Probabilistic aspects of character sums
Lecture 3: Moving intervals

Adam J Harper
University of Warwick

SSANT Paris, June 2021

Recall: r is a large prime, χ is a (non-principal) Dirichlet character mod r.

Plan of the talk:

- Introduction to the moving intervals problem
- First thoughts about a random model
- Previous work
- A negative and a positive result
- Open questions

The problem

Investigate the statistical behaviour of

$$S_{\chi,H}(x) := \sum_{x<n\leq x+H} \chi(n)$$

as $1 \leq x \leq r$ varies, where $H = H(r)$ is some length function.

Ideally, we would like to understand the behaviour for each fixed $\chi \neq \chi_0$.

We might make the problem easier by only seeking results for "almost all" χ mod r, or (even easier) varying χ in addition to varying x.

First thoughts

- As $1 \leq x \leq r$ varies, most of its values will be fairly large (e.g. larger than $\sqrt{r}$).
- So to understand the behaviour of $S_{\chi,H}(x)$, we will need to keep the Pólya Fourier expansion (PFE) in mind:

$$S_{\chi,H}(x) = \sum_{n \leq x+H} \chi(n) - \sum_{n \leq x} \chi(n)$$

$$= \frac{\tau(\chi)}{2\pi i} \sum_{0<|k|\leq r} \frac{\overline{\chi}(-k)}{k} (e(k(x+H)/r) - e(kx/r)) + O(\log r)$$

$$= \frac{\tau(\chi)}{2\pi i} \sum_{0<|k|\leq r} \frac{\overline{\chi}(-k)}{k} e(kx/r)(e(kH/r) - 1) + O(\log r)$$

$$\approx \frac{\tau(\chi)H}{r} \sum_{0<|k|\leq r/H} \overline{\chi}(-k) e(kx/r).$$

- $S_{\chi,H}(x) \approx \frac{\tau(\chi)H}{r} \sum_{0<|k|\leq r/H} \overline{\chi}(-k)e(kx/r)$.
- Compare with Lecture 2: here we have lost the denominator k.
- If $\sqrt{r} \leq H \leq r$, so that $r/H \leq \sqrt{r}$, then we might try modelling/investigating the values $\chi(-k)$ using random multiplicative functions. *This seems reasonable for "almost all" χ, but could there be some pathological χ that behave differently?*
- If $H \leq \sqrt{r}$, the random multiplicative model doesn't look so helpful.

Previous work

Theorem 1 (Davenport & Erdős, 1952)
If $\chi = \left(\frac{\cdot}{r}\right)$ is the Legendre symbol; and if the function H satisfies $H \to \infty$ but $(\log H)/\log r \to 0$ as the prime $r \to \infty$; and if $X \in \{0, 1, ..., r-1\}$ is uniformly random; then one has convergence in distribution to a standard Gaussian,

$$\frac{S_{\chi,H}(X)}{\sqrt{H}} \xrightarrow{d} N(0,1) \qquad \text{as } r \to \infty.$$

More explicitly, for any fixed $z \in \mathbb{R}$ we have

$$\mathbb{P}\left(\frac{S_{\chi,H}(X)}{\sqrt{H}} \le z\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt \qquad \text{as } r \to \infty.$$

What about non-real characters χ?

Mak & Zaharescu, 2011: if one chooses a non-real character χ modulo each prime r (in any way), then under the same conditions on H as Davenport and Erdős we have

$$\Re\frac{S_{\chi,H}(X)}{\sqrt{H}}, \quad \Im\frac{S_{\chi,H}(X)}{\sqrt{H}} \xrightarrow{d} N(0,1/2) \quad \text{as } r \to \infty.$$

Lamzouri, 2013: if one chooses a non-real character χ modulo each prime r (in any way), then under the same conditions on H as Davenport and Erdős we have

$$\frac{S_{\chi,H}(X)}{\sqrt{H}} \xrightarrow{d} Z_1 + iZ_2 \quad \text{as } r \to \infty,$$

where Z_1, Z_2 are independent $N(0,1/2)$ random variables.

