Odd Moments in the Distribution of Primes

Vivian Kuperberg

Stanford University
Let $\delta > 0$ and let $Q > 1/\delta$. Fix $k \in \mathbb{N}_{\geq 2}$. Let $S(k)$ be the number of $k$-tuples $(\frac{a_1}{q_1}, \ldots, \frac{a_k}{q_k})$ that satisfy:

- $q_i \in [Q, 2Q]$ for all $i$
- $\frac{a_i}{q_i} \in (0, 1)$ is a fraction in lowest terms with $\left\| \frac{a_i}{q_i} \right\| \leq \delta$
- $\sum_{i=1}^{k} \frac{a_i}{q_i} \in \mathbb{Z}$

How big is this set, in terms of $\delta$ and $Q$?

When $k = 2$, $\frac{a_1}{q_1} + \frac{a_2}{q_2} = 1$ implies that $q_1 = q_2$ and $a_1 = q_1 - a_2$, so there are $Q^2 \delta$ solutions.

When $k$ is even, the main term comes from pairing fractions, so that $\frac{a_1}{q_1} = 1 - \frac{a_2}{q_2}$, $\frac{a_3}{q_3} = 1 - \frac{a_4}{q_4}$, and so on, so that $S(k) \sim Q^k \delta^{k/2}$.

What about when $k$ is odd?
The distribution of primes in short intervals

Motivating Question
Consider intervals of size $h$, with $h = o(N)$ and $h/\log N \to \infty$ as $N \to \infty$. What is the distribution of $\pi(n + h) - \pi(n)$ for $n \leq N$? What is the distribution of $\psi(n + h) - \psi(n)$ for $n \leq N$?

Cramér model answer
If we model the primes by saying each $n$ is independently prime with probability $\frac{1}{\log n}$, then the distribution of $\psi(n + h) - \psi(n)$ would be Gaussian with mean $\sim h$ and variance $\sim h \log N$. 

Testing the Cramér model

Here’s the (normalized) distribution of $\psi(n + h) - \psi(n)$ for $1 \leq n \leq 10^7$, with $h = \sqrt{10^7}$. The red line is the Gaussian with mean $h$ and variance $h \log 10^7$. 
Hardy–Littlewood Conjecture

Let $\mathcal{D} = \{d_1, \ldots, d_k\}$ be a sequence of distinct integers. As $N \to \infty$,

$$\sum_{n \leq N} \prod_{i=1}^{k} \Lambda(n + d_i) = \mathcal{G}(\mathcal{D})N + o(N)$$

where

$$\mathcal{G}(\mathcal{D}) = \prod_{p} \frac{1 - \nu_{\mathcal{D}}(p)/p}{(1 - 1/p)^k}$$

for $\nu_{\mathcal{D}}(p)$ is the number of equivalence classes mod $p$ occupied by $\mathcal{D}$.

When $\mathcal{D} = \{0, 2\}$, Hardy–Littlewood predicts the asymptotic number of twin primes, via

$$\sum_{n \leq N} \Lambda(n)\Lambda(n + 2) \sim 2 \left( \prod_{p \geq 3} \frac{1 - 2/p}{(1 - 1/p)^2} \right) N$$

When $\mathcal{D} = \{0, 1\}$, $\mathcal{G}(\mathcal{D}) = 0$, since the factor at $p = 2$ is $\frac{1 - 2/2}{(1 - 1/2)^2} = 0$. “Either $n$ or $n + 1$ is even, so there are very few consecutive primes.”
Variance via Hardy–Littlewood

The Hardy–Littlewood conjectures tell us that the variance is smaller.

\[
\frac{1}{N} \sum_{n \leq N} \left( \sum_{\ell \leq h} \Lambda(n + \ell) - h \right)^2
\]

\[
\sim \frac{1}{N} \sum_{n \leq N} \sum_{\ell \leq h} \Lambda(n + \ell)^2 + \frac{2}{N} \sum_{\ell_1 < \ell_2 \leq h} \sum_{n \leq N} \Lambda(n + \ell_1) \Lambda(n + \ell_2) - h^2
\]

\[
\sim \frac{1}{N} \sum_{n \leq N} \sum_{\ell \leq h} \Lambda(n + \ell)^2 + 2 \sum_{\ell \leq h} (h - \ell) \mathcal{G}(\{0, \ell\}) - h^2
\]

\[
\sim h (\log N - 1)
\]

\[
\sim h \left( \log \frac{N}{h} + B - 1 \right)
\]

The Cramér guess, \( h \log N \), is bigger than this!
Here’s the (normalized) distribution of $\psi(n + h) - \psi(n)$ for $1 \leq n \leq 10^7$, with $h = \sqrt{10^7}$. The red line is the Gaussian with mean $h$ and variance $h \log 10^7$; the green line has mean $h$ and variance $\frac{1}{2} h \log 10^7 + h(B - 1)$. 
Moral: Sums of $\mathcal{G}$ give us information about the moments of the distribution. In 2004, Montgomery and Soundararajan showed, assuming Hardy–Littlewood, that the moments of the distribution of primes in short intervals converge to the moments of a Gaussian with variance $h(\log(N/h) + B - 1)$. Their work depends on the following key result:

**Theorem (Montgomery & Soundararajan, 2004)**

For $\mathcal{D} \subseteq \mathbb{N}$, let $\mathcal{G}_0(\mathcal{D}) = \sum_{\mathcal{J} \subseteq \mathcal{D}} (-1)^{|\mathcal{D} \setminus \mathcal{J}|} \mathcal{G}(\mathcal{J})$, and let

$$R_k(h) := \sum_{d_1, \ldots, d_k \atop 1 \leq d_i \leq h \atop \text{distinct}} \mathcal{G}_0(\mathcal{D}).$$

Then for any $k \in \mathbb{N}$,

$$R_k(h) = \mu_k(-h \log h + (B + 1)h)^{k/2} + O_{k, \varepsilon}(h^{k/2 - 1/(7k) + \varepsilon}).$$

For $k$ odd, we don’t know the asymptotic size of $R_k(h)$; we just have $R_k(h) = O_{k, \varepsilon}(h^{k/2 - 1/(7k) + \varepsilon})$. 
Conjecture (K., Lemke Oliver and Soundararajan)

For $k$ odd,

$$R_k(h) \asymp h^{(k-1)/2} (\log h)^{(k+1)/2}$$

Theorem (K.)

$$R_3(h) = O(h (\log h)^5).$$
Techniques: Adding Fractions

\[ R_k(h) \approx \sum_{q_1, \ldots, q_k \atop q_i > 1} \left( \prod_{i=1}^{k} \frac{\mu(q_i)}{\phi(q_i)} \right) \sum_{a_1, \ldots, a_k \atop 1 \leq a_i \leq q_i \atop (a_i, q_i) = 1} \sum_{\sum_i a_i/q_i \in \mathbb{Z}} E \left( \frac{a_1}{q_1} \right) \cdots E \left( \frac{a_k}{q_k} \right), \]

where \( E(\alpha) = \sum_{m=1}^{h} e(m\alpha). \)

\( E(\alpha) \) is about \( h \) if \( \|\alpha\| \leq \frac{1}{h} \) and small otherwise.

\[ R_k(h) \approx \sum_{q_1, \ldots, q_k \atop q_i > 1} \left( \prod_{i=1}^{k} \frac{\mu(q_i)}{\phi(q_i)} \right) h^k \# \left\{ 1 \leq a_i \leq q_i : \left\| \frac{a_i}{q_i} \right\| \leq \frac{1}{h}, (a_i, q_i) = 1, \sum_i \frac{a_i}{q_i} \in \mathbb{Z} \right\}. \]

Intuition for \( k = 3 \): about \( \phi(q_1)^{1/h} \) choices for \( a_1 \), and \( \phi(q_2)^{1/h} \) choices for \( a_2 \).
The problem in $F_q[t]$

<table>
<thead>
<tr>
<th>primes in $\mathbb{Z}$</th>
<th>irreducible polynomials in $F_q[t]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>n</td>
</tr>
<tr>
<td>interval $(n, n + h)$</td>
<td>$I(F(t), h) := {G(t) :</td>
</tr>
<tr>
<td></td>
<td>$h = q^\ell$</td>
</tr>
</tbody>
</table>

$$\mathcal{G}(D) = \prod_p \frac{1 - \nu_p(D)/p}{(1 - 1/p)^k}$$

$$\mathcal{G}(D) = \prod_p \frac{1 - \nu_p(D)/|P|}{(1 - 1/|P|)^k}$$

$$R_k(h) = \sum_{1 \leq d_1, \ldots, d_k \leq h \text{ distinct}} \mathcal{G}_0(d_1, \ldots, d_k)$$

$$R_k(h) = \sum_{D_1, \ldots, D_k \in F_q[t]} \mathcal{G}_0(D_1, \ldots, D_k)$$

$$R_k(h) \asymp h^{(k-1)/2} (\log h)^{(k+1)/2} \quad \text{for } k \text{ odd}$$

$$R_k(h) \asymp h^{(k-1)/2} (\log h)^{(k+1)/2} \quad \text{for } k \text{ odd}$$
In $\mathbb{F}_q[t]$,

$$R_k(h) \approx \sum_{Q_1, \ldots, Q_k \in \mathbb{F}_q[t], |Q_i| > 1} \prod_{\text{monic}}^k \frac{\mu(Q_i)}{\phi(Q_i)} \sum_{A_1, \ldots, A_k, |A_i| < |Q_i|, (A_i, Q_i) = 1, \sum_i A_i/Q_i = 0} E \left( \frac{A_1}{Q_1} \right) \cdots E \left( \frac{A_k}{Q_k} \right)$$

- For $\alpha(t) = \frac{F(t)}{G(t)} \in \mathbb{F}_q(t)$, $E(\alpha) = \sum_{M \in I(0, h)} e(M\alpha)$
- For $\text{res}(\alpha)$ the coefficient of $\frac{1}{t}$ in the Laurent series expansion of $\alpha$, $e(\alpha) = \exp(2\pi i \cdot \text{tr}(\text{res}(\alpha)))$

**Lemma (Hayes, 1966)**

*Let $\alpha \in \mathbb{F}_q(t)$ with $|\alpha| \leq \frac{1}{q}$. Then*

$$E(\alpha) = \begin{cases} h & \text{if } |\alpha| < \frac{1}{h} \\ 0 & \text{if } |\alpha| \geq \frac{1}{h}. \end{cases}$$
Theorem (K.)

