# On the upper bound of $\pi(x) - \mathrm{li}(x)$

David Stach
Comenius University
Faculty of Mathematics, Physics and Informatics
Mlynská dolina F1 842 48 Bratislava
Slovak Republic
E-mail: davsta@mathinstitute.sk

**Abstract**

Let $\pi(x)$ denote the prime counting function and let $\mathrm{li}(x)$ denote the logarithmic integral. We prove that

$$\pi(x) - \mathrm{li}(x) = \mathcal{O}(\sqrt{x}\log x).$$

The novelty of our approach lies in using new relations between $\pi(x)$ and the double sum of reciprocals of primes $\sum_{k=2}^{x}\sum_{p\le k}\frac{1}{p}$. By applying these relations, we develop a method on how to obtain a formula for $\pi(x) - \mathrm{li}(x)$ without using a complex analysis. Among all of the other interesting implications, our result implies a more regular distribution of primes than implied from the prime number theorem.

## 1 Introduction

Let $\pi(x) = \sum_{p\le x} 1$ and let

$$\mathrm{li}(x) = \lim_{\varepsilon\to 0^+}\left(\int_0^{1-\varepsilon}\frac{du}{\log u} + \int_{1+\varepsilon}^{x}\frac{du}{\log u}\right).$$

2020 *Mathematics Subject Classification*: Primary 11A41, 11N05, 11Y11; Secondary 40B05, 26A42, 65B15.

*Key words and phrases*: Primes, Primality, Prime counting function, Distribution of primes, Logarithmic Integral, Order of the function.

The beginnings of estimating $\pi(x)$ using $\mathrm{li}(x)$ date back to the 18th century to Gauss [3] who even suggested that $\pi(x) \sim \mathrm{li}(x)$. Although a direct calculation of $\mathrm{li}(x)$ showed a good estimation for $\pi(x)$, mathematicians were unable to rigorously prove what the real relation between them is. A major breakthrough came in 1859, when Riemann [9] introduced the explicit formula

$$\pi(x) = \sum_{k=1}^{\infty} \frac{\mu(k)}{k} \left( \mathrm{li}(x^{1/k}) - \sum_{\rho} \mathrm{li}(x^{\rho/k}) + \int_{x^{1/k}}^{\infty} \frac{dz}{(z^2-1)z\log z} \right)$$

where $\mu(k)$ denotes the Möbius function and $\rho$ denotes the nontrivial zeros of the zeta function. From the method Riemann used to derive this formula, it was obvious that $\mathrm{li}(x)$ is naturally related to $\pi(x)$. Although his formula does not prove $\pi(x) \sim \mathrm{li}(x)$, it is still remarkable for three reasons. First, it allows us to compute an exact value of $\pi(x)$ using a smooth function $\mathrm{li}(x)$. Second, it served as a basis for the proof that $\pi(x) - \mathrm{li}(x)$ changes sign infinitely often [7]. Third, it helped in obtaining a bound on the growth rate of $\pi(x)$ depending on the upper bound of the real parts of $\rho$. If $\beta$ denotes that upper bound, then $\pi(x) - \mathrm{li}(x) = \mathcal{O}(x^{\beta} \log x)$ [5]. From Riemann's work and the prime number theorem, which was proved by Hadamard [4] and de la Vallée Poussin [2] in 1896, it follows that $\beta \in [\frac{1}{2}, 1]$. We strengthen these results and prove that $\pi(x) - \mathrm{li}(x) = \mathcal{O}(\sqrt{x} \log x)$.

# 2 PRELIMINARIES

Throughout this article, $x$ denotes a real number, $n$ denotes a natural number, and $p$ denotes a prime number.

**Definition 2.1.** Let $\{a_n\}$ be a sequence of complex numbers. Define the partial sum

$$A(x) = \sum_{0 \le n \le x} a_n.$$

Let $f$ be a real valued and continuously differentiable function on $[y, x]$. Then, Abel's summation is defined by

$$\sum_{y < n \le x} a_n f(n) = A(x)f(x) - A(y)f(y) - \int_y^x A(u) f'(u) du.$$

**Definition 2.2.** If $g(x) > 0$ for all $x \geq x_o$, we write $f(x) = \mathcal{O}(g(x))$ [6] if there exists a constant $H$ such that for all $x \geq x_o$, we have

