TAO'S RESOLUTION
OF THE ERDŐS DISCREPANCY PROBLEM

K. SOUNDARARAJAN

ABSTRACT. This article gives a simplified account of some of the ideas behind Tao's resolution of the Erdős discrepancy problem.

The Erdős discrepancy problem is an easily stated question about arbitrary functions f from the positive integers to ± 1. It asks whether the signs ± 1 can be arranged evenly over all subsequences of the form kj for a given $k \in \mathbb{N}$ and as j varies. Precisely, must it always be the case that

$$
(1) \qquad \sup_{k,n \in \mathbb{N}} \left| \sum_{j=1}^{n} f(kj) \right| = \infty?
$$

The question appears in many of Erdős's lists of unsolved problems [3–5], and in [3] he dates the conjecture to the 1930s. Erdős highlights a striking special case of the problem: Suppose f is a completely multiplicative function (that is, $f(mn) = f(m)f(n)$ for all natural numbers m and n) taking the values ± 1. Then is it true that

$$
(2) \qquad \sup_{n \in \mathbb{N}} \left| \sum_{j=1}^{n} f(j) \right| = \infty?
$$

Since $f(kj) = f(k)f(j)$ in this situation, clearly a positive solution to problem (1) implies a resolution of problem (2) as well. So as not to keep the reader in suspense, let us state at once that the Erdős discrepancy problem was answered affirmatively by Tao [24]:

Theorem 1 (Tao). *For any function $f : \mathbb{N} \to \{-1, 1\}$, we have*

$$
\sup_{k,n \in \mathbb{N}} \left| \sum_{j=1}^{n} f(kj) \right| = \infty.
$$

In particular, the partial sums of a completely multiplicative function taking the values ± 1 are unbounded; that is, (2) holds.

The Erdős discrepancy problem asks whether every two-coloring of the natural numbers must exhibit some disorder when viewed along homogenous arithmetic

progressions jk (for $1 \le j \le n$). One may weaken the problem to allow all arithmetic progressions $a + jk$: Is it true that for all $f : \mathbb{N} \to \{-1, 1\}$,

$$(3) \qquad \sup_{a,k,n \in \mathbb{N}} \left| \sum_{j=1}^{n} f(a + jk) \right| = \infty?$$

Roth [20] established the existence of such irregularities of distribution. Indeed, more generally he showed that if A is a subset of the integers up to N with $|A| = \rho N$, then there exists an absolute positive constant c such that

$$\sup_{\substack{k \le \sqrt{N} \\ a \bmod k}} \left| \sum_{\substack{n \in A \\ n \equiv a \bmod k}} 1 - \frac{\rho N}{q} \right| \ge c \sqrt{\rho(1 - \rho)} N^{1/4}.$$

In other words, the only subsets of $[1, N]$ that are evenly distributed in all arithmetic progressions with moduli below $\sqrt{N}$ are (essentially) the empty set (with $\rho = 0$) or the whole set (with $\rho = 1$). The weak Erdős discrepancy problem (3) is only relevant for $\rho \approx 1/2$, and so Roth's theorem establishes (3), showing even a strong quantitative version. Matoušek and Spencer [17] have shown that the $N^{1/4}$ bound in Roth's result is best possible.

The Erdős discrepancy problem for homogeneous progressions (with one fewer degree of freedom) has proved much more difficult, in part because the discrepancy can be much smaller than one might expect. A random sequence of ± 1 would exhibit disorder in its partial sums up to N, typically on the scale of $\sqrt{N}$ (the central limit theorem) and occasionally on the scale of $\sqrt{N \log \log N}$ (the law of the iterated logarithm). In Roth's theorem, the discrepancy along all progressions is still of size a power of N, even if not $\sqrt{N}$. However, it is not hard to construct ± 1 sequences where the discrepancy along homogeneous progressions grows only logarithmically, and this slow rate of growth indicates why the Erdős discrepancy problem is so delicate.

Example 1. Consider the function

$$\chi(n) = \begin{cases} 1 & \text{if } n \equiv 1 \mod 3, \\ -1 & \text{if } n \equiv 2 \mod 3, \\ 0 & \text{if } n \equiv 0 \mod 3. \end{cases}$$

Then χ is a completely multiplicative function and is periodic with period 3 (that is, it is a Dirichlet character mod 3) and its partial sums are bounded (since every three consecutive terms sum to zero). Therefore, the left-hand sides of (1) and (2) are both finite. More generally, one can take any quadratic character mod q (generalizations of the Legendre symbol mod p), to get examples of completely multiplicative functions that are also periodic and which have bounded partial sums. Of course, this is not a counterexample to (1) and (2) since χ takes the value 0 in addition to ± 1.

