Distribution law for twin primes amongst naturals

Boris B. Benyaminov
6 Moa St, Belmont, North Shore,
Auckland, Postal Code: 0622, New Zealand
boris.b.b@hotmail.com

Abstract
A hypothesis is put forward regarding the function $\pi_2(x)$ which describes the distribution of twin primes in the set of natural numbers. The function $\pi_2(x)$ is tested by evaluation and an empirical $\pi^*_2(x)$ is arrived at, which is shown to be highly accurate. Several other questions are also addressed.

Keywords: twin prime, distribution, natural numbers.
Mathematics Subject Classification 2010: 11A41, 11M26.

1 Introduction

In 1923, Hardy and Littlewood proposed a hypothesis for the distribution of twin primes on the interval $(1, x)$ [1]:

\[ \pi_2(x) \sim 2 \prod_{p=3}^{\infty} \left(1 - \frac{1}{(p-1)^2}\right) \frac{x}{\ln x}. \] (1.1)

Later on, the following expression was put forward [2]:

\[ \pi_2(x) \sim 2 \prod_{p=3}^{\infty} \left(1 - \frac{1}{(p-1)^2}\right) \frac{x}{(\ln x)^2} \prod_{p=3}^{\infty} \frac{p-1}{p-2}. \] (1.2)

The asymptotic representations (1.1) and (1.2) of the very important function $\pi_2(x)$ are too complicated to be used in practice. In this article I propose a new law for the distribution of twin primes among the naturals in the form of a much simpler $\pi_2(x)$, based on composition with the function $\pi(x)$.

2 Results

Recall that $\pi(x)$ [or $\pi_2(x)$] is the number of primes (twin primes) not larger than $x$. The following hypothesis is proposed for the distribution of twin primes in the set of all naturals:
Hypothesis 2.1. Twin primes are distributed among prime numbers in the same way that primes are distributed among naturals. In other words,
\[ \pi_2(x) = \pi(\pi(x)). \] (2.3)

Table 1 gives some values of the functions \(\pi_2(x)\) and \(\pi(\pi(x))\) for \(x \leq 10^6\). The values of \(\pi_2(x)\) were computed according to Lehmer’s tables [3]. From the results in Table 1 it is safe to say that the ratio of \(\pi_2(x)\) to \(\pi(\pi(x))\) is either exactly one or differs from unity by some negligibly small amount.

| \(x\)    | \(\pi(x)\) | \(\pi_2(x)\) | \(\pi(\pi(x))\) | \(\frac{\pi_2(x)}{\pi(\pi(x)))}\) |
|----------|-------------|--------------|------------------|-----------------------------------|
| 25       | 9           | 4            | 4                | 1                                 |
| 50       | 15          | 6            | 6                | 1                                 |
| 75       | 21          | 8            | 8                | 1                                 |
| 125      | 30          | 10           | 10               | 1                                 |
| 150      | 35          | 12           | 11               | 1.091                             |
| 200      | 46          | 15           | 14               | 1.071                             |
| 300      | 62          | 19           | 18               | 1.056                             |
| 400      | 78          | 21           | 21               | 1                                 |
| 500      | 95          | 24           | 24               | 1                                 |
| 700      | 125         | 30           | 30               | 1                                 |
| 900      | 154         | 35           | 36               | 0.972                             |
| 1350     | 217         | 46           | 47               | 0.979                             |
| 1500     | 239         | 49           | 52               | 0.942                             |
| 2000     | 303         | 60           | 62               | 0.968                             |
| 3000     | 430         | 81           | 82               | 0.988                             |
| 4000     | 550         | 102          | 101              | 1.010                             |
| 5000     | 669         | 123          | 121              | 1.016                             |
| 10,000   | 1,226       | 201          | 201              | 1                                 |
| 15,000   | 1,754       | 268          | 273              | 0.982                             |
| 20,000   | 2,262       | 338          | 335              | 1.009                             |
| 25,000   | 2,762       | 403          | 402              | 1.002                             |
| 30,000   | 3,245       | 462          | 457              | 1.011                             |
| 40,000   | 4,203       | 585          | 575              | 1.017                             |
| 50,000   | 5,133       | 697          | 685              | 1.018                             |
| 100,000  | 9,592       | 1,224        | 1,184             | 1.034                             |
| 200,000  | 17,984      | 2,159        | 2,062             | 1.047                             |
| 500,000  | 41,538      | 4,343        | 4,343             | 1.035                             |
| 1,000,000| 78,498      | 7,902        | 7,902             | 1.033                             |

