VARIANCE OF DISTRIBUTION OF ALMOST PRIMES
IN ARITHMETIC PROGRESSIONS

À la mémoire de Jacques et Julie Elharrar

EMMANUEL KNAFO

Presented by John Friedlander, FRSC

Abstract. We give an effective lower bound for the variance of distribution of \(k\)-almost primes in arithmetic progressions. This lower bound approaches the expected asymptotic ‘exponentially fast’ as \(k\) goes to infinity.

Résumé. Nous donnons une borne inférieure effective pour la variance de la distribution des nombres presque premiers d’ordre \(k\) dans les progressions arithmétiques. Cette borne inférieure s’approche de la valeur asymptotique attendue ‘exponentiellement vite’ quand \(k\) tend vers l’infini.

In [1], Friedlander and Goldston considered the variance for the distribution of primes in arithmetic progressions. The mean square sum

\[
G(x, q) = \sum_{a \text{ (mod } q)} (E(x; q, a))^2 = \sum_{a \text{ (mod } q)} \left( \frac{x}{\phi(q)} \right)^2
\]

was studied. A lower bound for \(G(x, q)\) of the correct order of magnitude for \(q\) in the range \(x (\log x)^A \leq q \leq x\) was given and the following theorem was proven:

**Theorem 1.** Let \(A > 0, \epsilon > 0\). We have, for \(x\) sufficiently large,

\[
G(x, q) \geq \left( \frac{1}{2} - \epsilon \right) x \log q
\]

for \(\frac{x}{(\log x)^A} \leq q \leq x\).

Later, Hooley [2] extended this range to \(\frac{x}{\exp(A\sqrt{\log x})} < q \leq x\). The following theorem was then proven:

**Theorem 2.** Let \(\epsilon > 0\). Then, for some absolute constant \(A\),

\[
G(x, q) > \left( \frac{1}{2} - \epsilon \right) x \log q
\]

whenever \(\frac{x}{\exp(A\sqrt{\log x})} < q \leq x\) and \(x > x_0(\epsilon)\).
The author’s paper [4] generalizes the above work to consider the more complicated question of the variance for numbers with a restricted number of prime factors (or almost primes) in arithmetic progressions. To count the integers \( n \) up to \( x \) in a given arithmetic progression such that \( n \) has no more than \( k \) distinct prime factors, we use the function \( \psi_k(x; q, a) = \sum_{n \leq x \atop n \equiv a (\text{mod } q)} \Lambda_k(n) \) where \( \Lambda_k \) is the generalized von Mangoldt function defined by \( \Lambda_k = \mu * \log^k \). In other words, \( \Lambda_k \) is the arithmetical function defined by

\[
\Lambda_k(n) = \sum_{d \mid n} \mu(d) \log^k \frac{n}{d}
\]

where

\[
\mu(d) = \begin{cases} (-1)^r & \text{if } d \text{ is square-free and } d = p_1 \cdots p_r, \\ 0 & \text{otherwise}. \end{cases}
\]

Analogously, we consider the mean square sum

\[
G_k(x, q) = \sum_{a \equiv 0 (\text{mod } q)} \left( E_k(x; q, a) \right)^2
\]

where \( \omega(n) \) gives the number of distinct prime factors of \( n \) and \( EV(\cdot) \) stands for the expected value of the argument inside the parentheses. In determining the expected value of \( \psi_k(x; q, a) \), we encounter the constants \( a_n(q) \) defined by

\[
a_n(q) = \begin{cases} 1 & \text{if } n = 0, \\ (-1)^n(n+1)! \sum_{(\mu_1, \ldots, \mu_n) \in Q(n)} \frac{(c_0(q))^{\mu_1} \cdots (c_{n-1}(q))^{\mu_n}}{\mu_1!(1!)^{\mu_1} \cdots \mu_n!(n!)^{\mu_n}} & \text{if } n \geq 1.
\end{cases}
\]

