On the Gap Distribution of Prime Numbers.

Yasuo Yamasaki and Aiichi Yamasaki

Abstract. A "theoretical" distribution of prime number gaps is proposed and compared with the actual distribution. Some probabilistic discussions are given.

1. Introduction

Let $p_n$ be the $n$-th prime number and for $x > 0$, put $\pi(x) = \text{Max}\{n|p_n \leq x\}$. The prime number theorem tells us that $\pi(x) \sim \frac{x}{\log x}$, or equivalently $p_n \sim n \log n$.

We call $d_n = p_{n+1} - p_n$ the $n$-th prime gap. On the order of the growth of $d_n$, we have two conjectures.

(1.1) \[ \lim d_n = 2 \]

(1.2) \[ \lim \frac{d_n}{(\log p_n)^2} = 1, \]

or more weakly

\[ \lim \frac{d_n}{(\log p_n)^2} < \infty. \]

1991 Mathematics Subject Classification Primary 11N05
The conjecture (1.1) is the famous twin prime conjecture, which has long been believed to be true, though not yet proved. Put $\pi_2(x) = \#\{n|p_n \leq x \text{ and } d_n = 2\}$, then (1.1) is equivalent to $\lim_{x \to \infty} \pi_2(x) = \infty$. Much stronger Hardy-Littlewood conjecture\cite{[1]} says that

$$\pi_2(x) \sim 2c \frac{x}{(\log x)^2}$$

with

$$c = \prod_{n=2}^{\infty} \left(1 - \frac{1}{(p_n - 1)^2}\right) = 0.66016 \cdots.$$

Some experiments on counting twin prime numbers by computers\cite{[2]}\cite{[3]} seem to suggest that (1.3) is correct (at least up to $x = 10^{11}$).

Later we shall investigate (1.3) more closely.

Also the conjecture (1.2) has long been believed to be true\cite{[4]}\cite{[5]}\cite{[6]}, but the established results are much weaker: $d_n = O(p_n^\theta)$, $0 < \exists \theta < 1$. The best record at present is $\theta = \frac{11}{20} - \frac{1}{384} \approx 0.5473 \cdots$. Again by computers, (1.2) seems to be consistent with experiments up to $p_n \sim 10^{14}$.

Historically the studies on $d_n$ have concentrated on the following two points: namely the frequency of twin primes and the occurrences of large gaps. In this paper, we shall discuss the distribution of $d_n$ as a whole. There exists a belief (with no justification) that prime numbers distribute mutually independently except obvious inter-relations, such as $d_n(n \geq 2)$ must be even integers for instance. Under this "independence hypothesis", we can derive a "theoretical" distribution of $d_n$ and compare it with the actual distribution obtained by counting them by computers. This is the purpose of the present paper. Especially, we show that the conjectures (1.2) and (1.3) are true with probability 1 under our "theoretical" distribution hypothesis.

2. Exponential distribution

Discussions in this section are not rigorous mathematically, but the authors’ excuse
is that the aim of this section is to find a simple and plausible "theoretical" distribution of $d_n$, not to prove something.

Consider the exponential distribution on $\mathbb{R}_+ = [0, \infty)$. It is the probability measure $\mu$ given by $\mu([a, \infty)) = e^{-\alpha a}$, or equivalently by

$$
\mu(E) = \alpha \int_E e^{-\alpha t} dt
$$

for a Borel set $E$ of $[0, \infty)$. This is the distribution of the first occurrence time of the event which occurs with probability $\alpha \Delta t$ in an infinitesimal time interval $\Delta t$, independently of $t$.

Thanks to the "independence hypothesis", we shall assume that the exponential distribution can be applied to the gaps of prime numbers. But the gaps are always even integers, so do not distribute continuously on $[0, \infty)$. Our excuse is that the smallest gap $d_n = 2$ may be regarded infinitesimal compared with the mean value $<d_n> \sim \log n$ after $n$ primes. So, we shall apply the exponential distribution (of a continuous variable) to the gap distribution of prime numbers assumed to be sufficiently large.

