

On the subsequence of primes having prime subscripts.

Kevin A. Broughan and A. Ross Barnett

University of Waikato, Hamilton, New Zealand
E-mail: kab@waikato.ac.nz

We explore the subsequence of primes with prime subscripts, (q_n) , and derive its density and estimates for its counting function. We obtain bounds for the weighted gaps between elements of the subsequence and show that for every positive integer m there is an integer arithmetic progression $(an + b : n \in \mathbb{N})$ with at least m of the (q_n) satisfying $q_n = an + b$.

Key Words: prime-prime, prime-prime number theorem, prime-prime gaps, prime-primes in progressions.

MSC2000: 11A41, 11B05, 11B25, 11B83.

1. INTRODUCTION

There are a number of subsets of primes with a conjectured density of a constant times $x/\log^2 x$. These include the primes separated by a fixed even integer, Sophie Germain primes and the so-called “thin primes”, i.e. primes of the form $2^e q - 1$ where q is prime [3].

In order to gain some familiarity with sequences of this density we undertook an investigation of the set of primes of prime order, called here “prime-primes”, and report on the results of this investigation here. Some of the properties of this sequence are reasonably straight forward and the derivation follows that of the corresponding property for the primes themselves. Others appear to be quite deep and difficult. Of course the process of taking a prime indexed subsequence can be iterated, leading to sequences of primes of density $x/\log^3 x$, $x/\log^4 x$ etc, but these sequences are not considered here.

In Section 2 upper and lower bounds for the n^{th} prime-prime are derived, in Section 3 the prime-prime number theorem is proved with error bounds equivalent to that of the prime number theorem, in Section 4 an initial study of gaps between prime-primes is begun, and in Section 5 it is shown

that for each m there is an arithmetic progression containing m prime-primes.

DEFINITION 1.1. A **prime-indexed-prime** or **prime-prime** is a rational prime q such that when the set of all primes is written in increasing order $(p_1, p_2, \dots) = (2, 3, \dots)$, we have $q = p_n$ where n is prime also.

Here is a list of the primes up to 109 with the prime-primes in bold type so $q_1 = p_2 = 3$ and $q_{10} = p_{29} = 109$: 2, **3**, **5**, 7, **11**, 13, **17**, 19, 23, 29, **31**, 37, **41**, 43, 47, 53, **59**, 61, **67**, 71, 73, 79, **83**, 89, 97, 101, 103, 107, **109**.

DEFINITION 1.2. If $x > 0$ the number of prime-primes q up to x is given by $\pi_\pi(x) := \sum_{q \leq x} 1$. Then $\pi_\pi(x) = \pi(\pi(x)) = \sum_{n=1}^x \chi_{\mathbb{P}}(n) \cdot \chi_{\mathbb{P}}(\pi(n))$, where $\chi_{\mathbb{P}}(n) := \pi(n) - \pi(n-1)$, for $n \in \mathbb{N}$, is the characteristic function of the primes.

2. BOUNDS FOR THE SEQUENCE OF PRIME-PRIMES:

The following theorem and its corollary gives a set of useful inequalities for estimating the size of the n^{th} prime-prime and for comparing it with the n^{th} prime.

THEOREM 2.1. As $n \rightarrow \infty$:

$$\begin{aligned} q_n &< n(\log n + 2 \log \log n)(\log n + \log \log n) - n \log n + O(n \log \log n), \\ q_n &> n(\log n + 2 \log \log n)(\log n + \log \log n) - 3n \log n + O(n \log \log n). \end{aligned}$$

Proof. We use the inequalities of Rosser and Schoenfeld [14]. Namely

$$\begin{aligned} p_n &< n(\log n + \log \log n - 1/2) \text{ valid for } n \geq 20, \\ p_n &> n(\log n + \log \log n - 3/2) \text{ valid for } n \geq 2, \end{aligned}$$

by first substituting p_n for n , then using both the upper and lower bounds on each side, and finally simplifying. ■

COROLLARY 2.1. *The following inequalities are also satisfied by the (q_n) : as $n \rightarrow \infty$:*

