

On the Square-free Numbers in Shifted Primes

Zerui Tan

The High School Attached to The Hunan Normal University, China

Advisor : Yongxing Cheng

November 29, 2014

On the Square-free Numbers in Shifted Primes

Zerui Tan

The High School Attached to The Hunan Normal University

November 29, 2014

Abstract

For a fixed positive integer k , let $\mathcal{Q}_k(x)$ be the number of primes not exceeding x and greater than k such that $p - k$ is a square-free integer. In this paper, we prove that for any $A > 2$, we have

$$\mathcal{Q}_k(x) = \prod_{p|k} \left(1 + \frac{1}{p^2 - p - 1}\right) \prod_p \left(1 - \frac{1}{p(p-1)}\right) \text{li}x + O\left(\frac{x}{(\log x)^A}\right)$$

as $x \rightarrow +\infty$. Where the implied constant depends only on A and k .

⁰Keywords: shifted primes, square-free numbers

1 Introduction

Number theory is a branch of pure mathematics which studies the properties of integers. An integer is called square-free if it is not divisible by any perfect square except 1. A positive integer is called a prime number, if it has no more factors other than 1 and itself, with the convention that 1 is not a prime. In number theory, prime numbers and the properties of objects closely related to integers or defined as generalizations of the integers are studied. Prime numbers play a crucial role in number theory, and there are still a lot of unsolved problems surrounding them.

In [6], page 59, there is a table which presents all the primes $2 < p \leq 5000$, their least primitive root and the factorization of $p - 1$. Part of the table is here (unnecessary columns are removed).

p	$p - 1$	p	$p - 1$
3	2	37	$2^2 \times 3^2$
5	2^2	41	$2^3 \times 5$
7	2×3	43	$2 \times 3 \times 7$
11	2×5	47	2×23
13	$2^2 \times 3$	53	$2^2 \times 13$
17	2^4	59	2×29
19	2×3^2	61	$2^2 \times 3 \times 5$
23	2×11	67	$2 \times 3 \times 11$
29	$2^2 \times 7$	71	$2 \times 5 \times 7$
31	$2 \times 3 \times 5$	73	$2^3 \times 3^2$

By observation, we find that some of the prime numbers p like 7, 11, 23, their corresponding factorization of $p - 1$ only contains distinct prime factors, namely, their corresponding $p - 1$ are square-free numbers. Some primes do not have this property. So we are eager to know,

How many such primes are there in the interval $[1, x]$?

Do they have a positive density among all primes?

And what about the case $p - k$ for any given positive integer $k > 0$?

The book [6] did not mention these questions. Then we started searching these questions on the Internet (including Baidu, Google, etc.) in order to find out an answer. However, we did not find any of them. So we decided to study this question.

Definition 1. For a fixed positive integer k , we define $\mathcal{Q}_k(x)$ to be the number of primes not exceeding x and greater than k such that $p - k$ is square-free.

In this article, we prove that

Theorem 1. For any $A > 2$, we have

$$\mathcal{Q}_k(x) = C_k \operatorname{li} x + O\left(\frac{x}{(\log x)^A}\right) \quad (x \rightarrow +\infty)$$

where C_k is a constant depends on k , which can be written as

$$C_k = \prod_{p|k} \left(1 + \frac{1}{p^2 - p - 1}\right) \prod_p \left(1 - \frac{1}{p(p-1)}\right)$$

and the implied constant only depends on A and k .

The function $\text{li } x$ is the integral

$$\text{li } x = \int_2^x \frac{dt}{\log t}$$

and $f(x) = O(g(x))$ denotes $|f(x)| \leq C|g(x)|$ for some positive constant C . Notice that the constant C_k and C_l will have the same value if k has the same prime factors as l has, for example we have $C_6 = C_{18}$.