where

\[
Q(n) = \left\{ (\mu_1, \ldots, \mu_n) \mid \mu_1, \ldots, \mu_n \text{ are nonnegative integers such that } 0 \leq \sum_{i=1}^{n} i\mu_i \leq n \right\}
\]

for \( n \geq 1 \) and

\[
c_j(q) = g_j + \sum_{p \mid q} \frac{A_j(p) \log^{j+1} p}{(p-1)^{j+1}}
\]

for \( \sum_{p \mid q} \frac{\log^{j+1} p}{(p-1)^{j+1}} \geq 1 \).
for \( j \geq 0 \). We note that in (5), \( A_j(x) \) is the Eulerian polynomial of degree \( j \) and \( g_j \) is the constant such that

\[
g_j = (j + 1)! \sum_{(\mu_1, \ldots, \mu_{j+1}) \in \mathcal{P}(j+1)} \frac{(\mu_1 + \cdots + \mu_{j+1} - 1)!}{\mu_1! \mu_2! \cdots \mu_{j+1}! (j!)^{\mu_{j+1}}} \gamma_{\mu_1} \cdots \gamma_{\mu_j}
\]

where \( \gamma_i \) are the Stieltjes constants and

\[
P(m) = \left\{ (\mu_1, \ldots, \mu_m) \mid \mu_1, \ldots, \mu_m \text{ are nonnegative integers such that } \sum_{i=1}^{m} i \mu_i = m \right\}
\]

for \( m \geq 1 \). We also let

\[
C^*(n, m) = \left\{ (k_1, \ldots, k_m) \mid k_1, \ldots, k_m \text{ are positive integers with } n = k_1 + \cdots + k_m \right\}.
\]

In [4], we prove the following theorem:\footnote{For \( \frac{x}{\exp(A\sqrt{\log x})} \leq q \leq x \), the lower bound given in (7) is equivalent to the lower bound we get upon replacing \( \log q \) by \( \log x \).}

**THEOREM 3.** Let \( k \geq 1 \) and let

\[
G_k(x, q) = \sum_{\substack{0 < a \leq q \\
(a, q) = 1}} \left( \frac{\psi_k(x; q, a) - \frac{x}{\phi(q)} \sum_{j=1}^{k} a_j(q) \log^{k-j} x}{\phi(q)} \right)^2
\]

\[
+ \sum_{r=1}^{k-1} \sum_{\substack{0 < a \leq q \\
(a, q) = p_1^{i_1} \cdots p_r^{i_r} \\
p_1, \ldots, p_r \text{ are distinct primes} \\
i_1, \ldots, i_r \geq 1}} \left( \psi_k(x; q, a) - (-1)^r \frac{x}{\phi(q)} \sum_{j=0}^{k-r-1} \log^j x \sum_{(i_1, \ldots, i_r) \in C^*(k-j, r+1)} (-1)^{i_1 + \cdots + i_r} \left( \sum_{j=1}^{k} a_{j-1}(q) (\log^{i_1} p_1 \cdots \log^{i_r} p_r) \right)^2 \right).
\]

Let \( \epsilon > 0 \). Then, for some absolute constant \( A \),

\[
G_k(x, q) > \frac{k^2}{2^{k-1}} \left( 1 - \frac{1}{2^{k-1}} - \epsilon \right) x \log^{2k-1} q
\]

whenever \( \frac{x}{\exp(A\sqrt{\log x})} < q \leq x \) and \( x > x_0(\epsilon, k) \). This result is unconditional and effective, that is to say that given \( \epsilon \) and \( k \), one can assign a numerical value to \( x_0(\epsilon, k) \).
When \( k = 1 \), Theorem 3 yields the effective version of Theorem 2. We observe that the proofs of Theorems 1 and 2 are non-effective since they both use Siegel’s theorem; indeed, the proof of Theorem 1 relies on the Bombieri–Vinogradov theorem which itself depends on Siegel’s theorem, while the proof of Theorem 2 uses Siegel’s theorem to derive (38) of [2]. We conclude that Theorem 3 gives the first effective lower bound for the variance of distribution of primes in arithmetic progressions.