However, "obvious inter-relations" should be taken into account. We observe that $d_n = 6$ is twice as frequent as $d_n = 2$ or $d_n = 4$. The reason is as follows: if $d_n = 2$, then we must have $3 \nmid p_n$ and $3 \nmid p_n + 2$, thus $p_n \equiv 2 \pmod{3}$, while if $d_n = 6$, then $3 \nmid p_n$ assures automatically $3 \nmid p_n + 6$, whether $p_n \equiv 1$ or $\equiv 2$. Therefore, $d_n = 6$ is twice as probable as $d_n = 2$ or 4. Similar discussions can be applied to $d_n = 2k$, and we see that $d_n = 2k$ is $c_k$ times as probable as $d_n = 2$, where

$$
c_k = \prod_{p|k} \frac{p - 1}{p - 2},
$$

the product being taken over all odd primes dividing $k$.

How should we include this effect in the exponential distribution? Suppose that we are challenging to some trial with success probability $\alpha$. The probability that we succeed for the first time after $n$ trials is $\alpha(1 - \alpha)^n$. If a player is allowed to try twice after other person's $n$ trials, the probability that he becomes the first success is $\alpha(1 - \alpha)^n + \alpha(1 - \alpha)^{n+1}$. This consideration suggests that in the case of the gap distribution of prime numbers, in
order to include the above effect in the exponential distribution, it will suffice to take the time interval as $c_k \Delta t$ instead of $\Delta t$.

Thus, we obtain the following "theoretical" distribution of $d_n$.

\begin{equation}
\text{Prob}(d_n = 2k) = \exp(-\alpha t_{k-1}) - \exp(-\alpha t_k)
\end{equation}

where $t_k = \sum_{j=1}^{k} c_j$, $\alpha$: some constant $> 0$.

Now, we must determine the value of $\alpha$. The prime number theorem implies that the expectation value $<d_n>$ of $d_n$ under our "theoretical" distribution should be of the order of $\log n$. From (2.3), we have

\begin{equation}
<d_n> = 2\alpha \int_{0}^{\infty} k(t)e^{-\alpha t}dt
\end{equation}

where $k(t) = k$ for $t_{k-1} < t \leq t_k$.

We shall evaluate the order of $k(t)$, or equivalently the order of $t_k$. Again applying rough discussions, we shall suppose $t_k \sim ck$. Here $c$ is the mean of $c_j$. For a given $p$, $p \nmid j$ is $(p-1)$-times as probable as $p \mid j$, so that

$$c = \prod_{p} \left( \frac{p-1}{p} + \frac{1}{p} \frac{p-1}{p-2} \right) = \prod_{p} \frac{(p-1)^2}{p(p-2)},$$

the product being taken over all odd primes.

(The discussions of this part can be made rigorous, namely we can prove that

$$\lim_{k \to \infty} \frac{t_k}{k} = c.\)$$

From $t_k \sim ck$, we have $k(t) \sim \frac{t}{c}$, so that

$$<d_n> \sim \frac{2\alpha}{c} \int_{0}^{\infty} te^{-\alpha t}dt = \frac{2}{c\alpha}.$$

Combining this with $<d_n> \sim \log n$, we have $\alpha \sim \frac{2}{c \log n}$.

Note that

$$\frac{1}{c} = \prod_{p} \frac{p(p-2)}{(p-1)^2} = \prod_{p} \left( 1 - \frac{1}{(p-1)^2} \right).$$

Thus we have determined our "theoretical" distribution as follows:

\begin{equation}
\text{Prob}(d_n = 2k) = \exp(-\alpha_n t_{k-1}) - \exp(-\alpha_n t_k),
\end{equation}
\[ t_k = \sum_{j=1}^{k} c_j, \quad c_j = \prod_{p|j} \frac{p-1}{p-2}, \]
\[ \alpha_n = \frac{2c}{\log n}, \quad c = \prod_p \left(1 - \frac{1}{(p-1)^2}\right). \]

3. Conjectures (1.2) and (1.3)

Let \( X_n(n=1,2,\ldots) \) be mutually independent random variables whose distributions are given by (2.5), replacing \( d_n \) with \( X_n \). In this situation, we shall prove that both conjectures (1.2) and (1.3) are true with probability 1.

Theorem 1
\[ \varlimsup \frac{X_n}{(\log n)^2} = 1 \] almost surely.

