THE ORIGIN OF THE LOGARITHMIC INTEGRAL IN THE PRIME NUMBER THEOREM

KOLBJØRN TUNSTRØM

ABSTRACT. We establish why $\text{li}(x)$ outperforms $x/\log x$ as an estimate for the prime counting function $\pi(x)$. The result follows from subdividing the natural numbers into the intervals $s_k := \{p_{2k}^2, \ldots, p_{2k+1}^2 - 1\}$, $k \geq 1$, each being fully sieved by the $k$ first primes $\{p_1, \ldots, p_k\}$. Denoting the number of primes in $s_k$ by $\pi_k$, we show that $\pi_k \sim |s_k|/\log p_{2k+1}$ and that $\pi(x) \sim \text{li}(x)$ originates as a continuum approximation of the sum $\sum_k \pi_k$. In contrast, $\pi(x) \sim x/\log x$ stems from sieving repeatedly in regions already completed — explaining why $x/\log x$ underestimates $\pi(x)$. The explanatory potential arising from defining $s_k$ appears promising, evidenced in the last section where we outline further research.

1. Introduction

The prime number theorem states that as $x \to \infty$, the number of primes below $x$, denoted $\pi(x)$, can be approximated by either relations

$$\pi(x) \sim \frac{x}{\log x} \quad \text{or} \quad \pi(x) \sim \text{li}(x),$$

where $\text{li}(x)$ is the logarithmic integral defined by $\text{li}(x) := \int_2^x \frac{dt}{\log t}$. While the two estimates in (1) are equivalent — easily proved by a series expansion of $\text{li}(x)$, where $x/\log x$ appears as the leading term—they differ in performance. As is well known, established by proof already in 1899 by de la Vallée-Poussin [2], $\text{li}(x)$ is a superior guess of the number of primes up to $x$ compared to $x/\log x$. Nonetheless, the exact reason why $\text{li}(x)$ outcompetes $x/\log x$ as an estimator for $\pi(x)$ has remained unresolved. The absence of an explanation is evident in various literature surveys [3, 4, 5, 6], and perhaps most clearly stated by Goldston [6]: ‘The extraordinarily good fit between $\pi(x)$ and $\text{li}(x)$, far better than the first approximation $x/\log x$, has been the subject of intensive but largely unsuccessful investigation for the last one hundred years.’

In this paper, we introduce a slight shift of perspective on the sieve of Eratosthenes. Rather than focusing directly on the number of primes below a given number $x$, we examine what the effect is of sieving by the $k$th prime $p_k$. This eventually inspires a subdivision of the natural numbers in terms of the specific intervals $s_k := \{p_{2k}^2, \ldots, p_{2k+1}^2 - 1\}$, where each $s_k$ has the essential property that it is sieved entirely by the $k$ first primes. By writing the number of primes in $s_k$ as $\pi_k$, we have from the prime number theorem that $\pi_k \sim |s_k|/\log p_{2k+1}$. Summing these estimates up to the desired value of $x$ gives us $\pi(x)$, and it is easy to see that $\text{li}(x)$ arises as a continuum approximation of this sum. $x/\log x$, on the other hand,
derives from sieving the whole interval below \( x \) by every prime \( p \leq \sqrt{x} \), effectively overcounting the number of composites and underestimating ditto primes. As a consequence of these insights, we are able to explain the relation between \( x / \log x \) and \( \text{li}(x) \) in the prime number theorem.

The remaining parts of the paper consist of three sections. In the first, we expand in a slower pace the content of the previous paragraph, motivating the subdivision of the natural numbers from the sieve of Eratosthenes, and stating the definition of the intervals \( s_k \). From the definition we establish probabilistic estimates of the number of primes within each interval, and in turn, this naturally steers us towards what we could call a probabilistic prime number theorem. The insight we acquire from the probabilistic perspective motivates an alternative formulation of the prime number theorem—where the logarithmic integral now originates as a continuum approximation of the sum taken over the individual interval estimates. The second section contains a more compact overview of results and detailed proofs. In the third and last section we explore some of the implications following in the wake of defining \( s_k \). Most notably, we discuss how our results relate to the studies of primes in short intervals and the error term in the prime number theorem.

2. Deriving \( \text{li}(x) \) from a subdivision of the natural numbers

In 1849, the German astronomer Encke wrote a letter to his former academic advisor Gauss, in which he shared his thoughts on the frequency of primes. In his reply, Gauss explained that his first inquiries into the topic dated as far back as 1792 or 1793, when he had received the Lambert supplements to the logarithmic tables, containing lists of all primes up to 1 million—and also that he had kept on adding to that million: Gauss wrote to Encke that he 'frequently spent an idle quarter of an hour to count a chiliad\(^1\) here and there', which allowed him to calculate the average number of primes over short intervals, eventually conjecturing that\(^2\)

\[
\pi(x) \sim \text{li}(x).
\]

As we will see, Gauss’s approach of probing the distribution of primes across short intervals is key to understanding the emergence of the logarithmic integral. Essentially we will apply the same strategy, with the critical distinction that our variant of Gauss’s chiliads will be a set of precisely stated intervals that subdivides the natural numbers. Their definition is best motivated by a recapitulation of the sieve of Eratosthenes, which is a straightforward sieving procedure for finding all primes up to a given number \( x \) and a common starting point for understanding the distribution of primes.

Fairly typical of what is found in textbooks or other literature—see e.g. \([9, 6, 8]\)—we can describe the sieve of Eratosthenes as follows: First we remove all multiples of the first prime \( p_1 = 2 \) up to \( x \). The next remaining number is the second prime \( p_2 = 3 \), so now we remove all its surviving multiples up to \( x \). Thereafter, we repeat the process until no more composites can be removed by further sieving. This happens when we have sieved by the primes \( p \leq \sqrt{x} \), as any composite with prime factors all larger than \( \sqrt{x} \) must necessarily exceed \( x \). What’s more, the sieve of Eratosthenes lets us construct a probabilistic estimate of the number of primes up to \( x \) by the following reasoning: The probability that a randomly picked number

\(^1\)Gauss counted the primes within \textit{chiliads}; intervals of 1000 consecutive integers.

\(^2\)The historical account is found in \([10, p. 174]\).
less than or equal to \( x \) is prime equals the probability that the said number is not divisible by any of the primes \( p \leq \sqrt{x} \). Treating these events as independent, we arrive at the probability given by the Euler product

\[
\prod_{p \leq \sqrt{x}} \left(1 - \frac{1}{p}\right),
\]

and the expected number of primes up to \( x \) is therefore

\[
\tilde{\pi}(x) = x \cdot \prod_{p \leq \sqrt{x}} \left(1 - \frac{1}{p}\right),
\]

(2)

where \( \tilde{\pi}(x) \) is introduced as a probabilistic prime counting function.

What now if we wanted to guess the number of primes up to some \( y > x \), instead of \( x \)? Applying the same arguments, we could produce an estimate simply by replacing \( x \) by \( y \) in (2). However, by doing so, we would not factor in that we already have sieved the interval up to \( x \). Consequently, the estimate in (2) fails to take into account that every sieve step leaves behind it a completely sieved interval.

To remedy this shortcoming, let us dim the light on the upper limit \( x \) and shift focus towards what happens during one step of the sieve of Eratosthenes. For this purpose, let \( P_k := \{p_1, \ldots, p_k\} \) be the set of the \( k \) first primes, and assume we have already sieved the natural numbers by all primes up to and including \( p_{k-1} \). Necessarily, \( p_k^2 \) is the next composite that is not sieved, and therefore, when proceeding with step \( k \), the sieving has no effect on numbers smaller than \( p_k^2 \)—we are only removing elements from \( p_k^2 \) and upwards, and thereby locating all primes between \( p_k^2 \) and \( p_{k+1}^2 \). Evidently, for any \( k \geq 1 \), the interval \( \{p_k^2, \ldots, p_{k+1}^2 - 1\} \) can be sieved apart from any other interval by the \( k \) first primes. That being the case, what transpires as our version of Gauss’s chiliads is the set of intervals defined by

\[
s_k := \{p_k^2, \ldots, p_{k+1}^2 - 1\}.
\]

To summarize this insight:

**Remark 2.1.** A logical consequence of the sieve of Eratosthenes is that the natural numbers can be split into the set of intervals \( s_k \), \( k \geq 1 \), where \( s_k \) has the vital property that any one of its elements is either divisible by some \( p \in P_k \) or else is a prime \( p / \in P_k \). Therefore, each \( s_k \) can be sieved independently by the \( k \) first primes.