We are able to obtain an effective lower bound in Theorem 3 by eliminating the use of Siegel’s theorem. This is done by taking into account directly the effect of Siegel zeros through the use of the following:

**Proposition 1. (A Singular Series Average)**

Let \( u \geq 1 \). We have

\[
\sum_{0 < j < u} (u^\beta - j^\beta) \chi_r^*(j) \prod \frac{p - 1}{p - 2} \ll u^\beta \log u + 1) r^{\frac{1}{2}} (\log^2 r)(\log \log q).
\]

This proposition allows for a refinement of the following key estimate for even modulus \( q \) (see [2, p. 906]):

\[
J = \sum_{0 < j < u} (u^\beta - j^\beta) \chi_r^*(jq) \prod \frac{p - 1}{p - 2} = O(r^{\frac{1}{2} + \epsilon} u \log^2 x).
\]

Upon letting \( u = \frac{x}{q} \), the estimate (8) allows us to save (nearly) an additional factor of \((\log x)^{\frac{3}{2}}\) over the estimate (9) whenever \( \exp(A\sqrt{\log x}) < q \leq x \). This saving is sufficient to show that the effect of Siegel zeros is \( o(x \log q) \) without the need to appeal to Siegel’s bound.

We should emphasize that the proof of Theorem 3 requires the use of new truncated divisor sum approximations not considered before. They are intended to mimic the behaviour of the generalized von Mangoldt function on some averages. Inspired by Selberg’s upper bound sieve method, they are defined as follows:

\[
\Lambda_{k,R}(n) = \sum_{\substack{d|n \\text{d}| \leq R}} \mu(d) \frac{d}{\phi(d)} \sum_{\sigma \leq \frac{x}{\phi(d)}} \frac{\mu^2(\sigma)}{\phi(\sigma)} M_k(\sigma d)
\]

where \( M_k \) is the arithmetical function such that

\[
M_k(\rho) = \sum_{j=1}^{k} \binom{k}{j} a_{j-1}(\rho) \log^{k-j} \frac{x}{\rho}
\]

with \( a_n(\rho) \) as in (4). Alternatively, we may express our approximating function as

\[
\Lambda_{k,R}(n) = \sum_{\sigma \leq R} \frac{\mu^2(\sigma)}{\phi(\sigma)} M_k(\sigma) \sum_{d|\sigma \ \sigma|n} d \mu(d).
\]
We observe that when $k = 1$, the function $M_k$ is the unit function, i.e., $M_1(\rho) = 1$ for all $\rho$. In this case, the first form (10) yields $\Lambda_{1,R} = \Lambda_R$, i.e., the truncated divisor sum used in the proof of Theorem 2 [2], while the second form (11) gives $\Lambda_R = \lambda_R$ which is the truncated divisor sum used in the proof of Theorem 1 [1].

We conjecture the following:

**Conjecture 1.** Let $A > 0$. For $\frac{x}{\exp(A\sqrt{\log x})} \leq q \leq x$,

$$G_k(x, q) \sim \frac{k^2}{2k-1} x \log^{2k-1} q.$$

Support for this conjecture is obtained by adapting the methods described in Friedlander and Goldston [1] and Goldston [3] which deal with the case $k = 1$. The ratio of the lower bound of Theorem 3 to the expected asymptotic for $G_k(x, q)$ in the range $\frac{x}{\exp(A\sqrt{\log x})} \leq q \leq x$ is thus

$$(12) \quad \frac{k^2}{2k-1} \left(1 - \frac{1}{2k-1} - \epsilon\right) x \log^{2k-1} q$$

$$= \left(1 - \frac{2}{4k} - \epsilon\right).$$

We see that this ratio approaches $1 - \epsilon$ as $k$ goes to infinity. We stress that it does so ‘exponentially fast’. Therefore, we conclude that the lower bound of Theorem 3 approaches the expected asymptotic ‘exponentially fast’ as $k$ goes to infinity.

**References**