Nilsequences and the primes

(The lack of) hidden patterns in the prime numbers

Fields Medalists Symposium

April 26, 2007

Ben Green (Cambridge)

Terence Tao (UCLA)

Analytic prime number theory

Analytic prime number theory studies the distribution of, and patterns in, the prime numbers $2, 3, 5, 7, \ldots$. There are two main branches:

- Multiplicative prime number theory (e.g. expressing a number as the product of prime numbers; the residue class p mod q when dividing a prime p by a modulus q);
- Additive prime number theory (e.g. expressing a number as the sum or difference of prime numbers; arithmetic progressions of primes).

Some theorems from multiplicative prime number theory:
A large natural number n has...

- ...a probability about $\frac{6}{\pi^2}$ of having no square factors other than 1. (Euler, $\sim$ 1730)
- ...close to $\ln n$ factors on average. (Dirichlet, $\sim$ 1830)
- ...a probability about $\frac{1}{\ln n}$ of being prime. (Hadamard-de Vallée Poussin 1896)
- ...close to $\ln \ln n$ prime factors on average. (Erdős-Turán, 1935)

Some theorems and conjectures from additive prime number theory: A large natural number n...

- is the sum of three primes, if it is odd (Vinogradov, 1937)
- can be both prime, and two less than a prime, infinitely often (twin prime conjecture)
- is both prime, and two less than an almost prime, infinitely often (Chen, 1973)
- is the sum of two primes, if it is even (Goldbach conjecture)

To understand multiplicative problems (e.g. the distribution of products pq of primes), one needs to understand the distribution of the powers p^s where s is a complex number and p runs over primes (this is basically because of identities such as $(pq)^s = p^s q^s$). This leads one to the study of things such as the Riemann zeta function

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p (1 - \frac{1}{p^s})^{-1}$$

which is of course connected to the famous Riemann hypothesis.

To understand additive problems (e.g the distribution of sums $p+q$ of primes), one needs to understand the distribution of the exponentials $e(\alpha p) := e^{2\pi i \alpha p}$ where α is a real number and p runs over primes (this is basically because of identities such as $e(\alpha(p+q)) = e(\alpha p)e(\alpha q)$). This leads one to the study of things such as the prime exponential sum

$$\sum_{p<N} e(\alpha p)$$

which leads one to the Hardy-Littlewood-Vinogradov circle method.

A typical result in multiplicative prime number theory:

- (Primes in arithmetic progressions) Any infinite arithmetic progression $\{n : n = a \bmod q\}$ with a coprime to q (i.e. $a \in (\mathbf{Z}/q\mathbf{Z})^\times$) contains infinitely many primes. (Dirichlet 1837)

A typical result in additive prime number theory:

- (Arithmetic progressions in primes) The primes contain arbitrarily long arithmetic progressions. (Green-T. 2004)

Despite several similarities and connections, these two results are proven using very different types of mathematics!

It turns out that to prove the above qualitative results, one needs to first study their quantitative counterparts. We introduce the von Mangoldt function

$$\Lambda(n) := \begin{cases} \ln p & \text{if } n = p^j \text{ for some prime } p \text{ and } j \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

This is a convenient weight function for counting primes, and will serve as our quantitative proxy for the primes. It is also convenient to introduce the averaging notation

$$\mathbf{E}_{1 \leq n \leq N} f(n) := \frac{1}{N} \sum_{n=1}^{N} f(n).$$

The von Mangoldt function has two nice properties worth noting here. Firstly, we have the fundamental theorem of arithmetic

$$\ln n = \sum_{d|n} \Lambda(d) \text{ for all } n \geq 1$$

which gives rise to many important algebraic identities involving Λ. Secondly, we have the prime number theorem

$$\mathbf{E}_{1 \leq n \leq N} \Lambda(n) = 1 + o(1).$$

This is a fundamental result in number theory; an equivalent formulation is that the prime numbers from 1 to N have density $\frac{1+o(1)}{\ln N}$.

Quantitative versions of Dirichlet's theorem (primes in arithmetic progressions): If a is coprime to q, then

- $\mathbf{E}_{1 \leq n \leq N} 1_{n=a \bmod q} \Lambda(n) \geq c_q + o_q(1)$ as $N \to \infty$ for some $c_q > 0$. (Dirichlet, 1837)
- $\mathbf{E}_{1 \leq n \leq N} 1_{n=a \bmod q} \Lambda(n) = \frac{1}{\phi(q)} + O_A(\ln^{-A} N)$ for all $A > 0$. (Siegel-Walfisz, 1936)
- $\mathbf{E}_{1 \leq n \leq N} 1_{n=a \bmod q} \Lambda(n) = \frac{1}{\phi(q)} + O_\varepsilon(N^{-1/2+\varepsilon})$ for any $\varepsilon > 0$ (Generalised Riemann Hypothesis)

These quantitative versions of Dirichlet's theorem give quite precise information: for instance, it shows that the number of primes less than a large number N whose last digit is 3 is roughly $\frac{1}{4}\frac{N}{\log N}$.