- (a) $q_n = n \log^2 n + 3n \log n \log \log n + O(n \log n)$,
- (b) $q_n \sim n \log^2 n$,
- (c) $\frac{q_n}{q_{n+1}} \rightarrow 1$,
- (d) $q_n \sim \log n \cdot p_n$,
- (e) *for all $\epsilon > 0$ there is an $n_\epsilon \in \mathbb{N}$ such that for all $n \geq n_\epsilon$, $q_n \leq p_n^{1+\epsilon}$,*
- (f) *for $n > 1$, $q_n < p_n^{\frac{3}{2}}$,*
- (g) *For all $m, n \in \mathbb{N}$, $q_n \cdot q_m > q_{mn}$,*
- (h) *For real $a > 1$ and n sufficiently large, $q_{[an]} > aq_n$.*

Proof. The inequalities (a)-(e) follow from Theorem 2.1. Item (f) follows by an explicit computation up to $n = 2000$ and then by contradiction, assuming $n(\log n + \log \log n - 2) < p_n < n(\log n + \log \log n)$, for $n > 2000$. Item (g) follows from the corresponding theorem of Ishikawa [7] for primes and (h) from that of Giordano [6]. ■

3. THE PRIME-PRIME NUMBER THEOREM AND INEQUALITIES

Here we obtain two forms for the asymptotic order of the counting function of the prime-primes:

THEOREM 3.1. *As $x \rightarrow \infty$,*

$$\pi_\pi(x) \sim \frac{x}{\log^2 x}, \text{ and } \pi_\pi(x) = \text{Li}(\text{Li}(x)) + O(x \exp(-A \log^{\frac{3}{5}} x \log^{-\frac{1}{5}} x)),$$

for some absolute constant $A > 0$, where the implied “big- O ” constant is also absolute.

Proof. The first asymptotic relation follows from a substitution: As $x \rightarrow \infty$:

$$\begin{aligned}
\pi(\pi(x)) &= \frac{\pi(x)}{\log \pi(x)} + O\left(\frac{\pi(x)}{\log^2 \pi(x)}\right) \\
&= \frac{\frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right)}{\log\left(\frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right)\right)} + O\left(\frac{x \log \log x}{\log^4 x}\right) \\
&= \frac{x/\log x}{\log(x/\log x) + O(1/\log x)} + O\left(\frac{x \log \log x}{\log^4 x}\right), \\
&= \frac{x}{\log^2 x [1 + O(\frac{\log \log x}{\log x})]} + O\left(\frac{x \log \log x}{\log^4 x}\right), \\
&= \frac{x}{\log^2 x} + O\left(\frac{x \log \log x}{\log^3 x}\right).
\end{aligned}$$

Now let $\Delta(x) := O(x \exp(-A \log^{\frac{3}{5}} x \log \log^{-\frac{1}{5}} x))$ and use the equation [8, 15]

$$\pi(x) = \text{Li}(x) + \Delta(x).$$

By the Mean Value Theorem there is a real number θ with $|\theta| < 1$ such that

$$\begin{aligned}
\pi(\pi(x)) &= \text{Li}[\text{Li}(x) + \Delta(x)] + \Delta(\pi(x)), \\
&= \text{Li}(\text{Li}(x)) + \frac{\Delta(x)}{\log[\text{Li}(x) + \theta \Delta(x)]} + \Delta(x), \\
&= \text{Li}(\text{Li}(x)) + \Delta(x).
\end{aligned}$$

■

PROPOSITION 3.1. *The following inequalities are true for every integer $k > 1$ and real x, y sufficiently large:*

- (a) $\pi_\pi(kx) < k\pi_\pi(x)$,
- (b) $\pi_\pi(x + y) \leq \pi_\pi(x) + 4\pi_\pi(y)$,
- (c) $\pi_\pi(x + y) - \pi_\pi(x) \ll \frac{y}{\log^2 y}$.

Proof. (a) We apply the theorem of Panaitopol [11], namely that for all $k > 1$ and x sufficiently large $\pi(kx) < k\pi(x)$ to derive

$$\pi_\pi(kx) = \pi(\pi(kx)) \leq \pi(k\pi(x)) < k\pi(\pi(x)) = k\pi_\pi(x).$$

(b) Now apply the inequality of Montgomery and Vaughan [9] as well as the theorem of Panaitopol:

$$\begin{aligned}\pi_\pi(x+y) &= \pi(\pi(x+y)) \leq \pi(\pi(x) + 2\pi(y)) \\ &\leq \pi(\pi(x)) + 2\pi(2\pi(y)) < \pi_\pi(x) + 4\pi_\pi(y).\end{aligned}$$

(c) This follows directly from (b) and Theorem 3.1. \blacksquare

In part (c) compare the expression when π_π is replaced by π . [10, Page 34].

