Counting primes with polynomials

Orges Leka

September 24, 2025

Abstract

We define a family of integer polynomials $(f_n(x))_{n\geq 1)}$ and use three standard heuristic assumptions about Galois groups and Frobenius elements (H1–H3), together with the Inclusion–Exclusion principle (IE), to *heuristically* count: (1) primes up to N detected by irreducibility modulo a fixed prime p, and (2) primes in a special subfamily ("prime shapes") up to N. The presentation is self-contained and aimed at undergraduates.

Definition of the polynomials $f_n(x)$

Let $f_n(x) \in \mathbb{Z}[x]$ be defined recursively by

$$f_1(x) = 1, \qquad f_2(x) = x,$$
$$\text{if } n \text{ is prime:} \quad f_n(x) = 1 + f_{n-1}(x),$$
$$\text{if } n \text{ has the prime factorization } n = \prod_p p^{\nu_p(n)} : \quad f_n(x) = \prod_p \big(f_p(x)\big)^{\nu_p(n)}.$$

(All products are over primes p.) One checks that $\deg f_n$ grows logarithmically in n: there are fixed constants $0 < c_1 \leq c_2 < \infty$ such that

$$c_1 \log n \;\leq\; \deg f_n \;\leq\; c_2 \log n \qquad (n \geq 3),$$

e.g. $c_1 = 1/\log 3$ and $c_2 = 1/\log 2$ work.

Basic properties of $f_n(x)$

We collect elementary properties of the sequence $(f_n)_{n\geq 1}$ that follow immediately from the definition and simple inductions.

- **Multiplicativity.** For all $m, n \in \mathbb{N}$ one has

  $$f_{mn}(x) = f_m(x)\, f_n(x).$$

  Indeed this is built into the rule for composite n, and extends to all m, n by unique factorization.

- **Monicity, integral and nonnegative coefficients.** Since $f_2 = x$ is monic with integer coefficients and the rules are obtained from $f \mapsto f + 1$ and $(f, g) \mapsto fg$, it follows by induction that every f_n is monic in $\mathbb{Z}[x]$ and all coefficients are nonnegative. In particular the constant term is $f_n(0) = 1$ for all n (with $f_1(0) = 1$).

- **Evaluation at $x = 2$.** For all $n \geq 1$,

  $$f_n(2) = n.$$

Proof by strong induction on n: it holds for $n = 1, 2$. If n is prime, then

$$f_n(2) = 1 + \prod_{q|(n-1)} f_q(2)^{\nu_q(n-1)} = 1 + \prod_{q|(n-1)} q^{\nu_q(n-1)} = 1 + (n-1) = n.$$

If $n = \prod p^{\nu_p}$ is composite, then by multiplicativity

$$f_n(2) = \prod_p f_p(2)^{\nu_p} = \prod_p p^{\nu_p} = n.$$

- **Logarithmic degree growth.** There are absolute constants $0 < c_1 \leq c_2 < \infty$ such that for all $n \geq 3$,

  $$c_1 \log n \ \leq \ \deg f_n \ \leq \ c_2 \log n,$$

  e.g. $c_1 = 1/\log 3$ and $c_2 = 1/\log 2$. This follows by induction on n and the definiton of $f_n(x)$.

- **An equivalent characterization.** The sequence (f_n) is the unique family of nonzero polynomials satisfying: $f_2(x) = x$, $f_p(x) = f_{p-1}(x)+1$ for all primes $p > 2$, and $f_{mn} = f_m f_n$ for all m, n. (This tidy axiomatization was noted by Will Sawin.)[1]

Zeros lie in a left half-plane and irreducibility for prime indices

A key analytic observation is that, for every prime p, *all zeros of f_p lie in the half-plane* $\mathrm{Re}(z) < \frac{3}{2}$. From this, one can deduce irreducibility of f_p over $\mathbb{Z}[x]$. We include a self-contained proof adapted from Jonathan Love's MathOverflow answer.[2]

Lemma 1 (A root-location lemma). *Let $g(x) \in \mathbb{Z}[x]$ be a non-constant monic polynomial with constant term ± 1. If g is not a power of $(x+1)$, then g has a root θ with* $\mathrm{Re}(\theta) \geq -\frac{1}{2}$.