Proof
Since \( \text{Prob}(X_n > 2k) = \exp(-\alpha_n t_k) \), Borel-Cantelli’s lemma implies that
\[ \text{Prob}(\varlimsup \frac{X_n}{2k(n)} \geq 1) = 1 \text{ if } \sum_n \exp(-\alpha_n t_{k(n)}) = \infty \]
\[ \text{Prob}(\varlimsup \frac{X_n}{2k(n)} \leq 1) = 1 \text{ if } \sum_n \exp(-\alpha_n t_{k(n)}) < \infty. \]

Put \( k(n) = [\beta(\log n)^2] \), where \( \beta > 0 \) and \([\ ]\) is Gauss’ symbol. Since \( t_k \sim \frac{1}{c} k \), we have \( \alpha_n t_{k(n)} \sim \frac{2c}{\log n} \frac{1}{c} \beta(\log n)^2 = 2\beta(\log n) \) so that \( \exp(-\alpha_n t_{k(n)}) \sim n^{-2\beta} \). Thus if \( \beta < \frac{1}{2} \), then \( \varliminf \frac{X_n}{(\log n)^2} \geq 2\beta \) almost surely, and if \( \beta > \frac{1}{2} \), then \( \varlimsup \frac{X_n}{(\log n)^2} \leq 2\beta \) almost surely.

Combining these, we have \( \varlimsup \frac{X_n}{(\log n)^2} = 1 \) almost surely.

Theorem 2 (Probabilistic version of prime number theorem).

For \( x > 0 \), let \( \pi(x) \) be a random variable defined by \( \pi(x) = \text{Max}\{n|\sum_{k=1}^{n} X_k \leq x\} \). Then \( \pi(x) \sim \frac{x}{\log x} \) almost surely.
\( \pi(x) \sim \frac{x}{\log x} \) is equivalent to \( \sum_{k=1}^{n} X_k \sim n \log n \).

Proof

Put \( Y_n = \frac{X_n}{\log n} \), then \( Y_n(n = 1, 2, \ldots) \) are mutually independent random variables whose means and variances are bounded. So we can apply the strong law of large numbers. Since \( < Y_n > \sim 1 \), we have \( \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = 1 \) almost surely. But \( \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = 1 \) implies \( \sum_{k=1}^{n} X_k \sim n \log n \) as proved below.

Since \( Y_k = \frac{X_k}{\log k} \geq \frac{X_k}{\log n} \) for \( k \leq n \), we have \( \frac{1}{n} \sum_{k=1}^{n} Y_k \geq \frac{1}{n} \sum_{k=1}^{n} X_k \), so that

\[ \lim_{n \to \infty} \frac{1}{n \log n} \sum_{k=1}^{n} X_k \leq 1. \]

On the other hand, since \( Y_k \leq \frac{X_k}{\log n} \) for \( k \geq n \), we have

\[ \sum_{k=[n^\alpha]}^{n} Y_k \leq \frac{1}{\log [n^\alpha]} \sum_{k=1}^{n} X_k \]

for any \( 0 < \alpha < 1 \). The left hand side is equal to \( \sum_{k=1}^{n} Y_k - \sum_{k=1}^{[n^\alpha]-1} Y_k \), so that \( \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = 1 \), thus we have

\[ \lim_{n \to \infty} \frac{1}{\alpha n \log n} \sum_{k=1}^{n} Y_k \geq 1. \]

Letting \( \alpha \to 1 \), we obtain the desired result.

Theorem 3

For \( x > 0 \), let \( \pi_2(x) \) be a random variable defined by \( \pi_2(x) = \# \{ n \mid n \leq \pi(x), X_n = 2 \} \). Then \( \pi_2(x) \sim \frac{2cx}{(\log x)^2} \) almost surely.