Let $\varphi(n)$ be the number of positive integers not exceeding n that are relatively prime with n , we have that

$$\varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$$

where p runs over all distinct prime factors of n . The Möbius function $\mu(n)$ is defined as

$$\mu(n) = \begin{cases} 1 & n = 1 \\ 0 & \text{there is a prime } p \text{ that } p^2 | n \\ (-1)^{\omega(n)} & \text{otherwise} \end{cases}$$

where $\omega(n)$ is the number of distinct prime factors of n . To prove the theorem, we need the following basic ideas.

$$\sum_{d|n} \mu(d) = \delta(n)$$

where

$$\delta(n) := \begin{cases} 1 & n = 1 \\ 0 & \text{otherwise} \end{cases}.$$

Since any number $n \in \mathbb{N}$ can be written as the form $n = a^2 b$ uniquely, where b is a square-free number, n is square-free if and only if $a = 1$, it follows that

$$\mu^2(n) = \sum_{d|a} \mu(d) = \sum_{d^2|n} \mu(d)$$

where the sum $\sum_{d^2|n}$ is taken over all positive divisors d such that $d^2|n$. A well-known theorem due to Bombieri and A.I. Vinogradov, the Mertens' formula and the crude lower bound of the Euler function $\varphi(d)$ are also needed.

Theorem A. (Bombieri-Vinogradov) *Let $A > 0$, There is some constant $B = B(A)$ such that*

$$\sum_{q \leq x^{1/2}/(\log x)^B} \max_{y \leq x} \max_{(a,q)=1} \left| \pi(y; q, a) - \frac{\text{li}(y)}{\varphi(q)} \right| \ll \frac{x}{(\log x)^A} \quad (x \geq 2)$$

where the implied constant depends only on A .

Where $f \ll g$ means there is a positive constant C such that $|f| \leq C|g|$, the function $\pi(x; q, l)$ is defined as

$$\pi(x; q, l) = \sum_{\substack{p \leq x \\ p \equiv l \pmod{q}}} 1.$$

For the proof of the theorem, see [4].

Theorem B. (Mertens' formula) *For $x \geq 2$,*

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) = \frac{e^{-\gamma}}{\log x} \left(1 + O\left(\frac{1}{\log x}\right)\right)$$

where γ is the Euler-Mascheroni constant.

The Mertens formula is a classical result, and the proof can be found in [1].

Lemma 1.1. *If $d > 15$ is a square-free number, then we have*

$$\varphi(d) \gg \frac{d}{\log \log d}$$

where $c_2 > 0$ is some constant.

Proof. Let p_n be the n -th prime, by Mertens' formula we have that

$$\begin{aligned} \varphi(d) &= d \prod_{p|d} \left(1 - \frac{1}{p}\right) \geq d \prod_{p \leq p_{\omega(d)}} \left(1 - \frac{1}{p}\right) \\ &= \frac{de^{-\gamma}}{\log p_{\omega(d)}} \left(1 + O\left(\frac{1}{\log p_{\omega(d)}}\right)\right) \end{aligned}$$

since $\log p_N \ll \log N$ and $\omega(d) \ll \log d$, it follows that

$$\log p_{\omega(d)} \ll \log \log d.$$

Therefore we can get

$$\varphi(d) \gg \frac{d}{\log \log d}.$$

□

2 Proof of Theorem 1

Now we have that

$$\mathcal{Q}_k(x) = \sum_{k < p \leq x} \mu^2(p - k).$$

By the formula

$$\mu^2(n) = \sum_{d^2 | n} \mu(d)$$

and by changing the order of summation we obtain

$$\begin{aligned} \mathcal{Q}_k(x) &= \sum_{k < p \leq x} \sum_{d^2 | p - k} \mu(d) = \sum_{d < \sqrt{x}} \sum_{\substack{k < p \leq x \\ p \equiv k \pmod{d^2}}} \mu(d) + O\left(\sum_{\sqrt{x-k} < d < \sqrt{x}} \sum_{\substack{k < p \leq x \\ p \equiv k \pmod{d^2}}} 1\right) \\ &= \sum_{d < \sqrt{x}} \sum_{\substack{k < p \leq x \\ p \equiv k \pmod{d^2}}} \mu(d) + O\left(\frac{\sqrt{x}}{\log x}\right). \end{aligned}$$