Proof. If all roots of g had real part $< -\frac{1}{2}$, then $|\,\theta + 1\,| < |\theta|$ for each root θ. For any irreducible factor h of g we would have

$$|h(-1)| = \prod_{\theta: h(\theta)=0} |\theta + 1| < \prod_{\theta: h(\theta)=0} |\theta| = |h(0)| = 1,$$

forcing $h(-1) = 0$, hence $h(x) = x + 1$. Thus $g(x) = (x+1)^m$. □

Lemma 2 (Uniform bound on $|f_p(z)|$ away from a compact set). *For each prime p and each z with* $\mathrm{Re}(z) \geq \frac{3}{2}$*, one has $|f_p(z)| > 2$. Consequently, every root θ of f_p satisfies* $\mathrm{Re}(\theta) < \frac{3}{2}$.

Proof. The claim is evident for $p = 2$. For $p = 3$ and $p = 5$ one checks directly: if $z = a + bi$ with $a \geq \frac{3}{2}$, then

$$|f_3(z)| = |z+1| = |(a+1)+bi| \geq a+1 > 2, \qquad |f_5(z)|^2 = |z^2+1|^2 = (a^2+(b-1)^2)(a^2+(b+1)^2) \geq a^4 > 4.$$

For $p \geq 7$, write by definition

$$f_p(z) = 1 + \prod_{q|(p-1)} f_q(z)^{\nu_q(p-1)}.$$

If $p - 1$ has an odd prime divisor q, then $|f_2(z)| = |z| \geq \frac{3}{2}$ and by induction $|f_q(z)| > 2$, so

$$|f_p(z)| \ \geq \ |f_2(z)|\,|f_q(z)| - 1 \ > \ \tfrac{3}{2} \cdot 2 - 1 \ = \ 2.$$

If instead $p - 1 = 2^k$ with $k \geq 3$, then $|f_p(z)| \geq |f_2(z)|^k - 1 > (\frac{3}{2})^3 - 1 > 2$. This proves the claim. □

[1] See the MathOverflow discussion for details.

[2] MathOverflow question *"Polynomials for natural numbers and irreducible polynomials for prime numbers?"*, answer by Jonathan Love (Dec. 11, 2024).

Proposition 1 (Irreducibility for prime indices)**.** *For every prime p, the polynomial $f_p(x)$ is irreducible in $\mathbb{Z}[x]$.*

Proof. Assume $f_p = FG$ with non-constant $F, G \in \mathbb{Z}[x]$. Since $f_p(2) = p$, we may assume $F(2) = \pm 1$. Consider $g(x) := F(x+2)$; then g is monic with constant term ± 1. If g were a power of $(x+1)$, then $F(1) = 0$, contradicting $f_p(1) > 0$ (all coefficients are nonnegative). Thus, by Lemma 1, g has a root with real part $\geq -\frac{1}{2}$, i.e. F has a root with real part $\geq \frac{3}{2}$. By Lemma 2 this is impossible, because all roots of f_p lie strictly to the left of the line $\mathrm{Re}(z) = \frac{3}{2}$. Hence f_p is irreducible. □

Further remarks. The proof also shows that all zeros of f_p lie in a fixed compact region, e.g. the set $\{z : |z| \leq \frac{3}{2}\} \cup \{z : |z+1| \leq 2\} \cup \{z : |z^2+1| \leq 2\}$, which contains the zero sets of all f_p (see the MO discussion).

Heuristic assumptions (H1–H3)

Fix once and for all a prime p. For each prime q, write $d_q = \deg f_q$ and let $G_q \leq S_{d_q}$ be the Galois group of the splitting field of f_q over $\mathbb{Q}$. We adopt:

- **(H1) Large Galois group.** Typically $G_q \simeq S_{d_q}$ (or at least contains a d_q-cycle).
- **(H2) Random Frobenius at p.** The Frobenius class at p in G_q behaves like a uniformly random element of G_q.
- **(H3) Weak independence across q.** For different primes q, the events we consider are independent enough that expectations add and inclusion–exclusion behaves as in the random model.

