\[
\sum_{p \leq n} \frac{1}{p} = \ln(\ln(n)) + O(1): \text{ An Exposition}
\]
by William Gasarch and Larry Washington

1 Introduction

It is well known that \( \sum_{p \leq n} \frac{1}{p} = \ln(\ln(n)) + O(1) \) where \( p \) goes over the primes. We give several known proofs of this.

We first present a proof that \( \sum_{p \leq n} \frac{1}{p} \geq \ln(\ln(n)) + O(1) \). This is based on Euler’s proof that \( \sum \frac{1}{p} \) diverges. We then present three proofs that \( \sum_{p \leq n} \frac{1}{p} \leq \ln(\ln(n)) + O(1) \). The first one, essentially due to Mertens, does not use the prime number theorem. The second and third one do use the prime number theorem and hence are shorter.

For a complete treatment of Merten’s proof that \( \sum \frac{1}{p} \) diverges, and how it compares with modern treatments, see the scholarly work of Villarino [4].

2 Euler’s Proof that \( \sum_{p \leq n} \frac{1}{p} \geq \ln(\ln(n)) + O(1) \)

The proof here follows the one in [1].

**Lemma 2.1** For \( 0 \leq x \leq 1/2 \), \(-\ln(1-x) \leq x + x^2.\)

**Proof:** \(-\ln(1-x) = \int_0^x \frac{1}{1-t} dt.\) For \( 0 \leq t \leq 1/2, \frac{1}{1-t} \leq 1 + 2t.\) Hence

\[-\ln(1-x) = \int_0^x \frac{1}{1-t} dt \leq \int_0^x (1 + 2t) dt = x + x^2.\]

**Theorem 2.2** \( \sum_{p \leq n} \frac{1}{p} \geq \ln(\ln(n)) + O(1). \)

**Proof:** Clearly

\[
\sum_{j=1}^{\infty} \frac{1}{j} = (1 - \frac{1}{2} + \frac{1}{2^2} + \cdots)(1 - \frac{1}{3} + \frac{1}{3^2} + \cdots) \cdots = \frac{1}{1 - 2^{-1}} \times \frac{1}{1 - 3^{-1}} \times \cdots
\]

which we rewrite as

\[
\sum_{j=1}^{\infty} \frac{1}{j} = \prod_p (1 - p^{-1})^{-1}
\]

We need a finite version of this statement. Let \( S_n \) be the set of natural numbers whose prime factors \( p \) are all \( \leq n. \) Then

\[
\sum_{j \in S_n} \frac{1}{j} = \prod_{p \leq n} (1 - p^{-1})^{-1}.
\]
Clearly $\sum_{j \leq n} \frac{1}{j} \leq \sum_{j \in S_n} \frac{1}{j}$. By integration $\ln n \leq \sum_{j \leq n} \frac{1}{j}$. Hence we have

$$\ln(n) \leq \sum_{j \leq n} \frac{1}{j} \leq \sum_{j \in S_n} \frac{1}{j} = \prod_{p \leq n} (1 - p^{-1})^{-1}$$

$$\ln(\ln(n)) \leq \sum_{p \leq n} -\ln(1 - p^{-1}).$$

By Lemma 2.1

$$\sum_{p \leq n} -\ln(1 - p^{-1}) \leq \sum_{p \leq n} \frac{1}{p} + \frac{1}{p^2}.$$  

Putting this all together we get

$$\sum_{p \leq n} \frac{1}{p} \geq \ln(\ln(n)) - \sum_{p \leq n} \frac{1}{p^2}$$

Since the second sum is bounded by $\sum_{i=1}^{\infty} \frac{1}{i^2}$, which converges, we have

$$\sum_{p \leq n} \frac{1}{p} \geq \ln(\ln(n)) - O(1).$$

$\blacksquare$

**Note 2.3** If the above proof is done more carefully with attention paid to the constants you can obtain $\sum_{p \leq n} \frac{1}{p} \geq \ln(\ln(n)) - 0.48$. See [1].

### 3. Mertens Proof that Does Not Use the Prime Number Theorem

This is adapted from Landau’s book [2]. He works a little harder and gets $o(1)$ instead of $O(1)$.

We first need a weak form of the prime number theorem.

