From Szemerédi/Selberg to Green/Tao: Primes in arithmetic progressions

Christian Skau

Ben Green (1977 -)

Terence Tao (1975 -)

Atle Selberg (1917 - 2007)

Endre Szemerédi (1940 -)

Arithmetic progression (AP) of length L:

$$a, a+b, a+2b, a+3b, \cdots, a+(L-1)b$$
$$= \{a+kb \,|\, 0 \leq k \leq L-1\}.$$

Arithmetic progression (AP) of length L:

$$a, a+b, a+2b, a+3b, \cdots, a+(L-1)b$$
$$= \{a+kb \,|\, 0 \leq k \leq L-1\}.$$

Example

$$56211383760397 + k \cdot 44546738095860$$
$$0 \leq k \leq 22.$$

These are all primes! (Frind, Jobling, Underwood (2004)).

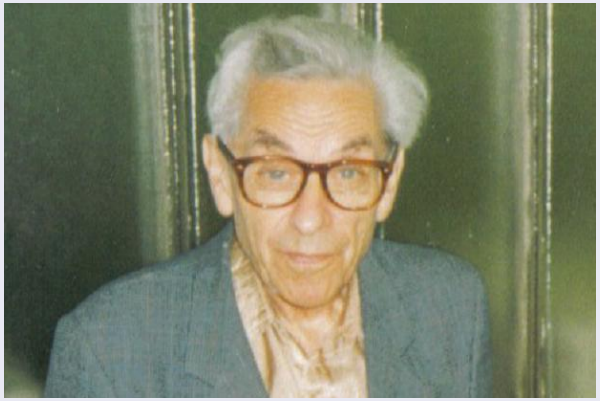
Erdős Conjecture (1973)

Let $a_1 < a_2 < a_3 < \cdots$ be an infinite set of natural numbers, such that

$$\sum_{n=1}^{\infty} \frac{1}{a_n} = \infty.$$

(In particular, the primes P satisfy this requirement.) Then $A = \{a_1, a_2, a_3, \cdots\}$ contains arithmetic progressions of arbitrary lengths.

Paul Erdős (1913 - 1996)



"Support" for Erdős Conjecture (as pertains to the primes):

The Prime Number Theorem implies that the density of primes around a large real number x is $\frac{1}{\log x}$. If we choose numbers in $\{1, 2, \cdots, N\}$ at random with probability $\frac{1}{\log N}$, then there ought to be approximately $\frac{N^2}{\log^L N}$ different AP's in $\{1, 2, \cdots, N\}$ of length L.

However, the primes are not randomly distributed: 2 is the only even prime, 3 is the only prime divisble by 3, etc.

Green and Tao (2004)

The primes P contains arithmetic progressions of arbitrary finite length.

Let L be the length of an AP in P. Green and Tao showed that there exists one where all the terms are less than the bound

$$2^{2^{2^{2^{2^{2^{2^{(100L)}}}}}}}$$

(The prime number theorem indicates that the bound should be $L! + 1$)

Extensions of the Green/Tao result

(i) (Green/Tao (2004)) There exist arithmetic progressions of arbitrary lengths in the set
$\{p \text{ prime} \big| p+2 \text{ prime or the product of two primes.}\}$. (Recall Chen's result from 1973).

(ii) (Tao/Ziegler (2006)) For $k \in \mathbb{N}$, let $F_1(x), \cdots, F_k(x)$ be any k polynomials over $\mathbb{Z}$ such that $F_i(0) = 0$, $i = 1, \cdots, k$. There exist $a, d \in \mathbb{N} = \{1, 2, 3, \cdots\}$ such that

$a + F_1(d), a + F_2(d), \cdots, a + F_k(d)$ are primes.

(Note that setting
$F_1(x) = 0, F_2(x) = x, F_3(x) = 2x, \cdots, F_k(x) = (k-1)x$
yields the original Green/Tao result.)

Theorem (Green and Tao)

Let $A \subseteq P$ be such that

$$\limsup_{N \to \infty} \frac{1}{\pi(N)} |A \cap \{1, 2, \cdots, N\}| > 0.$$

Then A contains AP's of any finite length.

$\pi(N) =$ number of primes in $\{1, 2, 3, \cdots, N\}$.

The prime number theorem (PNT)

$$\frac{\pi(x)}{\frac{x}{\log x}} \to 1$$

when $x \to \infty$.