Heuristic probability of irreducibility mod p

Fix a prime p. For each prime q let $d_q = \deg f_q$ and let $G_q \leq S_{d_q}$ be the Galois group of the splitting field of f_q over $\mathbb{Q}$. We keep the assumptions:

- **(H1) Large Galois group:** typically $G_q \simeq S_{d_q}$ (or at least contains a d_q-cycle).
- **(H2) Random Frobenius at p:** the Frobenius class at p in G_q behaves like a uniformly random element of G_q.

The key dictionary (Dedekind–Frobenius, used here heuristically) is:

factorization pattern of $f_q \bmod p$ in $\mathbb{F}_p[x]$ $\longleftrightarrow$ cycle type of a random element of $G_q \subseteq S_{d_q}$.

In particular,

$$f_q \bmod p \text{ is irreducible} \iff \text{the associated permutation is a single } d_q\text{-cycle.}$$

Counting d-cycles in S_d

We now compute the exact fraction of permutations in S_d that are a single d-cycle.

$$|S_d| = d!,$$
$$\#\{d\text{-cycles in } S_d\} = \frac{d!}{d} = (d-1)!.$$

Reason: a d-cycle is just an ordering of the d symbols on a circle; there are $d!$ linear orderings, but each cyclic order has d starting points, so we divide by d.

Therefore the exact proportion of d-cycles in S_d is

$$\frac{\#\{d\text{-cycles}\}}{|S_d|} = \frac{(d-1)!}{d!} = \frac{1}{d}.$$

Heuristic probability

Under (H1)–(H2) with $G_q \simeq S_{d_q}$ and a uniform random element,

$$\boxed{\mathbb{P}\big(f_q(x) \bmod p \text{ is irreducible over } \mathbb{F}_p\big) \;\approx\; \frac{1}{d_q}.} \tag{1}$$

Relating d_q to $\log q$

From the basic properties of the sequence (f_n) (degree multiplicativity and recursion), one has for all sufficiently large primes q the two-sided bound

$$\frac{\log q}{\log 3} \;\le\; d_q \;\le\; \frac{\log q}{\log 2}. \tag{2}$$

Equivalently, writing $d_q \approx c\,\log q$ with a constant c depending only on the sequence (and lying in the interval $[1/\log 3,\ 1/\log 2]$), the reciprocal satisfies the sandwich estimate

$$\frac{\log 2}{\log q} \;\le\; \frac{1}{d_q} \;\le\; \frac{\log 3}{\log q}. \tag{3}$$

Combining (1) and (3) yields the explicit approximation

$$\boxed{\mathbb{P}\big(f_q \bmod p \text{ irreducible}\big) \;\approx\; \frac{1}{d_q} \;\approx\; \frac{1}{c\,\log q}, \qquad c \in \left[\tfrac{1}{\log 3},\ \tfrac{1}{\log 2}\right],} \tag{4}$$

and in particular for all large q,

$$\frac{\log 2}{\log q} \;\lesssim\; \mathbb{P}\big(f_q \bmod p \text{ irreducible}\big) \;\lesssim\; \frac{\log 3}{\log q}.$$

Interpretation. Equation (4) says: for a fixed modulus p, each prime q independently "fires" (i.e. gives $f_q \bmod p$ irreducible) with chance on the order of $1/\log q$. This is the only input needed to derive the sums and inclusion–exclusion formulas used later to estimate

$$\sum_{q \le N} \mathbb{P}(f_q \bmod p \text{ irreducible}) \;\approx\; \sum_{q \le N} \frac{1}{c\,\log q},$$

and to show (heuristically) that the union over $p \le N$ hits almost all primes $\le N$.

Counting for a fixed p: primes $\le N$

Step 0. Setup and notation

Fix a prime modulus p. For each prime q let $d_q = \deg f_q$. Recall the heuristic from (H1)–(H2):

$$\mathbb{P}\big(f_q(x) \bmod p \text{ is irreducible over } \mathbb{F}_p\big) \;\approx\; \frac{1}{d_q}.$$

From the basic properties of (f_n) we have logarithmic degree growth, so there exists a constant

$$c \in \left[\tfrac{1}{\log 3},\ \tfrac{1}{\log 2}\right] \quad \text{with} \quad d_q \approx c\,\log q,$$

hence

$$\frac{1}{d_q} \;\approx\; \frac{1}{c\,\log q}.$$

Step 1. Define the random variables

For each prime $q \le N$, define the indicator variable

$$X_q = \begin{cases} 1, & \text{if } f_q(x) \bmod p \text{ is irreducible over } \mathbb{F}_p, \\ 0, & \text{otherwise.} \end{cases}$$