$$|f(x)| \leq Hg(x).$$

It is important to note that $f(u) = \mathcal{O}(g(u))$ for $u \geq x_o$ implies

$$\int_{x_o}^{x} f(u)du = \mathcal{O}\left(\int_{x_o}^{x} g(u)du\right).$$

**Definition 2.3.** Let $m$ and $n$ be natural numbers and let $f$ be a real valued and continuously differentiable function on $[m, n]$. Then, the Euler-Maclaurin summation formula [1] is defined by

$$\sum_{i=m}^{n} f(i) = \int_{m}^{n} f(x)dx + \frac{f(n) + f(m)}{2} + \frac{f'(n) - f'(m)}{12} + R_2$$

where

$$|R_2| \leq \frac{1}{12}\int_{m}^{n} \left|f^2(x)\right| dx.$$

Since the remainder term $R_2$ only has a negligible value for our purposes, this inequality is sufficient for us.

**Definition 2.4.** Let $M$ be the Mertens constant [8] defined by

$$M = \lim_{n\to\infty}\left(\sum_{p\leq n}\frac{1}{p} - \log\log n\right).$$

$M = 0.2614972128\ldots$

**Definition 2.5.** Let $E(x)$ be the error term defined by

$$E(x) = \sum_{p\leq x}\frac{1}{p} - \log\log x - M.$$

It is proved in [10] that $E(x)$ changes sign infinitely often. For $x > 1$, we have [11]

$$|E(x)| < \frac{1}{\log^2 x}.$$

# 3 MAIN RESULTS

**Lemma 3.1.** *For* $x \geq e^6$*, we have*

$$\mathrm{li}(x) - \frac{x}{\log x} \leq \frac{x}{\log^2 x} + \frac{4x}{\log^3 x} + C \tag{3.1}$$

*where*

$$C = -\frac{8}{\log^3 2} - \frac{2}{\log 2} - \frac{2}{\log^2 2} + \mathrm{li}(2) + \int_2^{e^6} \frac{12du}{\log^4 u}.$$

*Proof.* Recall that $\mathrm{li}(x) = \mathrm{li}(2) + \int_2^x \frac{du}{\log u}$. After two integrations by parts, we obtain

$$\mathrm{li}(x) = \frac{x}{\log x} + \frac{x}{\log^2 x} - \frac{2}{\log 2} - \frac{2}{\log^2 2} + \mathrm{li}(2) + \int_2^x \frac{2du}{\log^3 u}. \tag{3.2}$$

After integrations by parts of the last term, we have

$$\int_2^x \frac{2du}{\log^3 u} = \frac{2x}{\log^3 x} - \frac{4}{\log^3 2} + \int_2^x \frac{6du}{\log^4 u}.$$

Since $\frac{6}{\log^4 u} \leq \frac{1}{\log^3 u}$ holds for all $u \geq e^6$, we have

$$\int_2^x \frac{2du}{\log^3 u} - \frac{2x}{\log^3 x} + \frac{4}{\log^3 2} - \int_2^{e^6} \frac{6du}{\log^4 u} = \int_{e^6}^x \frac{6du}{\log^4 u} \leq \int_{e^6}^x \frac{du}{\log^3 u} \leq \int_2^x \frac{du}{\log^3 u}.$$

From this inequality for $x \geq e^6$, we have

$$\int_2^x \frac{2du}{\log^3 u} \leq \frac{4x}{\log^3 x} - \frac{8}{\log^3 2} + \int_2^{e^6} \frac{12du}{\log^4 u}.$$

Combining this with (3.2) we obtain (3.1). □

**Lemma 3.2.** *For* $n \geq 2$*, we have*

$$\sum_{p \leq n} \frac{1}{p} = \frac{\pi(n)}{n} + \int_2^n \frac{\pi(x)}{x^2} dx. \tag{3.3}$$

*Proof.* Let $A(x) = \pi(x)$ and let $f(x) = \frac{1}{x}$. Then, this equation directly follows from Definition 2.1, since $\pi(x) = 0$ for $x < 2$. □

**Lemma 3.3.** *For $n \geq 2$, we have*

$$\sum_{k=2}^{n} \sum_{p \leq k} \frac{1}{p} = (n+1) \sum_{p \leq n} \frac{1}{p} - \pi(n). \tag{3.4}$$