The Erdős-Turan Conjecture (1936)

Let $A \subseteq \mathbb{N} = \{1, 2, 3, \cdots\}$ have upper positive density, i. e.

$$\bar{d}(A) = \limsup_{N \to \infty} \frac{|A \cap \{1, 2, 3, \cdots, N\}|}{N} > 0.$$

Then A contains arbitrary long AP's.

Szemerédi proved the conjecture in 1975 (now called *Szemerédi's Theorem*) by what has been characterized as a masterpiece of combinatorial reasoning.

Note that Szemerédi's Theorem does not apply to the primes P:

$$\limsup_{N \to \infty} \frac{|P \cap \{1, 2, 3, \cdots, N\}|}{N} = \limsup_{N \to \infty} \frac{\pi(N)}{N} = \lim_{N \to \infty} \frac{1}{\log N} = 0.$$

Theorem A (Finitary version of Szemerédi's Theorem)

Let $L \in \mathbb{N}$ and let $0 < \delta \leq 1$. There exists a natural number $N_0(\delta, L)$ such that if $N \geq N_0(\delta, L)$ and $A \subseteq \{1, 2, 3, \cdots, N\}$ with $|A| \geq \delta N$, then A contains an AP of length L.

Note that if $A = P \cap \{1, 2, 3, \cdot, N\}$, then $|A| = \pi(N)$, and since $\pi(N) \sim \frac{N}{\log N}$, the inequality $|A| \geq \delta N$ is not obtainable.

Green-Tao strategy of the proof

1) Modify (the finitary version of) Szemerédi's theorem. They establish a certain "transference principle" whereby Szemerédi's theorem can be applied in a more general setting.

2) Use specific properties of the primes and their distribution based on the Selberg sieve.

$$\liminf_{N\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = 0; \text{Goldston, Pintz, Yildrim (2005)}$$

$$\liminf_{N\to\infty} (p_{n+1} - p_n) \leq 5414; \text{Zhang (2013)}$$

$$P = \{p_1 < p_2 < p_3 < \cdots\}$$

Measure-preserving dynamical system $(X, \mathcal{B}, m, T)$

$(X, \mathcal{B}, m)$ Lebesgue measure space; $m(X) = 1$. $T : X \to X$ measure-preserving map, i. e. $m(T^{-1}B) = m(B)$ for all $B \in \mathcal{B}$.

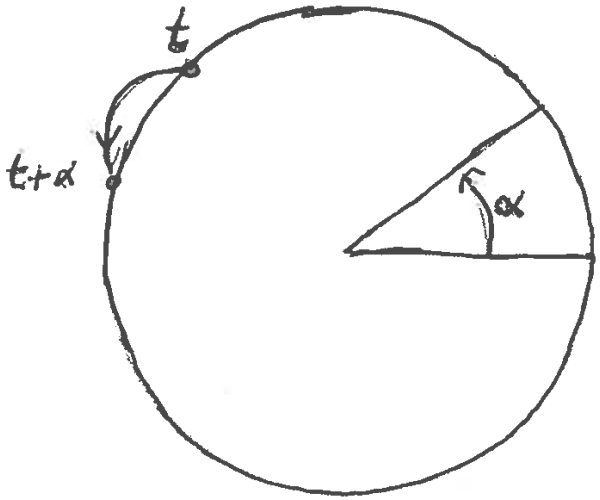
Example 1

$X = \mathbb{T}$, $T = \rho_\alpha : [0,1) \to [0,1) \pmod 1$, $t \to t + \alpha$.
$m =$ Lebesgue measure

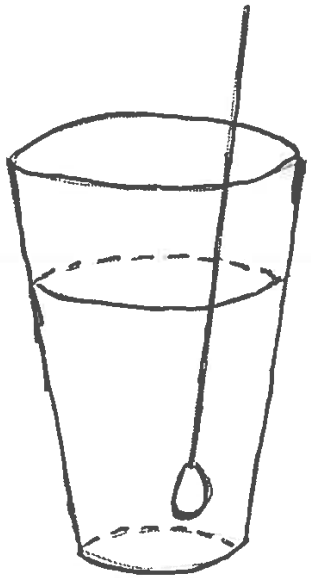
Example 2 (Bernoulli shift)

$X = \prod_{-\infty}^{\infty} \{0,1\}$, $T : X \to X$, $x(n) \to x(n+1)$, $m = \prod_{-\infty}^{\infty} \mu$,
$\mu(\{0\}) = \mu(\{1\}) = \frac{1}{2}$

