Almost Primes in Almost All Short Intervals

Joni Teräväinen

University of Turku

Contents

History of the Problem and Results

- Primes in All Short Intervals
- Primes in Almost All Short Intervals
- E_k Numbers in Almost All Short Intervals
- P_k Numbers in Almost All Short Intervals
- Results

Primes in All Short Intervals

- One would expect all intervals of the form $[x, x + (\log x)^{2+\varepsilon}]$ to contain a prime for x large (Cramér's conjecture).
- Even under the Riemann hypothesis, one only knows that $[x, x + C\sqrt{x}\log x]$ contains a prime.
- Best unconditional result due to Baker, Harman and Pintz (2001): $[x, x + x^{0.525}]$ contains a prime for all large x.

Primes in Almost All Short Intervals

We say that a property $P(x)$ holds for almost all x if the number of integers $x \leq X$ for which $P(x)$ fails is $o(X)$.

- Best known result on primes in almost all intervals is Jia's (1996): $[x, x + x^{\frac{1}{20}+\varepsilon}]$ contains a prime almost always.
- Conjecturally, primes should be exponentially distributed in short intervals so that $[x, x + \lambda \log x]$ should have a prime with probability $1 - e^{-\lambda}$.
- Therefore, for any $\psi(x)$ with $\psi(x) \to \infty$ as $x \to \infty$, the interval $[x, x + \psi(x) \log x]$ should have a prime almost always.

Primes in Almost All Short Intervals

- Gallagher (1976): If a certain uniform version of the Hardy–Littlewood prime k-tuple conjecture holds, the interval $[x, x + \lambda \log x]$ includes a prime with the anticipated probability.
- Goldston-Pintz-Yıldırım (2011): The interval $[x, x + \lambda \log x]$ contains a prime with positive probability for any $\lambda > 0$ (but we do not know what happens to the probability as $\lambda \to \infty$).
- Selberg (1949): Under the Riemann hypothesis, with $\psi(x)$ as before, the interval $[x, x + \psi(x) \log^2 x]$ contains a prime almost always.

E_k Numbers in Almost All Short Intervals

An E_k number is a product of exactly k primes.

- Wolke (1979): There exists c such that $[x, x + (\log x)^c]$ contains an E_2 number almost always ($c = 5 \cdot 10^6$).
- Harman (1982): One can take $c = 7 + \varepsilon$ above.
- Harman's and Wolke's proofs utilize sums over the zeros of the Riemann zeta function and the fact that the density hypothesis is known to hold in a nontrivial region ($\sigma \geq \frac{11}{14}$ in Harman's case).
- The best known density hypothesis region $\sigma \geq \frac{25}{32}$ by Bourgain (2000) improves Harman's exponent to $c = 6.86$.
- For E_k numbers with $k \geq 3$, the state of affairs used to be the same as for E_2 numbers.

P_k Numbers in Almost All Short Intervals

A P_k number is a product of at most k primes.

- Mikawa (1989): For any $\psi(x)$ such that $\psi(x) \to \infty$ as $x \to \infty$, the interval $[x, x + \psi(x)(\log x)^5]$ contains a P_2 number almost always.
- Friedlander&Iwaniec (Opera de Cribro): With $\psi(x)$ as above, $[x, x + \psi(x) \log x]$ contains a P_4 number (a P_3 number) almost always.
- The proofs of these results are based on combinatorial sieves, which are subject to the famous parity problem. One cannot distinguish between primes and products of two primes using these methods, so results on E_k numbers must instead be based on analytic arguments.
- The presence of the parity problem makes E_k numbers a much closer analogue of primes than P_k numbers.

Results

Theorem 1 (T., 2015)

The interval $[x, x + (\log x)^{1+\varepsilon}]$ contains an E_3 number almost always.

Theorem 2 (T., 2015)

The interval $[x, x + (\log x)^{3.51}]$ contains an E_2 number almost always.