*Proof.*

$$\begin{aligned}
\sum_{k=2}^{n} \sum_{p \leq k} \frac{1}{p} &= \sum_{p \leq 2} \frac{1}{p} + \sum_{p \leq 3} \frac{1}{p} + \sum_{p \leq 4} \frac{1}{p} + \ldots + \sum_{p \leq n} \frac{1}{p} \\
&= \frac{n-1}{p_1} + \frac{n-1-(p_2 - p_1)}{p_2} + \ldots + \frac{n-1-(p_{\pi(n)} - p_1)}{p_{\pi(n)}} \\
&= \sum_{i=1}^{\pi(n)} \frac{n-1-(p_i - p_1)}{p_i} \qquad (3.5)
\end{aligned}$$

where $p_i$ is the i-th prime. Substituting 2 for $p_1$ in (3.5), we obtain (3.4). □

**Corollary 3.4.** *Equating* (3.3) *and* (3.4)*, we obtain an elegant relation between the double sum and the integral involving prime numbers*

$$\sum_{k=2}^{n-1} \sum_{p \leq k} \frac{1}{p} = n \int_{2}^{n} \frac{\pi(x)}{x^2} dx.$$

**Lemma 3.5.** *For $n \geq 2$, we have*

$$\sum_{k=2}^{n} \sum_{p \leq k} \frac{1}{p} = \sum_{k=2}^{n} \log \log k + (n-1)M + \sum_{k=2}^{n} E(k). \tag{3.6}$$

*Proof.* This equation is easily derived in a straightforward manner using Definition 2.5. □

**Theorem 3.6.** *For $n > 2$, we have*

$$\pi(n) - \mathrm{li}(n) = nE(n) + \frac{\log \log n}{2} - \sum_{k=2}^{n-1} E(k) + C_1 \tag{3.7}$$

*where $-0.95171\ldots \geq C_1 \geq -1.07193\ldots$*

*Proof.* We start by equating (3.4) and (3.6)

$$(n+1)\sum_{p\le n}\frac{1}{p} - \pi(n) = \sum_{k=2}^{n}\log\log k + (n-1)M + \sum_{k=2}^{n}E(k). \tag{3.8}$$

Applying Definition 2.5, we have

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

Applying Definition 2.3, we have

$$\sum_{k=2}^{n}\log\log k = n\log\log n - \mathrm{li}(n) - 2\log\log 2 + \mathrm{li}(2) + \\ + \frac{\log\log n}{2} + \frac{\log\log 2}{2} + \frac{1}{12}\left(\frac{1}{n\log n} - \frac{1}{2\log 2}\right) + R_2 \tag{3.9}$$

where

$$|R_2| \le \frac{1}{12}\int_2^n \left|-\frac{\log x + 1}{x^2\log^2 x}\right| dx = \frac{1}{12}\left(\frac{1}{2\log 2} - \frac{1}{n\log n}\right).$$

Combining these results with (3.8) we obtain (3.7), where

$$C_1 = 2M + \frac{3\log\log 2}{2} - \mathrm{li}(2) - \frac{1}{12}\left(\frac{1}{n\log n} - \frac{1}{2\log 2}\right) - R_2.$$

□

**Theorem 3.7.** *If* $\pi(x) - \mathrm{li}(x) = \mathcal{O}(\sqrt{x}\log x)$*, then for* $n > 2657$ *we have*

$$\left|\sum_{k=2}^{n-1}E(k) - \frac{\log\log n}{2} - C_1 + nC_2\right| \le \frac{n}{8\pi}\left(\frac{2(\log 2657 + 2)}{\sqrt{2657}} - \frac{2(\log n + 2)}{\sqrt{n}}\right) \tag{3.10}$$

*where*

$$C_2 = M + \frac{\pi(2657)}{2657} + \log\log 2657 - \sum_{p\le 2657}\frac{1}{p} - \frac{\mathrm{li}(2657)}{2657} = -0.00701\ldots$$

*Proof.* Dividing (3.7) by $n$, we have

$$\frac{\pi(n)}{n} - \frac{\mathrm{li}(n)}{n} = E(n) + \frac{\log\log n}{2n} - \frac{\sum_{k=2}^{n-1} E(k)}{n} + \frac{C_1}{n}.$$