Table 1: Testing hypothesis \((2.3)\) for a few selected values of \(x \leq 10^6\).
Next, we will require the upper and lower bounds for \( \pi(x) \) given in [4]:

\[
\frac{2x}{3 \ln x} < \pi(x) < \frac{8x}{5 \ln x}.
\] (2.4)

**Theorem 2.1.** For all \( x \geq 5 \) for which (2.4) holds, we have

\[
A < \pi_2(x) = \pi(\pi(x)) < B,
\] (2.5)

where

\[
A = \frac{4x}{9 \ln x [\ln x - \ln(\ln x) - \ln 1.5]},
\]

\[
B = \frac{64x}{25 \ln x [\ln x - \ln(\ln x) + \ln 1.6]}.
\]

**Proof.** The function \( f(x) = x/\ln x \) is monotonically increasing for \( x \geq 3 \). From (2.4) we have

\[
\frac{2\pi(x)}{3 \ln [\pi(x)]} < \pi_2(x) = \pi(\pi(x)) < \frac{8\pi(x)}{5 \ln [\pi(x)]}.
\] (2.6)

If we now consider the expression \( \frac{\pi(x)}{\ln[\pi(x)]} \) as a function of \( \pi(x) \), we can see it is also monotonically increasing. In our case, \( \pi(x) \geq 3 \) as \( x \geq 5 \) (by theorem requirements). Taking this into consideration and using the right-hand side of inequality (2.4)

\[
\pi_2(x) = \pi(\pi(x)) < \frac{8\pi(x)}{5 \ln [\pi(x)]} < \frac{8 \cdot \frac{8x}{5 \ln x}}{5 \ln \left( \frac{8x}{5 \ln x} \right)} = \frac{64x}{25 \ln x [\ln x - \ln(\ln x) + \ln 1.6]} = B.
\]

Similarly, by using the left-hand side of inequality (2.4) we obtain the lower bound:

\[
\pi_2(x) = \pi(\pi(x)) > \frac{2\pi(x)}{3 \ln [\pi(x)]} > \frac{2 \cdot \frac{2x}{3 \ln x}}{3 \ln \left( \frac{2x}{3 \ln x} \right)} = \frac{4x}{9 \ln x [\ln x - \ln(\ln x) - \ln 1.5]} = A.
\]

Thus, from (2.6) we have

\[
A < \frac{2\pi(x)}{3 \ln [\pi(x)]} < \pi_2(x) = \pi(\pi(x)) < \frac{8\pi(x)}{5 \ln [\pi(x)]} < B,
\]

precisely inequality (2.5), as required. \( \square \)

In Table 2 we check inequality (2.5) for several values of \( x \).

Next let us look at the density of twin primes among the primes.

**Theorem 2.2.** Almost all primes are not twins, so

\[
\pi_2(x) = o(\pi(x)).
\] (2.7)
| $x$  | $A$ | $\pi_2(x)$ | $B$  |
|------|-----|------------|-----|
| 50   | 3   | 6          | 11  |
| 125  | 4   | 10         | 18  |
| 200  | 5   | 15         | 23  |
| 300  | 7   | 19         | 31  |
| 400  | 8   | 21         | 36  |
| 500  | 9   | 24         | 42  |
| 700  | 11  | 30         | 53  |
| 1,000| 14  | 35         | 67  |
| 5,000| 44  | 123        | 219 |
| 10,000| 73 | 201        | 372 |
| 25,000| 148| 403        | 762 |
| 50,000| 256| 697        | 1,328|
| 100,000| 445| 1,224      | 2,331|
| 500,000| 1,700| 4,494   | 8,853 |
| 1,000,000| 2,983| 8,164   | 15,887 |

Table 2: Testing inequality (2.5) for a few selected values of $x \leq 10^6$.