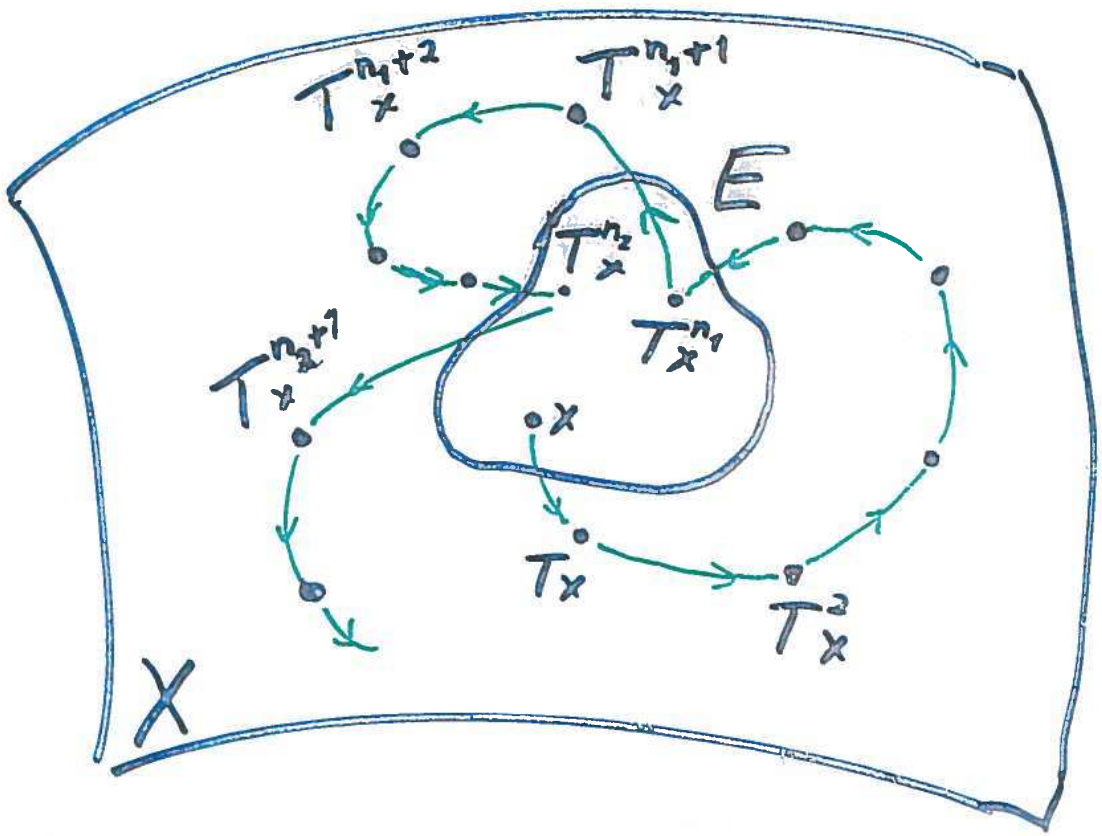
Example 1 exhibits the "rigid" case.



Example 2 exhibits the "mixing" (or "random") case.



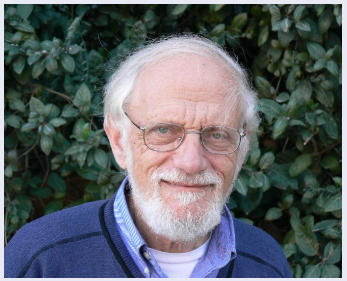
Recurrence



$x \in E \cap T^{-n_1}E \cap T^{-n_2}E$, and so $E \cap T^{-n_1}E \cap T^{-n_2}E \neq \emptyset$.
(Equivalently: $\chi_E \cdot \chi_E \circ T^{n_1} \cdot \chi_E \circ T^{n_2} \neq 0$)

In 1976 Furstenberg proved his *Multiple Recurrence Theorem*, which he showed was "equivalent" to Szemerédi's Theorem.

