Almost primes in almost all very short intervals

Kaisa Matomäki

University of Turku, Finland

MIT Number Theory Seminar, April 6th, 2021
Contents

1 Background and results
   • Primes
   • Primes in short intervals
   • Primes in almost all short intervals
   • Almost primes in (almost all) short intervals

2 Methods
   • The sieve method
   • Type I sums
   • To Kloosterman sums

3 Summary and further thoughts

Kaisa Matomäki
Almost primes in almost all very short intervals
1 Background and results
   - Primes
   - Primes in short intervals
   - Primes in almost all short intervals
   - Almost primes in (almost all) short intervals

2 Methods
   - The sieve method
   - Type I sums
   - To Kloosterman sums

3 Summary and further thoughts
Letter $p$ always denotes a prime, $p \in \{2, 3, 5, 7, 11, 13, \ldots \}$, i.e. a natural number $> 1$ that is only divisible by 1 and itself.
How many primes are there?

- Letter $p$ always denotes a prime, $p \in \{2, 3, 5, 7, 11, 13, \ldots \}$, i.e. a natural number $> 1$ that is only divisible by 1 and itself.
- Hadamard and de la Vallee Poussin showed independently in 1896 that the number of primes up to $x$ is

$$ (1 + o(1)) \int_2^x \frac{dx}{\log x} = (1 + o(1)) \frac{x}{\log x}. $$
How many primes are there?

- Letter \( p \) always denotes a prime, \( p \in \{2, 3, 5, 7, 11, 13, \ldots \} \), i.e. a natural number \( > 1 \) that is only divisible by 1 and itself.

- Hadamard and de la Vallee Poussin showed independently in 1896 that the number of primes up to \( x \) is

  \[
  (1 + o(1)) \int_{2}^{x} \frac{dx}{\log x} = (1 + o(1)) \frac{x}{\log x}.
  \]

  This is called the prime number theorem (PNT).

- It asserts that the "probability" that an integer \( n \) is prime is about \( 1/\log n \).
Letter $p$ always denotes a prime, $p \in \{2, 3, 5, 7, 11, 13, \ldots \}$, i.e. a natural number $> 1$ that is only divisible by 1 and itself.

Hadamard and de la Vallee Poussin showed independently in 1896 that the number of primes up to $x$ is

$$ (1 + o(1)) \int_2^x \frac{dx}{\log x} = (1 + o(1)) \frac{x}{\log x}. $$

This is called the prime number theorem (PNT).

It asserts that the "probability" that an integer $n$ is prime is about $1/\log n$.

PNT is equivalent to the fact that the Riemann zeta function does not have zeros with $\Re s = 1$. 
One wants to know about primes in short intervals: If we look at a "short" segment \((x, x + H]\) around \(x\), is the density of primes in that segment still \(1/\log x\)?
One wants to know about primes in short intervals: If we look at a ”short” segment $(x, x + H]$ around $x$, is the density of primes in that segment still $1/\log x$?

The smaller the $H$, the more difficult the problem.

Huxley’s prime number theorem from 1972 gives

$$\sum_{x < p \leq x + H} 1 = (1 + o(1)) \frac{H}{\log x}, \quad H \geq x^{\frac{7}{12} + \varepsilon}.$$ 

This is based on Huxley’s zero-density estimate for the zeta function and has resisted improvements, except Heath-Brown (1988) has shown this for $H \geq x^{\frac{7}{12} - o(1)}$. 

Kaisa Matomäki

Almost primes in almost all very short intervals
What about primes in short intervals?

- One wants to know about primes in short intervals: If we look at a "short" segment \((x, x + H]\) around \(x\), is the density of primes in that segment still \(1/\log x\)?
- The smaller the \(H\), the more difficult the problem.
- Huxley's prime number theorem from 1972 gives

\[
\sum_{x < p \leq x + H} 1 = (1 + o(1)) \frac{H}{\log x}, \quad H \geq x^{7/12 + \varepsilon}.
\]
What about primes in short intervals?

- One wants to know about primes in short intervals: If we look at a "short" segment \((x, x + H]\) around \(x\), is the density of primes in that segment still \(1/\log x\)?
- The smaller the \(H\), the more difficult the problem.
- Huxley’s prime number theorem from 1972 gives
  \[
  \sum_{x < p \leq x + H} 1 = (1 + o(1)) \frac{H}{\log x}, \quad H \geq x^{7/12+\varepsilon}.
  \]
- This is based on Huxley’s zero-density estimate for the zeta function and has resisted improvements, except Heath-Brown (1988) has shown this for \(H \geq x^{7/12-o(1)}\).
Baker-Harman-Pintz (2001) showed with a sieve method

\[ \sum_{x < p \leq x + H} 1 \geq \varepsilon \frac{H}{\log X}, \quad H \geq x^{0.525} \]

for some \( \varepsilon > 0 \).
Baker-Harman-Pintz (2001) showed with a sieve method

\[ \sum_{x < p \leq x + H} 1 \geq \varepsilon \frac{H}{\log X}, \quad H \geq x^{0.525} \]

for some \( \varepsilon > 0 \).

For shorter intervals one does not even know existence of primes!

Assuming RH one knows that \([x, x + x^{1/2} \log x]\) always contains primes.
Baker-Harman-Pintz (2001) showed with a sieve method

\[ \sum_{x < p \leq x + H} 1 \geq \varepsilon \frac{H}{\log X}, \quad H \geq x^{0.525} \]

for some \( \varepsilon > 0 \).

