WHAT IS...

the Bateman–Horn Conjecture?

Stephan Ramon Garcia

For a given family of univariate polynomials with integer coefficients, how often should we expect their values at positive integer arguments to be simultaneously prime? The Bateman–Horn conjecture, first formulated by Paul T. Bateman and Roger A. Horn in 1962 [BH62, BH65], proposes a complete answer to this question. It can be thought of as a successor to the First Hardy–Littlewood conjecture [HL23] (1923), which considers the asymptotic distribution of prime values assumed by tuples of linear polynomials, and Schinzel's hypothesis H [SS58] (1958), which conjectures the infinitude of simultaneously prime values assumed by certain tuples of polynomials.

To understand where the Bateman–Horn conjecture comes from, we start with the prime number theorem. The exposition below follows [AZFG20].

Prime number theorem. Let $\pi(x)$ denote the number of primes at most x. The prime number theorem, proved independently by Hadamard and de la Vallée Poussin in 1896, says that $\pi(x) \sim \operatorname{Li}(x)$, in which

$$\operatorname{Li}(x) = \int_2^x \frac{dt}{\log t} \tag{1}$$

is the *logarithmic integral* and $\sim$ is asymptotic equivalence; that is, $f \sim g$ means $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Stephan Ramon Garcia is W.M. Keck Distinguished Service Professor and Chair of the Department of Mathematics and Statistics at Pomona College. His email address is stephan.garcia@pomona.edu.

Communicated by Notices *Associate Editor William McCallum.*

DOI: https://doi.org/10.1090/noti3046

The prime number theorem suggests the possibility of a random model for the prime numbers: the probability that n is prime is about $1/\log n$. The Bateman–Horn conjecture follows by pursuing this to its logical extreme, while adjusting for congruence obstructions (for example, 2 is the only even prime).

A single polynomial. Let $\mathbb{Z}[x]$ denote the set of polynomials in x with coefficients in $\mathbb{Z}$, the set of integers. For $f \in \mathbb{Z}[x]$, define

$$Q(f;x) = \#\{n \le x : f(n) \text{ is prime}\},$$

in which $\#S$ denotes the cardinality of a set S and n is a natural number. What conditions must f satisfy if it generates infinitely many distinct primes?

First, f should be nonconstant and its leading coefficient must be positive. Second, f should be irreducible in $\mathbb{Z}[x]$. Less obvious is that f should not vanish identically modulo any prime. For example, $f(x) = x^3 - x + 3$ is irreducible, but $f(x) \equiv x^3 - x \equiv 0 \pmod 3$, so $f(n)$ is always divisible by 3.

Suppose $f \in \mathbb{Z}[x]$ is nonconstant, irreducible, and does not vanish identically modulo any prime. Let $d = \deg f$ and suppose that f has leading coefficient $c \ge 1$. Then $f(x) \sim cx^d$ and our heuristic suggests that the probability $f(n)$ is prime is

$$\frac{1}{\log f(n)} \sim \frac{1}{\log(cn^d)} \sim \frac{1}{d\log n}, \tag{2}$$

so we expect that

$$Q(f;x) \sim \sum_{n=2}^{|x|} \frac{1}{d\log n} \sim \frac{1}{\deg f}\int_2^x \frac{dt}{\log t}. \tag{3}$$

However, this is incorrect since we failed to take into account how likely it is that $f(n) \equiv 0 \pmod p$ (the letter p

will always denote a prime number). If we assume for the sake of our heuristic argument that divisibility by distinct primes are independent events, then we should weight our prediction by

$$C(f) = \prod_p \left(1 - \frac{1}{p}\right)^{-1} \left(1 - \frac{\omega_f(p)}{p}\right), \tag{4}$$

in which $\omega_f(p)$ is the number of solutions to $f(x) \equiv 0 \pmod{p}$, since $1 - \omega_f(p)/p$ is the probability that $f(n)$ is divisible by p and $1 - 1/p$ is the probability that a random integer is divisible by p. Thus, for a single polynomial f, we suspect that

$$Q(f; x) \sim \frac{C(f)}{\deg f} \int_2^x \frac{dt}{\log t}.$$

Multiple polynomials. Suppose $f_1, f_2, \dots, f_k \in \mathbb{Z}[x]$ are distinct, nonconstant, irreducible polynomials with positive leading coefficients. Although maybe no single f_i vanishes identically modulo a prime, the product $f = f_1 f_2 \cdots f_k$ might. For example, neither $f_1(x) = x$ nor $f_2(x) = x + 1$ vanish identically modulo any prime, but their product $x(x+1)$ vanishes identically modulo 2. This "congruence obstruction" prevents n and $n+1$ from being simultaneously prime infinitely often. Consequently, we must require that f does not vanish identically modulo any prime.