Now, let us return to the probabilistic prime counting function. How will our knowledge of \( s_k \) and the fact that it can be sieved by the \( k \) first primes affect \( \tilde{\pi}(x) \)? If we let the length of \( s_k \) be given by \( l_k := |s_k| = p_{k+1}^2 - p_k^2 \), and introduce \( \tilde{\pi}_k \) to be the expected number of primes in \( s_k \), it follows from the arguments leading to (2) that

\[
\tilde{\pi}_k = l_k \cdot \prod_{p \in P_k} \left(1 - \frac{1}{p}\right).
\]

(3)

If we further assume \( k \) to be the integer such that \( p_k^2 \leq x < p_{k+1}^2 \), then it follows that our improved estimate of \( \tilde{\pi}(x) \)—now incorporating that each interval \( s_k \) is fully sieved by the \( k \) first primes—is given by

\[
\tilde{\pi}(x) = \sum_{j=1}^{k-1} \tilde{\pi}_j + \frac{x - p_k^2}{l_k} \tilde{\pi}_k.
\]

(4)
A natural next step is to apply Merten’s product theorem to (2) and (4), which leads to the probabilistic estimates

\begin{equation}
\tilde{\pi}(x) \sim 2e^{-\gamma} \frac{x}{\log x} \quad \text{and} \quad \tilde{\pi}(x) \sim 2e^{-\gamma} \text{li}(x).
\end{equation}

The left estimate here corresponds to (2) and is previously stated in the literature, see e.g. [7]. The right—and more precise estimate—corresponds to (4), and is to the best of our knowledge an original result. The intermediate details necessary to arrive at this estimate are found in the proof of Theorem 3.4. Most importantly, we have arrived at two probabilistic estimates that each have their counterpart in the prime number theorem (1); except for a constant factor $2e^{-\gamma}$, (5) is identical to (1).

As we explain in more detail in Section 4.1, the probabilistic estimates in (5) can be understood as arising from finite sample spaces of possible prime positions within the specific intervals $s_k$. The positions of the actual primes naturally lie in these sample spaces and therefore the arguments we have laid out above also apply to the primes. In conclusion then, the reason why li(x) is a better estimate for \(\pi(x)\) than \(x/\log x\) is that it takes into account that each sieve step \(k\) completes the sieving of the interval \(s_k\), while \(x/\log x\) does not. Accordingly, this suggests that li(x) is the appropriate function to use for estimating the number of primes up to \(x\).

Rather than (5), it appears natural now to write

\begin{equation}
\tilde{\pi}_k \sim 2e^{-\gamma} \frac{l_k}{\log p_{k+1}^2} \quad \text{and} \quad \tilde{\pi}(x) \sim 2e^{-\gamma} \text{li}(x),
\end{equation}

where the left term—derived from (3) via Merten’s theorem (see Theorem 3.3)—provides the expected number of primes within the intervals \(s_k\), and the right term—obtained as a continuous approximation of \(\sum_k \tilde{\pi}_k\) (see Theorem 3.4)—describes the expected number of primes up to \(x\). Furthermore, the expression for \(\tilde{\pi}_k\) in (6) reflects the underlying sieving process, in that sieving by the \(k\) first primes removes all composites up to \(p_{k+1}^2\).

Following through to completion, (6) points towards the alternative formulation of the prime number theorem given by

\begin{equation}
\pi_k \sim \frac{l_k}{\log p_{k+1}^2} \quad \text{and} \quad \pi(x) \sim \text{li}(x),
\end{equation}

where \(\pi_k\) is the number of primes in \(s_k\) (see Theorem 3.5), and the natural interpretation of the logarithmic integral is that it originates from a continuous approximation of \(\sum_k \pi_k\) (see Theorem 3.6).
Proof. Each number in $s_k$ is either divisible by some $p \in \mathcal{P}_k$ or else is a prime $p \notin \mathcal{P}_k$. The probability of a number in $s_k$ being prime is therefore equivalent to the probability of the number not being divisible by any $p \in \mathcal{P}_k$. Assuming these events to be independent, the desired probability is given by the Euler product

$$\prod_{p \in \mathcal{P}_k} \left(1 - \frac{1}{p}\right),$$

and the expected number of primes in $s_k$ is found by multiplying with the length $l_k$ of the sequence, resulting in (8).

Lemma 3.2. Let $\tilde{\pi}(x)$ denote the expected number of primes less than or equal to $x$ and let $k$ be the integer such that $p_k^2 \leq x < p_{k+1}^2$. Then

$$\tilde{\pi}(x) = \sum_{j=1}^{k-1} \tilde{\pi}_j + \frac{x - p_k^2}{l_k} \tilde{\pi}_k.$$ (9)

Proof. The result follows by splitting $\tilde{\pi}(x)$ into a sum of two parts,

$$\tilde{\pi}(x) = \tilde{\pi}(p_k^2) + [\tilde{\pi}(x) - \tilde{\pi}(p_k^2)].$$ (10)

The first term on the right can be split further into a sum over the intervals $s_j$, $1 \leq j < k$,

$$\tilde{\pi}(p_k^2) = \sum_{j=1}^{k-1} [\tilde{\pi}(p_{j+1}^2) - \tilde{\pi}(p_j^2)] = \sum_{j=1}^{k-1} \tilde{\pi}_j.$$

The second term on the right in (10) represents the mean number of primes in the interval $[p_k^2, x]$. Since this interval is contained within $s_k$ it follows that

$$\tilde{\pi}(x) - \tilde{\pi}(p_k^2) = \left(x - p_k^2\right) \cdot \prod_{p \in \mathcal{P}_k} \left(1 - \frac{1}{p}\right)$$

$$= \frac{x - p_k^2}{l_k} \left[l_k \cdot \prod_{p \in \mathcal{P}_k} \left(1 - \frac{1}{p}\right)\right]$$

$$= \frac{x - p_k^2}{l_k} \tilde{\pi}_k.$$

Combining these two results produces (9). □

Theorem 3.3. Let $\tilde{\pi}_k$ denote the expected number of primes in the interval $s_k$. When $k \to \infty$ we have that

$$\tilde{\pi}_k \sim 2e^{-\gamma} \frac{l_k}{\log p_{k+1}}.$$ (11)

Proof. Assume that $x$ is a real number within the interval $s_k$, so that $p_k^2 \leq x < p_{k+1}^2$, and consider the situation when $k \to \infty$. Then we can state Merten’s product theorem [16] as

$$\prod_{p \in \mathcal{P}_k} (1 - \frac{1}{p}) = e^{-\gamma + \delta} \frac{1}{\log \sqrt{x}} = 2e^{-\gamma + \delta} \frac{1}{\log x},$$

where $\gamma$ is the Euler-Mascheroni constant and $\delta$ is a constant.
where $\gamma$ is the Euler–Mascheroni constant and $\delta$ is a measure of the uncertainty of the approximation, satisfying

$$|\delta| < \frac{4}{\log(\sqrt{x} + 1)} + \frac{2}{\sqrt{x}\log\sqrt{x}} + \frac{1}{2\sqrt{x}}.$$  

It now follows immediately from (8) that

$$\tilde{\pi}_k = 2e^{-\gamma + \delta} \frac{l_k}{\log x}$$  

for any $x$ such that $p_k^2 \leq x < p_{k+1}^2$. Additionally, (12) implies that in order to minimize the error we should choose $x$ as large as possible—that is, $x = p_{k+1}^2 - \epsilon$, where $\epsilon > 0$ is infinitesimal—by which we obtain

$$\tilde{\pi}_k \sim 2e^{-\gamma} \frac{l_k}{\log p_{k+1}^2}.$$  

Note that even without access to the error term (12), it would be natural to argue heuristically for the choice of $x = p_{k+1}^2 - 1$.

**Theorem 3.4.** Let $\tilde{\pi}(x)$ denote the expected number of primes less than or equal to $x$. When $x \to \infty$ we have that

$$\tilde{\pi}(x) \sim 2e^{-\gamma} \text{li}(x).$$

**Proof.** We start by defining $\text{li}_k$ as the logarithmic integral over the interval $s_k$,

$$\text{li}_k := \text{li}(p_{k+1}^2) - \text{li}(p_k^2) = \int_{p_k^2}^{p_{k+1}^2} \frac{dt}{\log t}.$$  

As $\log a < \log b$ whenever $a < b$ for any real numbers $a, b > 0$, $\text{li}_k$ satisfies

$$\frac{l_k}{\log p_{k+1}^2} < \text{li}_k < \frac{l_k}{\log p_k^2},$$

and since—from (13)—the following relations hold when $k \to \infty$,

$$\tilde{\pi}_k \sim 2e^{-\gamma} \frac{l_k}{\log p_{k+1}^2} \quad \text{and} \quad \tilde{\pi}_k \sim 2e^{-\gamma} \frac{l_k}{\log p_k^2},$$

we have that

$$\tilde{\pi}_k \sim 2e^{-\gamma} \text{li}_k.$$  

Now, let $k$ be the integer such that $p_k^2 \leq x < p_{k+1}^2$. Then it follows from (9) and (15) that

$$\tilde{\pi}(x) = \sum_{j=1}^{k-1} \tilde{\pi}_j + \frac{x - p_k^2}{l_k} \tilde{\pi}_k$$

$$\sim 2e^{-\gamma} \sum_{j=1}^{k-1} \text{li}_j + \frac{x - p_k^2}{l_k} 2e^{-\gamma} \text{li}_k$$

$$= 2e^{-\gamma} \text{li}(x).$$
Strictly, the last expression should be $2e^{-γ}\ln(x) - 2$, since $\ln(x)$ has lower integration limit $x = 2$, thereby including the 2 first primes 2 and 3. But that difference is insignificant as $x \to \infty$. □

**Remark 3.1.** Note that we can easily obtain an upper bound on the relative error

$$\eta_k := \frac{l_k/\log p_k^2 - l_k/\log p_{k+1}^2}{l_k/\log p_k^2} = \frac{\log p_{k+1}^2 - \log p_k^2}{\log p_k^2}.$$  

by applying Bertrand’s postulate. This states that for any integer $n > 1$ we will always find a prime $p$ such that $n < p < 2n$. Choosing $n$ to be $p_k$, we see that $p_k < p_{k+1} < 2p_k$ and therefore that $p_{k+1}^2 < 4p_k^2$. Then it follows from (16) that

$$\eta_k = \frac{\log p_{k+1}^2}{\log p_k^2} - 1 < \frac{4p_k^2}{\log p_k^2} - 1 = \frac{4}{\log p_k^2}.$$  

So, as $k \to \infty$, the relative error approaches 0. Naturally, the actual error approaches 0 much faster than what we found here, as Bertrand’s postulate is a very crude approximation to the average distance between primes. Applying the prime number theorem, the average gap length around $p_k$ is $\log p_k$, and hence $p_{k+1} \sim p_k + \log p_k$. The expected relative error should therefore satisfy

$$\eta_k \approx \frac{\log(p_k + \log p_k)}{\log p_k} - 1.$$  

Necessarily, since $\ln_k$ is bounded as shown in (14), the relative error of any numerical approximation to $\ln_k$ must obey (17) and approach 0 as $k \to \infty$.