The following integral expression shows that, approximately, the local density of the prime-primes is $dt/\log^2 t$:

PROPOSITION 3.2. As $x \rightarrow \infty$

$$\begin{aligned}\text{Li}(\text{Li}(x)) &= \int_2^x \left(\frac{1}{\log^2 t} + \frac{\log \log t}{\log^3 t} \right) dt + O\left(\frac{x(\log \log x)^2}{\log^4 t} \right) \\ &= \int_2^x \frac{dt}{\log^2 t} + O\left(\frac{x \log \log x}{\log^3 x} \right)\end{aligned}$$

Proof. Use the expression [5, Page 86]

$$\text{Li}(x) = \frac{x}{\log(x)} + \frac{x}{\log^2(x)} + \cdots + \frac{(n-1)!x}{\log^n(x)} + O\left(\frac{x}{\log^{n+1}(x)} \right).$$

in the case $n = 1$, so

$$\text{Li}(x) = \frac{x}{\log(x)} + O\left(\frac{x}{\log^2(x)} \right).$$

Then let

$$\begin{aligned}F(x) &:= \text{Li}(\text{Li}(x)) \\ &= \int_2^y \frac{1}{\log t} dt, \text{ where} \\ y &= \int_2^x \frac{1}{\log u} du.\end{aligned}$$

Then

$$\begin{aligned}
 F'(x) &= 1/[\log x \log(\text{Li}(x))] \\
 &= 1/[\log^2 x (1 - \frac{\log \log x}{\log x} + O(\frac{1}{\log^2 x}))] \\
 &= \frac{1}{\log^2 x} [1 + \frac{\log \log x}{\log x} + O(\frac{(\log \log x)^2}{\log^2 x})] \\
 &= \frac{1}{\log^2 x} + \frac{\log \log x}{\log^3 x} + O(\frac{(\log \log x)^2}{\log^4 x})
 \end{aligned}$$

so therefore

$$F(x) = \int_2^x (\frac{1}{\log^2 t} + \frac{\log \log t}{\log^3 t}) dt + O(\frac{x(\log \log x)^2}{\log^4 x}).$$

To derive the second expression, split the integral for the second term in the integrand of the first expression at \sqrt{x} . ■

4. EXTREME VALUES OF GAPS BETWEEN PRIME-PRIMES

Note that for $n > 1$ the number of primes between each pair of prime primes is always odd, so $q_{n+1} - q_n \geq 6$. It is natural to conjecture that this gap size of 6 is taken on an infinite number of times, as is every even gap size larger than 6.

Since for $n > 2$, $q_{n+1} \leq 2q_n$ [4], we have $q_{n+1} - q_n \leq q_n^\theta$ for $\theta = 1.0$. The same best current value for primes, due to Baker, Harman and Pintz [1, 2], namely

$$p_{n+1} - p_n \ll_\epsilon p_n^{\theta+\epsilon}, \quad \theta = 0.525$$

works for prime-primes. To see this replace x by $\pi(x)$ in their equation [2, Page 562]

$$\pi(x + x^{.525}) - \pi(x) \geq \frac{9}{10} \frac{x^{.525}}{\log x}$$

to derive the formula, for every $\epsilon > 0$ and x sufficiently large

$$\pi_\pi(x + x^{0.525+\epsilon}) - \pi_\pi(x) \geq \frac{9}{10} \frac{x^{0.525-\epsilon}}{\log^2 x}.$$

The first proposition below is modeled directly on the corresponding result for primes. The second is also closely related to the derivation for primes, [13, Page 155].

PROPOSITION 4.1. *For every integer $m > 1$ there exists an even number δ such that more than m prime-primes are at distance δ .*

Proof. Let $n \in \mathbb{N}$, $S := \{q_1, \dots, q_{n+1}\}$ be a finite initial subset of the ordered sequence of prime-primes, and let $D := \{q_{j+1} - q_j : 1 \leq j \leq n\}$ be the n differences of consecutive elements of S .

If $|D| \geq \lfloor \frac{n}{m} \rfloor$, then

$$\begin{aligned} q_{n+1} - q_1 &= (q_2 - q_1) + (q_3 - q_2) + \dots + (q_{n+1} - q_n) \\ &\geq 6 + 8 + \dots + 2 \lfloor \frac{n}{m} \rfloor \\ &\geq \frac{n^2}{m^2} + O(1). \end{aligned}$$

By Theorem 2.1, we can choose n sufficiently large so $q_{n+1} < 2n \log^2 n$ and the inequality for $q_{n+1} - q_1$ is not satisfied.