For shorter intervals one does not even know existence of primes!

Assuming RH one knows that \([x, x + x^{1/2} \log x]\) always contains primes.

Cramer made a probabilistic model based on ”probability of \( n \) being prime is \( 1/\log n \)”. Based on this, one expects that intervals \([x, x + (\log x)^{2+\varepsilon}]\) contain primes for all large \( x \).
Baker-Harman-Pintz (2001) showed with a sieve method
\[ \sum_{x < p \leq x + H} 1 \geq \varepsilon \frac{H}{\log X}, \quad H \geq x^{0.525} \]
for some \( \varepsilon > 0 \).

For shorter intervals one does not even know existence of primes!

Assuming RH one knows that \([x, x + x^{1/2} \log x]\) always contains primes.

Cramer made a probabilistic model based on "probability of \( n \) being prime is \( 1/\log n \). Based on this, one expects that intervals \([x, x + (\log x)^{2+\varepsilon}]\) contain primes for all large \( x \).

Huge gap between what's known and what's expected!
Primes in almost all short intervals

- Even under RH it is not known that \([x, x + x^{1/2}]\) always contains primes.
- What if one only requires that almost all intervals contain primes?
Even under RH it is not known that \([x, x + x^{1/2}]\) always contains primes.

What if one only requires that almost all intervals contain primes?

A variant of Huxley’s prime number theorem says that, for almost all \(x \in [X, 2X]\) (i.e. with \(o(X)\) exceptions),

\[
\sum_{x < p \leq x + H} 1 = (1 + o(1)) \frac{H}{\log X}, \quad H \geq x^{1/6+\epsilon}.
\]

This can be proved using the same zero-density estimates and has also resisted improvements.
Even under RH it is not known that \([x, x + x^{1/2}]\) always contains primes.

What if one only requires that almost all intervals contain primes?

A variant of Huxley’s prime number theorem says that, for almost all \(x \in [X, 2X]\) (i.e. with \(o(X)\) exceptions),

\[
\sum_{x < p \leq x + H} 1 = (1 + o(1)) \frac{H}{\log X}, \quad H \geq x^{1/6 + \varepsilon}.
\]

This can be proved using the same zero-density estimates and has also resisted improvements.

A lower bound has been shown for \(H \geq X^{1/20}\) by Jia.
Primes in almost all short intervals

- Even under RH it is not known that \([x, x + x^{1/2}]\) always contains primes.
- What if one only requires that almost all intervals contain primes?
- A variant of Huxley’s prime number theorem says that, for almost all \(x \in [X, 2X]\) (i.e. with \(o(X)\) exceptions),
  \[
  \sum_{x < p \leq x + H} 1 = (1 + o(1)) \frac{H}{\log X}, \quad H \geq x^{1/6+\varepsilon}.
  \]
- This can be proved using the same zero-density estimates and has also resisted improvements.
- A lower bound has been shown for \(H \geq X^{1/20}\) by Jia.
- One expects that, for any \(h \to \infty\) with \(X \to \infty\), the interval \((x, x + h \log x]\) contains primes for almost all \(x \in [X, 2X]\).
One expects that, for any $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains primes for almost all $x \in [X/2, X]$.

One can ask similar questions about almost-primes, i.e. $P_k$ numbers that have at most $k$ prime factors or $E_k$ numbers that have exactly $k$ prime factors.
One expects that, for any \( h \to \infty \) with \( X \to \infty \), the interval \((x - h \log X, x]\) contains primes for almost all \( x \in [X/2, X] \).

One can ask similar questions about almost-primes, i.e. \( P_k \) numbers that have at most \( k \) prime factors or \( E_k \) numbers that have exactly \( k \) prime factors.

Teräväinen has showed that, for almost all \( x \in [X/2, X] \), the interval \((x - (\log X)^{3.51}, x]\) contains an \( E_2 \)-number and the interval \((x - (\log \log X)^{6+\varepsilon} \log X, x]\) contains an \( E_3 \)-number.
Almost primes

- One expects that, for any $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains primes for almost all $x \in [X/2, X]$.
- One can ask similar questions about almost-primes, i.e. $P_k$ numbers that have at most $k$ prime factors or $E_k$ numbers that have exactly $k$ prime factors.
- Teräväinen has showed that, for almost all $x \in [X/2, X]$, the interval $(x - (\log X)^{3.51}, x]$ contain an $E_2$-number and the interval $(x - (\log \log X)^{6+\varepsilon} \log X, x]$ contains an $E_3$-number.
- Wu has shown that the interval $(x - x^{101/232}, x]$ contains $P_2$ numbers for all sufficiently large $x$. 
From now on we will concentrate on $P_k$ numbers in almost all short intervals.
From now on we will concentrate on $P_k$ numbers in almost all short intervals.

Following Friedlander, Friedlander and Iwaniec showed that as soon as $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains $P_{19}$-numbers for almost all $x \in [X/2, X]$. They used $\beta$-sieve with $\beta = 8$ and had level of distribution $D = X^{1/2}/(\log X)^A$. They say that if one was careful, one could use linear sieve instead and this would give $P_4$-numbers (with no prime factors $\leq X^{1/4-\varepsilon}$). Furthermore, they say that, using Duke-Friedlander-Iwaniec bounds on bilinear forms with Kloosterman fractions, one could slightly increase the level of distribution and obtain $P_3$ numbers. They write “It would be interesting to get integers with at most two prime divisors.”
From now on we will concentrate on $P_k$ numbers in almost all short intervals.