**Theorem 3.5.** Let $\pi_k$ denote the number of primes in the interval $s_k$. When $k \to \infty$ we have that

$$\pi_k \sim \frac{l_k}{\log p_{k+1}^2}.$$  

Proof. The truth of the statement follows directly from the original prime number theorem, since

$$\pi_k = \pi(p_{k+1}^2) - \pi(p_k^2) \sim \ln(p_{k+1}^2) - \ln(p_k^2) = \ln_k$$

$$= \int_{p_k^2}^{p_{k+1}^2} \frac{dt}{\log t} \approx \frac{l_k}{\log p_{k+1}^2}.$$  

The choice of the integral approximation

$$\frac{l_k}{\log p_{k+1}^2}$$  

rather than e.g.

$$\frac{l_k}{\log p_k^2} \text{ or } \frac{l_k}{\log \frac{1}{2}(p_k^2 + p_{k+1}^2)}$$  

is motivated by the proof of Theorem 3.3, while the accuracy of the approximation is discussed in Remark 3.1 above. □

**Theorem 3.6.** Let $\pi(x)$ denote the number of primes less than or equal to $x$. When $x \to \infty$ we have that

$$\pi(x) \sim \ln(x).$$  

(19)
Proof. This is already stated by the original prime number theorem, but here—following the proof of Theorem 3.4—we suggest that (19) should be interpreted as the continuous approximation of \( \sum k \pi_k \), where \( \pi_k \) is defined as in (18).

\[ \square \]

4. Possible directions

Initial investigations—theoretical and numerical—suggest that the subdivision of the natural numbers in terms of \( s_k \) provides a rich a structure from which to probe deeper in our understanding of the prime numbers; or even, reinterpreting old results in terms of the intervals \( s_k \). In the following, we briefly sketch out a few directions where we believe the insights of this paper could have impact, keeping in mind that there are also several additional routes to explore than those mentioned here. The format is mixed, in the sense that what we present range across rigorous results, heuristics and experimental insights. As such, this section is best read as a communication of ideas possibly worthy of exploration.

Accompanying this section is also a set of figures included as an appendix. The Mathematica software were used for generating and analyzing data, and creating the final figures.

4.1. Notation. Here we introduce notation that will be used in later sections. First, we construct an arithmetic function that describes which integers \( n \) are divisible by a given prime \( p_k \):

\[
\rho_k(n) := \begin{cases} 
    p_k & \text{if } p_k \mid n, \\
    1 & \text{otherwise}.
\end{cases}
\]

Now, let us apply this definition to construct a second arithmetic function that locates all \( n \) coprime to \( p_k# := \prod_{p \in P} p \):

\[
R_k(n) := \prod_{1 \leq i \leq k} \rho_i(n).
\]

We see that \( (n, p_k#) = 1 \) whenever \( R_k(n) = 1 \). Also, both \( \rho_k(n) \) and \( R_k(n) \) are periodic, satisfying for any integer \( m \) the equalities

\[
\rho_k(n + mp_k) = \rho_k(n) \quad \text{and} \quad R_k(n + mp_k#) = R_k(n).
\]

This notation is also useful for visualizing the positions of primes in \( s_k \), as exemplified here by \( s_3 \):

| \( s_3 \) | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
|------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| \( \rho_1 \) | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| \( \rho_2 \) | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| \( \rho_3 \) | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| \( R_3 \) | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |

Consider next an arbitrary interval \( A \). The number of coprimes to \( p_k# \) in \( A \) is equivalent to the count of 1s in \( R_k(n) \) across \( A \), in sieve notation denoted by

\[
S(A, p_k#) := |\{ n : n \in A, R_k(n) = 1 \}|.
\]

It follows from the periodicity of \( R_k(n) \) that also \( S(A, p_k#) \) is periodic; if \( A^n \) denotes \( A \) left-shifted \( n \) times, the equality

\[
S(A^mp_k#, p_k#) = S(A, p_k#)
\]
holds for any integer m. We can express $S(A, p_k \#)$ in terms of the Legendre identity as

$$S(A, p_k \#) = |A| - \sum_{q \in \mathcal{P}_k} \left\lfloor \frac{|A|}{q} \right\rfloor + \sum_{q_1, q_2 \in \mathcal{P}_k, q_1 < q_2} \left\lfloor \frac{|A|}{q_1 q_2} \right\rfloor - \cdots \pm \sum_{q_1, \ldots, q_k \in \mathcal{P}_k, q_1 < \cdots < q_k} \left\lfloor \frac{|A|}{q_1 \cdots q_k} \right\rfloor,$$

where $\lfloor x \rfloor$ is defined to mean that $x$ take either values $\lfloor x \rfloor$ or $\lceil x \rceil$. In the average case, the Legendre identity has the simpler form

$$E[S(A, p_k \#)] = |A| \cdot \prod_{p \in \mathcal{P}_k} \left( 1 - \frac{1}{p} \right).$$

Assume now the interval $s_k$—the focal object throughout this paper. By the sieve of Eratosthenes we know that in $s_k$, the primes are the only numbers coprime to $p_k \#$, so necessarily,

$$\pi_k = S(s_k, p_k \#).$$

In later sections, however, we approach the distribution of primes in $s_k$ probabilistically, in the sense that we consider all possible arrangements of coprimes to $p_k \#$ within an arbitrary interval of length $l_k$. Because of the periodicity of $R_k(n)$, there are only $p_k \#$ different arrangements, which we derive in terms of a shifted version of $s_k$,

$$s'_k := \{p_k^2 + j, \ldots, p_k^2 + 1 + j\}, \quad 0 \leq j < p_k \#.$$

In this notation, the probabilistic prime counting function $\tilde{\pi}_k$ takes the form

$$\tilde{\pi}_k = \frac{1}{p_k \#} \sum_{j=0}^{p_k \#-1} S(s'_k, p_k \#),$$

while $\pi_k$ coincides with the case $j = 0$,

$$\pi_k = S(s_k^0, p_k \#).$$

4.2. Primes in short intervals. There is a rich literature on primes in short intervals—see e.g. [17, 18, 8]—related to the question for which functions $\Phi(x)$, as $x \to \infty$,

$$\pi(x + \Phi(x)) - \pi(x) \sim \frac{\Phi(x)}{\log x}. \tag{20}$$

With respect to this matter, Selberg proved in 1943 [14]—assuming the Riemann hypothesis—that (20) holds for almost all $x$, required $\Phi(x)/(\log x)^2 \to \infty$ as $x \to \infty$. It was long thought that this would be true for all $x$ [8], but in 1985, Maier [12] published a proof demonstrating that there are infinitely many exceptions to Selberg’s result, even for functions $\Phi(x)$ growing much faster than $(\log x)^2$.

What Maier specifically proved—here in the formulation of Granville [8]—was that for any constant $\lambda > 2$, there exists a constant $\delta_\lambda > 0$ such that we find arbitrarily large integers $x$ and $X$ for which

$$\pi \left( x + (\log x)^\lambda \right) - \pi(x) \geq (1 + \delta_\lambda)(\log x)^{\lambda - 1}, \quad \text{and} \quad \pi \left( X + (\log X)^\lambda \right) - \pi(X) \leq (1 - \delta_\lambda)(\log X)^{\lambda - 1}. \tag{21}$$

How do our findings connect to these results? First of all, we recognize that the estimate in [13] of the number of primes in the interval $s_k$ is literally of the exact
Choosing \( x = p_k^2 \) and \( \Phi(x) = l_k \), and assuming \( x \to \infty \), we have that
\[
\pi(p_k^2 + l_k) - \pi(p_k^2) = \pi_k \sim \frac{l_k}{\log p_k^2},
\]
and it immediately follows that, as \( k \to \infty \), all intervals \( s_k \) satisfy \( (20) \).