Proof

Put

\[ Y_n = \begin{cases} 0, & \text{if } X_n \neq 2; \\ [1 - \exp(-\alpha_n)]^{-1}, & \text{if } X_n = 2. \end{cases} \]

Then \( Y_n(n = 1, 2, \ldots) \) are mutually independent random variables with means 1. Since \( < Y_n^2 > = [1 - \exp(-\alpha_n)]^{-1} \sim \frac{\log n}{2c} \), we can apply the strong law of large numbers to obtain \( \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = 1 \) almost surely. But \( \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = 1 \) implies \( \pi_2(x) \sim \frac{2cx}{(\log x)^2} \) as proved below.
Since $\lim_{x \to \infty} \pi(x) = \infty$ almost surely, we have
\[
\lim_{x \to \infty} \frac{1}{\pi(x)} \sum_{k \leq \pi(x) \atop X_k = 2} \left[ 1 - \exp \left( -\frac{2c}{\log k} \right) \right]^{-1} = 1 \text{ almost surely.}
\]

This can be rewritten as
\[
\lim_{x \to \infty} \frac{1}{\pi(x)} \int_0^x \left[ 1 - \exp \left( -\frac{2c}{\log \pi(t)} \right) \right]^{-1} d\pi_2(t) = 1
\]

Since $\pi(x) \sim \frac{x}{\log x}$ almost surely, we have
\[
\lim_{x \to \infty} \frac{\log x}{x} \int_0^x \frac{\log t}{2c} d\pi_2(t) = 1 \text{ almost surely.}
\]

Replacing $\log t$ with $\log x$, we get
\[
\lim \frac{(\log x)^2}{2cx} \int_0^x d\pi_2(t) = \lim \frac{(\log x)^2}{2cx} \pi_2(x) = 1.
\]

Replacing $\log t$ with $\log x^\alpha$, we get
\[
\int_{x^\alpha}^x \frac{\log t}{2c} d\pi_2(t) \geq \frac{\alpha \log x}{2c} (\pi_2(x) - \pi_2(x^\alpha)).
\]

The left hand side is equal to
\[
\int_0^x \frac{\log t}{2c} d\pi_2(t) - \int_0^{x^\alpha} \frac{\log t}{2c} d\pi_2(t),
\]
so that $\sim \frac{x}{\log x}$, thus we get
\[
\lim \frac{\alpha (\log x)^2}{2cx} (\pi_2(x) - \pi_2(x^\alpha)) \leq 0.
\]

Since $\pi_2(x) \leq \pi(x) \leq \frac{x}{2}$, we have $\lim_{x \to \infty} \frac{(\log x)^2}{x} \pi_2(x^\alpha) = 0$ for $\alpha < 1$.

Therefore $\lim \frac{\alpha (\log x)^2}{2cx} \pi_2(x) \leq 1$. Letting $\alpha \to 1$, we obtain the desired result.

4. Comparison with the actual distribution
In the following Table 1, $\pi_{2k}(x) = \#\{n| p_{n+1} \leq x, d_n = 2k\}$ is given for $x = 10^3, 10^4, 10^5, 10^6, 10^7$ and $10^8$. This is obtained by determining all prime numbers below $x$.

The corresponding expected value $\overline{\pi_{2k}(x)}$ under our "theoretical" distribution is given by

\[(4.1) \quad \overline{\pi_{2k}(x)} = \sum_{n=1}^{\pi(x)} \left[ \exp \left( -\frac{2ct_{k-1}}{\log n} \right) - \exp \left( -\frac{2ct_k}{\log n} \right) \right]. \]

But since the derivation of the distribution (2.5) is not so rigorous, it does not seem necessary to carry out this complicated summation to check the validity of our "theoretical" distribution. Instead, we shall assume that all $d_n(1 \leq n \leq \pi(x))$ follow the same distribution as that of $n = \pi(x)/2$, thus we get

\[(4.2) \quad \overline{\pi_{2k}(x)} = \pi(x) \left[ \exp \left( -\alpha t_{k-1} \right) - \exp \left( -\alpha t_k \right) \right], \]

where $\alpha = \frac{2c}{\log(\pi(x)/2)}$.

Hereafter $\overline{\pi_{2k}(x)}$ means the right hand side of (4.2). In the Table 1, we shall use the same notation $\overline{\pi_{2k}(x)}$ to denote its integral approximation, namely $\left[ \overline{\pi_{2k}(x)} + 0.5 \right]$. 
Table 1, Actual and Expected Number of Gaps.

