C. R. Math. Rep. Acad. Sci. Canada Vol. **28**, (1), 2006 pp. 12–16

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

$$G(x,q)=\sum_{\substack{a(\mathrm{mod}\ q)\\(a,q)=1}}\big(E(x;q,a)\big)^2=\sum_{\substack{a(\mathrm{mod}\ q)\\(a,q)=1}}\Big(\psi(x;q,a)-\frac{x}{\phi(q)}\Big)^2 \tag{1}$$

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

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

$$G(x,q)\ge\Big(\frac12-\epsilon\Big)x\log q$$

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

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

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

$$G(x,q)>\Big(\frac12-\epsilon\Big)x\log q$$

whenever $\frac{x}{\exp(A\sqrt{\log x})}<q\le x$ and $x>x_0(\epsilon)$.

Received by the editors on January 20, 2006.
AMS subject classification: 11N25, 11M20.
Keywords: effective lower bound, truncated divisor sum, generalized von Mangoldt function, almost primes.

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

$$\psi_k(x;q,a) = \sum_{\substack{n\le x \\ n\equiv a \pmod q}} \Lambda_k(n)$$

where Λ_k is the *generalized von Mangoldt function* defined by $\Lambda_k = \mu * \log^k$. In other words, Λ_k is the arithmetical function defined by

$$\Lambda_k(n) = \sum_{d\mid n} \mu(d) \log^k \frac{n}{d} \tag{2}$$

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

$$\begin{aligned} G_k(x,q) &= \sum_{\substack{a \pmod q \\ \omega((a,q))<k}} \big(E_k(x;q,a)\big)^2 \\ &= \sum_{\substack{a \pmod q \\ \omega((a,q))<k}} \big(\psi_k(x;q,a) - EV(\psi_k(x;q,a))\big)^2 \qquad (3) \end{aligned}$$

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

$$a_n(q) = \begin{cases} 1 & \text{if } n=0, \\ (-1)^n(n+1)! \sum_{(\mu_1,\ldots,\mu_n)\in Q(n)} \dfrac{(c_0(q))^{\mu_1}\cdots(c_{n-1}(q))^{\mu_n}}{\mu_1!(1!)^{\mu_1}\cdots\mu_n!(n!)^{\mu_n}} & \text{if } n\ge 1 \end{cases} \tag{4}$$

where

$$Q(n) = \Big\{(\mu_1,\ldots,\mu_n) \mid \mu_1,\ldots,\mu_n \text{ are nonnegative integers such that } 0 \le \sum_{i=1}^{n} i\mu_i \le n\Big\}$$

for $n\ge 1$ and

$$c_j(q) = g_j + \sum_{p\mid q} \frac{A_j(p)\log^{j+1} p}{(p-1)^{j+1}} \tag{5}$$

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

$$g_j = (j+1)! \sum_{(\mu_1,\ldots,\mu_{j+1}) \in P(j+1)} \frac{(\mu_1 + \cdots + \mu_{j+1} - 1)!}{\mu_1!\,(0!)^{\mu_1} \cdots \mu_{j+1}!\,(j!)^{\mu_{j+1}}} \gamma_0^{\mu_1} \cdots \gamma_j^{\mu_{j+1}}$$

where γ_i are the *Stieltjes constants* and

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

for $m \geq 1$. We also let

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

In [4], we prove the following theorem:[1]

THEOREM 3. *Let* $k \geq 1$ *and let*

$$\begin{aligned} (6) \quad G_k(x,q) = & \sum_{\substack{0<a\leq q \\ (a,q)=1}} \left(\psi_k(x;q,a) - \frac{x}{\phi(q)} \sum_{j=1}^{k} \binom{k}{j} a_{j-1}(q) \log^{k-j} x \right)^2 \\ & + \sum_{r=1}^{k-1} \sum_{\substack{0<a\leq q \\ (a,q)=p_1^{\alpha_1}\cdots p_r^{\alpha_r} \\ p_1\cdots p_r | q \\ \alpha_1,\ldots,\alpha_r \geq 1}} \Bigg(\psi_k(x;q,a) - (-1)^r \frac{x}{\phi(q)} \sum_{j=0}^{k-r-1} \log^j x \sum_{(i_1,\ldots,i_r,i)\in C^*(k-j,r+1)} \\ & \times (-1)^{i_1+\cdots+i_r} \binom{k}{i_1,\ldots,i_r,i,j} a_{i-1}(q) (\log^{i_1} p_1 \cdots \log^{i_r} p_r) \Bigg)^2 . \end{aligned}$$