Reasoning as above suggests the probability that all of the $f_i(n)$ are simultaneously prime is

$$\prod_{i=1}^k \frac{1}{\log f_i(n)} \sim \prod_{i=1}^k \frac{1}{d_i \log n} = \frac{1}{(\prod_{i=1}^k \deg f_i)(\log n)^k}.$$

Thus, the expected number of $n \le x$ such that $f_1(n), f_2(n), \dots, f_k(n)$ are prime is around

$$\int_2^x \frac{1}{(\prod_{i=1}^k \deg f_i)(\log n)^k} = \frac{1}{\prod_{i=1}^k \deg f_i} \int_2^x \frac{dt}{(\log t)^k}.$$

As before, this prediction is off by a constant factor. Instead of dividing by $1 - 1/p$ in (4), we now divide by $(1 - 1/p)^k$, the probability that a randomly selected k-tuple of integers has no element divisible by p.

The conjecture. Putting this all together yields the final conjecture (the convergence of the infinite product below is not obvious; see [AZFG20, Sect. 5] for a proof).

Bateman–Horn conjecture. *Let $f_1, f_2, \dots, f_k \in \mathbb{Z}[x]$ be distinct, nonconstant, irreducible polynomials with positive leading coefficients, and let*

$$Q(f_1, f_2, \dots, f_k; x) = \#\{n \le x : f_1(n), f_2(n), \dots, f_k(n) \text{ are prime}\}. \tag{5}$$

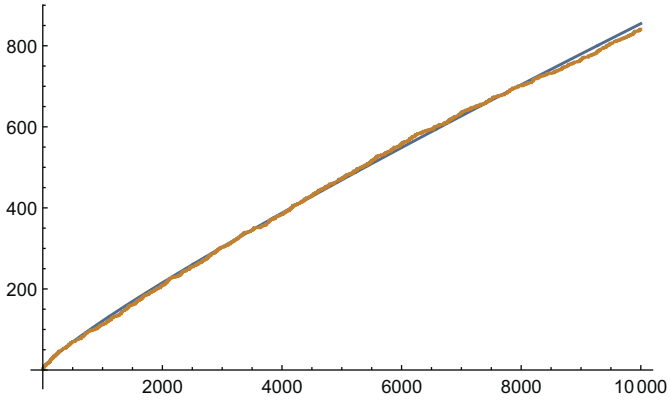

Figure 1. Landau's conjecture: $Q(f; x)$ (orange) versus the Bateman–Horn prediction $\frac{C(f)}{2}\,\mathrm{Li}(x)$ (blue). Although it is possible to plot such images at a larger scale, one loses sight of the discreteness of the underlying counting function.

Suppose that $f = f_1 f_2 \cdots f_k$ does not vanish identically modulo any prime. Then

$$Q(f_1, f_2, \dots, f_k; x) \sim \frac{C(f_1, f_2, \dots, f_k)}{\prod_{i=1}^k \deg f_i} \int_2^x \frac{dt}{(\log t)^k}, \tag{6}$$

in which

$$C(f_1, f_2, \dots, f_k) = \prod_p \left(1 - \frac{1}{p}\right)^{-k} \left(1 - \frac{\omega_f(p)}{p}\right) \tag{7}$$

and $\omega_f(p)$ is the number of distinct solutions modulo p to $f(x) \equiv 0 \pmod{p}$.

Only a few special cases of the conjecture, such as the prime number theorem for arithmetic progressions, are known to be true. However, an upper bound comparable to the conjectured asymptotic is provided by the Brun sieve [Ten15, Thm. 3, Sect. I.4.2]. Thus, the prediction afforded by the Bateman–Horn conjecture is not unreasonably large.

Applications. Landau asked if there are infinitely many primes of the form $x^2 + 1$. The Bateman–Horn conjecture with $f_1(x) = x^2 + 1$ suggests that the answer is yes. Indeed,

$$\omega_f(p) = \begin{cases} 1 & \text{if } p = 2, \\ 2 & \text{if } p \equiv 1 \pmod 4, \\ 0 & \text{if } p \equiv 3 \pmod 4, \end{cases}$$

where $f = f_1$, so $C(f) \approx 1.37281$ and we expect that $Q(f; x) \sim \frac{C(f)}{2}\,\mathrm{Li}(x)$; see Figure 1.