Now, if we substitute \( x \) for \( p_k^2 + 1 \) and \( g \) for \( g_k \) in
\[
l_k = 2p_k^2 + 1 - g_k - g_k^2,
\]
we find that each \( l_k \) lies on one of the curves given by the functions
\[
l_g(x) = 2\sqrt{xy} - g^2,
\]
where \( g = 2, 4, \ldots \). Obviously, all functions \( l_g(x) \) eventually outgrow any function \( (\log x)^\lambda \)—no matter what value of \( \lambda \) we start with—which demonstrate that the intervals \( s_k \) fulfill the requirement in Selberg’s theorem that
\[
\frac{\Phi(x)}{(\log x)^2} \to \infty \quad \text{as} \quad x \to \infty.
\]

Another consequence is that what we could call Maier’s interval, \([x, x + (\log x)^\lambda]\), will always be infinitesimal compared to \( s_k \) as \( k \) moves towards infinity. We can therefore place Maier’s interval within parts of the intervals \( s_k \) where the fluctuations in prime density deviates from the density predicted by the prime number by a multiplicative constant \( \delta_\lambda \). The example below should give a clear demonstration of this interpretation.

A final note is that in Maier’s theorem, \( \Phi(x) \) is restricted to grow logarithmically, while that is not the case in Selberg’s theorem. In some sense this suggests we can view the two theorems as describing the behavior of the distribution of the primes on different length scales. Of course—strictly—Selberg’s theorem also applies on the scale of Maier’s interval. This can even be understood from the fact that Mayer’s theorem guarantees that the density of primes in the interval \([x, x + \Phi(x)]\) crosses the mean value predicted by the prime number theorem infinitely many times as \( x \to \infty \). An alternative view is therefore that Selberg’s theorem is picking out intervals near these crossing points.

**Example.** Let us choose \( \lambda = 3 \). Then, according to \( (21) \), we should be able to find a \( \delta_3 \) such that the inequalities
\[
(22) \quad \pi(x + (\log x)^3) - \pi(x) \geq (1 + \delta_3)(\log x)^2 \quad \text{and}\n\pi(X + (\log X)^3) - \pi(X) \leq (1 - \delta_3)(\log X)^2
\]
hold for arbitrarily large \( x \) and \( X \). While arbitrarily large is beyond our reach, we will examine a few selected intervals—\( s_{500}, s_{750}, \) and \( s_{1000} \)—such that \( \Phi(x) \) is significantly smaller than \( l_k \) and fluctuations in prime density within the intervals \( s_k \) become apparent.

The following table provides an overview of how \( \Phi(x) \) compares to \( l_k \) for the chosen intervals. Note that as \( \Phi(x) \) is almost constant across an interval \( s_k \) for large \( k \), so is the ratio \( \Phi(x)/l_k \). The values of \( \Phi(x) \) are rounded to nearest integer and those of \( \Phi(x)/l_k \) are presented with two significant digits.

| \( k \) | \( g_k \) | \( l_k \) | \( \Phi(p_k^2) \) | \( \Phi(p_k^2 + 1) \) | \( \Phi(p_k^2)/l_k \) |
|---|---|---|---|---|---|
| 500 | 10 | 71250 | 4380 | 4384 | 0.061 |
| 750 | 8 | 91152 | 5172 | 5175 | 0.057 |
| 1000 | 8 | 126768 | 5787 | 5789 | 0.046 |

For a broader comparison, we can observe how \( \Phi(x) \) compares to \( l_g(x) \) across a range of values for \( x \). This is done in Figure \( 1 \) where we have plotted the values
of \( l_k \) together with \( l_q(x) = 2\sqrt{x} - q^2 \) and \( \Phi(x) = (\log x)^3 \). When \( x \) is within the shaded region, \( \Phi(x) \) is larger or equal to some \( l_k \), in particular all those \( l_k \) that lie on the curve \( l_2(x) \). But put \( x \) beyond shaded the region, and \( \Phi(x) \) is smaller than \( l_q(x) \) for all values of \( q \), and consequently also smaller than any interval length \( l_k \).

The behavior seen here is general; a different choice of lambda would stretch out the shaded region further, but no matter the size of \( \lambda \), when \( x \) crosses the threshold into the non-shaded region, \( \Phi(x) \) is left behind by every \( l_k \).

To illustrate how the prime counting function \( \pi(x) \) relates to \( \text{li}(x) \) when \( x \) is contained within an interval \( s_k \), we define

\[
\text{li}_k(x) := \text{li}(x) - \text{li}(p_k^2) \quad \text{and} \quad \pi_k(x) := \pi(x) - \pi(p_k^2),
\]

so that both functions are 0 at the start of the interval and equal to \( \text{li}_k \) and \( \pi_k \), respectively, at the end of the interval. We contrast the functions in (23) by plotting the error term

\[
\pi_k(x) - \text{li}_k(x)
\]

across the intervals \( s_{500}, s_{750}, \) and \( s_{1000} \), as shown in Figure 2. Visually, the wandering of the error term gives the impression of a random walk; it strays away from the theoretical estimate, sometimes smaller, sometimes larger, revealing regions with lower or higher density of primes than suggested by the prime number theorem.

Note that in the same plots we have also included the standard deviation \( \sigma_k(n) \) of a binomial distribution with probability of success \( 1 / \log p_{k+1}^2 \), and sequence length \( n \), where \( 1 \leq n \leq l_k \), such that

\[
\sigma_k(n) = \sqrt{\frac{n}{\log p_{k+1}^2} \left( 1 - \frac{1}{\log p_{k+1}^2} \right)}.
\]

The reason for doing this is discussed in Section 4.5.

Complementary to plotting \( \pi_k(x) - \text{li}_k(x) \) we also examine the distribution of gap sizes within an interval. Since the gap distribution is inversely related to the density of primes it helps give an impression of fluctuations in prime density. In Figure 3 we have plotted the prime gaps in each interval \( s_{500}, s_{750}, \) and \( s_{1000} \), together with measured average gap size and expected theoretical gap size \( g = \log p_{k+1}^2 \) (both estimates coincide on the scale of the figure). We also show the moving average of gap sizes with runs taken over 25 elements, demonstrating how the average gap size shifts across the intervals. Note that both Figure 2 and Figure 3 make obvious the presence of significant fluctuations in prime density within the chosen intervals \( s_k \).

Now, we are interested in the specific subintervals within \( s_k \) defined by Maier’s interval \([x, x + \Phi(x)]\), where \( p_k^2 \leq x < p_{k+1}^2 - \Phi(x) \). We can measure exactly how large the fluctuations in prime density are in these subintervals by calculating the ratio between the number of primes in \([x, x + \Phi(x)]\) and the estimate suggested by the prime number theorem, \( \Phi(x)/\log x \). More specifically, we calculate the ratios

\[
\frac{\pi \left( x + (\log x)^3 \right) - \pi(x)}{(\log x)^2},
\]

which we then use to locate an appropriate value of \( \delta_3 \). The result is displayed in Figure 4 where we have plotted the calculated ratios as \( x \) moves across each of the intervals \( s_{500}, s_{750}, \) and \( s_{1000} \). The largest value we can choose for \( \delta_3 \) in order to find examples of \( x \) and \( X \) in all three intervals that confirm to (22) is
\(\delta_3 \approx 0.064\) (since the minimal largest deviation from 1 across the intervals—found in the interval \(s_{500}\)—has this value). In the figure, we have chosen \(\delta_3 = 0.03\) as an illustration. In each plot, the shaded region contains the ratio values of Maier’s intervals where the density of primes is within a factor \(1 \pm \delta_3\) from that predicted by the prime number theorem, while all ratio values falling outside the shaded area correspond to Maier’s intervals where the prime density is \(1 \pm \delta_3\) times higher or lower than the theoretical density.

Also shown in Figure 4 is the ratio \(\pi_k/(l_k/\log p_{k+1}^2)\), demonstrating how much the prime density of the whole interval \(s_k\) varies from the prime number theorem estimate. Note that there is no connection between the increasing lengths of the intervals and the observed increasing ratio \(\pi_k/(l_k/\log p_{k+1}^2)\). In fact, if we plot \(\pi_k/(l_k/\log p_{k+1}^2)\) across a large range of \(k\), as done in Figure 5, we observe that the convergence of the ratios towards 1 is creepingly slow, with significant fluctuations which magnitude depends on the corresponding gap size \(y_k\); the smaller the gap size, the larger the fluctuations, and vice versa.

An important final note is that we do not show that this \(\delta_3\) is the proper choice in Maier’s theorem. However, Maier’s theorem guarantees that for any \(\lambda\), there must be a \(\delta_\lambda\) so that (22) holds for arbitrarily large \(x\) and \(X\). In theory therefore, we could for any given \(\lambda\) construct an infinite series of ratio plots similar to those in Figure 4, each corresponding to a different interval \(s_k\). By choosing the correct \(\delta_\lambda\) all plots would show the same shaded region, and importantly, all plots would have curves traveling outside this space.