$x = 10^3$

<table>
<thead>
<tr>
<th>$2k$</th>
<th>$\pi_{2k}(x)$</th>
<th>$\overline{\pi_{2k}(x)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>35</td>
<td>43</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>44</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
$x = 10^4$

<table>
<thead>
<tr>
<th>$2k$</th>
<th>$\pi_{2k}(x)$</th>
<th>$\overline{\pi_{2k}(x)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>205</td>
<td>228</td>
</tr>
<tr>
<td>4</td>
<td>202</td>
<td>186</td>
</tr>
<tr>
<td>6</td>
<td>299</td>
<td>275</td>
</tr>
<tr>
<td>8</td>
<td>101</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>119</td>
<td>105</td>
</tr>
<tr>
<td>12</td>
<td>105</td>
<td>113</td>
</tr>
<tr>
<td>14</td>
<td>54</td>
<td>48</td>
</tr>
<tr>
<td>16</td>
<td>33</td>
<td>32</td>
</tr>
<tr>
<td>18</td>
<td>40</td>
<td>48</td>
</tr>
<tr>
<td>20</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>22</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>24</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>26</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>28</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>30</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>36</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>38</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
\( x = 10^5 \)

<table>
<thead>
<tr>
<th>( 2k )</th>
<th>( \pi_{2k}(x) )</th>
<th>( \overline{\pi_{2k}(x)} )</th>
<th>( 2k )</th>
<th>( \pi_{2k}(x) )</th>
<th>( \overline{\pi_{2k}(x)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1224</td>
<td>1384</td>
<td>42</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>1215</td>
<td>1184</td>
<td>44</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>1940</td>
<td>1880</td>
<td>46</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>773</td>
<td>742</td>
<td>48</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>916</td>
<td>826</td>
<td>50</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>12</td>
<td>964</td>
<td>957</td>
<td>52</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>484</td>
<td>447</td>
<td>54</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>16</td>
<td>339</td>
<td>313</td>
<td>56</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>514</td>
<td>498</td>
<td>58</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>238</td>
<td>255</td>
<td>60</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>22</td>
<td>223</td>
<td>176</td>
<td>62</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>206</td>
<td>249</td>
<td>64</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>88</td>
<td>106</td>
<td>66</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>28</td>
<td>98</td>
<td>98</td>
<td>68</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>146</td>
<td>162</td>
<td>70</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>32</td>
<td>32</td>
<td>45</td>
<td>72</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>33</td>
<td>41</td>
<td>74</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>54</td>
<td>61</td>
<td>76</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>19</td>
<td>25</td>
<td>78</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>28</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\[ x = 10^6 \]

<table>
<thead>
<tr>
<th>(2k)</th>
<th>(\pi_{2k}(x))</th>
<th>(\overline{\pi_{2k}(x)})</th>
<th>(2k)</th>
<th>(\pi_{2k}(x))</th>
<th>(\overline{\pi_{2k}(x)})</th>
<th>(2k)</th>
<th>(\pi_{2k}(x))</th>
<th>(\overline{\pi_{2k}(x)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8169</td>
<td>9211</td>
<td>52</td>
<td>77</td>
<td>108</td>
<td>102</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>8143</td>
<td>8130</td>
<td>54</td>
<td>140</td>
<td>163</td>
<td>104</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>13549</td>
<td>13511</td>
<td>56</td>
<td>53</td>
<td>80</td>
<td>106</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>5569</td>
<td>5591</td>
<td>58</td>
<td>54</td>
<td>60</td>
<td>108</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>7079</td>
<td>6448</td>
<td>60</td>
<td>96</td>
<td>123</td>
<td>110</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>8005</td>
<td>7866</td>
<td>62</td>
<td>16</td>
<td>38</td>
<td>112</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>4233</td>
<td>3859</td>
<td>64</td>
<td>24</td>
<td>32</td>
<td>114</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>2881</td>
<td>2802</td>
<td>66</td>
<td>48</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>4909</td>
<td>4657</td>
<td>68</td>
<td>13</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2401</td>
<td>2518</td>
<td>70</td>
<td>22</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>2172</td>
<td>1801</td>
<td>72</td>
<td>13</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>2682</td>
<td>2674</td>
<td>74</td>
<td>12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>1175</td>
<td>1200</td>
<td>76</td>
<td>6</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>1234</td>
<td>1145</td>
<td>78</td>
<td>13</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1914</td>
<td>2006</td>
<td>80</td>
<td>3</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>550</td>
<td>596</td>
<td>82</td>
<td>5</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>557</td>
<td>559</td>
<td>84</td>
<td>6</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>767</td>
<td>867</td>
<td>86</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>330</td>
<td>378</td>
<td>88</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>424</td>
<td>411</td>
<td>90</td>
<td>4</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>476</td>
<td>587</td>
<td>92</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>202</td>
<td>218</td>
<td>94</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>155</td>
<td>179</td>
<td>96</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>196</td>
<td>284</td>
<td>98</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>106</td>
<td>153</td>
<td>100</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
$x = 10^7$