The condition $(\log H)/\log r \to 0$ arises because all these theorems are proved using the *method of moments*.

For example, Davenport and Erdős (with $\chi = \left(\frac{\cdot}{r}\right)$) calculated

$$\frac{1}{r}\sum_{0\le x\le r-1}\left(\frac{S_{\chi,H}(x)}{\sqrt{H}}\right)^j = \frac{1}{rH^{j/2}}\sum_{1\le h_1,\dots,h_j\le H}\sum_x \left(\frac{x+h_1}{r}\right)\cdots\left(\frac{x+h_j}{r}\right),$$

showing that for each fixed $j\in\mathbb{N}$ this converges to the standard normal moment $(1/\sqrt{2\pi})\int_{-\infty}^{\infty} z^j e^{-z^2/2}dz$ as $r\to\infty$.

This uses the Weil bound:

- given a tuple $(h_1,\dots,h_j)$ of shifts, if any shift h occurs with odd multiplicity then the sum over x is $\ll_j \sqrt{r}$;
- under the condition $(\log H)/\log r \to 0$, all these terms are

$$\ll_j \frac{1}{\sqrt{r}H^{j/2}} \sum_{\substack{1\le h_1,\dots,h_j\le H,\\ \text{a shift occurs with odd multiplicity}}} 1 \le \frac{H^{j/2}}{\sqrt{r}} \to 0 \quad \text{as } r\to\infty.$$

Theorem 2 (Chatterjee & Soundararajan, 2012)

If $f(n)$ is a Rademacher random multiplicative function, and $y = y(x)$ satisfies $x^{1/5} \log x \ll y = o(x/\log x)$, then

$$\frac{\sum_{x<n\leq x+y} f(n)}{\sqrt{\mathbb{E}\left(\sum_{x<n\leq x+y} f(n)\right)^2}} \xrightarrow{d} N(0,1) \quad \text{as } x \to \infty.$$

Motivated by Theorem 2, Lamzouri made the following conjecture.

Conjecture 1 (Lamzouri, 2013)

If we choose a non-real character χ modulo each prime r (in any way), then provided $H \to \infty$ but $H = o(r/\log r)$ we have

$$\frac{S_{\chi,H}(X)}{\sqrt{H}} \xrightarrow{d} Z_1 + iZ_2 \quad \text{as } r \to \infty,$$

where Z_1, Z_2 are independent $N(0,1/2)$ random variables.

Is Lamzouri's conjecture reasonable?

- At first look, it seems plausible.
- But as already discussed, this problem involves values $\chi(n)$ for large n (much bigger than $\sqrt{r}$). So we need to use the PFE (or something else that encodes periodicity) to properly analyse the situation.
- Also, a random multiplicative function is expected to model a randomly chosen character. Here the character is fixed, and the start point X of the interval varies randomly.

A negative result

Theorem 3 (H.)

Let $A > 0$ be arbitrary but fixed, and set $H(r) = r/\log^A r$. Then as r varies over large primes, there exists a corresponding sequence of non-real characters χ modulo r for which

$$\frac{S_{\chi,H}(X)}{\sqrt{H}} \overset{d}{\not\to} Z_1 + iZ_2 \quad \text{as } r \to \infty,$$

where Z_1, Z_2 are independent $N(0, 1/2)$ random variables.

Theorem 3 shows that Lamzouri's conjecture is false.

One can prove an analogous negative result for real characters $\left(\frac{\cdot}{r}\right)$.