Then the total number of such primes $q \le N$ is

$$U_p(N) := \sum_{\substack{q \le N \\ q \text{ prime}}} X_q.$$

By definition of expectation and linearity of expectation,

$$\mathbb{E}\, U_p(N) = \sum_{q \le N} \mathbb{E}\, X_q = \sum_{q \le N} \mathbb{P}(X_q = 1) \approx \sum_{q \le N} \frac{1}{d_q}.$$

Using $d_q \approx c \log q$ we obtain the *first-order* approximation

$$\mathbb{E}\, U_p(N) \;\approx\; \sum_{q \le N} \frac{1}{c \, \log q} \;=\; \frac{1}{c}\, S(N), \qquad S(N) := \sum_{q \le N} \frac{1}{\log q}. \tag{5}$$

Step 2. Estimating $S(N) = \sum_{q \le N} 1/\log q$ by summation by parts

Let $\pi(x)$ denote the prime-counting function. We write $S(N)$ as a Stieltjes integral with respect to $d\pi(x)$:

$$S(N) = \int_{2^-}^{N} \frac{1}{\log x}\, d\pi(x),$$

where 2^- indicates that if $N < 2$ the sum is empty (we will always take $N \ge 3$). Let

$$a(x) := \frac{1}{\log x} \quad \text{for } x \ge 3, \qquad A(x) := \pi(x).$$

By *summation by parts* (the discrete analogue of integration by parts),

$$\int_2^N a(x)\, dA(x) = a(N)\, A(N) \;-\; \int_2^N A(x)\, da(x).$$

We compute $da(x) = a'(x)\, dx$ with

$$a'(x) = -\,\frac{1}{x(\log x)^2}.$$

Hence

$$S(N) = \frac{\pi(N)}{\log N} \;+\; \int_2^N \frac{\pi(x)}{x(\log x)^2}\, dx. \tag{6}$$

Step 3. A Chebyshev-level upper bound for the integral term

We do *not* use the prime number theorem. Instead, we rely on the classical Chebyshev bounds (elementary) stating that for large x,

$$c_1 \frac{x}{\log x} \;\le\; \pi(x) \;\le\; c_2 \frac{x}{\log x}, \tag{7}$$

for some absolute constants $0 < c_1 \le c_2 < \infty$.

Plugging the upper bound from (7) into (6) gives

$$\int_2^N \frac{\pi(x)}{x(\log x)^2}\,dx \;\le\; c_2 \int_2^N \frac{1}{(\log x)^3}\,dx.$$

To estimate the last integral, set $u = \log x$ so $du = dx/x$ and $x = e^u$. Then

$$\int_2^N \frac{dx}{(\log x)^3} = \int_{\log 2}^{\log N} \frac{e^u}{u^3}\,du.$$

Integration by parts (or a simple comparison) shows that this integral grows like

$$\int_{\log 2}^{\log N} \frac{e^u}{u^3}\,du \;=\; O\Big(\frac{N}{(\log N)^3}\Big).$$

Therefore,

$$\int_2^N \frac{\pi(x)}{x(\log x)^2}\,dx \;=\; O\Big(\frac{N}{(\log N)^3}\Big). \tag{8}$$

Step 4. Dominant term and comparison of sizes

From (6) and (8) we obtain

$$S(N) = \frac{\pi(N)}{\log N} \;+\; O\Big(\frac{N}{(\log N)^3}\Big).$$

To see that the error term is genuinely smaller than the main term $\pi(N)/\log N$, compare sizes using the *lower* Chebyshev bound in (7):

$$\frac{\pi(N)}{\log N} \;\ge\; \frac{c_1 N}{(\log N)^2}.$$

Hence the ratio of error to main term is

$$\frac{O\big(N/(\log N)^3\big)}{\pi(N)/\log N} \;\ll\; \frac{N/(\log N)^3}{N/(\log N)^2} \;=\; \frac{1}{\log N} \;\longrightarrow\; 0.$$