4.3. Shrinking the divide between \(\tilde{\pi}_k\) and \(\pi_k\). Previously, we have encountered two similar estimates for the number of primes within an interval \(s_k\). One was the probabilistic estimate derived from Merten’s product theorem (6),

\[
\tilde{\pi}_k \sim 2e^{-\gamma} \frac{l_k}{\log p_{k+1}^2},
\]

the other was the estimate given by the prime number theorem (7),

\[
\pi_k = \frac{l_k}{\log p_{k+1}^2}.
\]

The question we ask here is whether it is possible to work from a probabilistic perspective, and then exploit properties of \(s_k\) to push \(\tilde{\pi}_k\) closer to \(\pi_k\). While analyzing the sequences \(s_k\) in terms of modern sieve theory would be a natural approach here, that is outside the scope of this paper. Rather, we go for a brief stab at the problem—in particular useful for building intuition—to arrive at a few interesting observations.

Let us first of all explain what we mean by a probabilistic perspective. In Section 4.1 we learned that the probabilistic prime counting function \(\tilde{\pi}_k\) can be stated as

\[
\tilde{\pi}_k = \frac{1}{p_k^\#} \sum_{j=0}^{p_k^\#-1} S(s_j^k, p_k^\#),
\]

while \(\pi_k\) is given by

\[
\pi_k = S(s_0^k, p_k^\#).
\]
As our starting point for counting the number of primes in the interval $s_k$, assume the only information available is the length $l_k$ of the interval. Then the only certainty we have is that $\pi_k$ takes one of the values $S(s_j^k, p_k\#)$, $0 \leq j < p_k\#$. This is the largest sample space possible and the expected number of primes in the interval is given by $\tilde{\pi}_k$. Now, clearly, to shrink the gap between $\tilde{\pi}_k$ and $\pi_k$, we need to reduce this sample space in such a way that it still contains $S(s_j^k, p_k\#)$.

Of course, $S(s_j^k, p_k\#)$ is completely determined by where in the natural numbers the interval $s_k$ is situated (allowing for shifts divisible by $p_k\#$), and in the end, the constraints that will push $\tilde{\pi}_k$ towards $\pi_k$ must reflect that. A very natural constraint is for example that when $n \in s_k$, all elements in $R_k(n)$ must be strictly smaller than $p_{k+1}^2$. This constraint is easily implemented into the Legendre identity by allowing only denominators smaller than $p_{k+1}^2$:

$$S(s_j^k, p_k\#) = \left\lfloor \frac{l_k}{q} \right\rfloor + \sum_{q_1, q_2 \in P_k} \left\lfloor \frac{l_k}{q_1 q_2} \right\rfloor + \cdots + \sum_{q_1 \cdots q_k \in P_k} \left( \prod_{q \in P_k \ q < p_{k+1}^2} \right) \left\lfloor \frac{l_k}{q_1 \cdots q_k} \right\rfloor.
$$

Numerically, we demonstrate in Figure 6 that this alone reduces the ratio $\frac{\tilde{\pi}_k}{l_k/\log p_{k+1}^2}$ from approximately $2e^{-\gamma} \approx 1.12$ to approximately 1.03.

Now, the Legendre identity is notoriously known for its untamable error term; the unconstrained Legendre identity contains $2^k$ terms, and since each can be rounded either up or down (neglecting the cases of exact divisibility), we arrive at an upper error bound of $2^k$, which rapidly grows out of control as we increase $k$. The result is that it is problematic to make straightforward use of the Legendre identity to count primes, and a look to modern sieve theory is necessary to find methods that work around this complication [15]. Nonetheless, the constraint we imposed here also influences the error bound in the Legendre identity; while still overwhelming, it shifts from exponential to polynomial growth. We show numerical evidence of this in Figure 7.

Another example of constraint is this: In the sequence $s_k$, the first appearance of a composite divisible by $p_i$, for $i < k$, is in one of the positions $p_i - m$, where $m$ is an element in the residue class of $p_j^2 \pmod{p_i}$, $j \neq i$. (The first appearance of a composite of $p_k$ is of course in position 1, since the first element in $s_k$ is $p_k^2$.) For the primes $p_1, \ldots, p_6$, these are the possible positions of first appearance:

| Prime | Position of first appearance |
|-------|-----------------------------|
| $p_1$ | 2                           |
| $p_2$ | 3                           |
| $p_3$ | 2, 5                        |
| $p_4$ | 4, 6, 7                     |
| $p_5$ | 3, 7, 8, 9, 11              |
| $p_6$ | 2, 4, 5, 10, 11, 13         |

Note that this restrains the sequences $p_i$, $1 \leq i \leq k$, effectively halving the sample space of values $S(s_j^k, p_k\#)$. Nonetheless, verified by numerical checks, the statistical properties of the reduced sample space do not change significantly from the original
one, which we can understand from the fact that the constraint does not incorporate positional information of the sequence $s_k$.

4.4. The bias of the error term in the prime number theorem. A well known observation is that the error term in the prime number theorem starts out negative:

$$\pi(x) - \text{li}(x) < 0.$$ 

As we demonstrate in Figure 8 this bias continues for the full stretch of our data set, which stops at $x \approx 7.3 \times 10^{13}$. Nonetheless, as Littlewood proved in 1914 [11], if we persist in moving $x$ towards infinity, the error term will eventually shift sign, and it will continue to do so infinitely many times over. When that happens for the first time, however, no one knows, though it has been proved that 1.39822 \times 10^{316} is an upper bound for the event [1]. A probabilistic measure of the bias was provided in 1994, when Rubinstein and Sarnak [13]—assuming the Riemann hypothesis—calculated the logarithmic density of those $x$ such that $\pi(x) - \text{li}(x) > 0$ to be around $0.00000026$.

So, why this bias? While the phenomenon has been widely studied, the literature provides no clear answer as to why there is a bias in the first place. We will not give a definitive answer here, but what we have seen so far strongly suggests that the bias derives from the underlying sieving process. Remember, in the proof of Theorem 3.3, the best approximation of the probabilistic prime counting function $\tilde{\pi}_k$—when applying Merten’s product theorem—is obtained by

$$\tilde{\pi}_k \sim 2e^{-\gamma} \frac{l_k}{\log p_k^2}.$$ 

This choice, rather than for example

$$\tilde{\pi}_k \sim 2e^{-\gamma} \frac{l_k}{\log p_k^2} \text{ or } \pi_k \sim 2e^{-\gamma} \frac{l_k}{\log \frac{1}{2}(p_k^2 + p_{k+1}^2)}$$

also reflects the fact that sieving by the $k$ first primes completes sieving up to $p_{k+1}^2$. Similarly, while the prime number theorem guarantees

$$\pi_k \sim \frac{l_k}{\log p_k^2} \sim \frac{l_k}{\log \frac{1}{2}(p_k^2 + p_{k+1}^2)} \sim \frac{l_k}{\log p_{k+1}^2},$$

the latter statement—that each sieve step $k$ sieves up to $p_{k+1}^2$—suggests that of these three estimates for the prime counting function $\pi_k$, the most accurate is

$$\pi_k \sim \frac{l_k}{\log p_{k+1}^2}.$$ 

Now, since $\log p_k^2 < \log t < \log p_{k+1}^2$ whenever $p_k^2 < t < p_{k+1}^2$, we have that

$$\frac{l_k}{\log p_{k+1}^2} < \int_{p_k^2}^{p_{k+1}^2} \frac{dt}{\log t} < \frac{l_k}{\log p_k^2}.$$ 

And then, summing over all intervals $s_j$, $j \geq 1$, we get

$$\sum_{j=1}^{k} \frac{l_j}{\log p_{j+1}^2} < \text{li}(p_{k+1}^2) < \sum_{j=1}^{k} \frac{l_j}{\log p_j^2}.$$
Therefore, under the assumption that

\[ \pi(p_{k+1}^2) \sim \sum_{j=1}^{k} \frac{l_j}{\log p_{j+1}^2} \]

is our most accurate estimate of the prime counting function, the reason why \( \pi(x) - \text{li}(x) \) is leaning towards negative becomes apparent: using \( \text{li}(p_{k+1}^2) \) to estimate \( \pi(p_{k+1}^2) \), rather than \( \sum_{j=1}^{k} \frac{l_j}{\log p_j^2} \), gives a negative expectation value for the error term:

\[ E[\pi(x) - \text{li}(x)] < 0. \]

Following this logic, it would be natural to suggest that

\[ E[\pi(p_{k+1}^2) - \sum_{j=1}^{k} \frac{l_j}{\log p_j^2}] = 0. \]

However, as we see in Figure 8, the statistical evidence seems to indicate that the truth lies somewhere in between,

\[ \sum_{j=1}^{k} \frac{l_j}{\log p_j^2} < E[\pi(p_{k+1}^2)] < \text{li}(p_{k+1}^2). \]

To gain a more accurate impression of this observation, we normalize the curves in Figure 8 by

\[ \Delta_k := \frac{1}{2} \sum_{j=1}^{k} \left( \frac{l_j}{\log p_j^2} - \frac{l_j}{\log p_j^2} \right), \]

which places the differences

\[ \sum_{j=1}^{k} \left( \frac{l_j}{\log p_j^2} - \text{li}_j \right) \quad \text{and} \quad \sum_{j=1}^{k} \left( \frac{l_j}{\log p_j^2} - \text{li}_j \right) \]

approximately at the constant lines -1 and 1 respectively. The resulting plot is shown in Figure 9. There we observe that on the scale of the data available, \( [\pi(p_{k+1}^2) - \text{li}(p_{k+1}^2)]/\Delta_k \) apparently fluctuates fairly stable around a mean value of \(-0.60\). Constructing an empirical probability density function, we see in Figure 10 that across the full interval \( x \), the fluctuations closely resembles a Gaussian distribution with mean and standard deviation given by \(-0.60\) and 0.12.

