A New Sieve for the Twin Primes
and how the number of twin primes is related to the number of primes

by

H.L. Mitchell

Department of mathematics CUNY-The City College
160 Convent avenue New York, NY 10031 USA
e-mail: hlmitchell88@yahoo.com
Abstract. We introduce a sieve for the number of twin primes less than $n$ by sieving through the set \( \{ k \in \mathbb{Z}^+ \mid 6k < n \} \). We derive formula accordingly using the Euler product and the Brun Sieve.

We then use the Prime Number Theorem and Mertens’ Theorem.

The main results are:

1) A sieve for the twin primes similar to the sieve of Eratosthenes for primes involving only the values of $k$, the indices of the multiples of 6, ranging over $k = p, 5 \leq p < \sqrt{n}$. It shows the uniform distribution of the pairs $(6k-1, 6k+1)$ that are not twin primes and the decreasing frequency of multiples of $p$ as $p$ increases.

2) A formula for the approximate number of twin primes less than $N$ in terms of the number of primes less than $n$.

3) The asymptotic formula for the number of twin primes less than $n$ verifying the Hardy Littlewood Conjecture.

2010 Mathematics Subject Classification: 11A07; 11A41; 11N05; 11N25; 11N36

Keywords and phrases: Primes, Twin Primes, Congruence, Sieve methods
1. Introduction

The twin primes have been studied by a number of mathematicians over the past 3 centuries and thus far it is not known whether there exist infinitely many of them.

Hardy and Littlewood proposed their famous conjecture in 1923 giving a formula for the number of twin primes less than a given integer \( n \).

We introduce a sieve for the twin primes less than \( n \) similar to the sieve of Eratosthenes for primes. It is applied twice to the set of all natural numbers \( k \) such that \( k < 6n \) and the range for the primes is \( p = 5 \) to \( p < \sqrt{n} \).

We consider the set of all pairs \((6k - 1, 6k + 1)\) which are less than \( n \) and delete the values of \( k \) such that \( 6k-1 \) is composite. This leaves us with the pairs for which \( 6k-1 \) is prime. From these we delete the values of \( k \) such that \( 6k+1 \) is composite and that leaves us with the twin primes less than \( N \).

Using the Euler product formula, The Brun Sieve, The Prime number theorem and Mertens’ 3rd Theorem, we derive a formula for the approximation of \( \pi_2(n) \) in terms of \( \pi(n) \) (the number of primes less than \( n \)) and the asymptotic formula for \( \pi_2(n) \) to verify the Hardy Littlewood Conjecture.
2. Deriving the formula and some Set Theory

All the twin primes except \{3, 5\} are of the form \{6k –1, 6k+1\}

Let \( T = \{(6k–1, 6k+1) \mid k = 1, 2, 3\ldots \} \)

Let \( u_k = 6k – 1 \) and let \( v_k = 6k + 1 \)

And define \( t_k = (u_k, v_k) \)

Listed below are the first few members of the set \( T \). (the composite numbers are underlined)


\( k \equiv \pm 1 \mod 5 \Rightarrow t_k \) contains a multiple of 5 and is therefore not a pair of twin primes.

Let \( S_p = \{ t_k \mid t_k \) contains a multiple of prime \( p \} \)

\( T_p = T \setminus S_p = \{ t_k \mid t_k \) does not contain a multiple of prime \( p \} \)

\( S_5 = \{ t_4, t_6, t_9, t_{11}, t_{14}, t_{16}, t_{19} \ldots \} = \{(23, 25), (35, 37), (53, 55), (65, 67)\ldots \} \)

\( T_5 = \{ t_1, t_2, t_3, t_5, t_7, t_8, t_{10}, t_{12} \ldots \} \)

The values of \( k \) in \( S_5 = \{4, 6, 9, 11, 14, 16, 19, 21\ldots \} = \{k \in \mathbb{Z}^+ \mid k \equiv \pm 1 \mod 5 \} \)

