

Counting arithmetic progressions of primes

Bogdan Grechuk • 23 Oct 2026

<https://functor.network/user/3333/entry/2025>

This post continues my series on favorite theorems of the twenty-first century. For an overview of the categories and my earlier selections, see this introductory post.

My choice for 2006 in Number Theory is a theorem of Ben Green and Terence Tao giving an asymptotic formula for the number of four-term arithmetic progressions of primes. More generally, their paper developed a method for counting solutions of systems of linear equations in which all unknowns are required to be prime. The paper was first posted online in June 2006 and appeared in the *Annals of Mathematics* in 2010 (Green and Tao 2010).

In a previous post, we discussed my favorite theorem of 2004: Green and Tao proved that the set of prime numbers contains arbitrarily long arithmetic progressions (Green and Tao 2008). Their result established the existence of infinitely many such progressions for every fixed length. However, it did not determine precisely how many of them there are.

For example, how many quadruples of primes

$$p_1 < p_2 < p_3 < p_4 \leq N$$

form an arithmetic progression, that is, satisfy

$$p_2 - p_1 = p_3 - p_2 = p_4 - p_3?$$

Equivalently, we want to count pairs of positive integers a, d such that

$$a, \quad a + d, \quad a + 2d, \quad a + 3d$$

are all prime and $a + 3d \leq N$.

There are approximately $N^2/6$ pairs of positive integers (a, d) satisfying $a + 3d \leq N$. On the other hand, the prime number theorem tells us that the probability that an integer of size approximately N is prime is about $1/\log N$. Here $\log$ denotes the natural logarithm.

If we naively assume that the four numbers behave independently, we might therefore predict approximately

$$\frac{N^2}{6(\log N)^4}$$

four-term arithmetic progressions of primes.

However, this heuristic ignores restrictions arising from divisibility by small primes. For example, in any four-term arithmetic progression of primes greater than 3, the common difference d must be divisible by 6. Indeed, if d were odd, some terms would be even, while if d were not divisible by 3, one of the first three terms would be divisible by 3.

Similar restrictions arise from other primes. Accounting for all of them leads to a correction factor expressed as an infinite product over primes.

Green and Tao proved that, after including these corrections, the heuristic gives the exact asymptotic answer.

Theorem 1 (Green–Tao) *Let $A_4(N)$ denote the number of quadruples of primes*

$$p_1 < p_2 < p_3 < p_4 \leq N$$

forming an arithmetic progression. Then

$$A_4(N) = (C_4 + o(1)) \frac{N^2}{(\log N)^4} \quad \text{as } N \rightarrow \infty,$$

where

$$C_4 = \frac{3}{4} \prod_{\substack{p \geq 5 \\ p \text{ prime}}} \left(1 - \frac{3p-1}{(p-1)^3} \right) \approx 0.4764.$$

Here $o(1)$ denotes a quantity tending to zero as $N \rightarrow \infty$.

The infinite product in the theorem converges to a positive number. Its factors describe the corrections to the naive probability arising from divisibility by each prime. Thus, the theorem does not merely prove that there are infinitely many four-term arithmetic progressions of primes, but determines their number asymptotically, including the exact leading constant.

In fact, the paper proves much more than this special case. Recall that a *linear polynomial* in n variables with integer coefficients is a polynomial of the form

$$P(x_1, \dots, x_n) = c_1 x_1 + \dots + c_n x_n + b,$$

where $c_1, \dots, c_n, b$ are integers.

Consider a collection of such polynomials

$$P_1(x_1, \dots, x_n), \dots, P_k(x_1, \dots, x_n).$$

We would like to know how often all of them take prime values simultaneously.

An obvious necessary condition is that, for every prime p , we can choose integer values of $x_1, \dots, x_n$ such that none of the numbers $P_1, \dots, P_k$ is divisible by p . Such a collection of polynomials is called *admissible*. Otherwise, at least one of the polynomial values is necessarily divisible by p , so simultaneous primality is impossible apart from exceptional cases where that value equals p itself.

We must also require that there are sufficiently many choices of the variables for which all the polynomial values are positive.

Finally, we exclude systems in which two polynomials are *affinely related*, meaning that

$$P_i = \alpha P_j + \beta$$

for some rational constants α, β . This condition excludes particularly difficult problems such as the twin prime conjecture, which asks whether n and $n + 2$ are simultaneously prime for infinitely many integers n .