Hillel Furstenberg (1935 -)



Multiple Recurrence Theorem

Let $(X, \mathcal{B}, m, T)$ be a measure-preserving system, and let $L \in \mathbb{N}$. For any $E \in \mathcal{B}$ with $m(E) > 0$ there exists $n \in \mathbb{N}$ such that

$$m(E \cap T^{-n}E \cap T^{-2n}E \cap \cdots \cap T^{-(L-1)n}E) > 0.$$

(Equivalently: There exists $B \subseteq E$, $m(B) > 0$, such that $T^nB \subseteq E, T^{2n}B \subseteq E, \cdots, T^{(L-1)n}B \subseteq E$)

A precursor of Furstenberg's Multiple Recurrence Theorem is *Poincaré's Recurrence Theorem*:

Poincaré's Recurrence Theorem (1890)

Let $(X, \mathcal{B}, m, T)$ be a measure-preserving dynamical system, and let $L \in \mathbb{N}$. For any $E \in \mathcal{B}$, $m(E) > 0$, there exists distinct $n_1, n_2, \cdots, n_{L-1}$ in $\mathbb{N}$ such that

$$m(E \cap T^{-n_1}E \cap T^{-n_2}E \cap \cdots \cap T^{-n_{L-1}}E) > 0.$$

Marc Kac: "There are many proofs of this theorem, all of which are almost trivial. We have here another example of an important and even profound fact whose purely mathematical content is very much on the surface."

Walter Gottschalk: "On occasions a mathematician will have an insight that is ahead of the time in the sense that the insight is not fully expressible in the mathematical theory and language developed at the moment. For example the Poincaré Recurrence Theorem as first stated and proved by Poincaré was strictly not meaningful. What was needed was the language of Lebesgue measure which came later."

Henri Poincaré (1854 - 1912)

Hérmite: "Poincaré est un voyant auquel apparaissent les verités dans une vive luminière". ("Poincaré is a seer for whom the truth appears in a sharp light.")

Poincaré: "C'est ainsi, c'est comme cela".
("It's like that, it just is like that.")

King Oscar II Prize Competition (1890). Committee: Hérmite, Weierstrass, Mittag-Leffler.

Poincaré's original formulation of his recurrence theorem. (Section 8 of "Sur le problème des trois corps et les équations de la dynamique", Acta Mathematica 13 (1890), 1-270):

Théorème I

Supposons que la point P reste à distance finie, et que la volume $\int dx_1 dx_2 dx_3$ soit un invariant intégrale; si l'on considère une région r_0 quelconque, quelque petite que soit cette région, il y aura des trajectoires qui la transverseront une infinité de fois.

"Wiederkehreinwand" – controversy (1896).
Ernst Zermelo $\leftrightarrow$ Ludwig Boltzmann
(Nietzsche!)

Classical Mechanics (n degrees of freedom)

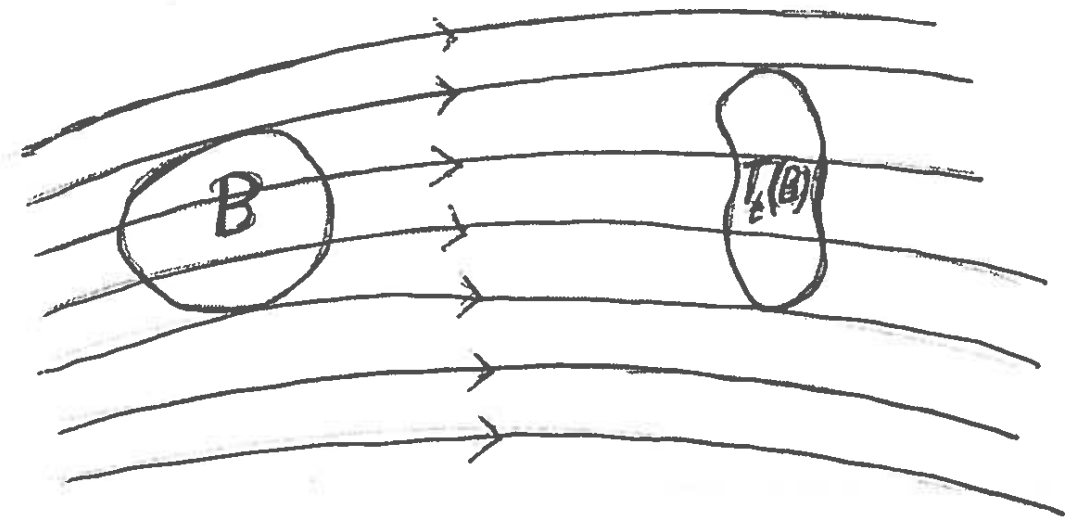
$q = (q_1, q_2, \cdots, q_n)$, $p = (p_1, p_2, \cdots, p_n)$
$x = (q, p) \in S$ (phase space)
$H = H(p, q) = K(p) + U(q) = E$ (energy, assumed constant)

$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}$, $\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}$ $(i = 1, 2, \cdots, n)$

$X = \{(q, p) \in S \big| H(q, p) = E\}$, $T_t : X \to X$, $T_{t_1+t_2}x = T_{t_1}(T_{t_2}x)$
$x(t) = T_t x$, $x(0) = x(= (q_0, p_0))$.