We will not attempt an explanation to why the bias of the error term lies where it does in these observations, or whether it will continue to do so. However, these directions of thoughts seems to add valuable clues towards a rigorous understanding of the error term.

4.5. **An upper bound for the error term in the prime number theorem.**

Despite the fact that the number of primes up to \( x \) is deterministic, let us for a moment assume that \( \pi(x) \) can be represented by a random variable \( \Pi(x) \) with mean value \( \text{li}(x) \) and variance \( \sigma^2(x) \). Then we could interpret the error term \( |\pi(x) - \text{li}(x)| \) in terms of the standard deviation

\[ \sigma(x) = \sqrt{E[(\Pi(x) - \text{li}(x))^2]}. \]

In this section we will walk through a strategy for obtaining an upper bound on the error term based on this view. We start with formulating a random model for the number of primes within \( s_k \).
From Section 4.1, we know that the number of coprimes to $p_k#$ within any interval of length $l_k$ takes one of the values $S(s^j_k, p_k#), 0 \leq j < p_k#$, and specifically, we have for $s_k$ that $\pi_k = S(s^0_k, p_k#)$. This naturally suggests a random model for $\tilde{\pi}_k$: Let the number of coprimes to $p_k#$ within any interval of length $l_k$ be described by the random variable $\tilde{\Pi}_k$, with sample space $S(s^j_k, p_k#), 0 \leq j < p_k#$. Using the approximation

$$l_k \cdot \prod_{p \in P_k} \left(1 - \frac{1}{p}\right) \sim 2e^{-\gamma} \frac{l_k}{\log p_k+1}$$

(and neglecting the error term) we can write the mean value of $\tilde{\Pi}_k$ as

$$\tilde{\mu}_k := E[\tilde{\Pi}_k] = \frac{1}{p_k#} \sum_{j=0}^{p_k#-1} S^j(s_k, p_k#) = 2e^{-\gamma} \frac{l_k}{\log p_k+1},$$

and likewise the variance

$$\tilde{\sigma}_k^2 := E[(\tilde{\Pi}_k - \tilde{\mu}_k)^2] = \frac{1}{p_k#} \sum_{j=0}^{p_k#-1} \left(S^j(s_k, p_k#) - 2e^{-\gamma} \frac{l_k}{\log p_k+1}\right)^2.$$

With a slight adjustment to this model—multiplying all elements in the sample space by $e^{\gamma}/2$—the mean value changes to that of the prime number theorem and we arrive at a random model for $\pi_k$. Therefore, let the number of primes in $s_k$ be represented by the random variable $\Pi_k$, with mean value

$$\mu_k = e^{\gamma} \tilde{\mu}_k = \frac{l_k}{\log p_k^2+1},$$

and variance

$$\sigma_k^2 = e^{2\gamma} \tilde{\sigma}_k^2 = \frac{1}{p_k#} \sum_{j=0}^{p_k#-1} \left(e^{\gamma} S^j(s_k, p_k#) - \frac{l_k}{\log p_k^2+1}\right)^2.$$

One thing to note is that the mean value is an explicit function of $l_k$ and $p_{k+1}$, while this is not the case with the variance. Rather than aiming for deriving an exact expression, we go about this by finding another distribution where both mean and variance are known functions, and which variance bounds $\sigma_k^2$. An obvious candidate is the binomial distribution with number of trials given by $l_k$ and probability of success $1/\log p_k^2+1$. Therefore, let $\Pi_{k,B}$ be the binomial random variable,

$$\Pi_{k,B} \sim B\left(l_k, \frac{1}{\log p_k^2+1}\right).$$

Intuitively, there are two reasons why $\Pi_{k,B}$ should have larger variance than $\Pi_k$; one is that the sample space of our $\pi_k$ random model is only a very small fraction of the sample space of the binomial model—$p_k#$ elements versus $2^{l_k}$; a second is that the $\pi_k$ random model has strict upper and lower bounds, while the binomial model allows $\Pi_{k,B}$ to take any values from 0 to $l_k$. These reasons are not proof of course, but for now we assume it is true that the variance of $\Pi_{k,B}$ bounds $\sigma_k^2$. (This seems a fairly simple problem, and we expect theory to exist that applies. It is possible to construct a version of Polya’s urn model that works heuristically, but in the end the approach mirrors an inclusion-exclusion process and does not provide proof).
Since we are mostly interested in the limit of $k \to \infty$, where the binomial distribution approaches a Poisson distribution, we can replace the binomial variable $\Pi_{k,B}$ by the Poisson random variable $\Pi_{k,Pois}$, with parameter $l_k / \log p_{2k+1}^2$:

$$\Pi_{k,Pois} \sim \text{Pois} \left( \frac{l_k}{\log p_{2k+1}^2} \right).$$

The mean and variance of $\Pi_{k,Pois}$ are identical and given by

$$\mu_{k,Pois} = \sigma^2_{k,Pois} = \frac{l_k}{\log p_{2k+1}^2},$$

and it follows by assumption that the variance of the $\pi_k$ random model satisfies

$$\sigma^2_k < \sigma^2_{k,Pois} = \frac{l_k}{\log p_{2k+1}^2}.$$

Now, let us turn to $\pi(x)$. Since $\pi(x)$ can be written as the sum

$$\pi(x) = \sum_{j=1}^{k-1} \pi_j + \frac{x - p_{2k}^2}{l_k} \pi_k,$$

where $k$ is the integer such that $p_{2k}^2 \leq x < p_{2k+1}^2$, it makes sense to define a random model for $\pi(x)$ in the same way. Hence, let $\Pi(x)$ be the random variable given by

$$\Pi(x) := \sum_{j=1}^{k-1} \Pi_j + \frac{x - p_{2k}^2}{l_k} \Pi_k.$$

Since $\Pi(x)$ is defined as a sum over random independent variables, its mean and variance are obtained simply by summing over the individual means and variances:

$$\mu(x) := E[\Pi(x)] = \sum_{j=1}^{k-1} E[\Pi_j] + \frac{x - p_{2k}^2}{l_k} E[\Pi_k]$$

$$= \sum_{j=1}^{k-1} \frac{l_j}{\log p_{2j+1}^2} + \frac{x - p_{2k}^2}{l_k} \frac{l_k}{\log p_{2k+1}^2}$$

$$\approx \text{li}(x),$$

and

$$\sigma^2(x) := E[(\Pi(x) - \mu(x))^2] = \sum_{j=1}^{k-1} E[(\Pi_j - \mu_j)^2] + \frac{x - p_{2k}^2}{l_k} E[(\Pi_k - \mu_k)^2]$$

$$< \sum_{j=1}^{k-1} \sigma^2_{j,Pois} + \frac{x - p_{2k}^2}{l_k} \sigma^2_{k,Pois}$$

$$= \sum_{j=1}^{k-1} \frac{l_j}{\log p_{2j+1}^2} + \frac{x - p_{2k}^2}{l_k} \frac{l_k}{\log p_{2k+1}^2}$$

$$\approx \text{li}(x).$$

Consequently, the standard deviation in the $\pi(x)$ random model satisfies

$$\sigma(x) = \sqrt{E[(\Pi(x) - \mu(x))^2]} < \text{li}(x).$$
Of course, the most important question remains: Does the properties of the $\pi(x)$ random model translate to $\pi(x)$ proper? For the answer to be positive, there are essentially two requirements that must be fulfilled. Firstly, we must be able to interpret the number of primes $\pi_k$ in each interval $s_k$ as a random variable with well defined mean and variance. Secondly, we must be able to view $\pi(x)$ as a random variable, with mean and variance derived by summing over the means and variances of the individual intervals $s_k$.

We have already seen that the first requirement is satisfied; $\pi_k$ is always in the sample space of $\tilde{\Pi}_k$ and can be viewed as drawn from this set (or alternatively using the $\pi_k$ model as we do above; the models differ only by a multiplicative constant). Regarding the second requirement; the distribution of prime numbers is deterministic, and therefore the assumption of independent random variables $\pi_k$ does not hold. While this does not affect constructing the mean of $\pi(x)$ as a sum of the individual means, we need to be more careful about the variance. Fortunately, it is enough to show that if the $\pi_k$s can be considered to be uncorrelated, i.e.,

$$\rho_{ij} := \frac{E[(\pi_i - l_i/\log p_{i+1})(\pi_j - l_j/\log p_{j+1})]}{E[(\pi_i - l_i/\log p_{i+1})^2]} = 0, \quad i \neq j,$$

that—according to the Bienaymé formula—will be enough for us to sum variances linearly. While we do not rigorously prove uncorrelation, we proceed with a few heuristic arguments for why it very likely holds.