Results

These are consequences of the following quantitative versions.

Theorem 1' (T., 2015)

Let $P_1 = (\log\log X)^{6+\varepsilon}$ and $P_2 = (\log X)^{\varepsilon^{-2}}$. For $h \geq P_1 \log X$, we have

$$\frac{1}{h} \sum_{\substack{x \leq p_1 p_2 p_3 \leq x+h \\ P_i \leq p_i \leq P_i^{1+\varepsilon},\, i \leq 2}} 1 - \frac{1}{X} \sum_{\substack{X \leq p_1 p_2 p_3 \leq 2X \\ P_i \leq p_i \leq P_i^{1+\varepsilon},\, i \leq 2}} 1 = o\left(\frac{1}{\log X}\right)$$

for almost all $x \leq X$.

Theorem 2' (T., 2015)

For $P_1 = (\log X)^{2.51}$ and $h \geq P_1 \log X$, we have

$$\frac{1}{h} \sum_{\substack{x \leq p_1 p_2 \leq x+h \\ P_1 \leq p_1 \leq P_1^{1+\varepsilon}}} 1 \geq \varepsilon \frac{1}{X} \sum_{\substack{X \leq p_1 p_2 \leq 2X \\ P_1 \leq p_1 \leq P_1^{1+\varepsilon}}} 1$$

for almost all $x \leq X$.

Results

- The dyadic sums in the theorems can be evaluated with the prime number theorem, and are $\gg \frac{1}{\log X}$, so we actually find $\gg \frac{h}{\log X}$ E_k numbers on almost all intervals $[x, x+h]$ with $h = (\log X)(\log\log X)^{6+\varepsilon}$ for $k = 3$ and $h = (\log X)^{3.51}$ for $k = 2$.
- We can find in general E_k numbers on $[x, x + (\log X)(\log_{k-1} X)^{C_k}]$ almost always with $\log_\ell$ the ℓth iterated logarithm.
- Nevertheless, we are not quite able to find find E_k numbers on $[x, x + \psi(x)\log x]$ for arbitrary $\psi(x) \to \infty$ for any fixed k.

Results

- The number of exceptional intervals is poor (but $o(X)$, of course!). Earlier arguments, such as Harman's, bound the number of exceptions by $\ll_A \frac{X}{(\log X)^A}$ for any A, while the mentioned theorems allow $\frac{X}{(\log X)^\varepsilon}$ exceptions or more.
- The numbers $p_1p_2p_3$ and p_1p_2 that we found must have a specific shape; namely they have very small prime factors. One could obtain almost equal prime factors if one considered $p_1p_2p_3p_4$ or $p_1p_2p_3$ numbers instead.
- The limit of the method for E_2 numbers is the exponent $c = 3 + \varepsilon$, which is the same that one gets from Harman's method under the density hypothesis for the Riemann zeta function.

The E_3 case

- Reduction to Dirichlet Polynomials
- Factorizing Dirichlet Polynomials
- Bounding Dirichlet Polynomials
- Decomposing the Integral

Reduction to Dirichlet Polynomials

- As in the works of Harman, Watt and Jia (and others) on primes in almost all short intervals, we reduce the problem to finding cancellation in the mean square of a Dirichlet polynomial.
- We also adapt the ideas of Matomäki and Radziwiłł on multiplicative functions in short intervals to the setting of almost primes.
- If $S_h(x)$ is the short sum in Theorem 1' and $S_X(X)$ is the dyadic sum, Theorem 1' boils down to showing

$$\frac{1}{X}\int_X^{2X}\left|\frac{1}{h}S_h(x)-\frac{1}{X}S_X(X)\right|^2 dx = o\left(\frac{1}{\log^2 X}\right).$$

Reduction to Dirichlet Polynomials

Denoting

$$F(s) = \sum_{\substack{p_1 p_2 p_3 \sim X \\ P_j \le p_j \le P_j^{1+\varepsilon}, j \le 2}} (p_1 p_2 p_3)^{-s}$$