Applying Lemma 3.2, we have

$$\sum_{p\le n} \frac{1}{p} - \int_2^n \frac{\pi(x)}{x^2} dx - \frac{\mathrm{li}(n)}{n} = E(n) + \frac{\log\log n}{2n} - \frac{\sum_{k=2}^{n-1} E(k)}{n} + \frac{C_1}{n}. \quad (3.11)$$

Applying Definition 2.5, we have

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

It is proved in [12] that if $\pi(x) - \mathrm{li}(x) = \mathcal{O}(\sqrt{x}\log x)$, then for all $x \ge 2657$, we have $|\pi(x) - \mathrm{li}(x)| \le \frac{\sqrt{x}\log x}{8\pi}$. For this reason, we modify the second term of the left-hand side of (3.11) as follows

$$\int_2^n \frac{\pi(x)}{x^2} dx = \int_2^{2657} \frac{\pi(x)}{x^2} dx + \int_{2657}^n \frac{\pi(x)}{x^2} dx.$$

Substituting $\mathrm{li}(x) + \mathcal{O}(\sqrt{x}\log x)$ for $\pi(x)$ in the last term, we have

$$\int_2^n \frac{\pi(x)}{x^2} dx = \int_2^{2657} \frac{\pi(x)}{x^2} dx + \int_{2657}^n \frac{\mathrm{li}(x)}{x^2} dx + \int_{2657}^n \frac{\mathcal{O}(\sqrt{x}\log x)}{x^2} dx.$$

Applying Lemma 3.2, we have

$$\int_2^{2657} \frac{\pi(x)}{x^2} dx = \sum_{p\le 2657} \frac{1}{p} - \frac{\pi(2657)}{2657}.$$

Integrating, we have

$$\int_{2657}^n \frac{\mathrm{li}(x)}{x^2} dx = \log\log n - \frac{\mathrm{li}(n)}{n} - \log\log 2657 + \frac{\mathrm{li}(2657)}{2657}$$

and

$$\int_{2657}^n \frac{\mathcal{O}(\sqrt{x}\log x)}{x^2} dx = \mathcal{O}\left(\frac{2(\log 2657 + 2)}{\sqrt{2657}} - \frac{2(\log n + 2)}{\sqrt{n}}\right).$$

So, we have

$$\int_2^n \frac{\pi(x)}{x^2}dx = \sum_{p\le 2657}\frac{1}{p} - \frac{\pi(2657)}{2657} + \log\log n - \frac{\mathrm{li}(n)}{n} - $$
$$- \log\log 2657 + \frac{\mathrm{li}(2657)}{2657} + \mathcal{O}\left(\frac{2(\log 2657+2)}{\sqrt{2657}} - \frac{2(\log n+2)}{\sqrt{n}}\right).$$

Combining these results with (3.11) we obtain (3.10).

□

We proof the validity of (3.10) in a sequence of the following lemmas.

**Lemma 3.8.** *For* $n-1>e^6$*, we have*

$$\left|\sum_{k=2}^{n-1} E(k)\right| < (n-1)\left(\frac{1}{\log^2(n-1)} + \frac{4}{\log^3(n-1)}\right) + C_3 \qquad (3.12)$$

*where* $16.51740\ldots \le C_3 \le 17.01786\ldots$

*Proof.* Applying Definition 2.5, for $n\ge 3$ we have

$$\left|\sum_{k=2}^{n-1} E(k)\right| < \sum_{k=2}^{n-1}\frac{1}{\log^2 k}. \qquad (3.13)$$

Applying Definition 2.3, we have

$$\sum_{k=2}^{n-1}\frac{1}{\log^2 k} = \mathrm{li}(n-1) - \frac{n-1}{\log(n-1)} - \mathrm{li}(2) + \frac{2}{\log(2)} + \frac{1}{2\log^2(n-1)} +$$
$$+ \frac{1}{2\log^2 2} - \frac{1}{6(n-1)\log^3(n-1)} + \frac{1}{12\log^3 2} + R_2 \qquad (3.14)$$

where

$$|R_2| \le \frac{1}{12}\int_2^{n-1}\left|\frac{2(\log x+3)}{x^2\log^4 x}\right|dx = \frac{1}{12}\left(\frac{1}{\log^3 2} - \frac{2}{(n-1)\log^3(n-1)}\right).$$