Following Friedlander, Friedlander and Iwaniec showed that as soon as $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains $P_{19}$-numbers for almost all $x \in [X/2, X]$.

They used $\beta$-sieve with $\beta = 8$ and had level of distribution $D = X^{1/2} / (\log X)^A$. 
From now on we will concentrate on $P_k$ numbers in almost all short intervals.

Following Friedlander, Friedlander and Iwaniec showed that as soon as $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains $P_{19}$-numbers for almost all $x \in [X/2, X]$.

They used $\beta$-sieve with $\beta = 8$ and had level of distribution $D = X^{1/2}/(\log X)^A$.

They say that if one was careful, one could use linear sieve instead and this would give $P_4$-numbers (with no prime factors $\leq X^{1/4-\varepsilon}$).
From now on we will concentrate on $P_k$ numbers in almost all short intervals.

Following Friedlander, Friedlander and Iwaniec showed that as soon as $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains $P_{19}$-numbers for almost all $x \in [X/2, X]$.

They used $\beta$-sieve with $\beta = 8$ and had level of distribution $D = X^{1/2}/(\log X)^A$.

They say that if one was careful, one could use linear sieve instead and this would give $P_4$-numbers (with no prime factors $\leq X^{1/4-\varepsilon}$).

Furthermore, they say that, using Duke-Friedlander-Iwaniec bounds on bilinear forms with Kloosterman fractions, one could slightly increase the level of distribution and obtain $P_3$ numbers.
From now on we will concentrate on $P_k$ numbers in almost all short intervals.

Following Friedlander, Friedlander and Iwaniec showed that as soon as $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains $P_{19}$-numbers for almost all $x \in [X/2, X]$.

They used $\beta$-sieve with $\beta = 8$ and had level of distribution $D = X^{1/2}/(\log X)^A$.

They say that if one was careful, one could use linear sieve instead and this would give $P_4$-numbers (with no prime factors $\leq X^{1/4-\varepsilon}$).

Furthermore, they say that, using Duke-Friedlander-Iwaniec bounds on bilinear forms with Kloosterman fractions, one could slightly increase the level of distribution and obtain $P_3$ numbers.

They write "It would be interesting to get integers with at most two prime divisors".
Theorem (M. (202?))

As soon as \( h \to \infty \) with \( X \to \infty \), the interval \((x - h \log X, x]\) contains \( P_2 \)-numbers for almost all \( x \in [X/2, X] \).
Theorem (M. (202?))

As soon as $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains $P_2$-numbers for almost all $x \in [X/2, X]$.

Write $\Omega(n)$ for the number of prime factors, counted with multiplicity. E.g. $\Omega(18) = \Omega(2 \cdot 3 \cdot 3) = 3$. 

$P_2$ numbers in almost all very short intervals
Theorem (M. (202?))

As soon as $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains $P_2$-numbers for almost all $x \in [X/2, X]$.

Write $\Omega(n)$ for the number of prime factors, counted with multiplicity. E.g. $\Omega(18) = \Omega(2 \cdot 3 \cdot 3) = 3$. We have the following more precise theorem.

Theorem (M. (202?))

Let $h \leq X^{1/100}$. There exist constants $c, C > 0$ such that

$$ch \leq \sum_{x-h \log X < n \leq x} 1_{\Omega(n) \leq 2} \leq Ch$$

for all $x \in [X/2, X]$ apart from an exceptional set of measure $O(X/h)$. 
Theorem (M. (202?))

Let \( h \leq X^{1/100} \). There exists constant \( c > 0 \) such that

\[
\sum_{x-h \log X \leq n \leq x} 1_{\Omega(n) \leq 2} \geq ch
\]

for all \( x \in [X/2, X] \) apart from an exceptional set of measure \( O(X/h) \).
**Theorem (M. (202?))**

Let $h \leq X^{1/100}$. There exists constant $c > 0$ such that

$$\sum_{x-h \log X < n \leq x} 1_{\Omega(n) \leq 2} \geq ch$$

for all $x \in [X/2, X]$ apart from an exceptional set of measure $O(X/h)$.

- We use Richert’s weighted sieve with well-factorability and Vaughan’s identity. We get level of distribution $D = X^{5/9}$ (not optimized) from Deshouillers-Iwaniec bounds for averages of Kloosterman numbers. Mikawa used similar strategy with Weil bound, but lost some logs in $h$. 
1 Background and results
   - Primes
   - Primes in short intervals
   - Primes in almost all short intervals
   - Almost primes in (almost all) short intervals

2 Methods
   - The sieve method
   - Type I sums
   - To Kloosterman sums

3 Summary and further thoughts
Theorem (M. (202?))

Let $h \leq X^{1/100}$. There exists constant $c > 0$ such that

$$
\sum_{x-h \log X < n \leq x \atop p|n \implies p > X^{1/8}} 1_{\Omega(n) \leq 2} \geq ch
$$

for all $x \in [X/2, X]$ apart from an exceptional set of measure $O(X/h)$.