Applying the Bateman–Horn conjecture to $f_1(x) = x$ and $f_2(x) = x + 2$ suggests the truth of the twin-prime conjecture. Indeed, $f_1(x)$ and $f_2(x)$ are simultaneously prime if and only if x is the least prime in a twin-prime pair. For $f = f_1 f_2$,

$$\omega_f(p) = \begin{cases} 1 & \text{if } p = 2, \\ 2 & \text{if } p \ge 3, \end{cases}$$

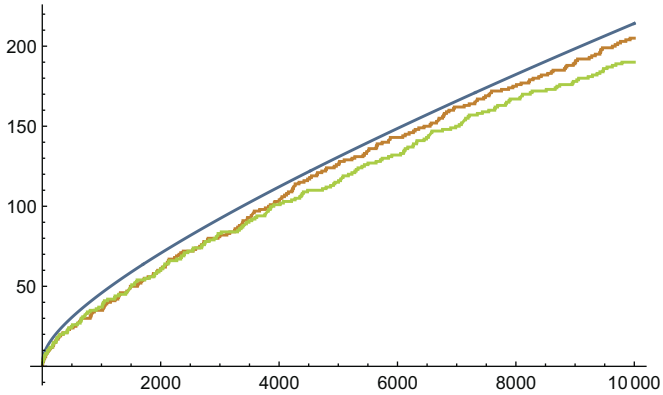

Figure 2. Counting functions of the twin primes (orange) and Sophie Germain primes (green) versus $2C_2 \int_2^x (\log t)^{-2}\, dt$ (blue). The Bateman–Horn conjecture asserts that these three functions are asymptotically equivalent.

so the corresponding Bateman–Horn constant is

$$C(f_1, f_2) = \prod_p \left(1 - \frac{1}{p}\right)^{-2} \left(1 - \frac{\omega_f(p)}{p}\right)$$
$$= 2 \prod_{p \geq 3} \frac{p(p-2)}{(p-1)^2} = 2C_2,$$

in which $C_2 \approx 0.660161815$ is the *twin primes constant*. The Bateman–Horn conjecture predicts that

$$Q(f_1, f_2; x) \sim 2C_2 \int_2^x \frac{dt}{(\log t)^2}.$$

In fact, we get the same prediction for $f_1(x) = x$ and $f_2(x) = x + 2^k$ with $k \geq 2$.

A *Sophie Germain prime* is a prime p such that $2p + 1$ is prime. The Bateman–Horn conjecture with $f_1(x) = x$ and $f_2(x) = 2x + 1$ yields the same prediction as in the twin-prime case; see Figure 2. The Bateman–Horn even explains the presence of curious patterns in the Ulam spiral [AZFG20, Sect. 6.6]!

References

[AZFG20] Soren Laing Aletheia-Zomlefer, Lenny Fukshansky, and Stephan Ramon Garcia, *The Bateman-Horn conjecture: heuristic, history, and applications*, Expo. Math. **38** (2020), no. 4, 430–479, DOI 10.1016/j.exmath.2019.04.005. MR4177951

[BH62] Paul T. Bateman and Roger A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. **16** (1962), 363–367, DOI 10.2307/2004056. MR148632

[BH65] Paul T. Bateman and Roger A. Horn, *Primes represented by irreducible polynomials in one variable*, Proc. Sympos. Pure Math., Vol. VIII, Amer. Math. Soc., Providence, RI, 1965, pp. 119–132. MR176966

[HL23] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio numerorum'; III: On the expression of a number as a sum of primes*, Acta Math. **44** (1923), no. 1, 1–70, DOI 10.1007/BF02403921. MR1555183

[SS58] A. Schinzel and W. Sierpiński, *Sur certaines hypothèses concernant les nombres premiers* (French), Acta Arith. **4** (1958), 185–208; erratum 5 (1958), 259, DOI 10.4064/aa-4-3-185-208. MR106202

[Ten15] Gérald Tenenbaum, *Introduction to analytic and probabilistic number theory*, 3rd ed., Graduate Studies in Mathematics, vol. 163, American Mathematical Society, Providence, RI, 2015. Translated from the 2008 French edition by Patrick D. F. Ion, DOI 10.1090/gsm/163. MR3363366

Stephan Ramon Garcia

Credits

Figures are courtesy of Stephan Ramon Garcia. Author photo is courtesy of Gizem Karaali.