Liouville's Theorem

The "Hamilton-flow" $T_t : X \to X$ preserves the measure $dm = \frac{dx}{||\text{grad } H||}$, i. e. $m(T_t B) = m(B)$ for all $t \in \mathbb{R}$, $B \in \mathcal{B}$.



$f \in L^1(X, m)$, $dm = \frac{dx}{||\text{grad } H||}$

Boltzmann's Ergodic Hypthesis:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(T_t x)\, dt = \int_X f\, dm$$

for a.a. $x \in X$.
Discretizing:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^n x) = \int_X f\, dm$$

Observation

Let $A \subseteq \mathbb{Z}$. Assume

$$A \cap (A - n_1) \cap (A - n_2) \cap \cdots \cap (A - n_L) \neq \emptyset$$

where $n_1, n_2, \cdots, n_L \in \mathbb{N}$. There exists $a \in A$ such that $a, a + n_1, a + n_2, \cdots, a + n_L$ lie in A. In particular, if $n_1 = b, n_2 = 2b, \cdots, n_L = Lb$ then $a, a + b, a + 2b, \cdots, a + Lb$ lie in A.
(Equivalently: $\chi_A(a)\chi_A(a + b)\chi_A(a + 2b) \cdots \chi_A(a + Lb) \neq 0$)

Erdős-Turan Conjecture (1936)

Let $A \subseteq \mathbb{Z}$ have upper positive density, i. e.

$$\bar{d}(A) = \limsup_{b_k - a_k \to \infty} \frac{|A \cap [a_k, b_k)|}{b_k - a_k} > 0.$$

Then A contains arbitrarily long arithmetic progressions.

(1952) Klaus Roth: Length 3
(1969) Endre Szemerédi: Length 4
(1975) Endre Szemerédi: Arbitrary length!
(1976) Hillel Furstenberg: Ergodic-theoretic proof.

Theorem (Furstenberg's Correspondence Principle)

Let $A \subseteq \mathbb{Z}$, where $\bar{d}(A) > 0$. There exists a measure-preserving dynamical system $(X, \mathcal{B}, m, T)$ and a set $E \in \mathcal{B}$ such that $m(E) = \bar{d}(A)$. Furthermore, for every $L \in \mathbb{N}$ and for arbitrary $n_1, n_2, \cdots, n_L \in \mathbb{Z}$, we have
$\bar{d}(A \cap (A - n_1) \cap \cdots \cap (A - n_L)) \geq m(E \cap T^{-n_1}E \cap \cdots \cap T^{-n_L}E)$.

Proof sketch. Let $x_0 \in \prod_{-\infty}^{\infty}\{0, 1\}$ be defined by

$$x_0(n) = \begin{cases} 1 & \text{if } n \in A \\ 0 & \text{if } n \notin A \end{cases}$$

Let $X = \overline{\{T^n x_0 \big| n \in \mathbb{Z}\}}$, where $T : X \to X$ is the shift map: $x(m) \to x(m+1)$.

Define $F : C(X) \to \mathbb{R}$ by

$$F(f) = \lim_{k\to\infty} \frac{1}{b_k - a_k} \sum_{n=a_k}^{b_k-1} f(T^n x_0)$$

where $[a_k, b_k)$ are the intervals that "determine" $\bar{d}(A)$. Then $F \geq 0, F(1) = 1, F(f \circ T) = F(f)$. By Riesz' representation theorem there exists a T-invariant probability measure m on X such that $F(f) = \int_X f\, dm$. Let $E = \{x \in X \,|\, x(0) = 1\}$. Define $\phi \in C(X)$ by $\phi(z) = z(0)$, where $z \in X$. Then

$$F(\phi) = \lim_{k\to\infty} \frac{1}{b_k - a_k} \sum_{n=a_k}^{b_k-1} (T^n x_0)(0) = \lim_{k\to\infty} \frac{|A \cap [a_k, b_k)|}{b_k - a_k} = \bar{d}(A)$$

We get:

$$m(E) = \int_X \phi\, dm = F(\phi) = \bar{d}(A).$$

Theorem B (Finitary version of Szemerédi's Theorem)

Let $0 < \delta \leq 1$ and let $L \geq 2$ be a natural number. If N is sufficiently large and $f : \mathbb{Z}_N \to \mathbb{R}$ is a function such that $0 \leq f(x) \leq 1$ for all $x \in \mathbb{Z}_N$ and $\mathbb{E}(f(x)\big|x \in \mathbb{Z}_N) \geq \delta$, then $\mathbb{E}(f(x)f(x+r)\cdots f(x+Lr)\big|x, r \in \mathbb{Z}_N) \geq c(L,\delta)$ where $c(L,\delta) > 0$ does not depend on f or N.