- We use Richert’s weighted sieve with well-factorability and Vaughan’s identity. We get level of distribution $D = X^{5/9}$ (not optimized) from Deshouillers-Iwaniec bounds for averages of Kloosterman sums.
Write $A(x) = (x - h \log X, x] \cap \mathbb{N}$ and $P(z) = \prod_{p < z} p$. Define $z := X^{5/36}$ and $y = X^{1/2}$. Study, for $x \in (X/2, X]$,

$$
\sum_{n \in A(x) \atop (n, P(z)) = 1} w_n := \sum_{n \in A(x) \atop (n, P(z)) = 1} \left( 1 - \sum_{p | n \atop z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \right)
$$
Write $\mathcal{A}(x) = (x - h \log X, x] \cap \mathbb{N}$ and $P(z) = \prod_{p < z} p$. Define $z := X^{5/36}$ and $y = X^{1/2}$. Study, for $x \in (X/2, X]$, 

$$\sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1}} w_n := \sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1}} \left(1 - \sum_{\substack{p | n \\ z \leq p < y}} \left(1 - \frac{\log p}{\log y}\right)\right)$$

$$\leq \sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1}} \left(1 - \sum_{\substack{p | n}} \left(1 - \frac{\log p}{\log y}\right)\right)$$

$$\leq \sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1}} (1 - \sum_{\substack{p | n}} (1 - \frac{\log p}{\log y}))$$

Hence it suffices to show that, with $O\left(X/h\right)$ exceptions, $\sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1}} w_n \gg h$. 

Kaisa Matomäki
Almost primes in almost all very short intervals
Write $\mathcal{A}(x) = (x - h \log X, x] \cap \mathbb{N}$ and $P(z) = \prod_{p < z} p$. Define $z := X^{5/36}$ and $y = X^{1/2}$. Study, for $x \in (X/2, X]$,

$$
\sum_{n \in \mathcal{A}(x) \cap \mathbb{N}, (n, P(z)) = 1} w_n := \sum_{n \in \mathcal{A}(x) \cap \mathbb{N}, (n, P(z)) = 1} \left( 1 - \sum_{p \mid n, z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \right)
$$

$$
\leq \sum_{n \in \mathcal{A}(x) \cap \mathbb{N}, (n, P(z)) = 1} \left( 1 - \sum_{p \mid n} \left( 1 - \frac{\log p}{\log y} \right) \right)
$$

$$
\asymp \sum_{n \in \mathcal{A}(x) \cap \mathbb{N}, (n, P(z)) = 1} \left( 1 - \Omega(n) + \frac{\log X}{\log y} \right). \quad \text{.}
$$
Write \( \mathcal{A}(x) = (x - h \log X, x] \cap \mathbb{N} \) and \( P(z) = \prod_{p \leq z} p \). Define \( z := X^{5/36} \) and \( y = X^{1/2} \). Study, for \( x \in (X/2, X] \),

\[
\sum_{n \in \mathcal{A}(x) \atop (n, P(z)) = 1} w_n := \sum_{n \in \mathcal{A}(x) \atop (n, P(z)) = 1} \left( 1 - \sum_{p \mid n \atop z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \right)
\]

\[
\leq \sum_{n \in \mathcal{A}(x) \atop (n, P(z)) = 1} \left( 1 - \sum_{p \mid n} \left( 1 - \frac{\log p}{\log y} \right) \right)
\]

\[
\geq \sum_{n \in \mathcal{A}(x) \atop (n, P(z)) = 1} \left( 1 - \Omega(n) + \frac{\log X}{\log y} \right) \leq 2 \sum_{n \in \mathcal{A}(x) \atop (n, P(z)) = 1} 1_{\Omega(n) \leq 2}.
\]
Setting up Richert’s weighted sieve

Write $\mathcal{A}(x) = (x - h \log X, x] \cap \mathbb{N}$ and $P(z) = \prod_{p < z} p$. Define $z := X^{5/36}$ and $y = X^{1/2}$. Study, for $x \in (X/2, X]$,

$$\sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1 \atop (n, P(z)) = 1}} w_n := \sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1 \atop (n, P(z)) = 1}} \left( 1 - \sum_{p \mid n \atop z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \right)$$

$$\leq \sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1 \atop (n, P(z)) = 1}} \left( 1 - \sum_{p \mid n} \left( 1 - \frac{\log p}{\log y} \right) \right)$$

$$\ll \sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1 \atop (n, P(z)) = 1}} \left( 1 - \Omega(n) + \frac{\log X}{\log y} \right) \leq 2 \sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1 \atop (n, P(z)) = 1}} 1_{\Omega(n) \leq 2}.$$

Hence it suffices to show that, with $O(X/h)$ exceptions,

$$\sum_{\substack{n \in \mathcal{A}(x) \\ (n, P(z)) = 1 \atop (n, P(z)) = 1}} w_n \gg h.$$
A sieve lower bound