We can change the summation condition $k < p \leq x$ to $p \leq x$, the error we get is $O(\sqrt{x})$, so

$$\mathcal{Q}_k(x) = \sum_{d < \sqrt{x}} \mu(d) \pi(x; d^2, k) + O(\sqrt{x}).$$

For the sake of simplicity, let $\mathcal{L} = \log x$. We divide the sum into two pieces by the intervals $[1, x^\alpha \mathcal{L}^{-B_0}] \cup [x^\alpha \mathcal{L}^{-B_0}, x^{1/2}]$, where $0 < \alpha < 1/2$ and $B_0 > 0$ are constants, their value will be determined later. We write

$$\mathcal{Q}_k(x) = S_\alpha + S_{\alpha, 1/2} + O(\sqrt{x})$$

where

$$S_\alpha = \sum_{d < x^\alpha \mathcal{L}^{-B_0}} \mu(d) \pi(x; d^2, k)$$

and

$$S_{\alpha, 1/2} = \sum_{x^\alpha \mathcal{L}^{-B_0} \leq d < x^{1/2}} \mu(d) \pi(x; d^2, k).$$

For the first sum, we can write $\pi(x; d^2, k)$ as the form

$$\pi(x; d^2, k) = \frac{\rho_k(d)}{\varphi(d^2)} \text{li}x + \left(\pi(x; d^2, k) - \frac{\rho_k(d)}{\varphi(d^2)} \text{li}x \right).$$

Where $\rho_k(n) := \delta((n, k))$ is the characteristic function of numbers n that are relatively prime with k . Namely, it equals to 1 if and only if n is relatively prime with k , otherwise it equals to 0. Inserting the above formula into S_α we can get

$$S_\alpha = \text{li}x \sum_{d < x^\alpha \mathcal{L}^{-B_0}} \frac{\rho_k(d) \mu(d)}{d \varphi(d)} + \sum_{d < x^\alpha \mathcal{L}^{-B_0}} \mu(d) \left(\pi(x; d^2, k) - \frac{\rho_k(d) \text{li}x}{\varphi(d^2)} \right).$$

Notice that $\mu(d) = 0$ if d is not square-free, we can assume d is square-free, in this case we have

$$\varphi(d^2) = d\varphi(d).$$

By Lemma 1.1, it is easy to check that the sum

$$\sum_{d=1}^{\infty} \frac{\rho_k(d)\mu(d)}{d\varphi(d)}$$

converges absolutely. Thus

$$\begin{aligned} S_\alpha &= \text{lix} \sum_{d < x^\alpha \mathcal{L}^{-B_0}} \frac{\rho_k(d)\mu(d)}{d\varphi(d)} + \sum_{d < x^\alpha \mathcal{L}^{-B_0}} \mu(d) \left(\pi(x; d^2, k) - \frac{\rho_k(d)\text{lix}}{\varphi(d^2)} \right) \\ &= \text{lix} \sum_{d=1}^{\infty} \frac{\rho_k(d)\mu(d)}{d\varphi(d)} + O \left(\text{lix} \sum_{d \geq x^\alpha \mathcal{L}^{-B_0}} \frac{\rho_k(d)\mu(d)}{\varphi(d^2)} \right) \\ &+ \sum_{d < x^\alpha \mathcal{L}^{-B_0}} \mu(d) \left(\pi(x; d^2, k) - \frac{\rho_k(d)}{\varphi(d^2)} \text{lix} \right) \\ &= C_k \text{lix} + O(S_1) + R_\alpha. \end{aligned}$$

Where

$$\begin{aligned} C_k &= \sum_{d=1}^{\infty} \frac{\rho_k(d)\mu(d)}{\varphi(d^2)} \\ &= \prod_p \left(1 + \frac{\rho_k(p)\mu(p)}{\varphi(p^2)} \right) \\ &= \prod_{p|k} \left(1 + \frac{1}{p^2 - p - 1} \right) \prod_p \left(1 - \frac{1}{p(p-1)} \right) \\ R_\alpha &= \sum_{d < x^\alpha \mathcal{L}^{-B_0}} \mu(d) \left(\pi(x; d^2, k) - \frac{\rho_k(d)}{\varphi(d^2)} \text{lix} \right) \end{aligned}$$

and

$$S_1 = \text{lix} \sum_{d \geq x^\alpha \mathcal{L}^{-B_0}} \frac{\rho_k(d)\mu(d)}{\varphi(d^2)}.$$