**Proof.** Assume that hypothesis (2.3) is true. Then, denoting $y = \pi(x)$, we have

$$0 \leq \frac{\pi_2(x)}{\pi(x)} = \frac{\pi(\pi(x))}{\pi(x)} = \frac{\pi(y)}{y}.$$  

We can find an upper bound for $\pi(y)/y$ by the sieve method, taking the set $\{y\}$ to contain no repeated values.

Let $\varphi(y, r)$ be the number of naturals no larger than $y$ and not divisible by any of the first $r$ primes $P_1, P_2, \ldots, P_r$. Then

$$\varphi(y, r) = \sum_{d \mid P_1P_2\ldots P_{\pi(\sqrt{y})}} \mu(d) \left\lfloor \frac{y}{d} \right\rfloor, \quad (2.8)$$

where $\mu(d)$ is the Mobius function and $d \mid P_1 \ldots P_r$ means all $d$ not divisible by $P_1$ to $P_r$. It is clear that

$$\pi(y) \leq \varphi(y, r) + r. \quad (2.9)$$

We next drop the floor operator in (2.8), and note that there are $2^r$ terms being summed. This means that the resulting expression has an error no larger than $2^r$, and by (2.9) we subsequently get

$$\pi(y) \leq \sum_{d \mid P_1P_2\ldots P_{\pi(\sqrt{y})}} \mu(d) \left\lfloor \frac{y}{d} \right\rfloor + r \leq y \times \sum_{d \mid P_1P_2\ldots P_{\pi(\sqrt{y})}} \frac{\mu(d)}{d} + r + 2^r = y \prod_{P \leq P_r} \left(1 - \frac{1}{P-1}\right) + r + 2^r < y \prod_{P \leq P_r} \left(1 - \frac{1}{P-1}\right) + 2^{r+1},$$
because $r < P_r < 2^r$. Furthermore, from the inequality
\[
\prod_{P \leq x} (1 - P^{-1})^{-1} > \ln x,
\]
we find
\[
\pi(y) < \frac{y}{\ln P_r} + 2^{r+1} < \frac{y}{\ln r} + 2^{r+1}.
\]
Choose $r = c \ln y$, $c < 1/\ln 2$. Then $2^r < y$ and
\[
\pi(y) < \frac{y}{\ln (c \ln y)} + 2y^{\ln 2} = \frac{y}{\ln c + \ln \ln y} + 2y^{\ln 2},
\]
where $c \ln 2 < 1$. Dividing through by $y$, we get
\[
0 \leq \frac{\pi(y)}{y} < \frac{1}{\ln c + \ln (\ln y)} + \frac{2}{y^{1-c \ln 2}}.
\]
As $y \to \infty$, the right-hand side of the above goes to zero, which implies the validity of equation (2.7). \hfill \square

**Corollary 2.1.** Since we know that $\pi(x) = o(x)$, from Theorem 2.2 it follows that $\pi_2(x) = o(x)$. In fact,
\[
\lim_{x \to \infty} \frac{\pi_2(x)}{x} = \lim_{x \to \infty} \frac{\pi_2(x)}{\pi(x)} \cdot \lim_{x \to \infty} \frac{\pi(x)}{x} = 0.
\]

Next, we move on to construct an empirical function for the law of distribution of twin primes.

Denote by $\eta_P$ the density of primes in the reals, and by $\eta_{PP}$ the density of twin primes in the primes, i.e. $\eta_P = \pi(x)/x$ and $\eta_{PP} = \pi_2(x)/\pi(x)$. Based on $\pi(x) = o(x)$ and (2.7), the densities $\eta_P$ and $\eta_{PP}$ go to zero as $x \to \infty$, but the ratio
\[
h = \frac{\eta_{PP}}{\eta_P}
\]
remains bounded in a well-defined, constant interval (see Table 3). We can obtain a rough estimate of an upper bound for $h > 0$; for this we need the inequality $\pi(x) > x/\ln x$ and the right-hand side of (2.5). We get
\[
h = \frac{x \pi_2(x)}{[\pi(x)]^2} < \frac{64x^2}{(\ln x)^2 \cdot 25 \ln x (\ln x - \ln(\ln x) + \ln 1.6)} \cdot \frac{2.56 \ln x}{\ln x - \ln(\ln x) + \ln 1.6} < \frac{2.56 \ln x}{\ln x - \ln(\ln x)} < 5.12.
\]
Thus, $0 < h < 5.12$. This fact allows one to construct an empirical function $\pi_2^*(x)$ for the number of twin primes on $(2, x)$. As is evident from Table 3, $\pi_2^*(x)$ defined below is rather accurate.
We obtain \( \pi^*_2(x)/\pi(x) = h_c \cdot \pi(x)/x \), leading to