($n \sim X$ means $X \le n < 2X$), we have the Parseval-type inequality

$$\frac{1}{X}\int_X^{2X} \left| \frac{1}{h} S_h(x) - \frac{1}{X'} S_X(X') \right|^2 dx \ll \frac{1}{T_0} + \int_{T_0}^{\frac{X}{h}} |F(1+it)|^2 dt + \max_{T \ge \frac{X}{h}} \frac{X}{Th} \int_T^{2T} |F(1+it)|^2 dt$$

with $T_0 = X^{0.01}, X' = X^{0.97}$.

Reduction to Dirichlet Polynomials

- Hence, the essence of the proof is showing that

$$\int_{T_0}^{\frac{X}{h}} |F(1+it)|^2 dt = o\left(\frac{1}{\log^2 X}\right).$$

- Note that we got rid of short sums, and are working with dyadic sums instead.
- We see that the smaller h is, the harder the mean square becomes to bound.
- The more prime variables we have (in this case 3), the more flexibility we have for factorizing $F(s)$ into a product of Dirichlet polynomials of various lengths.

Factorizing Dirichlet Polynomials

- When separating the variables in $F(s)$, we cannot even afford to lose a factor of $(\log X)^\varepsilon$ in some cases.
- Therefore, we perform the factorization with the following lemma that also splits the variables into short intervals.

Factorizing Dirichlet Polynomials

Factorization lemma

Let $H \geq 1$ and

$$F(s) = \sum_{\substack{mn \sim X \\ M \leq m \leq M'}} \frac{a_m b_n}{(mn)^s}$$

for some $M' > M \geq 1$ and for some complex numbers a_m, b_n. Let

$$A_{v,H}(s) = \sum_{e^{\frac{v}{H}} \leq m < e^{\frac{v+1}{H}}} \frac{a_m}{m^s}, \quad B_{v,H}(s) = \sum_{n \sim Xe^{-\frac{v}{H}}} \frac{b_n}{n^s}.$$

Then

$$\int_{-T}^{T} |F(1+it)|^2 dt \ll \left(H \log \left(\frac{M'}{M} \right) \right)^2 \int_{-T}^{T} |A_{v_0,H}(1+it) B_{v_0,H}(1+it)|^2 dt$$
$$+ \int_{-T}^{T} \left| \sum_{n \in [Xe^{-1/H}, Xe^{1/H}] \cup [2X, 2Xe^{1/H}]} c_n n^{-1-it} \right|^2 dt,$$

for some integer $v_0 \in [H \log M, H \log M']$, and with $|c_n| \leq \sum_{n=k\ell} |a_k b_\ell|$.

Bounding Dirichlet Polynomials

- There are three different approaches to bounding Dirichlet polynomials: pointwise, large value and mean value estimates.
- We will use all three types, and sometimes in forms that are specific to Dirichlet polynomials over primes or zeta sums.

Pointwise bound: If $F(s) = \sum_{m \sim X} \frac{a_m}{m^s}$ has as its coefficients one of the sequences $(1), (\mu(m))$ or $(1_{\mathbb{P}}(m))$, we have $|F(1+it)| \ll_A (\log X)^{-A}$ for all $A > 0$ in a wide range of t and X.

Bounding Dirichlet Polynomials

Mean value theorem

For $F(s) = \sum_{m \sim X} \frac{a_m}{m^s}$,

$$\int_{-T}^{T} |F(1+it)|^2 dt \ll \frac{T+X}{X} \sum_{n \sim X} \frac{|a_n|^2}{n}.$$

- In many cases this is not good enough for us, since if a_n is supported on the primes, it gives the bound $\ll \frac{1}{\log X}$ and not $\ll \frac{1}{\log^2 X}$.
- For that reason, we use an improved mean value theorem.