The values of \( k \) in \( T_5 = \{1, 2, 3, 5, 7, 8, 10, 12, 13, 15, 17, 18, 20, 22, 23, 25, 27, 28, 30\ldots \} \)

\( k \equiv \pm 1 \mod 7 \Rightarrow t_k \) contains a multiple of 7

The values of \( k \) in \( S_7 = \{6, 8, 13, 15, 20, 22, 27, 29, 36, 38\ldots \} = \{k \in \mathbb{Z}^+ \mid k \equiv \pm 1 \mod 7 \} \)

The values of \( k \) in \( T_7 = \{1, 2, 3, 4, 5, 7, 9, 10, 11, 12, 14, 16, 17, 18, 19 \ldots \} \)

Since \( 6(2) – 1 = 11 \), \( k \equiv \pm 2 \mod 11 \Rightarrow t_k \) contains a multiple of 11

The values of \( k \) in \( S_{11} = \{9, 13, 20, 24, 31, 35, 42, 46\ldots \} = \{k \in \mathbb{Z}^+ \mid k \equiv \pm 2 \mod 11 \} \)

The values of \( k \) in \( T_{11} = \{1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 14, 15, 16, 17, 18, 19, 21, 22, 23, 25\ldots \} \)

If \( p \) is prime and \( p = 6a \pm 1 \), \( t_k \) contains a multiple of \( p \Leftrightarrow k \equiv \pm a \mod p \)
Let $\mathcal{P}$ be the set of all primes.

Let $T_w = \{ t_k = (u_k, v_k) \mid u_k \in \mathcal{P} \text{ and } v_k \in \mathcal{P}, k \in \mathbb{Z}^+ \}$

$T_w = T \setminus \bigcup_{p \geq 5} S_p$; by De Morgan’s Law $T_w = \bigcap_{p \geq 5} T_p$

**Lemma 1:**

Define $k_p$ as the value of $k$ for primes $p = 6k + 1$ or $p = 6k - 1$.

and $T_w$ as the set of all twin prime pairs.

Given a large integer $N$ and $6k + 1 < N$,

$t_k \notin T_w \iff k \equiv \pm k_p \pmod{p}$ for some prime $p$, $5 \leq p < \sqrt{N}$.

As in the sieve of Eratosthenes, we delete $\{ k \mid k = np \pm k_p \text{ for primes } p = 6k \pm 1 \}$

$n = \{1, 2, 3\ldots \} \forall p \geq 5 \leq p < N$

Consider the set $K = \{k \in \mathbb{Z}^+ \mid k < 6N\}$.

In every interval $I \in K$ such that $I = \{np, (n+1)p\} \forall n \in \mathbb{Z}^+$ and $p$ is prime, $5 \leq p < \sqrt{N}$

$\exists$ exactly 2 values of $k$, (i.e. $k = np + k_p$ and $k = (n+1)p - k_p$), such that $t_k$ contains a multiple of $p$.

Let $\pi_2(N)$ be the number of primes $p$ less than $N$ such that $p + 2$ is also prime.

By the Brun Sieve we have:

\[
\pi_2(N) = \frac{N}{6} \prod_{5}^{V} (1 - \frac{2}{p}) + R_p \quad \text{where } R_p \text{ is the error term}
\]

and $V =$ maximum prime $p < \sqrt{N}$

**Example 1:**

$N = 529, V = 19$

$\pi_2(N) \approx \frac{529}{6} \prod_{5}^{19} (1 - \frac{2}{p}) = 20.6521\ldots$

Actually, $\pi_2(N) = 25$ so $R_p \approx 4.3$
Let $\pi_2 (n) = \text{the number of primes } p \text{ less than } n \text{ such that } p + 2 \text{ is also prime.}$