Example 2. A completely multiplicative function may be specified by its values on the primes. We tweak Example 1, defining the completely multiplicative function $\widetilde{\chi}$ by setting $\widetilde{\chi}(p) = \chi(p)$ if $p \ne 3$, and $\widetilde{\chi}(3) = 1$. Then for any k and n we have

$$\left| \sum_{j=1}^{n} \widetilde{\chi}(kj) \right| = \left| \sum_{j=1}^{n} \widetilde{\chi}(j) \right| = \left| \sum_{\substack{\ell \ge 0 \\ 3^\ell \le n}} \sum_{m \le n/3^\ell} \chi(m) \right| \le 1 + \frac{\log n}{\log 3}.$$

On the other hand, taking $k = 1$ and $n = 1 + 3 + 3^2 + \cdots + 3^r$, we find

$$\Big| \sum_{j=1}^{n} \widetilde{\chi}(n) \Big| = \Big| \sum_{\ell=0}^{r} \sum_{m \leq 3^{r-\ell}+1} \chi(m) \Big| = r + 1 = \Big\lceil \frac{\log n}{\log 3} \Big\rceil.$$

Thus, in this example, the discrepancy along homogeneous progressions does go to infinity, but only at a slow logarithmic pace. In [1], Borwein, Choi, and Coons carry out an analysis of examples of this type, settling the Erdős discrepancy problem for modified characters $\widetilde{\chi}_p$ defined to be the Legendre symbol mod p on primes $\ell \neq p$, and with $\widetilde{\chi}_p(p) = 1$.

Example 3. *Numerical examples.* The sequence $(1, -1, -1, 1, -1, 1, 1, -1, -1, 1, 1)$ of length 11 has discrepancy 1 along homogeneous progressions, and it is the longest such sequence [13]. In [12] it is shown that the longest sequence of ± 1 with discrepancy 2 along homogeneous progressions has size 1160 and that there is a sequence of length 13000 with discrepancy 3.

While the examples above indicate the delicate nature of the Erdős discrepancy problem, Tao's proof in fact gives the following more general result.

Theorem 2 (Tao). *Let $\mathcal{H}$ be a Hilbert space, and let $f : \mathbb{N} \to \mathcal{H}$ be a function with $\|f(n)\|_{\mathcal{H}} = 1$ for all n. Then*

$$\sup_{k,n \in \mathbb{N}} \Big\| \sum_{j=1}^{n} f(jk) \Big\|_{\mathcal{H}} = \infty.$$

Example 4. The Hilbert space in Theorem 2 can be real or complex. In particular, Theorem 2 applies to any function f from the natural numbers to the unit circle $\{|z| = 1\}$. To specialize once more, rather than looking at ± 1 completely multiplicative functions (as in (2)), we may consider any completely multiplicative function $f : \mathbb{N} \to \{|z| = 1\}$ and ask whether

$$\text{(4)} \qquad\qquad \sup_{n \in \mathbb{N}} \Big| \sum_{j=1}^{n} f(j) \Big| = \infty.$$

Of course, Theorem 2 provides a positive answer to this question.

Example 5. In Theorem 2 the discrepancy can grow as slowly as $\sqrt{\log n}$. Take $\mathcal{H}$ to be a Hilbert space with orthonormal basis $e_0, e_1, e_2, \ldots$. Write $n \in \mathbb{N}$ as $3^a b$, where a is a nonnegative integer and b is coprime to 3. Then set $f(n) = \chi(b) e_a$ where χ, the Dirichlet character mod 3, is as in Example 1. Now if $k = 3^c d$ (with c a nonnegative integer and d coprime to 3), then

$$\Big\| \sum_{j=1}^{n} f(kj) \Big\|_{\mathcal{H}} = \Big\| \sum_{\substack{a \geq 0 \\ 3^a \leq n}} e_{c+a} \sum_{b \leq n/3^a} \chi(b) \Big\|_{\mathcal{H}} \leq \sqrt{1 + \log n / \log 3}.$$

As in Example 2, this bound is attained if $n = 1 + 3 + \cdots + 3^r$.

The Erdős discrepancy problem remained dormant for a long time, with no promising avenues of attack. In December 2009, Gowers proposed the Erdős discrepancy problem as a possible Polymath project, continuing his idea of "massively collaborative mathematics" which began earlier in 2009 with a new (and quantitative) proof of the density Hales-Jewett theorem (originally due to Furstenberg and Katznelson). Another successful Polymath project is Polymath 8 from 2013,

on improving the bounds on small gaps between primes after the breakthroughs of Zhang, and Maynard and Tao. Polymath 5, the project on the Erdős discrepancy problem, began in January 2010 and was active until the end of 2012. Many interesting reformulations of the problem were found, and several of these are recounted in Gowers's article [7]. The website [19] documents the various Polymath discussions and blog posts centered around this problem.

For Tao's eventual resolution of the problem, two developments from the Polymath 5 project proved important: First, the general Erdős discrepancy problem (in the form of general ± 1 sequences as in (1) or in the more general Hilbert space situation as in Theorem 2) can be deduced from the special case of completely multiplicative functions taking values on the unit circle (as in (4)). Second, if such a completely multiplicative function correlates with a Dirichlet character (in a sense to be made precise later), then the discrepancy does tend to infinity. In other words, if we are close to the situation of Example 2, then (generalizing the work of [1]) it is possible to show that the discrepancy must grow, even if only very slowly. Tao himself was a participant in the Polymath 5 project and played a major role in both these developments.