Improved Mean Value Theorem

Let $F(s)$ be as above. We have

$$\int_{-T}^{T} |F(1+it)|^2 dt \ll \frac{T}{X} \sum_{n \sim X} \frac{|a_n|^2}{n} + \frac{T}{X} \sum_{1 \le h \le \frac{X}{T}} \sum_{\substack{m-n=h \\ m,n \sim X, m \ne n}} \frac{|a_m||a_n|}{X}. \quad (1)$$

In the first sum, we gain a crucial factor of $\log X$ since $\frac{T}{X} \le \frac{1}{\log X}$. In the second term, we count solutions to $p - q = h$ in primes on average over h, and hence get the bound $\ll (\log X)^{-2}$.

Bounding Dirichlet Polynomials

We also need a bound for the number of large values of very short Dirichlet polynomials with coefficients supported on primes, coming from the paper of Matomäki and Radziwiłł.

Large value estimate for Dirichlet polynomials over primes

Let $P \geq 1$, $\alpha \in \mathbb{R}$ and

$$F(s) = \sum_{p \sim P} \frac{a_p}{p^s}$$

with $|a_p| \leq 1$. Let $\mathcal{T} \subset [-T, T]$ be a well-spaced set of points (any two points have distance ≥ 1) such that $|F(1+it)| \geq P^{\alpha}$ for each $t \in \mathcal{T}$. Then we have

$$|\mathcal{T}| \ll T^{-2\alpha} P^{-2\alpha} \exp\left((1+o(1)) \frac{\log T}{\log P} \log\log T \right).$$

Decomposing the Integral

As in the result of Matomäki and Radziwiłł on multiplicative functions, a key ingredient in the proof is splitting the integration domain according to whether certain Dirichlet polynomials are small or large there. We write, for suitable $\alpha_1, \alpha_2 > 0$,

$$\begin{aligned}
\mathcal{T}_1 &= \{t \in [T_0, T] : |P_1(1+it)| \leq P_1^{-\alpha_1}\}, \\
\mathcal{T}_2 &= \{t \in [T_0, T] : |P_2(1+it)| \leq P_2^{-\alpha_2}\} \setminus \mathcal{T}_1, \\
\mathcal{T} &= [T_0, T] \setminus (\mathcal{T}_1 \cup \mathcal{T}_2),
\end{aligned}$$

where, roughly speaking,

$$P_1(s) = \sum_{p \sim P_1} \frac{1}{p^s}, \quad P_2(s) = \sum_{p \sim P_2} \frac{1}{p^s}.$$

We bound the mean square of $F(s)$ separately over $\mathcal{T}_1, \mathcal{T}_2$ and $\mathcal{T}$.

Decomposing the Integral

- In the $\mathcal{T}_1$ case, the improved mean value theorem produces a bound of $o\left(\frac{1}{\log^2 X}\right)$ for the mean square of $F(s)$.
- In the $\mathcal{T}_2$ case, we make use of the fact that $|P_1(1+it)P_1^{\alpha_1}|^{2\ell} \geq 1$, $|P_2(1+it)| \leq P_2^{-\alpha_2}$ to estimate

$$\int_{T_0}^{\frac{X}{h}} |F(1+it)|^2 dt \ll P_2^{-2\alpha_2} P_1^{2\alpha_1 \ell} \int_{T_0}^{\frac{X}{h}} |P_1(1+it)^{\ell} \tilde{F}(1+it)|^2 dt,$$

where, roughly speaking, $\tilde{F}(s) = \sum_{p \sim \frac{X}{P_1 P_2}} \frac{1}{p^s}$. This high moment can be bounded for $\ell = \lfloor \frac{\log P_2}{\log P_1} \rfloor$.