$\mathbb{E}(f(x)\big|x \in \mathbb{Z}_N) = \frac{1}{N}\sum_{x\in\mathbb{Z}_N} f(x)$.
$\mathbb{E}(f(x,y)\big|x,y \in \mathbb{Z}_N) = \frac{1}{N^2}\sum_{x,y\in\mathbb{Z}_N} f(x,y)$.

In particular, choose $f = \chi_A$, where $A \subseteq \mathbb{Z}_N$ satisfies $|A| \geq \delta N$. Then

$$\mathbb{E}(\chi_A(x)\big|x \in \mathbb{Z}_N) = \frac{1}{N}\sum_{x \in \mathbb{Z}_N} \chi_A(x) = \frac{|A|}{N} \geq \delta.$$

Then

$$\mathbb{E}(\chi_A(x)\chi_A(x+r)\cdots\chi_A(x+Lr)\big|x, r \in \mathbb{Z}_N) \geq c(L,\delta) > 0.$$

Hence there exist $x, r \in \mathbb{Z}_N$ such that

$$\chi_A(x)\chi_A(x+r)\cdots\chi_A(x+Lr) = 1.$$

This is equivalent to say that $x, x+r, x+2r, \cdots, x+Lr$ lie in A.

Optimistically, choose $f = \chi_P$, where P denotes the prime numbers. We get

$$\mathbb{E}(\chi_P(x)\big|x \in \mathbb{Z}_N) = \frac{1}{N}\sum_{x\in\mathbb{Z}_N}\chi_P(x) = \frac{|P \cap \{1, 2, \cdots, N\}|}{N}$$

However, $\frac{\pi(N)}{N} \sim \frac{1}{\log N} \to 0$ as $N \to \infty$, so the hypotheses of Theorem B are not satisfied.

The main idea of the Green/Tao proof is an ingenious way of getting around the difficulty that the primes less than N do not form a dense subset of $\{1, 2, \cdots, N\}$. They exploit the fact that one has a lot of control over *random* (or random-like, also called *quasirandom*) sets. In particular, there are various results that assert that, if X is a random-like set and Y is a subset of X that is dense in X (in the sense that $|Y|/|X|$ is bounded below by a positive constant), then Y behaves in a way that is analogous to how a dense set (in the ordinary sense) would behave; that is, sparse sets can be handled if you can embed them densely into random-like sets.

Proof idea of Green/Tao

Find a function $\nu : \mathbb{Z}_N = \{1, 2, \cdots, N\} \to \mathbb{R}_+$ that dominates P, i.e. $\chi_P(n) \leq \nu(n)$ for all $n \in \mathbb{Z}_N$, and such that P has positive density with respect to ν, i. e. $\sum_{n \leq N} \chi_P(n) \geq c \sum_{n \leq N} \nu(n)$, where $c > 0$.

The choice of ν is very subtle: One needs to establish specific properties of ν (which one is unable to do for χ_P). In particular, one needs asymptotic bounds on sums of the type

$$\sum_{n \leq N} \nu(n+k)\nu(n+2k)\cdots\nu(n+mk)$$

Selberg sieve and Selberg weights.

In the 1940's Selberg introduced a wonderfully simple, yet powerful, idea to analytic number theory. If R is any parameter and if $(\lambda_d)_{d\leq R}$ is a sequence of real numbers with $\lambda_1 = 1$, we have the pointwise inequality

$$\chi_p(n) \leq \Big(\sum_{\substack{d|n \\ d\leq R}} \lambda_d\Big)^2 = \nu(n)$$

provided that $n > R$. This gives an enormous number of potential ν's serving our need.
We will be interested in the set of primes less than some cutoff N, and then R will be some power N^α, $\alpha < 1$. In this situation the function ν majorize the primes between N^α and N, that is to say almost all primes less than N.