The missing link is to show that completely multiplicative functions taking values on the unit circle and with bounded partial sums must correlate with characters. This is the most subtle part of Tao's argument, and it is carried out in [25]. The starting point is a recent breakthrough of Matomäki and Radziwiłł [14] in understanding multiplicative functions in short intervals, with further important refinements due to Matomäki, Radziwiłł, and Tao [15]. On top of this, Tao brings to the problem several other novel ideas, some motivated by work in additive combinatorics.

In the rest of this article, I want to give an overview of the ideas behind Theorems 1 and 2, oversimplifying the situation to convey the flavor of the arguments.

Reducing the problem to completely multiplicative functions. Suppose $\mathcal{H}$ is a Hilbert space, and suppose $f(n)$ is a sequence of unit vectors in $\mathcal{H}$ with bounded discrepancy,

$$(5) \qquad \sup_{k,n} \Big\| \sum_{j=1}^{n} f(kn) \Big\| \le C.$$

From this we wish to construct completely multiplicative functions taking values on the unit circle and with small discrepancy; we will not exactly achieve this but something good enough.

Let X be large, and let M be an integer much larger than X. Let $p_1, \ldots, p_r$ denote the primes below X. Define $F : (\mathbb{Z}/M\mathbb{Z})^r \to \mathcal{H}$ by setting

$$F(a_1, \ldots, a_r) = f(p_1^{a_1} \cdots p_r^{a_r}),$$

provided $0 \le a_i \le M - 1$, and then extending the definition periodically in each coordinate. Given a natural number n below X, write its prime factorization as $n = p_1^{a_1} \cdots p_r^{a_r}$ so that the exponents a_i are nonnegative integers. Set $\pi(n) = (a_1, \ldots, a_r)$, which we view as an element of $(\mathbb{Z}/M\mathbb{Z})^r$.

If $j \le X$, then note that the coordinates of $\pi(j)$ are all smaller than $\lceil \log X / \log 2 \rceil$. Therefore if $x = (x_1, \ldots, x_r)$ is such that $0 \le x_i \le M - X$ for all i, then for all $n \le X$ we have

$$(6) \qquad \Big\| \sum_{j=1}^{n} F(x + \pi(j)) \Big\|_{\mathcal{H}} = \Big\| \sum_{j=1}^{n} f(j p_1^{x_1} \cdots p_r^{x_r}) \Big\|_{\mathcal{H}} \le C.$$

Now there are M^r vectors $x \in (\mathbb{Z}/M\mathbb{Z})^r$, and the bound (6) applies to the vast majority of such x; namely, for $(M - X + 1)^r = M^r + O(XM^{r-1})$ of them, and recall that M is much larger than X. For the few values of x not covered by (6), there holds the trivial bound

$$\Big\| \sum_{j=1}^{n} F(x + \pi(j)) \Big\|_{\mathcal{H}} \leq n \leq X.$$

Combining this trivial bound with the estimate (6), we conclude that for all $n \leq X$

$$(7) \qquad \frac{1}{M^r} \sum_{x \in (\mathbb{Z}/M\mathbb{Z})^r} \Big\| \sum_{j=1}^{n} F(x + \pi(j)) \Big\|_{\mathcal{H}}^2 \leq C^2 + 1,$$

assuming that M is large enough compared with X.

With $e(t) = e^{2\pi i t}$, recall the Fourier transform on $(\mathbb{Z}/M\mathbb{Z})^r$: for $\xi \in (\mathbb{Z}/M\mathbb{Z})^r$, put

$$\widehat{F}(\xi) = \frac{1}{M^r} \sum_{x \in (\mathbb{Z}/M\mathbb{Z})^r} F(x) e\Big(-\frac{x \cdot \xi}{M} \Big),$$

which is an element of $\mathcal{H}$. The Fourier inversion formula gives

$$F(x) = \sum_{\xi \in (\mathbb{Z}/M\mathbb{Z})^r} \widehat{F}(\xi) e\Big(\frac{x \cdot \xi}{M} \Big),$$

and Parseval's formula reads

$$(8) \qquad \sum_{\xi \in (\mathbb{Z}/M\mathbb{Z})^r} \|\widehat{F}(\xi)\|_{\mathcal{H}}^2 = \frac{1}{M^r} \sum_{x \in (\mathbb{Z}/M\mathbb{Z})^r} \|F(x)\|_{\mathcal{H}}^2 = 1.$$

Use the Fourier inversion formula

$$\sum_{j=1}^{n} F(x + \pi(j)) = \sum_{\xi \in (\mathbb{Z}/M\mathbb{Z})^r} \widehat{F}(\xi) e\Big(\frac{x \cdot \xi}{M} \Big) \Big(\sum_{j=1}^{n} e\Big(\frac{\pi(j) \cdot \xi}{M} \Big) \Big),$$

so that by Parseval and (7) we obtain

$$(9) \qquad \sum_{\xi \in (\mathbb{Z}/M\mathbb{Z})^r} \|\widehat{F}(\xi)\|_{\mathcal{H}}^2 \Big| \sum_{j=1}^{n} e\Big(\frac{\pi(j) \cdot \xi}{M} \Big) \Big|^2 \leq C^2 + 1.$$