<table>
<thead>
<tr>
<th>2k</th>
<th>$\pi_{2k}(x)$</th>
<th>$\overline{\pi_{2k}(x)}$</th>
<th>2k</th>
<th>$\pi_{2k}(x)$</th>
<th>$\overline{\pi_{2k}(x)}$</th>
<th>2k</th>
<th>$\pi_{2k}(x)$</th>
<th>$\overline{\pi_{2k}(x)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>58980</td>
<td>65554</td>
<td>56</td>
<td>1072</td>
<td>1304</td>
<td>110</td>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>58621</td>
<td>59087</td>
<td>58</td>
<td>1058</td>
<td>1003</td>
<td>112</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>99987</td>
<td>101265</td>
<td>60</td>
<td>1834</td>
<td>2135</td>
<td>114</td>
<td>11</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>42352</td>
<td>43270</td>
<td>62</td>
<td>543</td>
<td>681</td>
<td>116</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>54431</td>
<td>51129</td>
<td>64</td>
<td>559</td>
<td>593</td>
<td>118</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>65513</td>
<td>64568</td>
<td>66</td>
<td>973</td>
<td>1116</td>
<td>120</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>14</td>
<td>35394</td>
<td>32772</td>
<td>68</td>
<td>358</td>
<td>451</td>
<td>122</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>16</td>
<td>25099</td>
<td>24357</td>
<td>70</td>
<td>524</td>
<td>589</td>
<td>124</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>18</td>
<td>43851</td>
<td>41744</td>
<td>72</td>
<td>468</td>
<td>611</td>
<td>126</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>20</td>
<td>22084</td>
<td>23383</td>
<td>74</td>
<td>218</td>
<td>268</td>
<td>128</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>22</td>
<td>19451</td>
<td>17159</td>
<td>76</td>
<td>194</td>
<td>248</td>
<td>130</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>24</td>
<td>27170</td>
<td>26311</td>
<td>78</td>
<td>362</td>
<td>432</td>
<td>132</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>26</td>
<td>12249</td>
<td>12208</td>
<td>80</td>
<td>165</td>
<td>220</td>
<td>134</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>28</td>
<td>13255</td>
<td>11924</td>
<td>82</td>
<td>100</td>
<td>150</td>
<td>136</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>21741</td>
<td>21733</td>
<td>84</td>
<td>247</td>
<td>294</td>
<td>138</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>32</td>
<td>6364</td>
<td>6719</td>
<td>86</td>
<td>66</td>
<td>105</td>
<td>140</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>34</td>
<td>6721</td>
<td>6438</td>
<td>88</td>
<td>71</td>
<td>102</td>
<td>142</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>36</td>
<td>10194</td>
<td>10307</td>
<td>90</td>
<td>141</td>
<td>201</td>
<td>144</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>38</td>
<td>4498</td>
<td>4649</td>
<td>92</td>
<td>37</td>
<td>65</td>
<td>146</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>40</td>
<td>5318</td>
<td>5172</td>
<td>94</td>
<td>39</td>
<td>57</td>
<td>148</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>42</td>
<td>7180</td>
<td>7683</td>
<td>96</td>
<td>65</td>
<td>95</td>
<td>150</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>44</td>
<td>2779</td>
<td>2958</td>
<td>98</td>
<td>29</td>
<td>48</td>
<td>152</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>46</td>
<td>2326</td>
<td>2493</td>
<td>100</td>
<td>36</td>
<td>47</td>
<td>154</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>48</td>
<td>3784</td>
<td>4069</td>
<td>102</td>
<td>34</td>
<td>63</td>
<td>156</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>2048</td>
<td>2279</td>
<td>104</td>
<td>21</td>
<td>27</td>
<td>158</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>52</td>
<td>1449</td>
<td>1644</td>
<td>106</td>
<td>12</td>
<td>23</td>
<td>160</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>54</td>
<td>2403</td>
<td>2570</td>
<td>108</td>
<td>26</td>
<td>38</td>
<td>162</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
\[ x = 10^8 \]