**Lemma 3.1** $\pi(x) = O(x / \ln x)$.

**Proof:** Let $n$ be a positive integer. Clearly every prime $p$ with $n < p \leq 2n$ occurs in the prime factorization of the binomial coefficient $\binom{2n}{n}$. Therefore,

$$n^{\pi(2n) - \pi(n)} = \prod_{n < p \leq 2n} n \leq \prod_{n < p \leq 2n} p \leq \binom{2n}{n} \leq (1 + 1)^{2n} = 4^n.$$
Taking logs yields

\[ \pi(2n) - \pi(n) \leq \frac{n \ln 4}{\ln n} \leq 2 \ln 4 \left( \frac{2n}{\ln(2n)} - \frac{n}{\ln n} \right) \]

for \( n \geq 8 \). If \( y \geq 16 \) is a real number, let \( 2n \) be the largest even integer with \( 2n \leq y \). Then \( \pi(y) - \pi(2n) \leq 1 \) and \( |\pi(y/2) - \pi(n)| \leq 1 \). By increasing \( 2 \ln 4 \) to \( 4 \) we can absorb these errors and obtain

\[ \pi(y) - \pi(y/2) \leq 4 \left( \frac{y}{\ln y} - \frac{y/2}{\ln(y/2)} \right) \]

for \( y \geq 16 \). Adding up this inequality for \( y = x, x/2, x/4, \ldots \) yields

\[ \pi(x) - \pi(16) \leq 4 \left( \frac{x}{\ln x} \right). \]

This yields the lemma.

We now need a result that is interesting in its own right.

**Proposition 3.2**

\[ \sum_{p \leq x} \frac{\ln p}{p} = \ln x + O(1). \]

**Proof:** If \( n \) is a positive integer and \( p \) is a prime, the power of \( p \) dividing \( n! \) is

\[ \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots. \]

Therefore,

\[ \ln(n!) = \sum_{p \leq n} \ln p \left( \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots \right). \]

Changing \( \lfloor n/p \rfloor \) to \( n/p \) introduces an error of most 1, so we have

\[ \sum_{p \leq n} \ln p \left\lfloor \frac{n}{p} \right\rfloor = n \sum_{p \leq n} \frac{\ln p}{p} + O\left( \sum_{p \leq n} \ln p \right). \]

Since there are \( \pi(n) \) terms in the sum, Lemma 1 implies that

\[ O\left( \sum_{p \leq n} \ln p \right) = O(\pi(n) \ln n) = O(n). \]

Let’s treat the higher terms:

\[ \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots < \frac{n}{p^2} \left( 1 + p^{-1} + p^{-2} + \cdots \right) = \frac{n}{p^2 - p}. \]
Therefore,
\[ \sum_{p \leq n} \ln p \left( \left\lfloor \frac{n}{p^2} \right\rfloor + \frac{n}{p^2} + \cdots \right) \leq n \sum_{p \leq n} \frac{\ln p}{p^2 - p} = O(n) \]
since \( \sum \ln p/(p^2 - p) \leq \sum_{j \geq 2} \ln j/(j^2 - j) \), which converges.

Stirling’s formula says that
\[ \ln(n!) = n \ln n + O(n) \]
(this weak form can be proved by comparing \( \sum \ln j \) with \( \int \ln t \, dt \)). Putting everything together yields
\[ n \ln n + O(n) = n \sum_{p \leq n} \frac{\ln p}{p} + O(n). \]

Dividing by \( n \) yields the proposition for \( x = n \). The error introduced by changing from \( x \) to \( n = \lfloor x \rfloor \) is absorbed by \( O(x) \), so the proposition is proved.

The following lemma is well known. It is an analog of integration by parts for summations. It is easily proven by induction on \( n \).

**Lemma 3.3** Let both \( f_1, f_2, \ldots \) and \( g_1, g_2, \ldots \) be sequences of complex numbers. Then, for all \( m \leq n \),
\[ \sum_{i=m}^{n} f_i (g_{i+1} - g_i) = f_{n+1} g_{n+1} - f_{m} g_{m} - \sum_{i=m}^{n} g_{i+1} (f_{i+1} - f_i). \]

We can now prove the theorem.

**Theorem 3.4** \( \sum_{p \leq x} \frac{1}{p} = \ln \ln x + O(1). \)

**Proof:** We have
\[ f(x) = \sum_{p \leq x} \frac{\ln p}{p} = \ln x + r(x), \]
where \( r(x) = O(1) \). Then
\[ \sum_{p \leq x} \frac{1}{p} = \sum_{p \leq x} \frac{\ln p}{p} \ln p = \sum_{n=2}^{x} \frac{f(n) - f(n - 1)}{\ln n} = \sum_{n=2}^{x} \frac{\ln n - \ln(n - 1)}{\ln n} + \sum_{n=2}^{x} \frac{r(n) - r(n - 1)}{\log n}. \]

Since
\[ \ln n - \ln(n - 1) = -\ln \left( 1 - \frac{1}{n} \right) = \frac{1}{n} + O(1/n^2), \]

and
\[ \sum_{n=2}^{x} \frac{1}{n \ln n} = \ln \ln x + O(1), \]
we find that
\[ \sum_{n=2}^{x} \frac{\ln n - \ln(n-1)}{\ln n} = \ln \ln x + O(1). \]
Summation by parts yields
\[ \sum_{n=2}^{x} \frac{r(n) - r(n-1)}{\log n} = \sum_{n=2}^{x} r(n) \left( \frac{1}{\ln n} - \frac{1}{\ln(n+1)} \right) + \frac{r([x])}{\ln([x]+1)} \]
\[ = O \left( \sum_{n=2}^{x} \frac{1/n}{(\ln n)^2} \right) + O(1) = O(1). \]
Putting everything together yields the theorem.