We can now give a probabilistic interpretation of the estimate (9). We define a probability space of random completely multiplicative functions $\mathbf{f} : [1, X] \to \{|z| = 1\}$, by setting $\mathbf{f}(p_j) = e(\xi_j/M)$ for all $1 \leq j \leq r$ with probability $\|\widehat{F}(\xi)\|_{\mathcal{H}}^2$ where $\xi = (\xi_1, \ldots, \xi_r) \in (\mathbb{Z}/M\mathbb{Z})^r$. Parseval's formula (8) shows that this is indeed a probability space. Further, the estimate (9) may be recast as

$$(10) \qquad \mathbb{E}\Big(\Big| \sum_{j \leq n} \mathbf{f}(j) \Big|^2 \Big) \leq C^2 + 1,$$

for all $n \leq X$.

Thus for all large X, we have constructed a probability space (depending on X) of random completely multiplicative functions with values on the unit circle such that the expected value of partial sums up to n (for all $n \leq X$) is bounded. By a compactness argument, Tao shows that one can construct a space of random completely multiplicative functions such that (10) holds for all n instead of just n below X.

The estimate (10) forms the basis for Tao's proof of Theorem 2. Note that (10) is an average statement, and one cannot extract from it a (deterministic) completely multiplicative function with all partial sums up to X being bounded. For each n below X there clearly exists such a function, but perhaps no one function works for all $n \leq X$. If we average (10) over all $n \leq X$, then

$$\mathbb{E}\Big(\frac{1}{X}\sum_{n \leq X}\Big|\sum_{j \leq n}\mathbf{f}(j)\Big|^2\Big) \leq C^2 + 1,$$

so that there must exist completely multiplicative functions f, taking values on the unit circle, such that

$$\frac{1}{X}\sum_{n \leq X}\Big|\sum_{j \leq n}f(j)\Big|^2 \leq C^2 + 1.$$

We expect that this estimate cannot hold for large enough X, but this remains an open problem.

Let us highlight one interesting feature of this construction. Most values of $(a_1, \ldots, a_r)$ (with $0 \leq a_i \leq M-1$) will satisfy $a_i \geq A$ for all i and any fixed natural number A. Therefore, most of the values of F will be built out of f evaluated at multiples of $\prod_{p \leq X} p^A$. This shows the importance of needing all the values of f to be unit vectors; if f took small values along the multiples of some number (as in Example 1, or for any Dirichlet character) this will force F to be small almost all the time, and the construction leads nowhere.

Correlations of values of completely multiplicative functions. Given a sequence of unit vectors in a Hilbert space $\mathcal{H}$ violating Theorem 2, the argument above produces a probability space of random completely multiplicative functions with partial sums bounded in expectation. For the sake of simplicity, let us suppose that we actually have a completely multiplicative function with values on the unit circle, and bounded partial sums; that is, a violation to (4).

To motivate the ideas that follow, we begin with a brief discussion of partial sums of multiplicative functions. Apart from the constant function 1, and Dirichlet characters (homomorphisms from $(\mathbb{Z}/q\mathbb{Z})^* \to \mathbb{C}^*$), perhaps the best known multiplicative functions are the Möbius function $\mu(n)$ and the Liouville function $\lambda(n)$. The Liouville function is completely multiplicative and is defined on the primes by $\lambda(p) = -1$; thus $\lambda(n) = 1$ if n has an even number of prime factors (counted with multiplicity) and is -1 otherwise. The Möbius function $\mu(n)$ equals $\lambda(n)$ whenever n is square-free, and $\mu(n) = 0$ if n is divisible by the square of some prime; thus the Möbius function is merely multiplicative ($f(mn) = f(m)f(n)$ whenever m and n are coprime) rather than being completely multiplicative. For the Möbius and Liouville functions the first main goal was to exhibit cancellation in their partial sums, since the estimates

$$(11) \qquad \lim_{N \to \infty} \frac{1}{N}\sum_{n \leq N}\mu(n) = \lim_{N \to \infty} \frac{1}{N}\sum_{n \leq N}\lambda(n) = 0$$

can be elementarily shown to be equivalent to the prime number theorem. In fact we believe that $\lambda(n)$ behaves in many ways like a random collection of signs ± 1 (but there are limits to this belief), and that the partial sums above exhibit roughly square-root cancellation—this is related to the Riemann hypothesis. Of course our goal is to show that partial sums must get large sometimes; for the Möbius function

(and a similar result holds for $\lambda(n)$), let us remark that the partial sums are known to be larger than $\sqrt{N}$ infinitely often, disproving a conjecture of Mertens (see [18]).