Let's try to make $\sum_{n\leq N}\nu(n)$ as small as possible.

$$\sum_{n\leq N}\nu(n)=\sum_{n\leq N}\Big(\sum_{\substack{d|n\\ d\leq R}}\lambda_d\Big)^2=\sum_{d,d'\leq R}\lambda_d\lambda_{d'}\sum_{\substack{n\leq N\\ d|n,d'|n}}1 \qquad (*)$$

Now

$$\sum_{\substack{n\leq N\\ d|n,d'|n}}1=\frac{N}{\{d,d'\}}+O(1)$$

So the main term of (*) is $N\sum_{d,d'\leq R}\frac{\lambda_d\lambda_{d'}}{\{d,d'\}}$. Using standard technique for minimizing this (such that $\lambda_1=1$), one gets the "Selberg weights":

$$\boxed{\lambda_d^{\text{Sel}}=\mu(d)\log(R/d)}$$

($\mu=$ Möbius function)

It turns out that by choosing the Selberg weights, we get

$$\sum_{n \leq N} \nu(n) \leq c \frac{N}{\log N}$$

for some $c > 0$ independent of N. Hence primes have positive density with respect to the "measure" ν. This is in contrast to the uniform "measure" δ. (Recall that $\sum_{n \leq N} \delta(n) = N$.)

Theorem (Green/Tao ; "Transference theorem")

Let $0 < \delta \leq 1$ and let $L \geq 2$ be a natural number. If N is a sufficiently large natural number, and $\nu : \mathbb{Z}_N \to \mathbb{R}_+$ is a L-pseudorandom measure, and if $f : \mathbb{Z}_N \to \mathbb{R}$, $0 \leq f(x) \leq \nu(x)$ for all $x \in \mathbb{Z}_N$ and $\mathbb{E}(f(x)\big|x \in \mathbb{Z}_N) \geq \delta$ then

$$\mathbb{E}(f(x)f(x+r)\cdots f(x+Lr)\big|x, r \in \mathbb{Z}_N) \geq c(L,\delta) - o_{L,\delta}(1)$$

where the constant $c(L,\delta)$ is the same as in Theorem B.

Comments

Green/Tao's "transfer Theorem" can be considered to be a generalization of Furstenberg's recurrence theorem. In the latter case a natural choice for ν is the uniform type, i.e. each number k in $\{1, 2, \cdots, N\}$ has weight $\nu(k) = \frac{1}{N}$. The uniform measure is invariant with respect to the shift map $x \to x + 1 \pmod{N}$.
In the Green/Tao version the measure ν behaves pseudorandomly with respect to the shift.

L-pseudorandom measure (or rather, density function relative the uniform measure on $\mathbb{Z}_N$): $\nu : \mathbb{Z}_N \to \mathbb{R}_+$

(i) $\mathbb{E}(\nu(x)\big|x \in \mathbb{Z}_N) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{x\in\mathbb{Z}_N}\nu(x) = 1 + o(1)$

(ii) ν satisfies a k-pseudorandom condition and a k-correlation condition, for every k in $\{1, 2, \cdots, N\}$ that is less than a number which depends on L.

von Mangoldt function $\wedge : \mathbb{N} \to \mathbb{R}_+$

$$\wedge(n) = \begin{cases} \log p & \text{if } n = p^m, p \text{ prime} \\ 0 & \text{otherwise} \end{cases}$$

(i) $\log n = \sum_{d|n} \wedge(d)$

(ii) $\wedge(n) = \sum_{d|n} \mu(d) \log \frac{n}{d}$

(iii) $\frac{1}{N} \sum_{1 \leq n \leq N} \wedge(n) = 1 + o(1) \left(\Leftrightarrow \pi(x) \sim \frac{x}{\log x} \right)$

The truncated von Mangoldt function

$$\wedge_R(n) = \sum_{d|n, d \leq R} \mu(d) \log(\frac{R}{d}),$$

where $\mu =$ Möbius function.

(Observe that $\wedge_R(n) = \wedge(n)$ if $R > n$)

Definition

The level R almost primes $P_R(N)$ are defined to be the set of all numbers between 1 and N that contain no non-trivial factors less than or equal to R.
(Ex. $P_6(100) = \{\text{primes } p \big| 6 \leq p \leq 100\} \cup \{49, 77, 91\}$)

$P_R(N) \sim \frac{cN}{\log R}$ (Mertens (1874))
Combining this with the prime number theorem we get that the density of primes in $P_R(N)$ is $c\frac{\log R}{\log N}$ for some $c > 0$. Choosing $R = N^\alpha$ a small power of N, we get that the primes have positive density in $P_{N^\alpha}(N)$. In fact

$$\frac{|P \cap \{R, R+1, \cdots, N\}|}{|P_R(N)|} \approx \frac{\pi(N)}{\frac{cN}{\log N^\alpha}} \approx \frac{\frac{N}{\log N}}{\frac{cN}{\alpha \log N}} = \frac{\alpha}{c} > 0.$$