Therefore we can assume n is sufficiently large so $|D| < \lfloor \frac{n}{m} \rfloor$. Then one of the differences must appear more than m times. Call the size of this difference δ . ■

PROPOSITION 4.2.

$$\liminf_{n \rightarrow \infty} \frac{q_{n+1} - q_n}{\log^2 q_n} \leq 1.$$

Proof. Let $\epsilon > 0$ be given. Define two positive constants α, β with

$$\alpha = \frac{\beta + 1}{\beta - 1}$$

with $\beta > 3$ and so $1 < \alpha < 2$. Let $\epsilon > 0$ be another positive constant with $\epsilon < 1/\alpha - 1/2$. Let $L := 1 + 2\epsilon$ and let

$$\{q_m, \dots, q_{m+k}\}$$

be all of the prime-primes in the interval $[x, \beta x]$. Now suppose (to obtain a contradiction) that for all n with $m \leq n \leq m + k - 1$ we have

$$q_{n+1} - q_n \geq L \log^2 q_n.$$

Then

$$\begin{aligned} (\beta - 1)x &\geq q_{m+k} - q_m \\ &\geq L \sum_{n=m}^{m+k-1} \log^2 q_n \\ &\geq Lk \log^2 x. \end{aligned}$$

But by Theorem 3.1, for all x sufficiently large

$$\left(1 - \frac{\epsilon}{2}\right) \frac{x}{\log^2 x} < \pi_\pi(x) < \left(1 + \frac{\epsilon}{2}\right) \frac{x}{\log^2 x}$$

so

$$\begin{aligned} k &\geq \pi_\pi(\beta x) - \pi_\pi(x) - 1 \\ &\geq \frac{(1 - \frac{\epsilon}{2})\beta x}{\log^2 \beta x} - \frac{(1 + \frac{\epsilon}{2})x}{\log^2 x} - 1 \\ &\geq \frac{x(\beta - 1 - \epsilon(\beta + 1))}{\log^2 x} \text{ for } x \text{ sufficiently large.} \end{aligned}$$

Therefore $(\beta - 1) \geq L(\beta - 1 - \epsilon(\beta + 1))$, or in other words $1 \geq (1 + 2\epsilon)(1 - \alpha\epsilon)$, which is impossible.

Therefore there exists an n such that $q_n \in [x, \beta x]$ and

$$\frac{q_{n+1} - q_n}{\log^2 q_n} < 1 + 2\epsilon$$

so $\liminf_{n \rightarrow \infty} (q_{n+1} - q_n) / \log^2 q_n \leq 1$. ■

Using a similar approach one can show that $\limsup_{n \rightarrow \infty} \frac{q_{n+1} - q_n}{\log^2 q_n} \geq 1$. It is expected however that the limit infimum of the ratio should be zero and the limit supremum infinity. Fig. 1 is based on the normalized nearest neighbor gaps for the first two million prime-primes with a bin size of 0.025 and with the x-axis 160 corresponding to a normalized gap value of 4.0.

5. PRIME-PRIMES IN ARITHMETIC PROGRESSIONS

Using the prime number graph technique of Pomerance [12], applied to the sequences (n, q_n) and $(n, \log q_n)$, we are able to demonstrate the existence of infinite subsets of the q_n such as the following:

PROPOSITION 5.1. (a) *There exists an infinite set of $n \in \mathbb{N}$ with $2q_n \leq q_{n-i} + q_{n+i}$ for all $0 < i < n$.*