Recall $\mathcal{A}(x) = (x - h \log X, x] \cap \mathbb{N}$ and $P(z) = \prod_{p < z} p$. We need

$$\sum_{n \in \mathcal{A}(x), (n, P(z)) = 1} w_n = \sum_{n \in \mathcal{A}(x), (n, P(z)) = 1} 1 - \sum_{z \leq p < y} \left(1 - \frac{\log p}{\log y}\right) \sum_{np \in \mathcal{A}(x), (n, P(z)) = 1} 1 \gg h$$
Recall $\mathcal{A}(x) = (x - h \log X, x] \cap \N$ and $P(z) = \prod_{p < z} p$. We need

$$
\sum_{n \in \mathcal{A}(x) \atop (n,P(z)) = 1} w_n = \sum_{n \in \mathcal{A}(x) \atop (n,P(z)) = 1} 1 - \sum_{z \leq p < y} (1 - \frac{\log p}{\log y}) \sum_{np \in \mathcal{A}(x) \atop (n,P(z)) = 1} 1 \gg h
$$

By sieve theory we have nice $\alpha^+_d$ and $\alpha^-_{d,p}$ such that

$$
\sum_{d \mid (n,P(z)) \atop d \leq D} \alpha^-_d \leq 1(n,P(z)) = 1 \leq \sum_{d \mid (n,P(z)) \atop d \leq D/p} \alpha^+_d,
$$

where $D = X^{5/9}$,
A sieve lower bound

Recall \( \mathcal{A}(x) = (x - h \log X, x] \cap \mathbb{N} \) and \( P(z) = \prod_{p < z} p \). We need

\[
\sum_{n \in \mathcal{A}(x) \atop (n, P(z)) = 1} w_n = \sum_{n \in \mathcal{A}(x) \atop (n, P(z)) = 1} 1 - \sum_{z \leq p < y} (1 - \frac{\log p}{\log y}) \sum_{n p \in \mathcal{A}(x) \atop (n, P(z)) = 1} 1 \gg h
\]

By sieve theory we have nice \( \alpha_d^+ \) and \( \alpha_d^- \) such that

\[
\sum_{d \mid (n, P(z)) \atop d \leq D} \alpha_d^- \leq 1_{(n, P(z)) = 1} \leq \sum_{d \mid (n, P(z)) \atop d \leq D/p} \alpha_d^+
\]

where \( D = X^{5/9} \), so that, with \( \mathcal{B}_d := \{ n \in \mathbb{N} : dn \in \mathcal{B} \} \),

\[
\sum_{n \in \mathcal{A}(x) \atop (n, P(z)) = 1} w_n \geq \sum_{d \mid P(z) \atop d \leq D} \alpha_d^- |\mathcal{A}(x)_d| - \sum_{z \leq p < y} (1 - \frac{\log p}{\log y}) \sum_{d \mid P(z) \atop d \leq D/p} \alpha_d^+ |\mathcal{A}(x)_{dp}|,
\]
\[ \sum_{n \in A(x)} w_n \geq \sum_{d \mid P(z) \atop d \leq D} \alpha_d^- |A(x)_d| - \sum_{z \leq p < y} (1 - \frac{\log p}{\log y}) \sum_{d \mid P(z) \atop d \leq D/p} \alpha_{d,p}^+ |A(x)_{dp}| \]
A sieve lower bound

$$\sum_{n \in A(x) \atop (n, P(z)) = 1} w_n \geq \sum_{d \mid P(z) \atop d \leq D} \alpha_d^+ |A(x)_d| - \sum_{z \leq p < y} (1 - \frac{\log p}{\log y}) \sum_{d \mid P(z) \atop d \leq D/p} \alpha_{d, p}^+ |A(x)_{dp}|$$

Writing, for $e \in \{d, dp\}$, $|A(x)_e| = \frac{h \log X}{e} + \left( |A(x)_e| - \frac{h \log X}{e} \right)$,
A sieve lower bound

\[ \sum_{n \in A(x) \atop (n, P(z)) = 1 \atop (n, P(z)) = 1} w_n \geq \sum_{d | P(z) \atop d \leq D} \alpha_d^- |A(x)_d| - \sum_{z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \sum_{d | P(z) \atop d \leq D/p} \alpha_{d,p}^+ |A(x)_{dp}| \]

Writing, for \( e \in \{d, dp\}, |A(x)_e| = \frac{h \log X}{e} + \left( |A(x)_e| - \frac{h \log X}{e} \right), \)

\[ \sum_{n \in A(x) \atop (n, P(z)) = 1} w_n \geq h \log X \cdot M(z, y) + E^-(x, y, z) - E^+(x, y, z), \]

\[ M(z, y) := \sum_{d | P(z)} \alpha_d^- \frac{d}{d} - \sum_{z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \sum_{d | P(z)} \frac{\alpha_{d,p}^+}{dp} \gg \frac{1}{\log X} \]

\[ E^-(x, y, z) := \sum_{d | P(z)} \alpha_d^- \left( |A(x)_d| - \frac{h \log X}{d} \right) \]

\[ E^+(x, y, z) := \sum_{z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \sum_{d | P(z)} \frac{\alpha_{d,p}^+}{dp} \left( |A(x)_{dp}| - \frac{h \log X}{dp} \right). \]
A reduction to mean square estimates

\[\sum_{n \in A(x)} w_n \geq 3ch + E^{-}(x, y, z) - E^{+}(x, y, z),\]

where \(c > 0\),

\[E^{-}(x, y, z) := \sum_{d | P(z)} \alpha_{d}^{-} \left( |A(x)_{d}| - \frac{h \log X}{d} \right)\]

\[E^{+}(x, y, z) := \sum_{z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \sum_{d | P(z)} \alpha_{d, p}^{+} \left( |A(x)_{dp}| - \frac{h \log X}{dp} \right).\]
A reduction to mean square estimates

\[
\sum_{n \in A(x) \cap \{n, P(z)\} = 1} w_n \geq 3ch + E^-(x, y, z) - E^+(x, y, z),
\]

where \( c > 0 \),

\[
E^-(x, y, z) := \sum_{d \mid P(z)} \alpha_d^- \left( |A(x)_d| - \frac{h \log X}{d} \right)
\]

\[
E^+(x, y, z) := \sum_{z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \sum_{d \mid P(z)} \alpha^+_{d, p} \left( |A(x)_{dp}| - \frac{h \log X}{dp} \right).
\]