Where $V$ is maximum prime $p < \sqrt{n}$

**Table 1** ($\pi_2 (n)$ compared to the formula)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\pi_2 (n)$</th>
<th>$\frac{n}{6} \prod_{5}^{V} (1 - 2/p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Approx.</td>
<td></td>
</tr>
<tr>
<td>529</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>1000</td>
<td>35</td>
<td>31</td>
</tr>
<tr>
<td>2500</td>
<td>72</td>
<td>64</td>
</tr>
<tr>
<td>5000</td>
<td>126</td>
<td>111</td>
</tr>
<tr>
<td>7500</td>
<td>169</td>
<td>150</td>
</tr>
<tr>
<td>10000</td>
<td>205</td>
<td>191</td>
</tr>
<tr>
<td>15000</td>
<td>272</td>
<td>261</td>
</tr>
<tr>
<td>20000</td>
<td>342</td>
<td>328</td>
</tr>
<tr>
<td>25000</td>
<td>408</td>
<td>394</td>
</tr>
<tr>
<td>30000</td>
<td>467</td>
<td>456</td>
</tr>
<tr>
<td>35000</td>
<td>539</td>
<td>520</td>
</tr>
<tr>
<td>40000</td>
<td>591</td>
<td>570</td>
</tr>
<tr>
<td>50000</td>
<td>705</td>
<td>700</td>
</tr>
<tr>
<td>75000</td>
<td>958</td>
<td>968</td>
</tr>
</tbody>
</table>

This estimate exceeds the actual number of twin prime pairs for large values of $n$ because for some primes $p \notin T_w$ and elements in \{ $k : 1 \leq k < \frac{n}{6}$ \}, the number of elements in each of the sets \{ $k : p \mid 6k - 1$ \} \{ $k : p \mid 6k + 1$ \} = $\left\lfloor \frac{n}{6p} \right\rfloor + 1 \rfloor$ where $\left\lfloor . \right\rfloor$ is the greatest integer function, therefore some composites will not be sifted out by the product formula given above. The formula can be refined by rewriting it as a two-part sieve formula that represents the application of Eratosthenes’ Sieve first to $6k-1$ type numbers then to the $6k +1$ type.

See equation (2).
Except for 3, all lesser twin primes are of the form $6k - 1$.

Consider the set \( \{ u_k \mid u_k = 6k - 1, \; k \in \mathbb{Z}^+ \} \).

**Lemma 2:**

Given \( u_k < N, \; u_k \notin T_w \Rightarrow k \equiv \pm k_p \pmod p \) for some prime \( p, \; 5 \leq p < \sqrt{N} \).

i.e. \( (u_k, v_k) \) is not a pair of twin primes \( \iff k \equiv \pm k_p \pmod p \) for some prime \( p, \; 5 \leq p < \sqrt{N} \).

Out of every \( p \) elements in the set \( \{u_k\} \), \( (p \) prime and \( p \geq 5) \),

exactly one is a multiple of \( p \) and one precedes a \( 6k+1 \) multiple of \( p \).

If we list the elements of \( \{ u_k \mid u_k = 6k - 1, \; k \in \mathbb{Z}^+ \} \) and delete every \( u_k \) in which \( k \equiv \pm 1 \pmod{5} \)

or \( \pm 1 \pmod{7} \) or \( \pm 2 \pmod{11} \) or \( \pm 2 \pmod{13} \) or \( k \equiv \pm 3 \pmod{17} \) or \( \pm 3 \pmod{19} \) \( \ldots \pm k_p \pmod{p} \) up to \( p < \sqrt{N} \).

The remaining terms are all twin primes.