Key steps in the proof:

- When $H = r/\log^A r$, the PFE implies that

$$\frac{S_{\chi,H}(x)}{\sqrt{H}} \approx \frac{\tau(\chi)\sqrt{H}}{r} \sum_{0<|k|\le r/H} \overline{\chi}(-k)e(kx/r)$$
$$= \frac{\tau(\chi)\sqrt{H}}{r} \sum_{0<|k|\le \log^A r} \overline{\chi}(-k)e(kx/r).$$

- For any fixed $A > 0$, we can find non-real Dirichlet characters χ mod r for which $\chi(k)$ looks "sort of like" 1 for all $1 \le k \le \log^A r$.
- For such χ, our sum $S_{\chi,H}(x)/\sqrt{H}$ will look "sort of like" the scaled Dirichlet kernel $\frac{\tau(\chi)\sqrt{H}}{r}\sum_{0<|k|\le \log^A r} e(kx/r)$, which shouldn't have Gaussian behaviour.

Second part:
Granville and Soundararajan, 2001: for any $A > 0$ and any prime r that is large enough in terms of A, there exist (many) χ mod r such that

$$|\sum_{n \leq \log^A r} \chi(n)| \gtrsim \rho(A) \log^A r,$$

where $\rho(A) > 0$.

(Roughly speaking, the χ produced are such that $\chi(p) \approx 1$ for all $p \leq \log r$. The lower bound comes from the contribution from $\log r$-smooth numbers n, which are a positive proportion of all $n \leq \log^A r$.)

Third part:
To rigorously exploit our lower bound $|\sum_{n\leq \log^A r}\chi(n)| \gtrsim \rho(A)\log^A r$, we need:

Lemma 1
Let $0 \leq \tau < 1$, and suppose $(V_n)_{n=1}^{\infty}$ is a sequence of random variables satisfying $\mathbb{E}|V_n|^2 \leq \tau$ for all n. Then if Z is any random variable such that $\mathbb{E}|Z|^2 = 1$, we have

$$V_n \overset{d}{\nrightarrow} Z \quad \text{as } n \to \infty.$$

Proof of Lemma 1.

Choose $a \in \mathbb{R}$ such that $\mathbb{E}\min\{|Z|^2, a^2\} \geq (1+\tau)/2$ (such a exists by the monotone convergence theorem).

Since $v \mapsto \min\{|v|^2, a^2\}$ is a continuous *bounded* function on $\mathbb{C}$, if we had $V_n \xrightarrow{d} Z$ then we would have

$$\mathbb{E}\min\{|V_n|^2, a^2\} \to \mathbb{E}\min\{|Z|^2, a^2\} \quad \text{as } n \to \infty.$$

But this is impossible, since
$\mathbb{E}\min\{|V_n|^2, a^2\} \leq \mathbb{E}|V_n|^2 \leq \tau < (1+\tau)/2.$ □

Then if we let α be such that $\sum_{0<k\leq\log^A r}\overline{\chi}(-k) = \alpha\sum_{0<k\leq\log^A r} 1$, and set $G_{\chi,H}(x) := \frac{\alpha\tau(\chi)\sqrt{H}}{r}\sum_{1\leq k\leq\log^A r} e(kx/r)$, we have

$$\frac{S_{\chi,H}(x)}{\sqrt{H}} \approx \frac{\tau(\chi)\sqrt{H}}{r}\sum_{0<|k|\leq\log^A r}\overline{\chi}(-k)e(kx/r)$$

$$= \left(\frac{\tau(\chi)\sqrt{H}}{r}\sum_{0<|k|\leq\log^A r}\overline{\chi}(-k)e(kx/r) - G_{\chi,H}(x)\right) + G_{\chi,H}(x).$$

- Using the formula for the sum of a geometric progression, we see $G_{\chi,H}(x)$ is small whenever $\sqrt{rH} \leq x \leq r - \sqrt{rH}$.
- This is almost all values of x mod r, so if $\frac{S_{\chi,H}(X)}{\sqrt{H}}$ converges in distribution to $Z_1 + iZ_2$ then the same must be true for $(\frac{\tau(\chi)\sqrt{H}}{r}\sum_{0<|k|\leq\log^A r}\overline{\chi}(-k)e(kX/r) - G_{\chi,H}(X))$.