Given a completely multiplicative function f with $f(n) = \pm 1$, Erdős and Wintner asked whether

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} f(n)$$

always exists. If the limits in (11) exist, then it is not hard to show that they must be 0; so the Erdős–Wintner question offers another way of generalizing and thinking about the prime number theorem. Wintner [27] showed that if the function f is mostly like the constant function 1, then the limit above does exist; precisely, if

$$(12) \quad \sum_p \frac{1 - f(p)}{p} < \infty, \qquad \text{then} \qquad \frac{1}{N} \sum_{n \leq N} f(n) \to \prod_p \left(1 - \frac{f(p)}{p}\right)^{-1} \left(1 - \frac{1}{p}\right).$$

The harder direction (which includes the Liouville function) is to show that when the prime sum $\sum_p (1 - f(p))/p$ diverges, then the limiting average of $f(n)$ exists, and equals 0. This was beautifully resolved by Wirsing [28], who showed that

$$(13) \quad \sum_p \frac{1 - f(p)}{p} = \infty \qquad \text{implies} \qquad \frac{1}{N} \sum_{n \leq N} f(n) \to 0,$$

thereby settling this question of Erdős and Wintner.

For a complex valued completely multiplicative function f with $|f(n)| = 1$, the story is a little more complicated. The function $f(n) = n^{i\alpha}$ for a fixed real number α has no limiting average value: indeed, by comparing with the integral,

$$\frac{1}{N} \sum_{n \leq N} n^{i\alpha} \sim \frac{1}{N} \frac{N^{1+i\alpha}}{1 + i\alpha} = \frac{N^{i\alpha}}{1 + i\alpha},$$

which oscillates with N. Halász [9] found the right generalization of Wirsing's work and developed an ingenious analytic method to show that if

$$(14) \quad \sum_p \frac{1 - \mathrm{Re}\ f(p)p^{-i\alpha}}{p} = \infty \qquad \text{for all } \alpha \in \mathbb{R}, \text{ then} \qquad \frac{1}{N} \sum_{n \leq N} f(n) \to 0.$$

Given two multiplicative functions f and g taking values in the unit disc, it is convenient to define the distance between them (up to some point X) as

$$\mathbb{D}(f, g; X)^2 = \sum_{p \leq X} \frac{1 - \mathrm{Re}\ f(p)\overline{g(p)}}{p},$$

and this satisfies the properties of a pseudo-metric, notably the triangle inequality; see [8] for further discussion. Thus the hypotheses in (13) and (14) may be interpreted as f not being close to the function 1 or to any function of the form $n^{i\alpha}$; in the language of [8], this is stated as f not pretending to be 1 or $n^{i\alpha}$, or (abusing the English language) as f being "unpretentious".

Let us now return to the Erdős discrepancy problem. Suppose f is a completely multiplicative function with $|f(n)| = 1$ and bounded partial sums. Since the partial sums are bounded, the sums $\sum_{h=1}^{H} f(n + h)$ must also be bounded for any interval

$[n + 1, n + H]$. Therefore, one must have

$$\frac{1}{N} \sum_{n \leq N} \Big| \sum_{h=1}^{H} f(n + h) \Big|^2 \leq C$$

for some constant C. Expanding the left-hand side above gives

$$\frac{1}{N} \sum_{n \leq N} \sum_{h_1 = h_2 \leq H} |f(n + h_1)|^2 + \frac{1}{N} \sum_{n \leq N} \sum_{h_1 \neq h_2 \leq H} f(n + h_1) \overline{f(n + h_2)}$$

$$= H + \sum_{h_1 \neq h_2 \leq H} \frac{1}{N} \sum_{n \leq N} f(n + h_1) \overline{f(n + h_2)}.$$

If H is chosen to be large compared to C, then there must exist $h_1 \neq h_2 \leq H$ with

$$(15) \qquad \frac{1}{N} \Big| \sum_{n \leq N} f(n + h_1) \overline{f(n + h_2)} \Big| \geq \frac{1}{2H}.$$

Thinking of N as being very large compared to H, the above estimate says that there is a strong correlation between the values of f at n and $n + h$ with $h = h_2 - h_1 \neq 0$. This type of reasoning is often referred to as a *van der Corput argument*, originating in van der Corput's work on bounding exponential sums.

When can a multiplicative function f correlate with a shift of itself? Clearly the function $f(n) = 1$ does, and so does a function that pretends to be 1. More generally the function $n^{i\alpha}$, or anything pretending to be $n^{i\alpha}$, will correlate with its shifts: for large n there is not much difference between $n^{i\alpha}$ and $(n + 1)^{i\alpha}$. Less obviously, if f is a Dirichlet character $\bmod q$, then the periodicity $\bmod q$ implies that $f(n)\overline{f(n + q)} = 1$ whenever $(n, q) = 1$. One can generalize this by taking $\chi(n)n^{i\alpha}$ for any real number α, and still more generally one can take any function pretending to be $\chi(n)n^{i\alpha}$. Note that $\widetilde{\chi}$ in Example 2 is a function pretending to be the character χ. In analogy with the results of Wirsing and Halász mentioned earlier, we may hope that these examples exhaust all the possibilities for a multiplicative function correlating with a shift of itself. Indeed Elliott [2] formulated such a conjecture; it needs a small technical correction, which is made in [15].