Hence \( \sum w_n \geq ch \) with \( O(X/h) \) exceptions if \( |E^\pm(x, y, z)| \leq ch \) with \( O(X/h) \) exceptions.
A reduction to mean square estimates

\[ \sum_{n \in A(x)} w_n \geq 3ch + E^{-}(x, y, z) - E^{+}(x, y, z), \]

where \( c > 0 \),

\[ E^{-}(x, y, z) := \sum_{d \mid P(z)} \alpha^{-}_d \left( |A(x)_d| - \frac{h \log X}{d} \right) \]

\[ E^{+}(x, y, z) := \sum_{z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \sum_{d \mid P(z)} \alpha^{+}_{d,p} \left( |A(x)_{dp}| - \frac{h \log X}{dp} \right). \]

Hence \( \sum w_n \geq ch \) with \( O(X/h) \) exceptions if \( |E^{\pm}(x, y, z)| \leq ch \) with \( O(X/h) \) exceptions. This follows if

\[ \int_{X/2}^{X} |E^{\pm}(x, y, z)|^2 dx = O(hX). \]
We need to show that

\[
\int_{X/2}^{X} \left| \sum_{d \leq D} \lambda_d \left( |A(x)_{d}| - \frac{h \log X}{d} \right) \right|^2 dy = O(hX)
\]

with \( \lambda_d = \alpha_d^- \) in case of \( E^-(x, y, z) \) and with

\[
\lambda_d = \sum_{d = pe \atop z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \alpha_{e,p}^+
\]

in case of \( E^+(x, y, z) \).
The requirement

- We need to show that

\[ \int_{X/2}^X \left| \sum_{d \leq D} \lambda_d \left( |A(x)_d| - \frac{h \log X}{d} \right) \right|^2 dy = O(hX) \]

with \( \lambda_d = \alpha_d^- \) in case of \( E^- (x, y, z) \) and with

\[ \lambda_d = \sum_{d=p e \atop z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \alpha_{e,p}^+ \]

in case of \( E^+ (x, y, z) \).

- In other words, we need type I information for almost all very short intervals with level of distribution \( D = X^{5/9} \) and some useful bilinear structure in the coefficients.

Kaisa Matomäki  Almost primes in almost all very short intervals
Mean square of type I sums

Let \( g : \mathbb{R} \to \mathbb{R} \) be a smooth, supported on \([1/4, 2]\), \( H = h \log X \)

\[
\int_{-\infty}^{\infty} g \left( \frac{y}{X} \right) \left| \sum_{d \leq D} \lambda_d \left( |A(x)_d| - \frac{H}{d} \right) \right|^2 dy
\]
Let $g: \mathbb{R} \to \mathbb{R}$ be a smooth, supported on $[1/4, 2]$, $H = h \log X$

\[
\int_{-\infty}^{\infty} g \left( \frac{y}{X} \right) \left| \sum_{d \leq D} \lambda_d \left( |A(x)_d| - \frac{H}{d} \right) \right|^2 dy
\]

\[
\ll HX \sum_{d \leq D} d \left( \sum_{m \leq D, \ m \equiv 0 \mod d} \lambda_m \right)^2 + H^3 X^\varepsilon
\]

\[
+ \sum_{0 < |k| \leq H} (H - |k|) \sum_{d_1, d_2 \leq D, \ (d_1, d_2) \mid k} \lambda_{d_1} \lambda_{d_2} \left( \sum_{\substack{m_1, m_2 \ \mid\mid d_1 m_1 = d_2 m_2 + k \ \text{gcd}(d_1, d_2) = 1}} g \left( \frac{d_1 m_1}{X} \right) - \hat{g}(0) \frac{X}{[d_1, d_2]} \right)
\]

\[
+ H \sum_{n} g \left( \frac{n}{X} \right) \left( \sum_{d \mid n} \lambda_d \right)^2 + HX \frac{1}{X^{10}} \sum_{n \leq X^{10}} \left( \sum_{d \mid n} \lambda_d \right)^2.
\]
Mean square of type I sums

Let \( g : \mathbb{R} \to \mathbb{R} \) be a smooth, supported on \([1/4, 2]\), \( H = h \log X \)

\[
\int_{-\infty}^{\infty} g \left( \frac{y}{X} \right) \left| \sum_{d \leq D} \lambda_d \left( |A(x)_d| - \frac{H}{d} \right) \right|^2 dy
\]

\[
\ll HX \sum_{d \leq D} d \left( \sum_{m \leq D} \frac{\lambda_m}{m} \right)^2 + H^3 X^\varepsilon
\]

\[
+ \sum_{0 < |k| \leq H} (H - |k|) \sum_{d_1, d_2 \leq D} \lambda_{d_1} \lambda_{d_2} \left( \sum_{m_1, m_2} g \left( \frac{d_1 m_1}{X} \right) - \hat{g}(0) \frac{X}{[d_1, d_2]} \right)
\]

\[
+ H \sum_{n} g \left( \frac{n}{X} \right) \left( \sum_{d | n} \lambda_d \right)^2 + HX \frac{1}{X^{10}} \sum_{n \leq X^{10}} \left( \sum_{d | n} \lambda_d \right)^2.
\]