It will be helpful to start by considering how the values $\pi_k$ and $\pi_{k+1}$ of two neighboring intervals $s_k$ and $s_{k+1}$ relate when $k$ is large. The first thing to note is that we can safely assume that the lengths of the intervals, $l_k$ and $l_{k+1}$, are much smaller than the period $p_k \#$ of $R_k(n)$, since the latter grows much faster with $k$. Now, suppose $l_k = l_{k+1}$ and that both $s_k$ and $s_{k+1}$ are sieved by the $k$ first primes (of course, none of which are true). Then we would have

$$\pi_k = S(s_k^0, p_k \#) \quad \text{and} \quad \pi_{k+1} = S(s_{k+1}^1, p_{k+1} \#).$$

Under these assumptions we see that $\pi_k$ and $\pi_{k+1}$ each correspond to different elements of the full sample space of $\tilde{\Pi}_k$. What’s more, the two elements are derived by counting the coprimes to $p_k \#$ within two non-overlapping subintervals of the much longer periodic sequence $R_k(n)$. We would therefore expect the correlation between $\pi_k$ and $\pi_{k+1}$ to be weak already under these considerations; possibly could this be proved in terms of the autocorrelation of $R_k(n)$, which we would expect to drop off to zero at a distance of $l_k$. Naturally, $l_k$ and $l_{k+1}$ are never equal, and often differ significantly because of the distribution of gaps $g_k$. In addition, $s_{k+1}$ is sieved by the additional prime $p_{k+1}$. Both of these factors serve to decrease correlation, and in the case when we look at intervals further apart, the effects are magnified. Heuristically therefore, we expect the $\pi_k$s to behave as uncorrelated variables and we can apply the Bienaymé formula.

Taking what we have seen to its conclusion, we conjecture an upper bound for the error term in the prime number theorem:

**Conjecture 1.** As $x \to \infty$, the error term in the prime number theorem satisfies

$$|\pi(x) - \text{li}(x)| < \sqrt{\text{li}(x)}.$$

From the above—and from what we understand—it is also clear that the two remaining obstacles for a rigorous proof of the conjecture is 1) to prove that the
\( \pi_k \) can be considered uncorrelated; and 2) to prove that \( \sigma_k^2 < \sigma_{k, \text{Pois}}^2 \). We expect neither of these proofs to have strong barriers, nor that any surprises should arise.

4.5.1. **Empirical evidence.** As we have just seen, underlying the suggested conjecture is the interpretation that the \( \pi_k \)s behave as random, uncorrelated variables, with well defined means and variances. And additionally, that we can employ the variance of a binomial distribution to bound the variance of \( \pi_k \). To round off the paper, we here present empirical results supporting this picture, starting plainly with the set of \( \pi_k \)s.

From the prime number theorem we know that
\[
\pi_k \sim \frac{l_k}{\log p_{k+1}^2} = \frac{2p_{k+1}g_k - g_k^2}{\log p_{k+1}^2}.
\]
Therefore, by substituting \( x = p_{k+1}^2 \), we would expect all values of \( \pi_k \) to lie close to one of the curves
\[
\pi(x, g) := \frac{2\sqrt{x}g - g^2}{\log x}, \quad g = 2, 4, \ldots,
\]
which is demonstrated to hold in Figure 11. Now, to gain an impression of how the \( \pi_k \)s fluctuate around each curve, we plot all values \( \pi_k - \text{li}_k \). At first eyesight then, Figure 12 reveals that these values are seemingly evenly distributed around 0, confirming to the behavior of a random variable. By sampling across values of \( \pi_k - \text{li}_k \) we arrive at a set of empirical probability density functions and find that the fluctuations are well described by Gaussian distributions, both for samples taken across \( \pi_k \)s corresponding to different curves \( \pi(x, g) \) and for samples taken along the curve \( \pi(x, 6) \) across increasing \( k \); all shown in Figure 13.

We continue with the correlation between the number of primes \( \pi_k \) and \( \pi_{k+j} \) in intervals that differ by \( j \) sieve steps. In Figure 14A we plot the empirical correlation as calculated from
\[
\rho_{ij} := \frac{E[(\pi_i - \text{li}_i/\log p_{i+1})(\pi_j - \text{li}_j/\log p_{j+1})]}{E[(\pi_i - \text{li}_i/\log p_{i+1})^2]} = 0, \quad i \neq j,
\]
with the average taken over all values \( 1 \leq k \leq p_{k+1}^2 \). What we observe is that for \( j > 50 \), the correlation fluctuates evenly around 0. However, for smaller values of \( j \), and in particular for \( j = 1 \)—corresponding to neighboring intervals—we note a slight anti-correlation. This clearly expresses itself in Figure 14B, where the correlation is calculated for intervals of 10000 non-overlapping samples of \( k \). Here we see that for increasing interval number and hence larger \( k \), there is a visible trend towards stronger anti-correlation of close neighbors. We do not currently have an explanation for this observation. Nonetheless, if the effect of this behavior is significant, it will in fact reduce the variance of \( \pi(x) \), since anti-correlated variables contribute negatively to the total variance, as seen from the identity
\[
\sigma^2(p_{k+1}^2) = \sum_{i=1}^{k} \sum_{j=1}^{k} \sigma_i \sigma_j = \sum_{i=1}^{k} \sigma_i^2 + 2 \sum_{1 \leq i < j \leq k} \sigma_i \sigma_j.
\]

Consider now the assumption that the variance of the binomial distribution \( \text{B}(l_k, \text{li}_k/\log p_k + 1^2) \) can bound the variance of \( \pi_k \). We describe this assumption by three different examples, on somewhat different length scales. To start with the smallest scale, we view in Figure 2 how the standard deviation of the binomial distribution compares to \( \pi_k(x) - \text{li}_k(x) \) across the three intervals \( s_{500}, s_{750}, \) and
While the fluctuations of $\pi_k(x) - \text{li}_k(x)$ are large, the steady growth of the binomial standard deviation ensures it ends up with the highest value at $x = p_{k+1}^2$.

Moving up in scale, we now take the point of view of probability density functions (PDF). For this purpose, we have simulated the $\pi_k$ random model for the interval $s_{50}$, and plotted the measured PDF of $\Pi_{s_{50}} - \text{li}_{s_{50}}$ together with the theoretical binomial PDF for that interval (Figure 15A). In parallel, we have sampled values of $\pi_k - \text{li}_k$ for intervals $s_k$ with $g_k = 6$. The resulting PDF is shown in Figure 15B together with the two binomial PDFs corresponding to the start and end of the sample interval. In both the simulated and real case, we see that the binomial distribution has the widest PDF, and hence greatest variance.

Finally, we compare standard deviations of binomial distributions with standard deviations calculated from sampled data, across all intervals $s_k$, $1 \leq k \leq 575200$. This is shown in Figure 16 for those $s_k$ such that $g_k = 6$. Not only are the standard deviations of the binomial distributions always larger than those of the $\pi_k$s, but the gap increases as $k \to \infty$.

What we find then is convincing numerical evidence of our claim that the $\pi_k$s can be interpreted as a set of random, uncorrelated variables. This implies, of course, that we should expect to observe empirically that $|\pi(x) - \text{li}(x)| < \sqrt{\text{li}(x)}$ as $x \to \infty$. This is apparent in Figure 8, where $\pi(x) - \text{li}(x)$ wiggles towards infinity, well within the range of the conjectured error bound.