Decomposing the Integral

- In the $\mathcal{T}$ case, we first use Heath-Brown's identity to see that, essentially $\tilde{F}(s)$ is a sum of polynomials of the form $N_1(s)N_2(s)$ or $M_1(s)M_2(s)M_3(s)$, where $N_1(s)$ and $N_2(s)$ are zeta sums and $M_1(s)$, $M_2(s)$ and $M_3(s)$ have essentially Möbius function as their coefficients.
- If the zeta sums dominate, we may use Watt's twisted fourth moment of the Riemann zeta function to see that

$$\int_{\mathcal{T}} |F(1+it)|^2 dt \ll P_1^{2\alpha_1 \ell} \int_{T_0}^{\frac{X}{h}} |P_1(1+it)^{\ell} N_1(1+it) N_2(1+it)|^2 dt \ll T^{-\varepsilon}.$$

Decomposing the Integral

- If the polynomials $M_i(s)$ dominate, we may use the pointwise saving in these polynomials and the Halász-Montgomery inequality to see that

$$\begin{aligned}\int_{\mathcal{T}} |F(1+it)|^2 dt &\ll \sum_{t\in\mathcal{U}} |M_1(1+it)M_2(1+it)M_3(1+it)|^2 \\ &\ll (\log X)^{-100} \sum_{t\in\mathcal{U}} |M_1(1+it)M_2(1+it)|^2 \\ &\ll (\log X)^{-90} \left(1 + \frac{|\mathcal{U}|T^{\frac{1}{2}}}{X^{\frac{2}{3}-\varepsilon}}\right)\end{aligned}$$

  for a suitable well-spaced $\mathcal{U} \subset \mathcal{T}$.
- We inspect that for $\alpha_1 = 100\varepsilon, \alpha_2 = \frac{1}{12} - 10\varepsilon$ the mean value theorem for short Dirichlet polynomials over primes yields $|\mathcal{U}| \ll T^{\frac{1}{6}-\varepsilon}$ for our choices of P_1 and P_2, and we arrive at the result for E_3 numbers.

The E_2 case

- Application of Buchstab's Identity
- Sums $\Sigma_1(h)$ and $\Sigma_2(h)$
- Sum $\Sigma_3(h)$
- Reversing Buchstab's Identity
- Buchstab Integrals

Application of Buchstab's Identity

- If we define

$$F(s) = \sum_{\substack{p_1 p_2 \sim X \\ P_1 \leq p_1 \leq P_1^{1+\varepsilon}}} (p_1 p_2)^{-s},$$

  Theorem 2' reduces to bounding the mean square of $F(s)$.
- However, we have too few variables to use the same method as for E_3 numbers, so we need to create some additional variables.
- If we use Heath-Brown's identity, we get one additional variable, and going through the proof of Theorem 1' we arrive at the exponent $c = 5 + \varepsilon$ for E_2 numbers which is the same as Mikawa's exponent for P_2 numbers. To outperform the E_3 approach, we need several additional tools.

Application of Buchstab's Identity

- We use the theory of exponent pairs, which gives us small pointwise power savings in zeta sums of various lengths.
- We use Buchstab's Identity to extract an additional variable, arriving at

$$\begin{aligned}
S_h(x) &= \sum_{\substack{x \le p_1 n \le x+h \\ P_1 \le p_1 \le P_1^{1+\varepsilon} \\ (n,\mathcal{P}(w))=1 \\ n>1}} 1 - \sum_{\substack{x \le p_1 q_1 n \le x+h \\ P_1 \le p_1 \le P_1^{1+\varepsilon} \\ w \le q_1 < \sqrt{x} \\ (n,\mathcal{P}(q_1))=1 \\ n>1}} 1 \\
&= \sum_{\substack{x \le p_1 n \le x+h \\ P_1 \le p_1 \le P_1^{1+\varepsilon} \\ (n,\mathcal{P}(w))=1 \\ n>1}} 1 - \sum_{\substack{x \le p_1 q_1 n \le x+h \\ P_1 \le p_1 \le P_1^{1+\varepsilon} \\ w \le q_1 < \sqrt{x} \\ (n,\mathcal{P}(w))=1 \\ n>1}} 1 + \sum_{\substack{x \le p_1 q_1 q_2 n \le x+h \\ P_1 \le p_1 \le P_1^{1+\varepsilon} \\ w \le q_2 < q_1 < \sqrt{x} \\ (n,\mathcal{P}(q_2))=1 \\ n>1}} 1 \\
&=: \Sigma_1(h) - \Sigma_2(h) + \Sigma_3(h)
\end{aligned}$$

with $w = X^{o(1)}$.

