$. We will return to this point later.

### 3.2 Conclusion, second proof

We have established that

$$\sum_{k\le N} v_p(k) \sim \frac{N}{p-1},$$

and hence that

$$S(N) \sim \sum_{p\in\mathbb{P}(N)} p\frac{N}{p-1} = N \sum_{p\in\mathbb{P}(N)} \left(1+\frac{1}{p-1}\right).$$

The sum $S(N)$ is asymptotic to $N\pi(N)$. Using the Prime Number theorem, we once more have $S(N) \sim N^2/\ln N$, and can conclude the proof as we did in the first version.

# 4 Third proof, or the second simplification

This simplification results from the observation that we can do without the Mellin-Perron formula in the evaluation of $\sum_{k\le N} v_p(k)$, because we only need the first term of the asymptotic expansion of the sum. We proceed as follows.

$$\sum_{k\le N} v_p(k) = \sum_{r=1}^{\lfloor \log_2 N\rfloor} \sum_{k\le N, r=v_p(k)} r = \sum_{r=1}^{\lfloor \log_2 N\rfloor} \sum_{k\le N, r<v_p(k)} 1 = \sum_{r=1}^{\lfloor \log_2 N\rfloor} \left\lfloor \frac{N}{p^r} \right\rfloor$$

We certainly have

$$\left| \sum_{r=1}^{\lfloor \log_2 N\rfloor} \left\lfloor \frac{N}{p^r} \right\rfloor - \sum_{r=1}^{\lfloor \log_2 N\rfloor} \frac{N}{p^r} \right| < \log_2 N.$$

We switch to the exact sum and obtain

$$\sum_{r=1}^{\lfloor \log_2 N\rfloor} \frac{N}{p^r} = -N + N\frac{1-p^{-(\lfloor \log_2 N\rfloor+1)}}{1-p^{-1}} = -N + N\frac{p-p^{-\lfloor \log_2 N\rfloor}}{p-1} = N\frac{1-p^{-\lfloor \log_2 N\rfloor}}{p-1}.$$

The dominant term is $N/(p-1)$, and we may neglect the error term $\log_2 N$ in the limit. From this point on we may continue as in the last section.

# 5 Fourth proof, without the Prime Number theorem

We have successively eliminated Euler-MacLaurin summation and the Mellin-Perron formula from the proof. The only advanced theorem that remains is the Prime Number theorem. The current version of the proof does not use it until the last step. We shall see that the Prime Number theorem provides more information than we need. Recall our observation, based entirely on elementary operations, that the "expected value" of $s(k)$ is $\pi(k)$. Hence we have

$$\begin{aligned} S_1(N) &\sim \sum_{2k+1\le N, 2k+1\notin\mathbb{P}(N)} \pi(2k+1) = \sum_{2k+1\le N} \pi(2k+1) - \sum_{2k+1\in\mathbb{P}(N)} \pi(2k+1) \\ &= \sum_{2k+1\le N} \pi(2k+1) - \sum_{m=1}^{\pi(N)} m = \sum_{2k+1\le N} \pi(2k+1) - \frac{1}{2}\pi(N)(\pi(N)+1) \end{aligned}$$

and

$$S_2(N) \sim \sum_{2k\le N, 2k\notin\mathbb{P}(N)} \pi(2k) = \sum_{2k\le N} \pi(2k).$$

This yields

$$\lim_{N\to\infty} \frac{S_1(N)}{S_2(N)} = 1 - \lim_{N\to\infty} \frac{1}{2} \frac{\pi(N)(\pi(N)+1)}{\sum_{2k\le N} \pi(2k)}.$$

If we can show that $\pi(N) \in \omega(1)$ and $\pi(N) \in o(N)$, we have the result, because then we would have

$$\frac{\pi(N)(\pi(N)+1)}{\sum_{2k \leq N} \pi(2k)} \sim \frac{\pi(N)(\pi(N)+1)}{c_0 N \pi(N)} = \frac{\pi(N)+1}{c_0 N}$$

and, using $\pi(N) \in o(N)$ a second time, we could conclude

$$1 - \lim_{N\to\infty} \frac{S_1(N)}{S_2(N)} = 1 - \lim_{N\to\infty} \frac{\pi(N)+1}{c_0 N} = 1.$$

We need an elementary proof that $\pi(N) \in \omega(1)$ and $\pi(N) \in o(N)$. The first part is easy, because there is an infinite number of primes. Suppose the second part did not hold, i.e. that there existed positive integer $r$ such that

$$\pi(N) \geq \frac{1}{r} N$$

for all $N$ from some $N_0$ on. Counting, on the one hand, the primes, and on the other, the integers with exactly two distinct prime factors, we find the following inequality.

$$N > \pi(N) + \sum_{p \in \mathbb{P}(\sqrt{N})} \left( \pi\left( \left\lfloor \frac{N}{p} \right\rfloor \right) - \pi(p) \right)$$

Our assumption that $\pi(N)$ is not $o(N)$ then yields a lower bound for the right side.

$$\begin{aligned}
\geq\ & \frac{1}{r}N + \frac{1}{r} \sum_{p\in\mathbb{P}(\sqrt{N})} \left\lfloor \frac{N}{p} \right\rfloor - \sum_{p\in\mathbb{P}(\sqrt{N})} \pi(p) \geq \frac{1}{r}N + \frac{1}{r} \sum_{p\in\mathbb{P}(\sqrt{N})} \left(\frac{N}{p} - 1\right) - \sum_{k=1}^{\pi(\sqrt{N})} k \\
\geq\ & \frac{1}{r}N + \frac{1}{r}N \sum_{p\in\mathbb{P}(\sqrt{N})} \frac{1}{p} - \frac{1}{r}\pi(\sqrt{N}) - \frac{1}{2}\sqrt{N}(\sqrt{N}+1) \\
\geq\ & \frac{1}{r}N \sum_{p\in\mathbb{P}(\sqrt{N})} \frac{1}{p} - \left(\frac{1}{2} - \frac{1}{r}\right) N - \frac{1}{r^2}\sqrt{N} - \frac{1}{2}\sqrt{N} = \frac{1}{r}N \sum_{p\in\mathbb{P}(\sqrt{N})} \frac{1}{p} - \left(\frac{1}{2} - \frac{1}{r}\right) N - \left(\frac{1}{2} + \frac{1}{r^2}\right)\sqrt{N}
\end{aligned}$$

Recall the asymptotic expansion of the harmonic numbers, introduced earlier. Continuing under the assumption that $\pi(N)$ is not $o(N)$, we thus have

$$\sum_{p\in\mathbb{P}(\sqrt{N})} \frac{1}{p} \sim \Omega\left( \frac{1}{r} \sum_{t=1}^{\sqrt{N}/r} \frac{1}{t} \right) \sim \frac{1}{2r} \ln N.$$

Hence the original inequality is equivalent to

$$\frac{1}{2r} N \ln N < N$$

asymptotically, and for all $N$ from some $N_0$ on, a contradiction. We conclude that $\pi(N)$ is $o(N)$. This completes the proof; we point out that this last version is entirely elementary, in the sense that it does not use any theorems other than those up to basic calculus. (Recall that there are only two steps to the current version of the proof: we show that the expected value of $s(k)$ is $\pi(k)$, and then we use the fact that $\pi(k)$ is $o(k)$.)