We use this method to find twin primes in the table below by deleting all \( k \equiv \pm 1 \pmod{5} \)

or \( \pm 1 \pmod{7} \), \( \pm 2 \pmod{11} \) or \( \pm 2 \pmod{13} \), since \( 13 = \max p < \sqrt{179} \)

<table>
<thead>
<tr>
<th>( k )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_k )</td>
<td>5</td>
<td>11</td>
<td>17</td>
<td>23</td>
<td>29</td>
<td>35</td>
<td>41</td>
<td>47</td>
<td>53</td>
<td>59</td>
</tr>
<tr>
<td>( k )</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>( u_k )</td>
<td>65</td>
<td>71</td>
<td>77</td>
<td>83</td>
<td>89</td>
<td>95</td>
<td>101</td>
<td>107</td>
<td>113</td>
<td>119</td>
</tr>
<tr>
<td>( k )</td>
<td>24</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>( u_k )</td>
<td>125</td>
<td>131</td>
<td>137</td>
<td>143</td>
<td>149</td>
<td>155</td>
<td>161</td>
<td>167</td>
<td>173</td>
<td>179</td>
</tr>
</tbody>
</table>

The \( u_k \)'s that correspond to the undeleted values of \( k \) are the lesser of twin primes

i.e.: 5, 11, 17, 29, 41, 59, 71, 101, 107, 137, 149, 179
We can demonstrate this sieve method by the following procedure:

first we cross out all values of k such that $k \equiv k_p \mod p$ if $p \equiv -1 \mod 6$ (i.e. $p = 6k_p - 1$) and all values of k such that $k \equiv -k_p \mod p$ if $p \equiv 1 \mod 6$ (i.e. $p = 6k_p + 1$) up to $p < \sqrt{n}$ so that we are left with the set \{ $k \in \mathbb{Z}^+ \mid k < \frac{n}{6}$ and $(6k-1)$ is prime \}.

Table 3  Values of k such that 6k-1 is a prime (not deleted)

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>−6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_k$</td>
<td>5</td>
<td>11</td>
<td>17</td>
<td>23</td>
<td>29</td>
<td>35</td>
<td>41</td>
<td>47</td>
<td>53</td>
<td>59</td>
</tr>
<tr>
<td>$k$</td>
<td>14</td>
<td>12</td>
<td>43</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>$u_k$</td>
<td>65</td>
<td>71</td>
<td>77</td>
<td>83</td>
<td>89</td>
<td>95</td>
<td>101</td>
<td>107</td>
<td>113</td>
<td>119</td>
</tr>
<tr>
<td>$k$</td>
<td>24</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>$u_k$</td>
<td>125</td>
<td>131</td>
<td>137</td>
<td>143</td>
<td>149</td>
<td>155</td>
<td>161</td>
<td>167</td>
<td>173</td>
<td>179</td>
</tr>
</tbody>
</table>

We then cross out the elements of the set \{ $(k \in \mathbb{Z}^+ \mid k < \frac{n}{6}$ and $(6k+1)$ is composite \} i.e.

\{ $k \mid k \equiv -k_p \mod p$ if $p \equiv -1 \mod 6$ \} ∪ \{ $k \mid k \equiv k_p \mod p$ if $p \equiv 1 \mod 6$ \}. This leaves us with the set of all twin primes less than $N$. (Table 2)

This can be expressed as approximation formula that follows:

$$\pi_2(N) \approx \pi(N) \approx \frac{\pi(N)}{2} \prod_{p < V} \frac{p-2}{p-1} \approx \frac{\pi(N)}{2} \prod_{p < V} \frac{p(p-2)}{(p-1)^2} \frac{p-1}{p} , V= \max p < \sqrt{N}$$

By the Prime Number Theorem $\pi (N) \sim \frac{N}{\ln N}$ and by Mertens’ Theorem:

$$\prod_{p < V} \frac{p-1}{p} \sim 2e^{-\gamma} \frac{\ln N}{\ln \ln N} = 1.122... \frac{\ln N}{\ln \ln N}$$

which overestimates the true ratio $\frac{\pi(N)}{N}$

and $\gamma = 5772156649...$ is the Euler-Mascheroni constant. [4] (Polya)