References

[1] C Bays and R H Hudson. A new bound for the smallest $x$ with $\pi(x) > \text{li}(x)$. Mathematics of Computation, 69:1285–1296, 2000.
[2] C J de la Vallée-Poussin. Recherches analytiques sur la théorie des nombres premiers. Annales de la Société Scientifique de Bruxelles, 20:183–256, 1899.
[3] H G Diamond. Elementary Methods in the Study of the Distribution of Prime-Numbers. Bulletin of the American Mathematical Society, 7:553–589, 1982.
[4] B Fine and G Rosenberger. An epic drama: the development of the prime number theorem. Scientia. Series A. Mathematical Sciences. New Series, 20:1–26, 2010.
[5] D Goldfeld. The Elementary Proof of the Prime Number Theorem: an Historical Perspective. In Number Theory, pages 179–192. Springer, 2004.
[6] D A Goldston. Are There Infinitely Many Twin Primes? arXiv preprint arXiv:0710.2123, 2007.
[7] A Granville. Harald Cramér and the distribution of prime numbers. Scandinavian Actuarial Journal, 1995(1):12–28, January 1995.
[8] A Granville. Different approaches to the distribution of primes. Milan journal of mathematics, 78(1):65–84, 2010.
[9] G Greaves. Sieves in Number Theory. Springer, 2001.
[10] J Havil. Gamma: exploring Euler’s constant. Princeton University Press, 2003.
[11] J W Littlewood. Distribution des Nombres Premiers. C. R. Acad. Sci. Paris, 158:1869–1872, 1914.
[12] H Maier. Primes in short intervals. The Michigan Mathematical Journal, 32(2):221–225, 1985.
[13] M Rubinstein and P Sarnak. Chebyshev’s bias. Experimental Mathematics, 3(3):173–197, 1994.
[14] A Selberg. On the normal density of primes in small intervals and the difference between consecutive primes. Archiv for matematik og naturvidenskab, 47:87–105, 1943.
[15] T Tao. Open question: The parity problem in sieve theory. http://terrytao.wordpress.com/2007/06/05/open-question-the-parity-problem-in-sieve-theory 2007.
[16] M B Villarino. Mertens’ Proof of Mertens’ Theorem. arXiv preprint arXiv:math/0504289, 2005.
[17] C Y Yildirim. A survey of results on primes in short intervals. Lecture notes in pure and applied mathematics, pages 307–343, 1999.
[18] C Y Yildirim. The distribution of primes: conjectures vs. hitherto provables. In *Further progress in analysis*, pages 75–108. World Sci. Publ., Hackensack, NJ, 2009.
Figure 1. The lengths $l_k = 2p_k + 1 - g_k^2$ of the intervals $s_k$ plotted at the values $x = p_k^2 + 1$, $1 \leq k \leq 136$ (black circles). All points lie on the curves $l(x) = 2\sqrt{x}g - g^2$ (dashed lines), where $g = 2, 4, \ldots$. The points on the lowest line all correspond to values of $k$ such that $g_k = 2$; the points on the second lowest to those $k$ such that $g_k = 4$, and so on. In addition, $\Phi(x) = (\log x)^3$—the length of Maier’s interval with $\lambda = 3$—is plotted as a continuous curve (red line). In the shaded region $\Phi(x)$ is large enough to cover some intervals $s_k$ completely, while beyond this region, $\Phi(x)$ is strictly smaller than any $l_k$. 
Figure 2. $\pi_k(x) - \text{li}_k(x)$ (blue) plotted across the interval $s_k$ for three values of $k$: A) $k = 500$, B) $k = 750$, and C) $k = 1000$. Also shown is the standard deviation $\sigma_k(n)$ (red, dashed) of a binomial distribution with probability of success $p = 1/\log p_{k+1}^2$, and sequence length $n$, where $1 \leq n \leq l_k$. The different intervals have lengths $l_{500} = 71250$, $l_{750} = 91152$, and $l_{1000} = 126768$. The intervals correspond to the gaps $g_{500} = 10$, $g_{750} = 8$, and $g_{1000} = 8$, where $g_k := p_{k+1}^2 - p_k$. 
Figure 3. The prime gaps $g_i := p_{i+1} - p_i$ within $s_k$, plotted at $p_i$ for all $i$ such that $p_k^2 \leq p_i < p_{k+1} < p_{k+1}^2$ (gray). Also shown are the measured average gap length (red) and the expected gap length in the interval given by the prime number theorem (blue, dashed), in addition to a moving average taken over runs of 25 gap lengths (black). Each plot corresponds to a specific value value of $k$: A) $k = 500$, B) $k = 750$, and C) $k = 1000$. 
Figure 4. The ratio between the actual number of primes within a subinterval \([x, x + \Phi(x)]\) of \(s_k\) and the estimate suggested by the prime number theorem, plotted for different values of \(x\) (blue, rugged lines). The lengths of the subintervals are given by \(\Phi(x) = (\log x)^3\). The ratio values are plotted at \(x\), with \(p_k^2 \leq x < p_{k+1}^2 - \Phi(x)\), which explains the last empty stretch in each plot. All ratio values climbing above the shaded area correspond to intervals where the density is 1.003 times higher than that predicted by the prime number theorem. Likewise, the values falling below the shaded area reveal the intervals where the density is 0.97 times lower than what the prime number theorem predicts. The red line shows the ratio obtained when using the length \(l_k\) of the whole interval \(s_k\) rather than \(\Phi(x)\). The different plots are for the intervals \(s_k\), where A) \(k = 500\), B) \(k = 750\), and C) \(k = 1000\).
Figure 5. The ratio between the number of primes $\pi_k$ within the interval $s_k$, and the estimate $l_k / \log p_{k+1}^2$ suggested by the prime number theorem. The different curves show the ratio values of the intervals $s_k$ corresponding to the prime gaps $g_k := p_{k+1} - p_k = g$, where $g$ take the values 2 (blue), 6 (red), and 20 (black). $k$ runs from 1 to 575200, but only the values of $k$ corresponding to gaps 2, 6, and 20 are shown.

Figure 6. The ratios between the values of three different prime counting functions over the interval $s_k$ and the number of primes in $s_k$ estimated by the prime number theorem, $l_k / \log p_{k+1}^2$. The values are plotted at $x = p_{k+1}^2$, $1 \leq k \leq 1000$. The prime counting functions are 1) $\pi_k$, the number of primes in $s_k$ (gray); 2) $\tilde{\pi}_k$, the expected number of primes based on the Euler product (red); 3) a version of the latter where only denominators below $p_{k+1}^2$ are included in the expansion of the Euler product (blue). The horizontal lines illustrate the limits tended to by the two probabilistic prime counting functions.
Figure 7. The number of terms remaining in the Legendre identity after truncating every term equal to or larger than $p_{k+1}^2$ plotted as a function of number of sieve steps $k$ (blue). Corresponding to this curve is a best polynomial fit (black, dashed) and for comparison also the interval lengths $l_k$ are shown (dots).

Figure 8. $\pi(x) - \text{li}(x)$ plotted at the values $x = p_{k+1}^2$ for $1 \leq k \leq 575200$ (black). Likewise, the differences $\sum_l (l_k/\log p_k^2 - \text{li}_k)$ (red) and $\sum_l (l_k/\log p_{k+1}^2 - \text{li}_k)$ (blue) are plotted for the same values of $k$. The gray, dashed lines are the conjectured theoretical bounds for the error term, $\pm \sqrt{\text{li}(x)}$. 
28 THE ORIGIN OF LI(X) IN THE PNT

Figure 9. Log-linear plot of $\frac{\sum (\pi_k - li_k)}{\Delta_k}$, plotted at the values $p_{k+1}^2$ for $1 \leq k \leq 575200$ (black). Likewise, the normalized differences $\frac{\sum (l_k / \log p_{k+1}^2 - li_k)}{\Delta_k}$ (red) and $\frac{\sum (l_k / \log p_k^2 - li_k)}{\Delta_k}$ (blue) are plotted for the same values of $k$. Also shown is the mean value $\langle \frac{\sum (\pi_k - li_k)}{\Delta_k} \rangle$ taken across the whole interval (gray, dashed).

Figure 10. The distribution of the normalized error term $\frac{\sum (\pi_k - li_k)}{\Delta_k}$ (red), sampled across the values $1 \leq k \leq 575200$. The mean and the standard deviation are $\mu = -0.60$, and $\sigma = 0.12$, respectively, both presented with two significant digits. A Gaussian distribution using these parameters is included in the plot (black).
Figure 11. The number of primes $\pi_k$ in each interval $s_k$ plotted at the values $x = p_{k+1}^2$ (black dots). The emerging curves relates to the specific gap values $g_k$; the lower curve corresponds to the set of intervals $s_k$ where $g_k = 2$, the second lowest $g_k = 4$, etc. The red curves are the theoretical estimates $(2\sqrt{x} - g)/\log x$, where $g$ takes the gap values 2, 4, ….

Figure 12. $\pi_k - \text{li}_k$ plotted at the values $x = p_{k+1}^2$, $1 \leq k \leq 575200$. 
Figure 13. Empirical and fitted probability distributions for $\pi_k - li_k$. A) Distributions of $\pi_k - li_k$ for the sets of intervals $s_k$ where 1) $g_k = 2$, 2) $g_k = 6$, 3) $g_k = 10$, and 4) $g_k = 14$. All distributions are build up from sampling 4000 consecutive values of $\pi_k - li_k$ for each set of intervals, starting at the 20000th value in each. B) Distributions of $\pi_k - li_k$ for the set of intervals $s_k$ where $g_k = 6$, obtained by sampling 4000 consecutive values starting at value numbers 1) 20000, 2) 40000, 3) 60000, and 4) 80000. In both plots, the black curves are Gaussian distributions with mean and standard deviation identical to the colored curves they fit.

Figure 14. Measured average correlation between $\pi_k - li_k$ and $\pi_{k+j} - li_{k+j}$, plotted for $1 \leq j \leq 500$. A) Average is taken over the whole range of $k$, $1 \leq k \leq 575200$, and the correlation is plotted as a function of $j$. B) Average is taken over non-overlapping intervals of 10000 values of $k$. The correlation is plotted for each value of $j$ as a function of increasing interval number.
Figure 15. Empirical (red) and fitted (black) probability distributions of $\pi_k - \log_{\alpha_k}$ for A) 100000 samples of the $\pi_k$ random model for the interval $s_{50}$. The blue curve shows the binomial distribution $B(l_{50}, l_{50}/\log p_{51})$; B) 4000 consecutive samples of $\pi_k - \log_{\alpha_k}$ from the set of intervals $s_k$ with $g_k = 6$, starting at sample number 20000. The blue curve shows the binomial distribution for the interval $s_{k1}$, corresponding to the first of the 4000 samples. The dashed black curve similarly shows the binomial distribution for the interval $s_{k2}$, corresponding to the last of the 4000 values.

Figure 16. Measured standard deviations of $\pi_k - \log_{\alpha_k}$ for the set of intervals $s_k$ where $g_k = 6$ (red curve). The horizontal stretches spans 4000 values of $\pi_k - \log_{\alpha_k}$ (except the first stretch; spans between 4000 and 8000 values), and the standard deviations are calculated over these. The standard deviation of the binomial distribution for each interval $s_k$ in the same set is shown as the blue curve.