2.1 Upper bound for $S_{\alpha,1/2}$, S_1 and R_α

Now, we are going to find the upper bound for

$$S_{\alpha,1/2} = \sum_{x^\alpha \mathcal{L}^{-B_0} \leq d < x^{1/2}} \mu(d)\pi(x; d^2, k).$$

We only need to consider the trivial bound

$$\pi(x; d^2, k) \ll \frac{x}{d^2}.$$

So we obtain

$$\begin{aligned} |S_{\alpha,1/2}| &= \left| \sum_{x^\alpha \mathcal{L}^{-B_0} \leq d < x^{1/2}} \mu(d)\pi(x; d^2, k) \right| \\ &\ll x \sum_{x^\alpha \mathcal{L}^{-B_0} \leq d < x^{1/2}} d^{-2} \\ &\ll x \int_{x^\alpha \mathcal{L}^{-B_0}}^{x^{1/2}} \frac{dt}{t^2} \\ &\ll x^{1-\alpha}(\log x)^{B_0} - x^{1/2} \end{aligned}$$

provided x sufficiently large. Recall that $1/2 > \alpha > 0$, namely $|S_{\alpha,1/2}| \ll x^{1-\alpha}(\log x)^{B_0}$. As for the sum

$$S_1 = \text{lix} \sum_{n \geq x^\alpha \mathcal{L}^{-B_0}} \frac{\rho_k(n)\mu(n)}{\varphi(n^2)}$$

it is not difficult to get an upper bound, since $\varphi(d^2)$ is approximately equals to d^2 for a square-free number d .

Lemma 2.1. *For sufficiently large x , we have*

$$\text{lix} \sum_{n \geq x^\alpha \mathcal{L}^{-B_0}} \frac{\rho_k(n)\mu(n)}{\varphi(n^2)} \ll x^{1-\alpha} \mathcal{L}^{B_0}.$$

Proof. Because $\mu(n) = 0$ if n is not square-free, we can assume n is square-free. We have

$$\varphi(n^2) = n\varphi(n)$$

and

$$\sum_{n \geq x} \frac{1}{n^2} \leq \int_{x-1}^{\infty} \frac{dt}{t^2} \ll \frac{1}{x}$$

and by Lemma 1.1 we have

$$\left| \sum_{n \geq x^\alpha \mathcal{L}^{-B_0}} \frac{\rho_k(n)\mu(n)}{\varphi(n^2)} \right| \leq \sum_{n \geq x^\alpha \mathcal{L}^{-B_0}} \frac{1}{n\varphi(n)} \ll \frac{\log \log x}{x^\alpha (\log x)^{-B_0}}.$$

By the fact $\text{lix} \ll x/\log x$, we have

$$\text{lix} \left| \sum_{n \geq x^\alpha \mathcal{L}^{-B_0}} \frac{\rho_k(n)\mu(n)}{\varphi(n^2)} \right| \ll x^{1-\alpha} \log \log x (\log x)^{B_0-1}$$

which completes the proof of the lemma. □

Now we are going to consider the R_α .

$$R_\alpha = \sum_{d < x^\alpha \mathcal{L}^{-B_0}} \mu(d) \left(\pi(x; d^2, k) - \frac{\rho_k(d)}{\varphi(d^2)} \text{li}x \right)$$

where the constant $B_0 = B(A)/2$, $B(A)$ is the constant in the Bombieri-Vinogradov Theorem. Now we let $\alpha = 1/4$, by the Bombieri-Vinogradov Theorem we obtain

$$\begin{aligned} \left| R_{\frac{1}{4}} \right| &= \left| \sum_{d < x^{1/4} / (\log x)^{B_0}} \mu(d) \left(\pi(x; d^2, k) - \frac{\rho_k(d)}{\varphi(d^2)} \text{li}x \right) \right| \\ &\ll \left| \sum_{\substack{d^2 < x^{1/2} \mathcal{L}^{-2B_0} \\ \rho_k(d)=1}} \pi(x; d^2, k) - \frac{\text{li}x}{\varphi(d^2)} \right| + \left| \sum_{\substack{d^2 < x^{1/2} \mathcal{L}^{-2B_0} \\ \rho_k(d)=0}} \pi(x; d^2, k) \right| \\ &\ll \frac{x}{(\log x)^A}. \end{aligned}$$

Where the last inequality is true since the function $\pi(x; d^2, k)$ in the second sum is bounded due to the fact $(d, k) > 1$.