- But if we compute

$$\mathbb{E}\left|\frac{\tau(\chi)\sqrt{H}}{r}\sum_{0<|k|\le\log^A r}\overline{\chi}(-k)e(kX/r) - G_{\chi,H}(X)\right|^2,$$

(where $X \in \{0, 1, ..., r-1\}$ is uniformly random), we find this is $\approx 1 - |\alpha|^2$.

- We have $\mathbb{E}|Z_1 + iZ_2|^2 = \mathbb{E}Z_1^2 + \mathbb{E}Z_2^2 = 1$.
- And by our choice of χ, we know $|\alpha|$ is bounded away from zero (it is $\gtrsim \rho(A)$).
- So Theorem 3 follows from Lemma 1.

□

A positive result

The characters χ that we use to disprove Lamzouri's conjecture are rather pathological, so we might hope the conjecture could at least be true for "almost all" characters mod r.

Theorem 4 (H.)

Let $H = H(r)$ be a function satisfying $\frac{\log(r/H)}{\log r} \to 0$ but $H = o(r)$ as $r \to \infty$. Then there exist sets $\mathcal{G}_r$ of characters mod r, satisfying $\frac{\#\mathcal{G}_r}{r-1} \to 1$, such that for $\chi \in \mathcal{G}_r$ we have

$$\frac{S_{\chi,H}(X)}{\sqrt{H}} \xrightarrow{d} Z_1 + iZ_2 \quad \text{as } r \to \infty.$$

The condition that $H = o(r)$ is natural (so the number $\approx r/H$ of terms in the PFE tends to infinity).
The condition that $\frac{\log(r/H)}{\log r} \to 0$ is (presumably) just an artefact of the proof.

Key steps in the proof:

- Prove the analogous result for $\frac{\sqrt{H}}{\sqrt{r}} \sum_{0<|k| \leq r/H} f(-k) e(kX/r)$, where $f(k)$ is a Steinhaus random multiplicative function.
- This can be done using *martingale theory*. Another treatment, via moments and a non-trivial point counting problem, is in a 2020 paper of Benatar, Nishry and Rodgers.
- Compare the moments of $\frac{\sqrt{H}}{\sqrt{r}} \sum_{0<|k| \leq r/H} f(-k) e(kX/r)$ with those of $\frac{\sqrt{H}}{\sqrt{r}} \sum_{0<|k| \leq r/H} \overline{\chi}(-k) e(kX/r)$. *This is where the condition* $\frac{\log(r/H)}{\log r} \to 0$ *is used.*

Open questions

As a replacement for Lamzouri's conjecture, I tentatively propose:

Conjecture 2

If we choose a non-real character χ *modulo each prime* r *(in any way), then provided* $H \to \infty$ *and* $\frac{\log(r/H)}{\log\log r} \to \infty$ *we have*

$$\frac{S_{\chi,H}(X)}{\sqrt{H}} \xrightarrow{d} Z_1 + iZ_2 \quad \text{as } r \to \infty,$$

where Z_1, Z_2 *are independent* $N(0, 1/2)$ *random variables.*

(And the analogous conjecture for real characters.)

Open question: Prove Conjecture 2.

But beware!

To prove the conjecture, we need to know that the construction used in Theorem 3 cannot be extended, so we need to know (at least) that

$$\sum_{n \le r/H} \chi(n) = o(r/H) \ \forall \ \chi \ne \chi_0 \bmod r, \quad \text{provided } \frac{\log(r/H)}{\log\log r} \to \infty.$$

We only know how to prove this assuming GRH, so a proof of the conjecture will (probably!) need to be conditional.

Open question: Prove the conjecture for "almost all" χ on a range of H where it is unknown, e.g. when $H = \sqrt{r}$.

This will require finding an alternative to the method of moments (or perhaps a much cleverer application of it).