Unfortunately, very little is understood about correlations of multiplicative functions. For example, take the Liouville function $\lambda(n)$, which we said earlier is expected to look like a random sequence of signs ± 1. Furthermore, the Liouville function is known not to correlate with Dirichlet characters; this is essentially the prime number theorem in arithmetic progressions. Thus we should definitely expect that as $N \to \infty$,

$$(16) \qquad \frac{1}{N} \sum_{n \leq N} \lambda(n)\lambda(n + 1) \to 0,$$

which is also equivalent to saying that all four sign patterns $(+, +)$, $(+, -)$, $(-, +)$, $(-, -)$ occur roughly equally often among consecutive values of $\lambda(n)$. More generally, Chowla conjectured that if $a_j n + b_j$ (for $1 \leq j \leq k$) are affine functions with no two proportional to each other ($a_i b_j \neq b_i a_j$ for all $i \neq j$), then as $N \to \infty$,

$$\frac{1}{N} \sum_{n \leq N} \lambda(a_1 n + b_1) \cdots \lambda(a_k n + b_k) \to 0.$$

Chowla's conjecture implies that among k consecutive values of λ, all 2^k possible patterns of signs occur equally frequently.

The first breakthrough toward these problems came with the work of Matomäki and Radziwiłł [14]. They showed that the Liouville function exhibits cancellation in almost all intervals $[n, n + h]$, as soon as $h \to \infty$ (however slowly). This was a remarkable advance on earlier results which required h to grow like a power of n. Toward (16) they established that for all large N

$$\left| \sum_{n \leq N} \lambda(n)\lambda(n+1) \right| \leq (1 - \delta)N,$$

for some $\delta > 0$; from this it follows that all four sign patterns of $(\lambda(n), \lambda(n + 1))$ appear a positive proportion of the time. Following this breakthrough, Matomäki, Radziwiłł, and Tao [16] showed that all eight patterns of signs for $(\lambda(n), \lambda(n + 1), \lambda(n + 2))$ occur a positive proportion of the time, and moreover in [15] they showed an average version of the Chowla conjecture. While we have discussed just the Liouville function, the results in [14] and [15] apply more generally to all multiplicative functions. For an account of these papers, see the author's Bourbaki exposition [23].

These breakthroughs are still insufficient to show estimates like (16). However, Tao [25] realized that a weaker logarithmic version of (16) could be established. Namely, he showed that as $N \to \infty$,

$$\frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \lambda(n)\lambda(n+1) \to 0.$$

More generally, Tao showed a logarithmic version of the (corrected) Elliott conjecture: If f is a multiplicative function with $|f(n)| \leq 1$ for all n and such that, for two affine functions $a_1n + b_1$ and $a_2n + b_2$ with $a_1b_2 \neq a_2b_1$,

$$\frac{1}{\log N} \left| \sum_{n \leq N} \frac{1}{n} f(a_1n + b_1)\overline{f(a_2n + b_2)} \right|$$

is bounded away from 0, then there is a character χ to small modulus, and a small real number α such that $f(n)$ pretends to be $\chi(n)n^{i\alpha}$. This is the key result underlying Tao's proof of Theorems 1 and 2. It builds on the work of Matomäki, Radziwiłł, and Tao—most notably their work in [15]—and adding several other novel ideas, particularly an "entropy decrement argument" reminiscent of ideas from additive combinatorics. The subject is currently undergoing a dramatic and rapid transformation (see [6, 10, 11, 26]) and we simply mention two striking recent results. From the work of Frantzikinakis and Host [6], it follows that for any k integers $h_1, \ldots, h_k$ and any irrational number α, one has

$$\frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \lambda(n + h_1) \cdots \lambda(n + h_k)e^{2\pi in\alpha} \to 0, \qquad \text{as } N \to \infty.$$

This is a step toward the Möbius *disjointness conjectures* of Sarnak [21, 22]. The work of Tao and Teräväinen [26] takes a further step toward the logarithmic Chowla and Elliott conjectures. They show that for any odd natural number k and any k integers $h_1, \ldots, h_k$, one has

$$\frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \lambda(n + h_1) \cdots \lambda(n + h_k) \to 0, \qquad \text{as } N \to \infty.$$

As a consequence, they are able to show that all 16 possible sign patterns of $(\lambda(n), \lambda(n+1), \lambda(n+2), \lambda(n+3))$ occur a positive proportion of the time, extending the result of Matomäki, Radziwiłł, and Tao mentioned earlier.

Coming back again to the Erdős discrepancy problem, Tao's logarithmic version of the Chowla and Elliott conjectures suffices to show that a completely multiplicative function f with $|f(n)| = 1$ and bounded partial sums must correlate with functions of the form $\chi(n)n^{i\alpha}$. Indeed given a completely multiplicative f with bounded partial sums, we can carry out a logarithmic van der Corput argument:

$$\frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \Big| \sum_{h=1}^{H} f(n+h) \Big|^2 \leq C,$$

and expanding the above we obtain (in place of (15))

$$\frac{1}{\log N} \Big| \sum_{n \leq N} \frac{1}{n} f(n+h_1) \overline{f(n+h_2)} \Big| \geq \frac{1}{2H},$$

for some $h_1 \neq h_2 \leq H$. Now invoking Tao's work [25], it follows that f must correlate with (or pretend to be) $\chi(n)n^{i\alpha}$ for a suitable character χ and a real number α.