The general prediction is that, under these assumptions, the number of choices

$$-N \leq x_1, \dots, x_n \leq N$$

for which all $P_1, \dots, P_k$ are prime should be asymptotic to

$$C_P \frac{N^n}{(\log N)^k},$$

where $C_P > 0$ is a constant determined by the polynomials. Like the constant in Theorem 1, it incorporates geometric restrictions and corrections for divisibility by primes.

In their 2006 paper, Green and Tao established this prediction unconditionally for a large class of systems, including those described by at most two independent nondegenerate linear equations. The general result was conditional on two conjectures concerning the apparent randomness of certain number-theoretic sequences.

The paper also obtained some surprising consequences beyond ordinary arithmetic progressions. For example, Green and Tao determined asymptotically how many three-term arithmetic progressions of primes have a common difference that is one more, or one less, than another prime. They also obtained results concerning representations of integers as sums of three primes with additional linear restrictions.

Let us briefly discuss the proof. The main difficulty is that prime numbers are not distributed completely randomly: their divisibility properties impose strong correlations. The authors first separate out these unavoidable restrictions by considering residue classes modulo products of small primes.

They then use a technique called the *transference principle*, developed in their earlier work. Roughly speaking, this technique allows one to replace certain problems about the sparse set of prime numbers by analogous problems about more regular, pseudorandom sequences. A sequence is called *pseudorandom* if, despite being defined deterministically, it behaves like a random sequence with respect to a suitable collection of statistical tests.

The required tests are related to *Gowers norms*, which measure the presence of structured additive patterns in numerical sequences. The proof also uses estimates involving the Möbius function, defined by

$$\mu(n) = \begin{cases} (-1)^r, & \text{if } n \text{ is a product of } r \text{ distinct primes,} \\ 0, & \text{if } n \text{ is divisible by a prime square.} \end{cases}$$

The main idea is that, once the necessary divisibility restrictions have been removed, the primes behave sufficiently randomly for the predicted asymptotic formulas to hold.

At the time, the required pseudorandomness results were known only for certain relatively simple configurations, including four-term arithmetic progressions. Consequently, the general theorem remained conditional.

The story was completed in 2012. Green and Tao (Green and Tao 2012) proved the required conjecture concerning the Möbius function, while Green, Tao, and Ziegler (Green et al. 2012) established the remaining conjecture about Gowers norms. Together with the 2006 paper, these results established the predicted asymptotic for every fixed admissible collection of pairwise affine-independent linear polynomials, provided that the region where all polynomials are positive has nonzero limiting density.

In particular, for every fixed integer $k \geq 3$, the number of k -term arithmetic progressions of primes

$$p_1 < p_2 < \cdots < p_k \leq N$$

is now known to be

$$(C_k + o(1)) \frac{N^2}{(\log N)^k},$$

where $C_k > 0$ is an explicitly computable constant depending only on k .

Thus, the work initiated by Green and Tao in 2006 eventually provided a remarkably general solution to the problem of counting nondegenerate linear patterns in prime numbers. It not only established that such patterns exist, but also determined their asymptotic frequencies, revealing an extraordinary degree of regularity in the distribution of primes.

References

- Green, Ben, and Terence Tao. 2008. "The Primes Contain Arbitrarily Long Arithmetic Progressions." *Annals of Mathematics*, 2nd series, vol. 167 (2): 481–547. <https://doi.org/10.4007/annals.2008.167.481>.
- Green, Ben, and Terence Tao. 2010. "Linear Equations in Primes." *Annals of Mathematics*, 2nd series, vol. 171 (3): 1753–850. <https://doi.org/10.4007/annals.2010.171.1753>.
- Green, Ben, and Terence Tao. 2012. "The Möbius Function Is Strongly Orthogonal to Nilsequences." *Annals of Mathematics*, 2nd series, vol. 175 (2): 541–66. <https://doi.org/10.4007/annals.2012.175.2.3>.
- Green, Ben, Terence Tao, and Tamar Ziegler. 2012. "An Inverse Theorem for the Gowers $U^{s+1}[N]$ -Norm." *Annals of Mathematics*, 2nd series, vol. 176 (2): 1231–372. <https://doi.org/10.4007/annals.2012.176.2.11>.