Applying Lemma 3.1, for $n-1>e^6$ we have

$$\mathrm{li}(n-1) - \frac{n-1}{\log(n-1)} < \frac{n-1}{\log^2(n-1)} + \frac{4(n-1)}{\log^3(n-1)} + C.$$

Combining these results with (3.13) we obtain (3.12), where

$$C_3 = C - \mathrm{li}(2) + \frac{2}{\log(2)} + \frac{1}{2\log^2 2} + \frac{1}{12\log^3 2} - \\ - \frac{1}{6(n-1)\log^3(n-1)} + \frac{1}{2\log^2(n-1)} + R_2.$$

□

Let $R_{10}$ denote the right-hand side of (3.10) and let $R_{12}$ denote the right-hand side of (3.12).

**Lemma 3.9.** *If* $\sum_{k=2}^{n-1} E(k) > 0$*, then* (3.10) *holds for* $n \geq 18863$*.*

*Proof.* Substituting $R_{12}$ for $\sum_{k=2}^{n-1} E(k)$ in (3.10), we have

$$-R_{10} \leq R_{12} - \frac{\log\log n}{2} - C_1 + nC_2 \leq R_{10}. \tag{3.15}$$

For this proof, it is sufficient if we replace $C_1, C_3$ by their bounding values, adequately for each inequality, so that the inequality certainly holds. Let $C_1 = -0.95171$, let $C_3 = 16.51740$. Then, the first inequality of (3.15) holds for $n \geq 12141$. Let $C_1 = -1.07193$, let $C_3 = 17.01786$. Then, the second inequality of (3.15) holds for $n \geq 18863$. □

**Lemma 3.10.** *If* $\sum_{k=2}^{n-1} E(k) \leq 0$*, then* (3.10) *holds for* $n \geq 878649$*.*

*Proof.* Let $\sum_{k=2}^{n-1} E(k) = -R_{12}$. Then, this result can be easily obtained using the same approach as used in the previous lemma. □

It can be verified numerically that if $2657 < n < 18863$, then (3.10) holds, and since for $n < 878649$ we have $\sum_{k=2}^{n-1} E(k) > 0$, we conclude that (3.10) holds for $n > 2657$.

# References

[1] T. M. Apostol, An Elementary View of Euler's Summation Formula, *Amer. Math. Monthly* **106** (1999), 409–418. MR 1699259. Zbl 1076.41509. doi.org.

[2] Ch. J. de la Vallée Poussin, Recherches analytiques sur la théorie des nombres premiers, *Ann. Soc. Sci. Bruxelles* (1896) 20 B, 21 B: 183–256, 281–352, 363–397, 351–368. JFM 27.0155.05. archive.org.

[3] C. F. Gauss, Werke, *Königl. Gesellschaft d. Wissenschaften zu Göttingen* **2** (1863), 444–447. archive.org.

[4] J. Hadamard, Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques, *Bull. Soc. Math. Fr.* **24** (1896), 199–220. MR 1504264. JFM 27.0154.01. eudml.org.

[5] A. E. Ingham, The Distribution of Prime Numbers, *Cambridge Univ. Press* **30** (1932). MR 1074573. Zbl 0006.39701.

[6] E. Landau, Handbuch der Lehre von der Verteilung der Primzahlen, *B. G. Teubner* (1909), 59-61. JFM 40.0232.08. archive.org.

[7] J. E. Littlewood, Sur la distribution des nombres premiers, *C. R. Acad. Sci. Paris* **158** (1914), 1869–1872. JFM 45.0305.01.

[8] F. Mertens, Ein Beitrag zur analytischen Zahlentheorie, *J. für die Reine und Angew. Math.* **78** (1874), 46–62. JFM 06.0116.01. eudml.org.

[9] B. Riemann, Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse, *Cambridge Library Collection - Mathematics, Cambridge Univ. Press* (2013), 136-144. doi.org.

[10] G. Robin, Sur l'ordre maximum de la fonction somme des diviseurs, *Progr. Math.* **38** (1983), 233–244. Zbl 0521.10040.

[11] J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6** (1962), 64-94. MR 137689. Zbl 0122.05001. doi.org.

[12] L. Schoenfeld, Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. II. *Math. Comp.* **30** (1976), 337–360. MR 457374. Zbl 0326.10037. doi.org.