First and third lines \( \ll hX \) utilizing definition of sieve coefficients.
Need to bound, for $H = h \log X$,

$$\sum_{0 < |k| \leq H} (H - |k|) \sum_{d_1, d_2 \leq D \mid (d_1, d_2) \mid k} \lambda_{d_1} \lambda_{d_2} \left( \sum_{m_1, m_2 \mid d_1 m_1 = d_2 m_2 + k} g \left( \frac{d_1 m_1}{X} \right) - \hat{g}(0) \frac{X}{[d_1, d_2]} \right).$$

with $\lambda_d = \alpha_d^-$ in case of $E^-(x, y, z)$ and with

$$\lambda_d = \sum_{\substack{d = pe \leq y \mid z \leq p \leq y}} \left( 1 - \frac{\log p}{\log y} \right) \alpha_{e, p}^+.\$$

in case of $E^+(x, y, z)$. 
The critical terms

Need to bound, for \( H = h \log X \),

\[
\sum_{0 < |k| \leq H} (H - |k|) \sum_{d_1, d_2 \leq D} \lambda_{d_1} \lambda_{d_2} \left( \sum_{d_1 m_1 = d_2 m_2 + k} g \left( \frac{d_1 m_1}{X} \right) - \hat{g}(0) X \frac{1}{[d_1, d_2]} \right).
\]

with \( \lambda_d = \alpha_d^- \) in case of \( E^-(x, y, z) \) and with

\[
\lambda_d = \sum_{d = pe \atop z \leq p < y} \left( 1 - \frac{\log p}{\log y} \right) \alpha_{e, p}^+.
\]

in case of \( E^+(x, y, z) \). Note that in both cases \( \lambda_d \) can be factored to type I and II sums since the linear sieve weights are well-factorable and Vaughan’s identity applicable to \( p \).
Need to bound, for \( H = h \log X \),

\[
\sum_{0 < |k| \leq H} (H - |k|) \sum_{d_1, d_2 \leq D} \lambda_{d_1} \lambda_{d_2} \left( \sum_{m_1, m_2} g \left( \frac{d_1 m_1}{X} \right) - \hat{g}(0) \frac{X}{[d_1, d_2]} \right)
\]
Need to bound, for $H = h \log X$,

$$
\sum_{0 < |k| \leq H} (H - |k|) \sum_{d_1, d_2 \leq D} \lambda_{d_1} \lambda_{d_2} \left( \sum_{m_1, m_2 \atop d_1 m_1 = d_2 m_2 + k} g \left( \frac{d_1 m_1}{X} \right) - \hat{g}(0) \frac{X}{[d_1, d_2]} \right).
$$

Concentrate on $(d_1, d_2) = 1$. The sum is over $m_1 \equiv \overline{d_1} k \pmod{d_2}$.
To Kloosterman sums

Need to bound, for $H = h \log X$,

$$
\sum_{0 < |k| \leq H} (H - |k|) \sum_{d_1, d_2 \leq D} \lambda_{d_1} \lambda_{d_2} \left( \sum_{\substack{m_1, m_2 \in \mathbb{Z} \mid d_1 m_1 = d_2 m_2 + k \frac{d_1 m_1}{X}}} g \left( \frac{d_1 m_1}{X} \right) - \hat{g}(0) \frac{X}{[d_1, d_2]} \right).
$$

Concentrate on $(d_1, d_2) = 1$. The sum is over $m_1 \equiv \overline{d_1} k \pmod{d_2}$ and by Poisson this is

$$
\leq HX \sum_{0 < |k| \leq H} \left| \sum_{d_1, d_2 \leq D} \frac{\lambda_{d_1} \lambda_{d_2}}{d_1 d_2} \sum_{\substack{\ell \in \mathbb{Z} \mid \ell \neq 0 \frac{\ell X}{d_1 d_2}} \sum_{d_1, d_2 \leq D \mid (d_1, d_2) = 1}} \hat{g} \left( \frac{\ell}{d_1 d_2} \right) e \left( - \frac{k \ell \overline{d_1}}{d_2} \right) \right|
$$

which is an average of incomplete Kloosterman sums.
Suffices to show that, for some $\varepsilon > 0$, 

$$
\sum_{0 < |k| \leq H} \left| \sum_{d_1, d_2 \leq D, (d_1, d_2) = 1} \lambda_{d_1} \lambda_{d_2} \sum_{\ell \in \mathbb{Z}, \ell \neq 0} \hat{g} \left( \frac{\ell X}{d_1 d_2} \right) e \left( -\frac{k \ell d_1}{d_2} \right) \right| \ll X^{-\varepsilon}.
$$
Suffices to show that, for some $\varepsilon > 0$,

$$
\sum_{0 < |k| \leq H} \left| \sum_{d_1, d_2 \leq D \atop (d_1, d_2) = 1} \lambda_{d_1} \lambda_{d_2} \sum_{\ell \in \mathbb{Z} \atop \ell \neq 0} \widehat{g} \left( \frac{\ell X}{d_1 d_2} \right) e \left( - \frac{k \ell d_1}{d_2} \right) \right| \ll X^{-\varepsilon}.
$$