By using $\frac{1}{\ln N}$, which is a lower bound for $\frac{\pi(N)}{N}$ [5] (Rosser and Schoenfeld)

and a little bit of algebra, we obtain:

$$\pi_2 (N) \sim \frac{N}{2\ln N} \times \frac{4}{3} C_2 \times 3 \times \frac{1}{\ln N} ,$$

where

$$C_2 = \prod_{p \geq 3} \frac{p(p-2)}{(p-1)^2} = 0.6601618...$$

is the twin prime constant and

$$\pi_2(N) \sim 2C_2 \frac{N}{(\ln N)^2}$$

which is the Hardy-Littlewood Conjecture.
Hardy and Littlewood [2] also conjectured a better approximation:

\[(6) \quad \pi_2(N) \sim 2C_2 \int_2^N \frac{1}{(\ln t)^2} \, dt, \text{ also based on PNT}\]

Formula (4) is naturally equivalent to (3) but

\[(7) \quad \int_2^n \frac{1}{(\ln t)^2} \, dt = \frac{n}{(\ln n)^2} \left(1 + \frac{2!}{\ln n} + \frac{3!}{(\ln n)^2} + \cdots\right),\]

and the second factor on the right hand side is (for the values of \(n\) that we have to consider) far from negligible. [2] (Hardy and Wright)

This suggests \(2C_2 \frac{N}{(\ln N)^2} < \pi_2(N)\), for large integers \(N\)

From equation (2) and the fact that \(\frac{N}{6} \prod_{p \in T} \frac{p-1}{p}\) is an over-approximation of \(\frac{\pi(N)}{2}\),

(because for some primes \(p \notin T_w, 5 \leq p < \sqrt{N}, \lfloor\{k: p \nmid (6k-1) \setminus k_p\}\rfloor\)

\[= \lfloor\frac{N}{6} \left(\frac{p-1}{p} - \frac{6}{N}\right)\rfloor\text{ and likewise for }\{k: p \nmid (6k+1) \setminus k_p\}\).\]

After multiplying the right side of equation (2) by \(\frac{N}{6}\) and \(\frac{6}{N}\) we obtain:

\[(8) \quad \pi_2(N) \approx \frac{\pi(N)}{2} \cdot \frac{4}{3} \cdot C_N \cdot \frac{6}{N} \cdot \frac{\pi(N)}{2} \quad \text{where} \quad C_N = \prod_{2 < p < \sqrt{N}} \frac{p(p-2)}{(p-1)^2} \]

which includes \(\lim_{N \to \infty} \left(\frac{N(p-2)-6(p-1)}{N(p-1)}\right)\left(\frac{Np}{N(p-1)-6p}\right)\) for some primes \(p, 5 \leq p < \sqrt{N}\).

\[(9) \quad \pi_2(N) \approx 2C_N \frac{[\pi(N)]^2}{N}.\]

\(\lim_{N \to \infty} C_N = C_2 \land \pi(N) \sim \frac{N}{\ln N} \Rightarrow\)

\[(10) \quad \pi_2(N) \sim 2C_2 \frac{N}{(\ln N)^2}\]

As shown by Rosser and Schoenfeld [5],

\[\frac{N}{\ln N} < \pi(N) \quad \forall \quad N \geq 17 \quad \text{which gives us:}\]

\[(11) \quad 2C_2 \frac{N}{(\ln N)^2} < 2C_2 \frac{[\pi(N)]^2}{N} \quad \text{for large enough values of } N.\]
Table 4