Finishing the proof. We are now at the last stage of the proof. Suppose for simplicity that f is a completely multiplicative function taking values ± 1 and with bounded partial sums. As described above, f must correlate with some function of the form $\chi(n)n^{i\alpha}$, and since f is real, one can ensure that $\alpha = 0$ and χ is real as well. Let q denote the conductor of the character χ, and define $\widetilde{\chi}$ to be a completely multiplicative function with $\widetilde{\chi}(p) = \chi(p)$ for all $p \nmid q$, and $\widetilde{\chi}(p) = f(p)$ for $p | q$. Put $f(n) = \widetilde{\chi}(n)g(n)$ for a completely multiplicative function g taking values ± 1. Assume that the correlation of f and χ takes the simplified form

$$(17) \qquad \sum_{p} \frac{1 - f(p)\widetilde{\chi}(p)}{p} = \sum_{p} \frac{1 - g(p)}{p} \leq C,$$

for some constant C.

The idea is to decouple the character-like function $\widetilde{\chi}$ from g (which is close to the function 1). Let H be a large integer, and let k be such that $2^k > 2H$. Let X be much larger than q^k. Since the partial sums of f are bounded, it follows that

$$\Big| \sum_{n \equiv 0 \bmod q^k} \frac{1}{n^{1+1/\log X}} \sum_{h=1}^{H} f(n+h) \Big| \leq A \frac{\log X}{q^k},$$

for some constant A. If $n \equiv 0 \bmod q^k$, then for all $h \leq H$ we see that $(n+h, q^k) = (h, q^k)$ must be a divisor of q^{k-1}, and therefore $f(n+h) = \widetilde{\chi}(n+h)g(n+h) = \widetilde{\chi}(h)g(n+h)$. Using this above, we obtain

$$(18) \qquad \Big| \sum_{h=1}^{H} \widetilde{\chi}(h) \sum_{n \equiv 0 \bmod q^k} \frac{g(n+h)}{n^{1+1/\log X}} \Big| \leq A \frac{\log X}{q^k}.$$

The hypothesis (17) says that g is close to the constant function 1 and, rather as in Wintner's result (12), it is not hard to show that

$$\sum_{n \equiv 0 \bmod q^k} \frac{g(n+h)}{n^{1+1/\log X}} \approx \frac{\log X}{q^k} \prod_{p} \Big(1 - \frac{1}{p^{1+1/\log X}}\Big) \Big(1 - \frac{g(p)}{p^{1+1/\log X}}\Big)^{-1} \approx \mathfrak{S} \frac{\log X}{q^k},$$

for some constant $\mathfrak{S} > 0$. Roughly, this says that g is equidistributed in residue classes mod q^k, and in establishing this, one uses that $(h, q^k)|q^{k-1}$, and that $g(p) = 1$ (by construction) for all $p|q$. Inserting this in (18) we conclude that

$$(19) \qquad \Big| \sum_{h=1}^{H} \widetilde{\chi}(h) \Big| \leq B,$$

for some constant B.

Thus, from knowing that the partial sums of f are bounded, we have passed to knowing that the partial sums of the character-like function $\widetilde{\chi}$ are bounded. At this stage, let us simplify our task once more and assume that the modulus q is a prime number. We are now back to the situation of Example 2! Write H in base q as $H = h_0 + h_1 q + \cdots + h_r q^r$, where $0 \leq h_j \leq q - 1$ and $h_r \geq 1$. Then, a small calculation shows

$$\sum_{h=1}^{H} \widetilde{\chi}(h) = \sum_{j=0}^{r} f(q)^j \Big(\sum_{n \leq h_j} \chi(n) \Big).$$

Choose $h_j = 0$ if $f(q)^j = -1$ (which happens only if $f(q) = -1$ and j is odd) and $h_j = 1$ if $f(q)^j = 1$ (which happens whenever j is even, and if $f(q) = 1$, then for all j). With this choice the above is at least $\lfloor \log H/(2 \log q) \rfloor$, which goes to infinity with H. This contradicts (19), and completes our (oversimplified) proof sketch.

About the author

K. Soundararajan is professor of mathematics at Stanford University. His interests are number theory, analysis, probability, and combinatorics.

References

[1] P. Borwein, S. K. K. Choi, and M. Coons, *Completely multiplicative functions taking values in* $\{-1, 1\}$, Trans. Amer. Math. Soc. **362** (2010), no. 12, 6279–6291, DOI 10.1090/S0002-9947-2010-05235-3. MR2678974

[2] P. D. T. A. Elliott, *On the correlation of multiplicative functions*, Notas Soc. Mat. Chile **11** (1992), no. 1, 1–11. MR1292619

[3] P. Erdős, *Some unsolved problems*, Michigan Math. J. **4** (1957), 291–300. MR0098702

[4] P. Erdős, *Some of my favourite unsolved problems*, A tribute to Paul Erdős, Cambridge Univ. Press, Cambridge, 1990, pp. 467–478. MR1117038

[5] P. Erdős and R. L. Graham, *Old and new problems and results in combinatorial number theory*, Monographies de L'Enseignement Mathématique [Monographs of L'Enseignement Mathématique], vol. 28, Université de Genève, L'Enseignement Mathématique, Geneva, 1980. MR592420

[6] N. Frantzikinakis and B. Host. *The logarithmic Sarnak conjecture for ergodic weights.* `arXiv:1708.00677`, 2017.