Decompose $\lambda_d$ to type I and II sums and use Deshouillers-Iwaniec bounds for averages of Kloosterman sums.
The Kloosterman sums

Suffices to show that, for some \( \varepsilon > 0 \),

\[
\sum_{0 < |k| \leq H} \left| \sum_{d_1, d_2 \leq D} \frac{\lambda_{d_1} \lambda_{d_2}}{d_1 d_2} \sum_{\substack{\ell \in \mathbb{Z} \\ \ell \neq 0}} \hat{g} \left( \frac{\ell X}{d_1 d_2} \right) e \left( - \frac{k \ell d_1}{d_2} \right) \right| \ll X^{-\varepsilon}.
\]

Decompose \( \lambda_d \) to type I and II sums and use Deshouillers-Iwaniec bounds for averages of Kloosterman sums. They imply e.g.

**Lemma (Type II estimate)**

Assume that \( \alpha_n, \beta_n \) and \( \gamma_n \) are bounded complex coefficients. Let \( H \leq X^{1/60} \) and \( N \leq M \leq X^{21/50} \) and \( \max\{MN, Q\} \leq X^{14/25} \). Let \( g \) be smooth with compact support. Then

\[
\sum_{|k| \leq H} \left| \sum_{m \sim M \atop n \sim N} \frac{\alpha_m \beta_n}{mn} \sum_{q \sim Q} \frac{\gamma_q}{q} \sum_{\substack{\ell \in \mathbb{Z} \\ \ell \neq 0}} \hat{g} \left( \frac{\ell X}{mnq} \right) e \left( - \frac{k \ell mn}{q} \right) \right| \ll X^{-\frac{1}{1000}}
\]
Outline

1 Background and results
   - Primes
   - Primes in short intervals
   - Primes in almost all short intervals
   - Almost primes in (almost all) short intervals

2 Methods
   - The sieve method
   - Type I sums
   - To Kloosterman sums

3 Summary and further thoughts
Showed

**Theorem (M. (202?))**

Let $h \leq X^{1/100}$. There exist constants $c, C > 0$ such that

$$ch \leq \sum_{x-h \log X < n \leq x \atop p|n \implies p > X^{1/8}} 1_{\Omega(n) \leq 2} \leq Ch$$

for almost $x \in [X/2, X]$ apart from an exceptional set of measure $O(X/h)$. 

We used Richert's weighted sieve with well-factorability and Vaughan's identity. We got level of distribution $D = X^{5/9}$ (not optimized) from Deshouillers-Iwaniec bounds for averages of Kloosterman sums.
Showed

**Theorem (M. (202?))**

Let \( h \leq X^{1/100} \). There exist constants \( c, C > 0 \) such that

\[
ch \leq \sum_{x-h \log X < n \leq x, \ p|n \Rightarrow p > X^{1/8}} 1_{\Omega(n) \leq 2} \leq Ch
\]

for almost \( x \in [X/2, X] \) apart from an exceptional set of measure \( O(X/h) \).

We used Richert’s weighted sieve with well-factorability and Vaughan’s identity. We got level of distribution \( D = X^{5/9} \) (not optimized) from Deshouillers-Iwaniec bounds for averages of Kloosterman sums.
We have optimized neither the sieve weights or the level of distribution. Rather we have used a very simple sieve and worked out a sufficient level of distribution for that.
Now that we have shown that as soon as $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains $P_2$-numbers for almost all $x \in [X/2, X]$, it is natural to ask, what about primes?
Further thoughts — primes and $E_k$ numbers

- Now that we have shown that as soon as $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains $P_2$-numbers for almost all $x \in [X/2, X]$, it is natural to ask, what about primes?

- Unfortunately, there are no chances to replace $P_2$ by $P_1$ since we only use type I information. Due to the parity barrier, type I information never suffices for finding primes.

- Furthermore, our type I information is new only when the intervals are extremely short. In particular it does not help when trying to improve on Jia’s result that almost all intervals $(x - x^{1/20}, x]$ contain primes. Same issue for $E_k$ numbers.
Further thoughts — primes and $E_k$ numbers

- Now that we have shown that as soon as $h \to \infty$ with $X \to \infty$, the interval $(x - h \log X, x]$ contains $P_2$-numbers for almost all $x \in [X/2, X]$, it is natural to ask, what about primes?

- Unfortunately, there are no chances to replace $P_2$ by $P_1$ since we only use type I information. Due to the parity barrier, type I information never suffices for finding primes.

- Furthermore, our type I information is new only when the intervals are extremely short. In particular it does not help when trying to improve on Jia’s result that almost all intervals $(x - x^{1/20}, x]$ contain primes. Same issue for $E_k$ numbers.

- In an on-going work with J. Merikoski we are showing that if there are infinitely many exceptional characters, then there are many scales $X$ such that $(x - h \log X, x]$ contains primes for almost all $x \in (X/2, X]$ as soon as $h \to \infty$ with $X \to \infty$. 
Thank you!