\[
\pi^*_2(x) = \left[ \frac{h_c \pi^2(x)}{x} \right],
\]

(2.11)

where \( h_c = 1.325067 \ldots \) – the mean value of \( h \) for \( x \leq 10^6 \) and we round the right-hand side of (2.11) to the nearest integer.

In Table 3 I test the accuracy of \( \pi^*_2(x) \) for \( 50 \leq x \leq 10^6 \). Nevertheless, (2.11) is applicable for \( x \geq 10^6 \), too. For example, there are 183,728 twin primes less than or equal to \( x = 37 \cdot 10^6 \), while \( \pi^*_2(x) = 183,463 \) which gives a relative error of \( \delta = 0.0014 \) (see Table 3).

| \( x \)  | \( h \)    | \( \pi_2(x) \) | \( \pi^*_2(x) \) | \( |\Delta| \) | \( \frac{|\Delta|}{\pi_2(x)} \) |
|---------|-----------|----------------|-----------------|--------|----------------|
| 50      | 1.333336  | 6              | 6               | 0      | 0              |
| 150     | 1.346938  | 11             | 11              | 0      | 0              |
| 500     | 1.329639  | 24             | 24              | 0      | 0              |
| 1,500   | 1.286742  | 49             | 50              | 1      | 0.0204         |
| 2,000   | 1.307061  | 60             | 61              | 1      | 0.0167         |
| 3,000   | 1.314223  | 81             | 82              | 1      | 0.0123         |
| 4,000   | 1.348760  | 102            | 100             | 2      | 0.0196         |
| 5,000   | 1.374114  | 123            | 119             | 4      | 0.0325         |
| 10,000  | 1.330737  | 201            | 200             | 1      | 0.0050         |
| 15,000  | 1.306672  | 268            | 274             | 6      | 0.0224         |
| 20,000  | 1.321178  | 338            | 339             | 1      | 0.0030         |
| 25,000  | 1.320680  | 403            | 404             | 1      | 0.0025         |
| 30,000  | 1.316236  | 462            | 465             | 3      | 0.0065         |
| 40,000  | 1.324637  | 585            | 585             | 0      | 0              |
| 50,000  | 1.322696  | 697            | 698             | 1      | 0.0014         |
| 100,000 | 1.330341  | 1,224          | 1,219           | 5      | 0.0041         |
| 200,000 | 1.335088  | 2,159          | 2,143           | 16     | 0.0074         |
| 500,000 | 1.302302  | 4,494          | 4,573           | 79     | 0.0176         |
| 1,000,000 | 1.342908 | 8,164          | 8,165           | 1      | 0.0001         |

Table 3: Testing expression (2.11) for a few selected values of \( x \leq 10^6 \).

**Acknowledgements**

The author would like to thank Sophie S. Shamailov for translating and typing the article.
References

[1] G.H. Hardy and J.E. Littlewood, *Some problems of ‘Partitio numerorum’; III: On the expression of a number as a sum of primes*, Acta Mathematica, 44, 1, pp. 1-70 (1923).

[2] A.F. Lavrik, *On the theory of distribution of primes based on the method of trigonometric sums of I.M. Vinogradov*, Proceedings of the Mathematical Institute, USSR Academy of Sciences, 64, pp. 90-125 (1961).

[3] D.N. Lehmer, *List of primes numbers from 1 to 10,006,721*, Carnegie Institution Washington, D.C., 1914.

[4] E. Trost, *Primzahlen*, Verlag Birkhäuser, Basel-Stuttgart (1953).