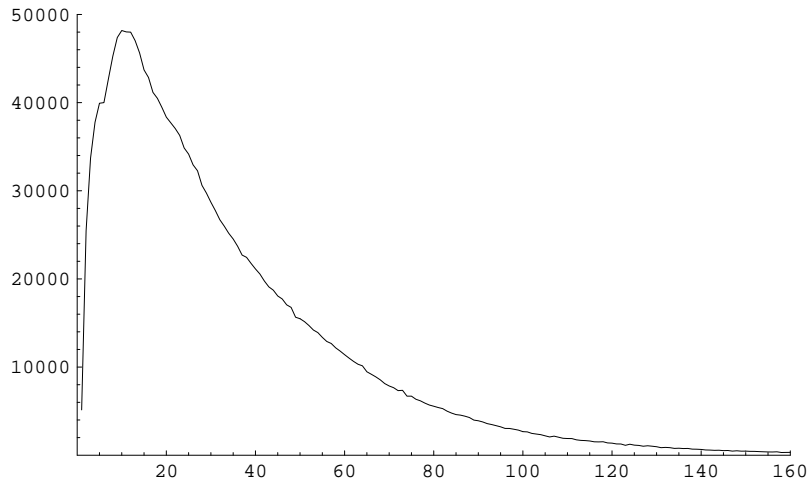


FIG. 1. Normalized gap frequencies.

(b) *There exists an infinite set of $n \in \mathbb{N}$ with $q_n^2 > q_{n-i}q_{n+i}$ for all $0 < i < n$.*

Now we show that there are arithmetic progressions $(an + b : n \in \mathbb{N})$ containing specified numbers of prime-primes, not necessarily consecutive, satisfying $q_n = an + b$. Although modelled on the technique of Pomerance, [12, Theorem 4.1], the key step is introducing functions, named $f(u)$ and $g(u)$, but delaying their explicit definitions until sufficient information becomes available.

THEOREM 5.1. *For every integer $m > 1$ there exists an arithmetic progression $(an + b : n \in \mathbb{N})$, with $a, b \in \mathbb{N}$, with at least m prime-primes satisfying $q_n = an + b$.*

Proof. 1. **Definitions:** Let $u > 0$ be a real variable and $f(u) > 0$ and $g(u) > 0$ two real decreasing functions both tending to zero as $u \rightarrow \infty$, to be chosen later. Let $v = u + u \cdot f(u)$ so

$$\log v = \log u \left(1 + O\left(\frac{f(u)}{\log u}\right)\right), \text{ and} \quad (1)$$

$$\log \text{Li}(v) = \log \text{Li}(u) \left(1 + O\left(\frac{f(u)}{\log u}\right)\right). \quad (2)$$

Let T be a parallelogram in the first quadrant of the x-y plane bounded by the lines $x = u$, $x = v$,

$$y = \text{Li}(\text{Li}(u)) + 2ug(u) + \frac{x-u}{k}, \text{ and}$$

$$y = \text{Li}(\text{Li}(u)) - 2ug(u) + \frac{x-u}{k},$$

where $k = \log \text{Li}(u) \log u$.

Claim: If $|y - \text{Li}(\text{Li}(x))| < ug(u)$ then $(x, y) \in T$. This is demonstrated in 2.-4. below.

2. The upper bound: since $y < \text{Li}(\text{Li}(x)) + ug(u)$ we have, for some $u < \xi < x$,

$$y < \text{Li}(\text{Li}(u)) + \frac{x-u}{\log \text{Li}(\xi) \log \xi} + 2ug(u),$$

$$\leq \text{Li}(\text{Li}(u)) + \frac{x-u}{\log \text{Li}(u) \log u} + 2ug(u).$$

3. The lower bound: We have $\text{Li}(\text{Li}(x)) - ug(u) < y$ and need $\text{Li}(\text{Li}(u)) + \frac{x-u}{\log \text{Li}(u) \log u} - 2ug(u) < y$ so it is sufficient to have

$$\frac{x-u}{\log \text{Li}(u) \log u} - \frac{x-u}{\log \text{Li}(\xi) \log \xi} \leq ug(u) \text{ or}$$

$$\frac{v-u}{\log \text{Li}(u) \log u} - \frac{v-u}{\log \text{Li}(v) \log v} \leq ug(u), \text{ which is equivalent to}$$

$$f(u) \left[\frac{1}{\log \text{Li}(u) \log u} - \frac{1}{\log \text{Li}(v) \log v} \right] \leq g(u), \text{ or by (1) and (2)}$$

$$\frac{f(u)}{\log \text{Li}(u) \log u} \left[1 - \frac{1}{1 + O(f(u)/\log u)} \right] = O\left(\frac{f(u)^2}{\log^3 u}\right) \leq M \cdot \frac{f(u)^2}{\log^3 u} \leq g(u) \quad (3)$$

for some $M > 0$.