Let $A \subseteq \mathbb{Z}_N = \{0, 1, 2, 3, \cdots, N-1\}$. Let $\hat{\chi}_A$ be the Fourier transform of χ_A. Then for $r \in \mathbb{Z}_N$:

$$\hat{\chi}_A(r) = \sum_{s \in \mathbb{Z}_N} e^{\frac{2\pi i}{N} rs}$$

Observe that $\hat{\chi}_A(0) = |A|$.
If $|\hat{\chi}_A(r)|$ is significantly smaller than $\alpha^2 N$ for every non-zero r, then A behaves in many ways like a random subset of $\mathbb{Z}_N$ (where every element is chosen with probability α, for some $0 < \alpha < 1$).

$P : 2, 3, 5, 7, 11, 13, 17, 19, \cdots$ Bias toward odd numbers.
$x \to \frac{x-1}{2} : 1, 2, 3, 5, 6, 8, 9, 11, 14, \cdots$ No bias toward odd/even.
(If this linear rescaling of the primes has L-term APs, then so has the primes.) The new sequence has bias (mod 3). By taking the multiples of 3 and rescaling $x \to \frac{x}{3}$ we get a new sequence that is well distributed (mod 6):

$$1, 2, 1, 5, 2, 8, 3, 11, 14, \cdots$$

Repeat for primes $5, 7, \cdots, p \leq w(N) = \log\log N$, say. The sequence one winds up with has no bias in any class $a \pmod q$, $q \leq \log\log N$.

Let N and L ($L << N$) be given. We define a function $\nu : \mathbb{Z}_N \to \mathbb{R}_+$:
Choose $R = N^{L^{-1} \cdot 2^{-L-4}}$ (Think of this as $R = N^{\epsilon}$)

$$\nu(n) = \begin{cases} \frac{\Phi(w)}{w} \cdot \frac{(\wedge_R(wn+1))^2}{\log R} & \text{if } \frac{N}{[2^L(L+4)]!} \leq n \leq \frac{2N}{[2^L(L+4)]!} \\ 1 & \text{otherwise.} \end{cases}$$

Here $w = \prod_{p<\omega(N)} p$, where $\omega(N) \to \infty$ "sufficiently slow" ; for example $\omega(N) = \log(\log N)$.

What weights λ_d should one choose? This depends on the application, but a very basic application is to the estimation of $\pi(x+y)-\pi(x)$, the number of primes in the interval $(x, x+y]$ (the Brun-Titchmarsh problem).

$$\pi(x+y)-\pi(x) \leq \sum_{n=x+1}^{x+y} \Big(\sum_{\substack{d|n \\ d\leq R}} \lambda_d \Big)^2$$

$$= \sum_{d\leq R} \sum_{d'\leq R} \lambda_d \lambda_{d'} \sum_{n=x+1}^{x+y} \chi_{d|n}(n)\chi_{d'|n}(n) \qquad (**)$$

$$= y \sum_{d\leq R} \sum_{d'\leq R} \frac{\lambda_d \lambda_{d'}}{[d,d']} + O\Big(\sum_{d\leq R} \sum_{d'\leq R} |\lambda_d||\lambda_{d'}| \Big)$$

Let us imagine that the weights $|\lambda_d|$ are chosen to be $<< y^{\epsilon}$ (this is always the case in practice). Then the second term is $O(R^2 y^{2\epsilon})$. If $R \leq y^{1/2-2\epsilon}$ then this is $O(y^{1-\epsilon})$ and may be thought of as an error term. This is why it is advantageous (indeed essential) to work with a majorant taken over a truncated range of divisors, and not with χ_P itself.

The first term in (**): $y \sum_{d \leq R} \sum_{d' \leq R} \frac{\lambda_d \lambda_{d'}}{[d,d']}$ is a quadratic form. It may be explicitly minimized subject to the condition $\lambda_1 = 1$, giving optimal weights which are independent of x and y. It turns out that a nearly optimal choice of the λ_d's are:

$$\lambda_d^{\mathrm{SEL}} = \mu(d) \frac{\log(R/d)}{\log R},$$

($\mu =$ Möbius function).

Goldston, Pintz, Yildrim (2005)

$$\liminf_{N \to \infty} \frac{p_{n+1} - p_n}{\log n} = 0$$

NB. By the prime number theorem $\frac{p_{n+1} - p_n}{\log n}$ has average value 1.