Quantitative versions of the Green-Tao theorem (arithmetic progressions in primes): If $k \geq 1$ and $N \to \infty$, then

- $\mathbf{E}_{1 \leq n,r \leq N} \Lambda(n)\Lambda(n+r)\ldots\Lambda(n+(k-1)r) \geq c_k + o_k(1)$ for some $c_k > 0$. ($k = 1, 2$ Chebyshev 1850; $k = 3$ van der Corput, 1939; $k > 3$ Green-T., 2004)
- $\mathbf{E}_{1 \leq n,r \leq N} \Lambda(n)\Lambda(n+r)\ldots\Lambda(n+(k-1)r) = \mathfrak{G}_k + o_k(1)$ ($k = 1, 2$ Hadamard-de Vallée Poussin 1896; $k = 3$ van der Corput, 1939; $k = 4$ Green-T. 2006; $k > 4$ work in progress)

The singular series $\mathfrak{G}_k$ is defined as

$$\mathfrak{G}_k := \prod_p \mathbf{E}_{n,r \in \mathbf{Z}/p\mathbf{Z}} \Lambda_p(n)\Lambda_p(n+r)\ldots\Lambda_p(n+(k-1)r)$$

where for each prime p, Λ_p is the local von Mangoldt function at p:

$$\Lambda_p(n) := \frac{p}{\phi(p)} 1_{n \neq 0 \bmod p}.$$

This strange series is predicted by a much more general conjecture known as the Hardy-Littlewood prime tuples conjecture. (This conjecture also implies the twin prime and Goldbach conjectures, among others.)

$$
\begin{aligned}
\mathfrak{G}_1 &= 1 \\
\mathfrak{G}_2 &= 1 \\
\mathfrak{G}_3 &= 2\prod_{p\geq 3}\left(1-\frac{1}{(p-1)^2}\right) \\
&= 1.320\ldots \\
\mathfrak{G}_4 &= \frac{9}{2}\prod_{p\geq 5}\left(1-\frac{3p-1}{(p-1)^3}\right) \\
&= 2.858\ldots \\
&\ldots
\end{aligned}
$$

Again, these results give fairly precise information on the distribution of patterns in primes; for instance we now know that the number of arithmetic progressions of primes of length 4 less than N is about $0.476\frac{N^2}{\ln^4 N}$.

Results in multiplicative prime number theory tend to rely on algebraic identities, for instance

$$1_{n=a \bmod q}\Lambda(n) = \frac{1}{\phi(q)} \sum_{\chi \bmod q} \overline{\chi(a)}\Lambda(n)\chi(n)$$

$$\sum_{n=1}^{\infty} \frac{\Lambda(n)\chi(n)}{n^s} = -\frac{L'(s,\chi)}{L(s,\chi)}$$

$$\Lambda(n)\chi(n) \text{ `}='\ 1_{\chi=\chi_0} - \sum_{L(\rho,\chi)=0} n^{\rho-1} + \dots$$

$$L(1,\chi) = \begin{cases} \frac{2\pi h}{w\sqrt{q}} & \text{if } \chi(-1) = -1 \\ \frac{2h \ln \varepsilon}{w\sqrt{q}} & \text{if } \chi(-1) = 1 \end{cases}$$

In contrast, results in additive prime number theory rely more on analytic correlations or discorrelations between the primes and other, more additively structured, objects. A good example are the correlations with linear phases $e(\alpha n)$, where $e(x) := e^{2\pi i x}$ and $\alpha \in \mathbf{R}$:

$$\lim_{N\to\infty} \mathbf{E}_{1\leq n\leq N} \Lambda(n) e(\alpha n) = \begin{cases} s(\frac{a}{q}) & \text{if } \alpha = \frac{a}{q} \\ 0 & \text{if } \alpha \text{ irrational} \end{cases}$$

where $s(\frac{a}{q})$ is the Ramanujan sum

$$s(\frac{a}{q}) := \mathbf{E}_{b\in(\mathbf{Z}/q\mathbf{Z})^\times} e(\frac{ab}{q}).$$

Notice the dichotomy here between rational α and irrational α. The dichotomy is ultimately best described using ergodic theory (the theory of dynamical systems): the circle shift map $x \mapsto x + \alpha \bmod 1$ on the unit circle $\mathbf{R}/\mathbf{Z}$ is periodic when α is rational, but totally ergodic when α is irrational.