4. Counting points and lines: by 2. and 3. each point (q_n, n) with $u \leq q_n \leq v$ is in T , so the number of such points is, by Theorem 3.1, bounded below by the numerator in the expression below. The number of lines with slope $1/k$ passing through the integer lattice points of T is bounded above by the denominator of this expression. Therefore for u sufficiently large:

$$\frac{\# \text{points}}{\# \text{lines}} \geq \frac{\frac{uf(u)}{2} \cdot \frac{1}{\log^2 u}}{(\log \text{Li}(u) \log u) 4ug(u)}$$

$$\geq \frac{f(u)/8}{(\log u)^4 g(u)} \quad (4)$$

Now let $f(u) := \frac{1}{\log^\alpha u}$ and $g(u) := \frac{1}{\log^\beta u}$ where $\alpha, \beta > 0$ are chosen so that

$$\alpha + 4 < \beta < 2\alpha + 3.$$

For example $\alpha = 2$, $\beta = 6.5$. Then in (4)

$$\frac{f(u)}{(\log u)^4 g(u)} = \log^{\beta - \alpha - 4} u \rightarrow \infty$$

and in (3)

$$\frac{f(u)^2}{(\log u)^3 g(u)} = \log^{\beta - 2\alpha - 3} u \rightarrow 0+,$$

so choose u sufficiently large that $f(u)^2/(\log^3 u g(u)) \leq 1/M$, thus ensuring the validity of the lower bound.

With these choices, the number of points divided by the number of lines tends to positive infinity, so for every natural number m there is at least one line on the graph of (q_n, n) with m of these points. Finally we note that since k varies continuously with u , we can choose $k \in \mathbb{N}$ and so $q_n = an + b$ with $a, b \in \mathbb{N}$ since $a = k$. ■

6. EPILOG

Leading on from Section 5, a natural aim is to show that there are an infinite number of prime-primes congruent, say, to 1 modulo 4, or some other explicit arithmetic progression. Then show that every arithmetic progression $(an + b : n \in \mathbb{N})$, with $(a, b) = 1$, contains an infinite number of prime-primes. This has been demonstrated numerically, with the number of prime-primes falling approximately evenly between the equivalence classes modulo a .

This problem appears to have considerable more depth than the results given here. For example it is, on the face of it, more difficult than the corresponding theorem of Dirichlet for primes, because the primes in residue classes given by that theorem do not appear in any particular order.

7. ACKNOWLEDGEMENTS

We acknowledge the assistance given by a helpful referee, especially in making us aware of the result in the paper of Baker, Harman and Pintz. The figure was produced using Mathematica.

REFERENCES

1. R. C. Baker and G. Harman, The difference between consecutive primes, *Proc. London Math. Soc.* (3) **72** (1996), 261–280.
2. R. C. Baker, G. Harman and J. Pintz, The difference between consecutive primes, II, *Proc. Lond. Math. Soc.* (3) **83** (2001), 532–562.
3. K. A. Broughan and Q. Zhou, Flat primes and thin primes, (preprint).
4. R. E. Dressler and S. T. Parker, Primes with a prime subscript, *J. ACM* **22** (1975), 380–381.
5. H. M. Edwards, *Riemann's zeta function*, Dover, 2001.
6. G. Giordano, The existence of primes in small intervals, *Indian J. Math.* **31** (1989), 105–110.
7. H. Ishikawa, Über die Verteilung der Primzahlen, *Sci. Rep. Tokyo Univ. Lit. Sci. Sect.d A* **2** (1934), 27–40.
8. N. M. Korobov, Estimates of trigonometric sums and their applications. *Uspehi Mat. Nauk* **13** (1958), 185–192.
9. H. L. Montgomery and R. C. Vaughan, The large sieve, *Mathematika* **20** (1973), 119–134.
10. H. L. Montgomery, Topics in Multiplicative Number Theory, *Springer Lecture Notes in Mathematics*, **227**, (1971).
11. L. Panaitopol, Eine Eigenschaft der Funktion über die Verteilung der Primzahlen, *Bull. Math. Soc. Sci. Math. R. S. Roumanie*, **23** (1979), 189–194.
12. C. Pomerance, The prime number graph, *Math. Comp.* **33** (1979), 399–408.
13. K. Prachar, *Primzahlverteilung*, Springer-Verlag, 1957.
14. J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6** (1962), 64–94.
15. I. M. Vinogradov, A new estimate of the function $\zeta(1 + it)$. *Izv. Akad. Nauk SSSR. Ser. Mat.* **22** (1958), 161–164.

On-line Encyclopedia of Integer Sequences: sequence number a006450.