\begin{tabular}{|c|c|c|c|c|}
\hline
2k & \( \pi_{2k}(x) \) & \( \overline{\pi_{2k}(x)} \) & 2k & \( \pi_{2k}(x) \) & \( \overline{\pi_{2k}(x)} \) \\
\hline
2 & 440312 & 489399 & 56 & 16595 & 17659  \\
4 & 440257 & 447828 & 58 & 14611 & 13817  \\
6 & 768752 & 784766 & 60 & 28439 & 30204  \\
8 & 334180 & 343127 & 62 & 8496 & 9922  \\
10 & 430016 & 412597 & 64 & 8823 & 8763  \\
12 & 538382 & 534172 & 66 & 15579 & 16900  \\
14 & 293201 & 277833 & 68 & 6200 & 7002  \\
16 & 215804 & 209958 & 70 & 8813 & 9334  \\
18 & 384738 & 367927 & 72 & 8453 & 9950  \\
20 & 202922 & 211398 & 74 & 4316 & 4469  \\
22 & 175945 & 158024 & 76 & 3580 & 4194  \\
24 & 257548 & 247981 & 78 & 6790 & 7492  \\
26 & 119465 & 117814 & 80 & 3281 & 3913  \\
28 & 129567 & 117075 & 82 & 2362 & 2710  \\
30 & 222847 & 219546 & 84 & 4668 & 5456  \\
32 & 68291 & 69824 & 86 & 1597 & 1997  \\
34 & 71248 & 67955 & 88 & 1637 & 1971  \\
36 & 114028 & 111305 & 90 & 3337 & 4007  \\
38 & 51756 & 51397 & 92 & 1083 & 1332  \\
40 & 60761 & 58215 & 94 & 971 & 1186  \\
42 & 86637 & 88902 & 96 & 1641 & 2031  \\
44 & 34881 & 35167 & 98 & 851 & 1056  \\
46 & 29327 & 30126 & 100 & 878 & 1049  \\
48 & 49824 & 50285 & 102 & 1059 & 1440  \\
50 & 27522 & 28892 & 104 & 494 & 638  \\
52 & 20595 & 21223 & 106 & 404 & 543  \\
54 & 33593 & 33953 & 108 & 711 & 932  \\
\hline
\end{tabular}
<table>
<thead>
<tr>
<th>$2k$</th>
<th>$\pi_{2k}(x)$</th>
<th>$\overline{\pi_{2k}(x)}$</th>
<th>$2k$</th>
<th>$\pi_{2k}(x)$</th>
<th>$\overline{\pi_{2k}(x)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>454</td>
<td>591</td>
<td>166</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>112</td>
<td>330</td>
<td>425</td>
<td>168</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>114</td>
<td>487</td>
<td>648</td>
<td>170</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>116</td>
<td>191</td>
<td>276</td>
<td>172</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>118</td>
<td>181</td>
<td>247</td>
<td>174</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>120</td>
<td>433</td>
<td>550</td>
<td>176</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>122</td>
<td>131</td>
<td>178</td>
<td>178</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>124</td>
<td>145</td>
<td>165</td>
<td>180</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>126</td>
<td>204</td>
<td>329</td>
<td>182</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>128</td>
<td>76</td>
<td>118</td>
<td>184</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>130</td>
<td>78</td>
<td>154</td>
<td>186</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>132</td>
<td>132</td>
<td>200</td>
<td>188</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>134</td>
<td>50</td>
<td>79</td>
<td>190</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>136</td>
<td>40</td>
<td>76</td>
<td>192</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>138</td>
<td>93</td>
<td>129</td>
<td>194</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>140</td>
<td>57</td>
<td>84</td>
<td>196</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>142</td>
<td>30</td>
<td>47</td>
<td>198</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>144</td>
<td>51</td>
<td>82</td>
<td>200</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>146</td>
<td>22</td>
<td>36</td>
<td>202</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>148</td>
<td>34</td>
<td>33</td>
<td>204</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>150</td>
<td>37</td>
<td>74</td>
<td>206</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>152</td>
<td>20</td>
<td>25</td>
<td>208</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>154</td>
<td>13</td>
<td>28</td>
<td>210</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>156</td>
<td>23</td>
<td>39</td>
<td>212</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>158</td>
<td>10</td>
<td>16</td>
<td>214</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>160</td>
<td>11</td>
<td>19</td>
<td>216</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>162</td>
<td>8</td>
<td>24</td>
<td>218</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>164</td>
<td>5</td>
<td>11</td>
<td>220</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
From the Table 1, we observe that the expected number $\overline{\pi_{2k}(x)}$ is in good accordance with the actual one $\pi_{2k}(x)$, at least qualitatively. Though the accordance is not good numerically for some $k$, the tendencies of both distributions coincide, thus we ascertain the exponential feature of the actual distribution of $d_n$. The number of twin primes $\pi_2(x)$ is about 10% smaller than $\overline{\pi_2(x)}$. The number of twin primes $\pi_2(x)$ is about 10% smaller than $\overline{\pi_2(x)}$. The maximum gap $\max_{n \leq \pi(x)} d_n$, which is given in the Table 2, shows remarkably good accordance, intensifying the belief that the conjecture (1.2) should be true.