(Values of $\pi_2(n)$ compared to logarithmic integral and ratio formulas) [1] (Caldwell)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\pi_2(n)$</th>
<th>$2C_2 \text{li}_2(n)$</th>
<th>$2C_2 \frac{n}{(\ln n)^2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^6$</td>
<td>8169</td>
<td>8248</td>
<td>6917</td>
</tr>
<tr>
<td>$10^7$</td>
<td>58980</td>
<td>58754</td>
<td>50822</td>
</tr>
<tr>
<td>$10^8$</td>
<td>440312</td>
<td>440368</td>
<td>389107</td>
</tr>
<tr>
<td>$10^9$</td>
<td>3424506</td>
<td>3425308</td>
<td>3074425</td>
</tr>
<tr>
<td>$10^{10}$</td>
<td>27412679</td>
<td>27411417</td>
<td>24902848</td>
</tr>
<tr>
<td>$10^{11}$</td>
<td>224376048</td>
<td>224368865</td>
<td>205808661</td>
</tr>
<tr>
<td>$10^{12}$</td>
<td>1870585220</td>
<td>1870559867</td>
<td>1729364449</td>
</tr>
<tr>
<td>$10^{13}$</td>
<td>15834664872</td>
<td>15834598305</td>
<td>14735413063</td>
</tr>
<tr>
<td>$10^{14}$</td>
<td>135780321665</td>
<td>135780264894</td>
<td>127055347335</td>
</tr>
<tr>
<td>$10^{15}$</td>
<td>1177209242304</td>
<td>1177208491861</td>
<td>1106793247903</td>
</tr>
</tbody>
</table>

Let $W(n) = 2C_2 \frac{[\pi(n)]^2}{n}$

Table 5

(limit $\frac{W(n)}{\pi_2(n)}$ approaching 1 as $n$ increases)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\pi(n)$</th>
<th>$W(n)$</th>
<th>$\frac{W(n)}{\pi_2(n)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^6$</td>
<td>78498</td>
<td>8136</td>
<td>0.9959603…</td>
</tr>
<tr>
<td>$10^7$</td>
<td>664579</td>
<td>58314</td>
<td>0.99251114…</td>
</tr>
<tr>
<td>$10^8$</td>
<td>5761455</td>
<td>438273</td>
<td>0.995242615…</td>
</tr>
<tr>
<td>$10^9$</td>
<td>50847534</td>
<td>3413659</td>
<td>0.9968325…</td>
</tr>
<tr>
<td>$10^{10}$</td>
<td>455052511</td>
<td>27340309</td>
<td>0.99735998…</td>
</tr>
<tr>
<td>$10^{11}$</td>
<td>4118054813</td>
<td>223905433</td>
<td>0.99790256…</td>
</tr>
<tr>
<td>$10^{12}$</td>
<td>37607912018</td>
<td>1867406346</td>
<td>0.998300599…</td>
</tr>
<tr>
<td>$10^{13}$</td>
<td>346065536839</td>
<td>15812374441</td>
<td>0.99859230168…</td>
</tr>
<tr>
<td>$10^{14}$</td>
<td>3204941750802</td>
<td>135619040528</td>
<td>0.999881219…</td>
</tr>
<tr>
<td>$10^{15}$</td>
<td>29844570422669</td>
<td>1176010096499</td>
<td>0.998981365…</td>
</tr>
</tbody>
</table>
Let \( t_k = (6k - 1, 6k + 1) \) and \( T_w = \) the set of all twin prime pairs.

The occurrence of twin primes may be summarized as follows:

\[ \forall k > 3 \text{ and primes } p, \]
\[ t_k \in T_w \iff k \equiv 0, 2 \text{ or } 3 \pmod{5} \land k \not\equiv \pm k_p \pmod{p} \forall p > 5 \]

Where \( k_p \) is the value of \( k \) for the primes \( p = 6k + 1 \) or \( p = 6k - 1 \).
References:

   utm.edu/staff/caldwell/preprints/Heuristics.pdf


3. Hardy, G. H., Littlewood, J. E.: Some Problems of ‘Partitio Numerorum’ III: On the


6. Rosser, J. Barkley, Schoenfeld, Lowell: Approximate formulas for some functions of prime
   numbers. Illinois J. Math. 6, (1962) p. 69