2.2 Completion

Combining the results we got, for a given positive integer $k > 0$, any $A > 2$, and sufficiently large x , we have

$$\begin{aligned} \mathcal{Q}_k(x) &= C_k \text{li}x + O(\sqrt{x}) + O(S_1) + S_{\frac{1}{4}, \frac{1}{2}} + R_{\frac{1}{4}} \\ &= C_k \text{li}x + O\left(x^{1/2} + x^{3/4}(\log x)^{B_0} + x^{3/4}(\log x)^{B_0} + \frac{x}{(\log x)^A}\right) \\ &= C_k \text{li}x + O\left(\frac{x}{(\log x)^A}\right). \end{aligned}$$

Inserting the expression of C_k , we can get the formula in Theorem 1

$$\mathcal{Q}_k(x) = \text{li}x \prod_{p|k} \left(1 + \frac{1}{p^2 - p - 1}\right) \prod_p \left(1 - \frac{1}{p(p-1)}\right) + O\left(\frac{x}{(\log x)^A}\right).$$

This is the result we desired. As a consequence, we have that

$$\lim_{x \rightarrow \infty} \frac{\mathcal{Q}_k(x)}{\text{li}x} = C_k > 0$$

Which means the primes p such that $p - k$ is square-free have a positive density among all primes.

3 Related Questions

Now we will discuss some interesting questions that appeared directly in the research of $\mathcal{Q}_k(x)$.

Question 1. *Is there any positive integer k , such that for sufficiently large x , we have*

$$\mathcal{Q}_k(x) < C_k \operatorname{li} x \quad ?$$

(See appendix for more details.)

It seems that the bound for $\left| R_{\frac{1}{4}} \right|$ is too crude that it lost some information. We used the Bombieri-Vinogradov Theorem to get this result. This theorem is quite important in number theory since it is essentially of the same strength with the Generalized Riemann Hypothesis (usually written as GRH.) on average, which hypothesis has a consequence that

$$\left| \pi(x; q, k) - \frac{\operatorname{li} x}{\varphi(q)} \right| \ll \sqrt{x} \log x$$

provided $(q, k) = 1$. So, if we assume GRH and $\alpha \leq \frac{1}{4}$ we can get

$$\left| \sum_{\substack{d < x^\alpha \mathcal{L}^{-B_0} \\ (k, d) = 1}} \pi(x; d^2, k) - \frac{\operatorname{li} x}{\varphi(d^2)} \right| \ll \sum_{\substack{d < x^\alpha \mathcal{L}^{-B_0} \\ (k, d) = 1}} \sqrt{x} \log x \ll x^{1/2+\alpha} \mathcal{L}^{1-B_0}$$

therefore

$$\text{all remainder terms} \ll x^{1/2+\alpha} \mathcal{L}^{1-B_0} + x^{1-\alpha} \mathcal{L}^{B_0}.$$

we know that we should choose

$$\alpha = \frac{1}{4} \quad \text{and} \quad B_0 = \frac{1}{2}$$

to get an optimal remainder term

$$\text{all remainder terms} \ll x^{3/4} \sqrt{\log x}.$$

If $\frac{1}{4} < \alpha \leq \frac{1}{2}$, we can just do as what we did before, divide the sum into two pieces by the intervals $[1, x^{1/4} \mathcal{L}^{-B_1}) \cup [x^{1/4} \mathcal{L}^{-B_1}, x^\alpha \mathcal{L}^{-B_0})$, where B_1 is some constant. Then use GRH to estimate the first sum and use an integral to bound the second, one gets

$$\left| \sum_{\substack{d < x^\alpha \mathcal{L}^{-B_0} \\ (k, d) = 1}} \pi(x; d^2, k) - \frac{\operatorname{li} x}{\varphi(d^2)} \right| \ll x^{3/4} \mathcal{L}^{1-B_1} + x^{3/4} \mathcal{L}^{B_1}$$

which is exactly the same as the case $\alpha \leq 1/4$, also notice that the above upper bound is independent with B_0 and α . The above arguments suggests the following two questions.