Sums $\Sigma_1(h)$ and $\Sigma_2(h)$

- In $\Sigma_1(h)$ and $\Sigma_2(h)$, we have an integer variable n with the restriction $(n, \mathcal{P}(w)) = 1$.
- The condition on n can essentially be disposed of with some sieve theory, so n becomes a free integer variable, and we can apply the same method as in $\mathcal{T}_1$ and $\mathcal{T}$ cases for E_3 numbers.
- We conclude that $\frac{1}{h}\Sigma_1(h) = \frac{1}{X}\Sigma_1(X) + o(\frac{1}{\log X})$, $\frac{1}{h}\Sigma_2(h) = \frac{1}{X}\Sigma_2(X) + o(\frac{1}{\log X})$ almost always.

Sum $\Sigma_3(h)$

- The sum $\Sigma_3(h)$, in turn, splits into several sums depending on the sizes of q_1 and q_2.
- If q_1 is a small power of X, we can use the method for E_3 numbers, applying in the end Jutila's large value theorem to bound the size of the exceptional set. However, we can no longer take $c = 1 + \varepsilon$, but we must take $c = 1 + \frac{1}{2(\alpha_2 - \alpha_1)}$, so that we want α_2 to be as large and α_1 as small as possible. With this value of c, we see that this part of $\frac{1}{h}\Sigma_3(h)$ is asymptotic to its dyadic version almost always.
- If q_1 and q_2 are relatively small, but not as small as above, and q_1 and q_2 satisfy certain dependencies, we may use Watt's theorem to deal with that part. We see that this part of $\frac{1}{h}\Sigma_3(h)$ is asymptotic to its dyadic version almost always.
- The part of $\Sigma_3(h)$ where neither of the above holds can only be discarded.

Reversing Buchstab's Identity

- Denote by $\Sigma_3'(h)$ the parts of $\Sigma_3(h)$ that can be evaluated asymptotically, and denote by $\Sigma_3''(h)$ the remaining part.
- We now find that

$$
\begin{aligned}
\frac{1}{h}S_h(x) &= \frac{1}{h}(\Sigma_1(h) - \Sigma_2(h) + \Sigma_3'(h) + \Sigma_3''(h)) \\
&= \frac{1}{X}(\Sigma_1(X) - \Sigma_2(X) + \Sigma_3'(X)) + \frac{1}{h}\Sigma_3''(h) + o\left(\frac{1}{\log X}\right) \\
&= \frac{1}{X}S_X(X) + \frac{1}{h}\Sigma_3''(h) - \frac{1}{X}\Sigma_3''(X) + o\left(\frac{1}{\log X}\right) \\
&\geq \frac{1}{X}S_X(X) - \frac{1}{X}\Sigma_3''(X) + o\left(\frac{1}{\log X}\right) \\
&\geq \varepsilon \cdot \frac{1}{X}S_X(X),
\end{aligned}
$$

provided that $\Sigma_3''(X) \leq (1-\varepsilon)S_X(X)$.

Buchstab Integrals

- The dyadic sum $\Sigma_3''(X)$ can be computed as a Buchstab integral.
- There is a tradeoff between the size of the integration domain (and hence the size of the Buchstab integral) and the efficiency of Jutila's large value theorem (and hence the size of $\alpha_2 - \alpha_1$).
- Optimizing all the parameters involved, we indeed have $\Sigma_3''(X) \leq (1-\varepsilon) \cdot S_X(X)$ for $c = 3.51$, and this finishes the E_2 case.