Therefore

$$S(N) \;=\; \frac{\pi(N)}{\log N}\Big(1+o(1)\Big) \qquad \text{(no PNT needed; Chebyshev suffices).} \tag{9}$$

Step 5. Conclusion for the expectation

Returning to (5) and substituting (9) gives

$$\mathbb{E}\,U_p(N) \;\approx\; \frac{1}{c}\,S(N) \;=\; \frac{1}{c}\cdot\frac{\pi(N)}{\log N}\Big(1+o(1)\Big).$$

In boxed form:

$$\boxed{\mathbb{E}\,U_p(N) \;\approx\; \frac{1}{c}\cdot\frac{\pi(N)}{\log N}, \qquad c \in \Big[\tfrac{1}{\log 3},\, \tfrac{1}{\log 2}\Big].}$$

Equivalently, solving for $\pi(N)$ we get the heuristic relation

$$\pi(N) \;\approx\; c\,\mathbb{E}\,U_p(N)\,\log N.$$

Remarks

- The constant c comes from the growth law $d_q \approx c \log q$ for the degrees $\deg f_q$. Any choice of c in the interval $[\frac{1}{\log 3}, \frac{1}{\log 2}]$ is consistent with the proven degree bounds; numerically c can be estimated from data by averaging $d_q/\log q$ over primes $q \le N$.
- We never used the full Prime Number Theorem. Chebyshev's inequalities are enough to show the integral term is smaller by a factor $1/\log N$.
- Linearity of expectation needs no independence. We used (H1)–(H2) only to model $\mathbb{P}(X_q = 1) \approx 1/d_q$.

Inclusion–Exclusion over many p: near full coverage

Step 0. Fix q and define the events

Fix a prime q. For each prime p we consider the event

$$E_p(q) \;:=\; \{\, f_q(x) \bmod p \text{ is irreducible in } \mathbb{F}_p[x] \,\}.$$

Under (H1)–(H2), the Frobenius class at p behaves like a uniformly random element of a group $G_q \simeq S_{d_q}$ (heuristically), hence

$$\mathbb{P}\big(E_p(q)\big) \;\approx\; \kappa_q \;:=\; \frac{1}{d_q}.$$

We also assume (H3) that for distinct primes $p \neq p'$ the events $E_p(q)$ and $E_{p'}(q)$ are "independent enough" (we model them as independent Bernoulli trials with success probability κ_q). Thus the entire family $\{E_p(q)\}_{p \le N}$ is modeled as i.i.d. Bernoulli(κ_q).

Step 1. Inclusion–Exclusion for the union probability

We want the probability that *at least one* prime $p \le N$ makes $f_q \bmod p$ irreducible, i.e.

$$\mathbb{P}\Big(\bigcup_{p \le N} E_p(q) \Big).$$

The *inclusion–exclusion* (IE) identity states, for finitely many events $A_1, \dots, A_m$,

$$\mathbb{P}\Big(\bigcup_{j=1}^{m} A_j\Big) = \sum_{r=1}^{m} (-1)^{r+1} \sum_{1 \le j_1 < \cdots < j_r \le m} \mathbb{P}\big(A_{j_1} \cap \cdots \cap A_{j_r}\big).$$

Here $m = \pi(N)$ and A_j runs over $E_p(q)$ with $p \le N$. Under our independence model and with all single-event probabilities equal to κ_q,

$$\mathbb{P}\big(E_{p_1}(q) \cap \cdots \cap E_{p_r}(q)\big) \;\approx\; \kappa_q^{\,r}.$$

There are $\binom{\pi(N)}{r}$ such r-fold intersections, so IE becomes the binomial series

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{p \le N} E_p(q)\Big) &\approx \sum_{r=1}^{\pi(N)} (-1)^{r+1} \binom{\pi(N)}{r} \kappa_q^{\,r} = 1 - \sum_{r=0}^{\pi(N)} \binom{\pi(N)}{r} (-\kappa_q)^{\,r} \\ &= 1 - (1 - \kappa_q)^{\pi(N)}.\end{aligned}$$

Thus we obtain the closed form

$$\boxed{\mathbb{P}\Big(\bigcup_{p \le N} E_p(q)\Big) \;\approx\; 1 - (1 - \kappa_q)^{\pi(N)}.} \tag{10}$$

Step 2. Elementary bounds for $1-(1-\kappa)^m$

For $0 \le \kappa \le 1$ and $m \ge 1$ we have the standard inequalities

$$1 - e^{-m\kappa} \;\le\; 1-(1-\kappa)^m \;\le\; \min\{\, m\kappa,\ 1\,\}. \tag{11}$$

The upper bound $1-(1-\kappa)^m \le m\kappa$ is the union bound (Boole's inequality) or the first Bonferroni term. The lower bound follows from $(1-\kappa)^m \le e^{-m\kappa}$ (since $\log(1-\kappa) \le -\kappa$).