For instance, if you start at a point on the circle, and move forward by quarter-rotations, you will simply visit four points on the circle periodically; but if you instead move forward by $\frac{1}{2\pi}$-rotations (one radian at a time) you will eventually visit nearby every point on the circle in an evenly distributed manner.

In the case of quarter-rotations, if you look at what the prime points of the orbit (i.e. the 2^{nd} point, the 3^{rd} point, the 5^{th} point, etc. do, they concentrate on two of the four points of the orbit; but in the case of $\frac{1}{2\pi}$-rotations, it turns out that the prime points are just as uniformly distributed as all the other points. Thus the primes "correlate with" or "conspire with" the quarter-rotation dynamical system, but do not conspire with the $\frac{1}{2\pi}$-rotation dynamical system.

Philosophy and heuristics

- The primes from 1 to N have density approximately $1/\ln N$ (the Prime Number Theorem).
- The primes from 1 to N "want" to behave like a random sequence with this density. If they did, then many statistics in additive prime number theory would be easy to compute (e.g. the number of twin primes $p, p+2$ from 1 to N would be roughly $N/\ln^2 N$).

- However, there are a number of "patterns" or "conspiracies" that the primes could have which would significantly distort the statistics to be different from the random count. (e.g. most primes are odd, which drastically reduces the number of *adjacent* primes $p, p+1$ but presumably increases the number of *twin* primes $p, p+2$.)
- Thus, one can hope to enumerate all the possible conspiracies that could affect a given statistic, work out which of these conspiracies are actually obeyed by the primes, and use all this information to compute the statistic to high accuracy.

- General belief: the only patterns the primes exhibit are those arising from simple algebraic considerations (e.g. primes are usually coprime to q for any fixed q). There should be no other conspiracies of consequence.
- This belief underpins many of the conjectures we have about the primes (e.g. generalised Riemann hypothesis, twin-primes and Goldbach conjectures, etc.). This general belief has been confirmed for specific types of statistics (particularly those with lots of "averaging"), and for specific types of conspiracies (particularly those of an algebraic nature).

The circle method

A classic way of implementing the above philosophy is the Hardy-Littlewood-Vinogradov circle method, based on Fourier analysis. In this case the "conspiracies" are the possible correlations that the primes (or whatever other object is being studied) has with the linear characters $e(\alpha n)$.

This method is useful for detecting some patterns in primes but not others - roughly speaking, it can only count patterns whose statistics can only be distorted by "linear" conspiracies.

For instance, in 1937, Vinogradov used the circle method to show that every sufficiently large odd number is the sum of three primes, thus solving (most of) the odd Goldbach conjecture.

In 1939, van der Corput used the same method to count the number of arithmetic progressions of primes $p, p+r, p+2r$ less than some large number N; he computed this number as

$$\frac{1}{4}(\mathfrak{G}_3 + o(1))\frac{N^2}{\ln^3 N} \approx 0.330\ldots\frac{N^2}{\ln^3 N}.$$

In particular, there are infinitely many arithmetic progressions of primes of length three.

What are "linear" conspiracies?

van der Corput's problem is essentially equivalent to that of computing the average

$$\mathbf{E}_{1\leq n,r\leq N}\Lambda(n)\Lambda(n+r)\Lambda(n+2r).$$

All other things being equal, given three functions f, g, h one expects

$$\mathbf{E}_{1\leq n,r\leq N}f(n)g(n+r)h(n+2r) \approx (\mathbf{E}f)(\mathbf{E}g)(\mathbf{E}h)$$

where $\mathbf{E}f$ is the mean value of f, etc.

Unfortunately, because of the identity

$$\alpha n - 2\alpha(n+r) + \alpha(n+2r) = 0 \bmod 1$$

the above heuristic fails if we have the "linear conspiracy"

$$f(n) \approx e(\alpha n); \quad g(n) \approx e(-2\alpha n); \quad h(n) \approx e(\alpha n).$$

To put it another way: if you know the value of a linear function at two points of an arithmetic progression, you can extrapolate to find the value at the third point of the progression. This is why linear phases play a key role in the theory of such patterns as progressions of length three.

The conspiracy has an ergodic theory interpretation using the circle shift $T : x \to x + \alpha \bmod 1$. Even if α is irrational (so that T is totally ergodic), there is enough algebraic structure in this system that the behaviour of an arithmetic progression $T^n x, T^{n+r} x, T^{n+2r} x$ in this dynamical system is highly constrained (indeed the position of the third point can be determined algebraically from the position of the first two).