4 A Proof that uses Summation by Parts

In this section we give the standard way to estimate \( \sum 1/p \) using the Prime Number Theorem.

**Theorem 4.1** \( \sum_{p \leq n} \frac{1}{p} = \ln(\ln(n)) + O(1). \)

**Proof:** Let \( \pi(i) \) be the number of primes \( \leq i \). Let \( g(i) = \pi(i) - 1 \) and \( f(i) = \frac{1}{i} \).

Let \( m = 2 \). Plugging these into Lemma 3.3 yields
\[ \sum_{i=2}^{n} \frac{1}{i} (\pi(i) - \pi(i-1)) = \frac{1}{n+1} \pi(n) - \frac{1}{2} \pi(1) - \sum_{i=2}^{n} \pi(i) \left( \frac{1}{i+1} - \frac{1}{i} \right). \]

We need:

- \( \pi(i) - \pi(i-1) \) is 1 if \( i \) is prime but 0 otherwise.
- \( \pi(n) = \frac{n}{\ln n} + O \left( \frac{n}{\ln^2 n} \right) \) by the Prime Number Theorem (when it is proved with an error term).

We have
\[ \pi(i) \left( \frac{1}{i+1} - \frac{1}{i} \right) = \frac{\pi(i)}{i(i+1) \ln i} + O \left( \frac{1}{(i+1) \ln^2 i} \right) \]
by the Prime Number Theorem. But this equals
\[ \frac{1}{i \ln i} - \frac{1}{i(i+1) \ln i} + O \left( \frac{1}{(i+1) \ln^2 i} \right) = \frac{1}{i \ln i} + O \left( \frac{1}{(i+1) \ln^2 i} \right). \]
Therefore,
\[
\sum_{p \leq n} \frac{1}{p} = \sum_{i=2}^{n} \frac{1}{i \ln i} + O\left(\frac{1}{(i+1) \ln^2 i}\right) = \ln(\ln(n)) + O(1),
\]
where we have used
\[
\sum_{i=2}^{n} \frac{1}{i \ln i} = \int_{2}^{n} \frac{1}{x \ln x} \, dx + O(1) = \ln(\ln(x)) + O(1)
\]
and
\[
\sum_{i=2}^{n} \frac{1}{(i+1) \ln^2 i} = O(1)
\]
by the Integral Test.

5 A Proof that uses Integration by Parts

This is the same as the previous proof, with the summation by parts replaced by integration by parts in a Stieltjes integral.

**Theorem 5.1** \( \sum_{p \leq n} \frac{1}{p} = \ln(\ln(n)) + O(1) \).

**Proof:** The preceding proof can be rewritten using Stieltjes integrals:
\[
\sum_{p \leq x} \frac{1}{p} = \int_{1.9}^{x} \frac{1}{t} \, d\pi(t).
\]
Integration by parts yields
\[
\frac{\pi(x)}{x} + \int_{1.9}^{x} \frac{\pi(t)}{t^2} \, dt.
\]
We use the Prime Number Theorem approximation \( \pi(x) = \frac{x}{\ln x} + O\left(\frac{x}{\ln^2 x}\right) \) to obtain
\[
\frac{1}{\ln x} + \int_{1.9}^{x} \frac{1}{t \ln t} + O\left(\int_{1.9}^{x} \frac{1}{t \ln^2 t}\right) = \ln(\ln(x)) + O(1).
\]
6 What Else is Known

Rosser and Schoenfeld [3] have shown that, when $n \geq 286,$

$$\ln(\ln n) - \frac{1}{2(\ln n)^2} + B \leq \sum_{p \leq n} \frac{1}{p} \leq \ln(\ln n) + \frac{1}{(2\ln n)^2} + B,$$

where $B = 0.261497212847643.$

Even though the sum $\sum_{p \leq n} \frac{1}{p}$ diverges, it grows very slowly:

- $\sum_{p \leq 10} \frac{1}{p} = 1.176$
- $\sum_{p \leq 10^6} \frac{1}{p} = 2.887$
- $\sum_{p \leq 10^9} \frac{1}{p} = 3.293$
- $\sum_{p \leq 10^{100}} \frac{1}{p} \sim 5.7$

References