Let $\epsilon > 0$. *Then, for some absolute constant* A,

$$(7) \qquad G_k(x,q) > \frac{k^2}{2k-1} \Big(1 - \frac{1}{2^{2k-1}} - \epsilon \Big) x \log^{2k-1} q$$

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* ϵ *and* k*, one can assign a numerical value to* $x_0(\epsilon, k)$.

---

[1] For $\frac{x}{\exp(A\sqrt{\log x})} < q \leq x$, the lower bound given in (7) is equivalent to the lower bound we get upon replacing $\log q$ by $\log x$.

When $k = 1$, Theorem 3 yields the effective version of Theorem 2. We observe that the proofs of Theorems 1 and 2 are *noneffective* 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*

$$\sum_{0<j<u} (u^\beta - j^\beta)\chi_r^*(j) \prod_{\substack{p|j \\ p\nmid 2q}} \left(\frac{p-1}{p-2}\right) \ll u^\beta(\log u + 1) r^{\frac{1}{2}} (\log^2 r)(\log\log q). \tag{8}$$

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

$$J = \sum_{0<j<u} (u^\beta - j^\beta)\chi_r^*(jq) \prod_{\substack{p|jq \\ p>2}} \left(\frac{p-1}{p-2}\right) = O(r^{\frac{1}{2}+\epsilon} u \log^2 x). \tag{9}$$

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 $\frac{x}{\exp(A\sqrt{\log x})} < q \leq x$. 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:

$$\Lambda_{k,R}(n) = \sum_{\substack{d|n \\ d\leq R}} \mu(d)\frac{d}{\phi(d)} \sum_{\substack{\sigma\leq \frac{R}{d} \\ (\sigma,d)=1}} \frac{\mu^2(\sigma)}{\phi(\sigma)} M_k(\sigma d) \tag{10}$$

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

$$\Lambda_{k,R}(n) = \sum_{\sigma\leq R} \frac{\mu^2(\sigma)}{\phi(\sigma)} M_k(\sigma) \sum_{\substack{d|\sigma \\ d|n}} d\mu(d). \tag{11}$$

We observe that when $k = 1$, the function M_k is the *unit function*, *i.e.*, $M_1(\rho) = 1$ for all ρ. 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* $\frac{x}{\exp(A\sqrt{\log x})} \le q \le x$,

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

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 $\frac{x}{\exp(A\sqrt{\log x})} \le q \le x$ is thus

$$\frac{\frac{k^2}{2k-1}\left(1 - \frac{1}{2^{2k-1}} - \epsilon\right) x \log^{2k-1} q}{\frac{k^2}{2k-1} x \log^{2k-1} q} = \left(1 - \frac{2}{4^k} - \epsilon\right). \tag{12}$$

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

1. J. B. Friedlander and D. A. Goldston, *Variance of Distribution of Primes in Residue Classes.* Quart. J. Math. Oxford (2) **47** (1996), 313–336.
2. C. Hooley, *On the Barban–Davenport–Halberstam theorem: XII.* In: Number Theory in Progress (Zakopane, Poland, 1997), Walter de Gruyter, Berlin, 1999, 893–910.
3. D. A. Goldston, *On Hardy and Littlewood's Contribution to the Goldbach Conjecture.* Proc. Amalfi Conf. Analytic Number Theory (E. Bombieri *et al.*, eds.), University of Salerno, 1992, 115–155.
4. E. R. Knafo, *Variance of Distribution of Almost Primes in Arithmetic Progressions.* PhD thesis, University of Toronto, 2006.

Department of Mathematics
University of Toronto
Toronto, Ontario
M5S 2E4
email: knafo@math.toronto.edu