<table>
<thead>
<tr>
<th>$x$</th>
<th>actual</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^3$</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>$10^4$</td>
<td>36</td>
<td>42</td>
</tr>
<tr>
<td>$10^5$</td>
<td>72</td>
<td>78</td>
</tr>
<tr>
<td>$10^6$</td>
<td>114</td>
<td>114</td>
</tr>
<tr>
<td>$10^7$</td>
<td>154</td>
<td>162</td>
</tr>
<tr>
<td>$10^8$</td>
<td>220</td>
<td>216</td>
</tr>
<tr>
<td>$10^9$</td>
<td>282</td>
<td>282</td>
</tr>
<tr>
<td>$10^{10}$</td>
<td>354</td>
<td>360</td>
</tr>
<tr>
<td>$10^{11}$</td>
<td>464</td>
<td>450</td>
</tr>
<tr>
<td>$10^{12}$</td>
<td>540</td>
<td>546</td>
</tr>
<tr>
<td>$10^{13}$</td>
<td>674</td>
<td>660</td>
</tr>
<tr>
<td>$10^{14}$</td>
<td>804</td>
<td>762</td>
</tr>
</tbody>
</table>

(Actual value for $x \geq 10^9$ is cited from [8]. Expected value is computed by (4.2).)
Our “theoretical” distribution (2.5) is rather simple. More elaborate and complicated “theoretical” distribution may be useful to obtain better accordance. From the Table 1, we observe that for small $k$ and large $k$, we have $\pi_{2k}(x) \geq \overline{\pi_{2k}(x)}$, while for the medium value of $k$, we have $\pi_{2k}(x) \geq \overline{\pi_{2k}(x)}$. This does not seem to be accidental, and suggests that we should replace the exponential distribution with some other one to get better accordance.

The following Figure 1 is the graphs of $\pi_{2k}(x)/\overline{\pi_{2k}(x)}$ as the functions of $\alpha(x)^2 t_k$ for $x = 10^5, 10^6, 10^7$ and $10^8$, where $\alpha(x) = 2c/\log(\pi(x)/2)$

![Figure 1. $\pi_{2k}(x)/\overline{\pi_{2k}(x)}$ as functions of $\alpha^2 t_k$.](image)

We observe that these graphs seem to converge to some curve as $x \to \infty$, but the limit is apparently not the constant 1. Note also that the abscissa is $\alpha^2 t_k$, not $\alpha t_k$. The reason for this deviation is not known, but some effect depending on $(\log \pi(x))^2$ seem to exist.
REFERENCES


Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606, JAPAN

(Y. Yamasaki)

Department of Mathematics, Kyoto University, Kyoto 606, JAPAN (A. Yamasaki)