The primes will exhibit this conspiracy if the prime orbit $\{T^p x : p \text{ prime}\}$ is not as uniformly distributed as the full orbit $\{T^n x : n \in \mathbf{Z}\}$.

Fortunately, one can show in this case that linear conspiracies are the *only* conspiracies that can distort this type of statistic. Indeed, from Fourier analysis we have (modulo some cheating) the identity

$$\mathbf{E}_{1\leq n,r\leq N} f(n)g(n+r)h(n+2r) = N\int_{\mathbf{R}/\mathbf{Z}} \hat{f}(\alpha)\hat{g}(-2\alpha)\hat{h}(\alpha)\,d\alpha$$

where $\hat{f}(\alpha) := \mathbf{E}_{1\leq n\leq N} f(n)e(-n\alpha)$.

Vinogradov and van der Corput established their results using identities like these, and by computing how the primes conspire with various linear characters.

More complex patterns

For arithmetic progressions of length 4, the circle method (Fourier analysis) is insufficient. This is because new "quadratic" conspiracies emerge, which are undetectable by the circle method but still bias the statistic one seeks to compute.

For instance, when computing

$$\mathbf{E}_{1\leq n,r\leq N} f(n)g(n+r)h(n+2r)k(n+3r),$$

the existence of the identity

$$\alpha n^2 - 3\alpha(n+r)^2 + 3\alpha(n+2r)^2 - \alpha(n+3r)^3 = 0 \bmod 1$$

means that we have to do something about the possible conspiracy

$$f(n) \approx e(\alpha n^2); g(n) \approx e(-3\alpha n^2);$$

$$h(n) \approx e(3\alpha n^2); k(n) \approx e(-\alpha n^2).$$

This conspiracy relates to Lagrange interpolation: the values of a quadratic at three points of an arithmetic progression determine the value at a fourth point.

More generally, a conspiracy can arise from any dynamical system $T : X \to X$ in which there is a non-trivial constraint between a four-term arithmetic progression $T^n x, T^{n+r} x, T^{n+2r} x, T^{n+3r} x$ in an orbit. (The previous quadratic example can be essentially associated to the skew shift $T : (x, y) \to (x + \alpha, y + x)$ on the 2-torus $(\mathbf{R}/\mathbf{Z})^2$.)

The space of all such dynamical "conspiracies" has been classified recently (Host-Kra 2005, Ziegler 2007). For instance, all the conspiracies which could bias four-term progressions are ultimately generated by 2-step nilflows, or more precisely a group action $T : x \to gx$ on a 2-step nilmanifold G/Γ (i.e. a quotient of a 2-step nilpotent Lie group G by a discrete co-compact subgroup Γ). The identity

$$(g^n x)(g^{n+r} x)^{-3}(g^{n+2r} x)^3(g^{n+3r} x) = \mathrm{id},$$

which holds for g, x in a 2-step nilpotent group, is ultimately the reason why these nilflows are an essential aspect of the theory.

With a significant amount of effort (combining ideas from the ergodic theory literature with the "higher order Fourier analysis" of Gowers, which involves additive combinatorics, and the transference principle of Green and myself) one can now compute statistics such as the number of arithmetic progressions of length four in the primes less than N (which is $(\frac{1}{6}\mathfrak{G}_4 + o(1))\frac{N^2}{\ln^4 N}$). The main number-theoretic ingredient is the correlation estimates between the prime numbers and 2-step nilsequences $F(g^n x)$, where $F : G/\Gamma \to \mathbf{C}$ is a smooth function.

Ratner's theorem

We mentioned earlier that a circle shift $x \mapsto x + \alpha$ is either periodic or totally ergodic. In either case, the orbits are uniformly distributed inside of closed (translates of) subgroups of the unit circle. This phenomenon is in fact rather general:

Ratner-type theorem (Ratner 1991, Shah 1998) Let T be a unipotent action on a compact symmetric space G/Γ. Then every orbit $\{T^n x : n \in \mathbf{Z}\}$ is uniformly distributed inside of some closed sub-symmetric space of G/Γ.

It turns out that quantitative versions of this theorem are decisive in establishing the required correlation estimates between primes and higher-step nilsequences, which in turn can be used to count many types of additive patterns in the primes. This is work currently in progress with Ben Green.

It seems clear, though, that ideas from ergodic theory - in particular, understanding the distribution of orbits of dynamical systems - will play an increasingly important role in the future development of analytic prime number theory.