[7] W. T. Gowers, *Erdős and arithmetic progressions*, Erdős Centennial, Bolyai Soc. Math. Stud., vol. 25, János Bolyai Math. Soc., Budapest, 2013, pp. 265–287, DOI 10.1007/978-3-642-39286-3_8. MR3203599

[8] A. Granville and K. Soundararajan, *Pretentious multiplicative functions and an inequality for the zeta-function*, Anatomy of Integers, CRM Proc. Lecture Notes, vol. 46, Amer. Math. Soc., Providence, RI, 2008, pp. 191–197. MR2437976

[9] G. Halász, *On the distribution of additive and the mean values of multiplicative arithmetic functions*, Studia Sci. Math. Hungar. **6** (1971), 211–233. MR0319930

[10] O. Klurman. *Correlations of multiplicative functions and applications*. Compositio Math., **153** (2017), 1622–1657.

[11] O. Klurman and A. Mangerel. *Rigidity theorems for multiplicative functions.* `arXiv:1707.07817`, 2017.

[12] B. Konev and A. Lisitsa, *A SAT attack on the Erdős discrepancy conjecture*, Theory and Applications of Satisfiability Testing—SAT 2014, Lecture Notes in Comput. Sci., vol. 8561, Springer, Cham, 2014, pp. 219–226, DOI 10.1007/978-3-319-09284-3_17. MR3252248

[13] A. R. D. Mathias, *On a conjecture of Erdős and Čudakov*, Combinatorics, Geometry and Probability (Cambridge, 1993), Cambridge Univ. Press, Cambridge, 1997, pp. 487–488. MR1476466

[14] K. Matomäki and M. Radziwiłł, *Multiplicative functions in short intervals*, Ann. of Math. (2) **183** (2016), no. 3, 1015–1056, DOI 10.4007/annals.2016.183.3.6. MR3488742

[15] K. Matomäki, M. Radziwiłł, and T. Tao, *An averaged form of Chowla's conjecture*, Algebra Number Theory **9** (2015), no. 9, 2167–2196, DOI 10.2140/ant.2015.9.2167. MR3435814

[16] K. Matomäki, M. Radziwiłł, and T. Tao, *Sign patterns of the Liouville and Möbius functions*, Forum Math. Sigma **4** (2016), e14, 44, DOI 10.1017/fms.2016.6. MR3513734

[17] J. Matoušek and J. Spencer, *Discrepancy in arithmetic progressions*, J. Amer. Math. Soc. **9** (1996), no. 1, 195–204, DOI 10.1090/S0894-0347-96-00175-0. MR1311824

[18] A. M. Odlyzko and H. J. J. te Riele, *Disproof of the Mertens conjecture*, J. Reine Angew. Math. **357** (1985), 138–160, DOI 10.1515/crll.1985.357.138. MR783538

[19] Polymath5. *The Erdős discrepancy problem*. Website at `http://michaelnielsen.org\penalty\z@/polymath1`, 2012.

[20] K. F. Roth, *Remark concerning integer sequences*, Acta Arith. **9** (1964), 257–260. MR0168545

[21] P. Sarnak, *Three lectures on the Möbius function: randomness and dynamics*, `https://www.math.ias.edu/files/wam/2011/PSMobius.pdf`, 2011.

[22] P. Sarnak, *Möbius randomness and dynamics*, Not. S. Afr. Math. Soc. **43** (2012), no. 2, 89–97. MR3014544

[23] K. Soundararajan, *The Liouville function in short intervals* [after Matomäki and Radziwiłł]. Seminaire Bourbaki, `arXiv:16016.08021`, 2016.

[24] T. Tao, *The Erdős discrepancy problem*, Discrete Anal., posted on 2016, Paper No. 1, 29, DOI 10.19086/da.609. MR3533300

[25] T. Tao, *The logarithmically averaged Chowla and Elliott conjectures for two-point correlations*, Forum Math. Pi **4** (2016), e8, 36, DOI 10.1017/fmp.2016.6. MR3569059

[26] T. Tao and J. Teräväinen, *The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures*, `arXiv:1708.02610`, 2017.

[27] A. Wintner, *The Theory of Measure in Arithmetical Semi-Groups*, publisher unknown, Baltimore, Md., 1944. MR0015083

[28] E. Wirsing, *Das asymptotische Verhalten von Summen über multiplikative Funktionen. II* (German), Acta Math. Acad. Sci. Hungar. **18** (1967), 411–467, DOI 10.1007/BF02280301. MR0223318

Department of Mathematics, Stanford University, 450 Serra Mall, Bldg. 380, Stanford, California 94305-2125

E-mail address: `ksound@math.stanford.edu`