Fix $q$. As $h \to \infty$, 

$$R_3(h) = O(h(\log h)^{19/2})$$

and for all $\varepsilon > 0$, 

$$R_5(h) = O_\varepsilon(h^{2+\varepsilon}).$$
Techniques: Fifth moment bound

We want to bound

\[ R_5(h) \approx h^5 \sum_{Q_1, \ldots, Q_k} \prod_{i=1}^k \frac{\mu(Q_i)}{\phi(Q_i)} \# \left\{ A_i \bmod Q_i : \left| \frac{A_i}{Q_i} \right| < -h, (A_i, Q_i) = 1, \sum_i \frac{A_i}{Q_i} = 0 \right\}. \]

The adaptation of the work of Montgomery–Soundararajan and Montgomery–Vaughan gives a sharp enough bound for the sum restricted over all terms \( Q_1, \ldots, Q_k \) except those such that:

- \( |Q_i| \geq h \) for all \( i \)
- no three \( Q_i \)'s are equal
- for any \( i, j \), either \( Q_i = Q_j \) or \( \frac{Q_i}{(Q_i, Q_j)} \geq h \) and \( |(Q_i, Q_j)| < h/2 \)
For a tuple $Q_1, ... Q_5$,
- $|Q_i| \geq h$ for all $i$
- no three $Q_i$’s are equal
- for any $i, j$, either $Q_i = Q_j$ or $\left| \frac{Q_i}{Q_i, Q_j} \right| \geq h$ and $|(Q_i, Q_j)| < h/2$

implies that, possibly after reordering,
- $\left| \frac{Q_2}{(Q_2, Q_1)} \right| \geq h$ and
- $\left| \frac{Q_3}{(Q_3, Q_1 Q_2)} \right| \geq h/2$.

Strategy: Count options for $A_1/Q_1$, then remaining options for $A_2/Q_2$ after accounting for what has already been determined, then remaining options for $A_3/Q_3$. 
Lemke Oliver and Soundararajan (2016) conjecture that consecutive primes in arithmetic progressions exhibit biases. With $\pi(x_0) = 10^8$, they have the following data:

<table>
<thead>
<tr>
<th>$a$</th>
<th>$b$</th>
<th>$\pi(x_0; 10, (a, b))$</th>
<th>$a$</th>
<th>$b$</th>
<th>$\pi(x_0; 10, (a, b))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4623042</td>
<td>7</td>
<td>1</td>
<td>6373981</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>429438</td>
<td>3</td>
<td>3</td>
<td>6755195</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>504612</td>
<td>7</td>
<td>7</td>
<td>4439355</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>5442345</td>
<td>9</td>
<td>9</td>
<td>7431870</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>6010982</td>
<td>9</td>
<td>1</td>
<td>7991431</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4442562</td>
<td>3</td>
<td>3</td>
<td>6372941</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7043695</td>
<td>7</td>
<td>7</td>
<td>6012739</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>7502896</td>
<td>9</td>
<td>9</td>
<td>4622916</td>
</tr>
</tbody>
</table>

Showing that sums of three-term singular series are small improves their heuristic. It may be possible to show that this behavior follows from Hardy–Littlewood.
Further questions

Question 1.

Let $\delta > 0$ and $Q > 1/\delta$. For $k$ odd, what is

$$\# \left\{ q_1, \ldots, q_k \in [Q, 2Q], 1 \leq a_i \leq q_i : \left\| \frac{a_i}{q_i} \right\| \leq \delta, \sum_i \frac{a_i}{q_i} \in \mathbb{Z} \right\}?$$

Question 2.

For $\delta > 0, Q > 1/\delta$, let $J_1, \ldots, J_k \subseteq [0, 1]$ be intervals with $|J_i| \geq \delta$. What is

$$\# \left\{ q_1, \ldots, q_k \in [Q, 2Q], 1 \leq a_i \leq q_i : \frac{a_i}{q_i} \in J_i, \sum_i \frac{a_i}{q_i} \in \mathbb{Z} \right\}?$$

Question 3.

Let $h \in \mathbb{R}_{>0}$. For arithmetic progressions $a_1 \mod q_1, \ldots, a_k \mod q_k$, what is

$$\sum_{h_1, \ldots, h_k \leq h \atop h_i \equiv a_i \mod q_i} \mathcal{S}_0(\{h_1, \ldots, h_k\})?$$
Thank you!