Question 2. *The question is, can we prove that for any $A > 0$, there is a constant $B = B(A)$ that*

$$\sum_{d < x^\alpha} \max_{y \leq x} \max_{\substack{(k,d)=1 \\ 1 \leq k < d^2}} \left| \pi(y; d^2, k) - \frac{\text{li}y}{\varphi(d^2)} \right| \ll x^{1/2+\alpha} \mathcal{L}^{-A}$$

for some $0 < \alpha \leq 1/4$ (especially for the case $\alpha = 1/4$.) without assuming GRH?

Question 3. *Is there an absolute constant B such that for sufficiently large x , we have*

$$\sum_{d \geq 1} \max_{y \leq x} \max_{\substack{(k,d)=1 \\ 1 \leq k < d^2}} \left| \pi(y; d^2, k) - \frac{\text{li}y}{\varphi(d^2)} \right| \ll x^{3/4} (\log x)^B \quad ?$$

It is not difficult to check that the above sum converges for every $x > 0$.

Remark. The above two questions can be regarded as the analogy of the Bombieri-Vinogradov Theorem. As for Question 1, in fact, we guess that it might be wrong. To answer these questions, we still need more research.

Appendix Tables of $\mathcal{Q}_k(x)$ For $k = 1, 2, 3$

We made a C++ program on our computer to calculate some numerical value of $\mathcal{Q}_k(x)$ (for $k = 1, 2, 3$ and $x \leq 10^7$). The exact value of $\mathcal{Q}_k(x)$, $C_k \text{lix}$ and their ratio $\mathcal{Q}_k(x) : (C_k \text{lix})$ are presented as follows.

x	$\mathcal{Q}_1(x)$	$C_1 \text{lix}$	$\mathcal{Q}_1(x) : (C_1 \text{lix})$
10	3	1.9148	1.5667
50	8	6.5156	1.2278
100	13	10.875	1.1954
500	40	37.676	1.0617
1000	68	66.027	1.0299
5000	255	255.50	0.99804
1×10^4	467	465.61	1.0030
5×10^4	1943	1931.7	1.0058
1×10^5	3599	3600.7	0.99953
5×10^5	15602	15559	1.0028
1×10^6	29397	29403	0.99980
5×10^6	130391	130375	1.0001
1×10^7	248518	248650	0.99947
x	$\mathcal{Q}_2(x)$	$C_2 \text{lix}$	$\mathcal{Q}_2(x) : (C_2 \text{lix})$
50	11	13.0312	0.8441
100	20	21.750	0.9195
500	74	75.352	0.9821
1000	127	132.05	0.9618
5000	506	511.00	0.99022
1×10^4	925	931.22	0.99332
5×10^4	3841	3863.4	0.99420
1×10^5	7175	7201.4	0.99639
5×10^5	31020	31118	0.99685
1×10^6	58653	58806	0.99740
5×10^6	260381	260750	0.99858
1×10^7	496848	497300	0.99909
x	$\mathcal{Q}_3(x)$	$C_3 \text{lix}$	$\mathcal{Q}_3(x) : (C_3 \text{lix})$
50	6	7.8187	0.76739
100	10	13.050	0.76628
500	41	45.211	0.90686
1000	74	79.232	0.93386
5000	295	306.60	0.96217
1×10^4	548	558.73	0.98079
5×10^4	2280	2318.0	0.98359
1×10^5	4292	4320.8	0.99333
5×10^5	18603	18671	0.99637
1×10^6	35153	35284	0.99630
5×10^6	156249	156450	0.99872
1×10^7	298075	298380	0.99898

References

- [1] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory (translated version)*, Higher Education Press, 2011
- [2] Melvyn B. Nathanson, *Elementary Methods in Number Theory*, Springer-Verlag, 1999
- [3] Kenneth Ireland and Michael Rosen, *A Classical Introduction to Modern Number Theory (Second Edition)*, Springer, 1990
- [4] Dimitris Koukoulopoulos, *Sieve Methods*, Preprint, 2012
- [5] Karatsuba, *Basic Analytic Number Theory (translated version)*, Harbin Institute of Technology Press, 2012
- [6] Luogeng Hua, *Introduction to Number Theory (in Chinese)*, Science Press, 1957