Applying (11) to (10) with $\kappa = \kappa_q$ and $m = \pi(N)$ gives the sandwich

$$1 - e^{-\pi(N)\kappa_q} \;\lesssim\; \mathbb{P}\Big(\bigcup_{p\le N} E_p(q)\Big) \;\lesssim\; \min\{\pi(N)\kappa_q,\, 1\}. \tag{12}$$

Step 3. Insert the size of κ_q and of $\pi(N)$

From the degree growth we have $d_q \asymp \log q$, hence for some absolute $C > 0$,

$$\kappa_q = \frac{1}{d_q} \;\gtrsim\; \frac{1}{C\log q}.$$

Also, by Chebyshev's elementary bounds, for large N there exists an absolute $c > 0$ with

$$\pi(N) \;\ge\; c\,\frac{N}{\log N}.$$

Therefore, uniformly for all $q \le N$,

$$\pi(N)\kappa_q \;\gtrsim\; \frac{N}{\log N}\cdot\frac{1}{C\log q} \;\ge\; \frac{N}{C(\log N)^2}.$$

Insert this in the *lower* bound of (12):

$$\mathbb{P}\Big(\bigcup_{p\le N} E_p(q)\Big) \;\gtrsim\; 1 - \exp\Big(-\frac{N}{C(\log N)^2}\Big) \;=\; 1 - o(1). \tag{13}$$

Thus, for each fixed $q \le N$, the probability that *no* prime $p \le N$ makes $f_q \bmod p$ irreducible is exponentially small in $N/(\log N)^2$.

Step 4. Expected size of the union over $p \le N$ and all $q \le N$

Define the random set of "hit" primes

$$\mathcal{H}(N) \;:=\; \big\{\, q \le N \text{ prime } : \exists\, p \le N \text{ prime with } E_p(q)\,\big\}.$$

Its (random) size is

$$|\mathcal{H}(N)| \;=\; \sum_{q\le N} \mathbf{1}_{\{\exists p\le N:\, E_p(q)\}}.$$

Taking expectations and using linearity,

$$\mathbb{E}\,|\mathcal{H}(N)| \;=\; \sum_{q\le N} \mathbb{P}\Big(\bigcup_{p\le N} E_p(q)\Big).$$

By (13), each summand is $1 - o(1)$ (with the same small $o(1)$ for all $q \le N$), hence

$$\begin{aligned}\mathbb{E}\,|\mathcal{H}(N)| &= \sum_{q\le N}\Big(1 - o(1)\Big) = \big(\pi(N)\big)\cdot\Big(1 - o(1)\Big)\\ &= \pi(N) - o\big(\pi(N)\big).\end{aligned}$$

In particular,

$$\boxed{\mathbb{E}\,\#\big\{\, q \le N :\ \exists p \le N,\ f_q \bmod p \ \text{ irreducible}\,\big\} \;=\; \pi(N) - o\big(\pi(N)\big).} \tag{14}$$

Step 5. Interpretation and robustness

- **"Near full coverage".** Equation (14) says that, under (H1)–(H3), the union over $p \le N$ *hits almost every prime* $q \le N$. The expected number of "misses" is at most of order

$$\sum_{q \le N} \exp\Big(-\Omega\big(N/(\log N)^2\big)\Big) \;\le\; \pi(N) \cdot \exp\Big(-\Omega\big(N/(\log N)^2\big)\Big),$$

  which is tiny compared to $\pi(N)$.

- **Why inclusion–exclusion matters.** If we kept only the first term (union bound), we would get the coarse estimate

$$\mathbb{P}\Big(\bigcup_{p \le N} E_p(q)\Big) \;\le\; \pi(N)\kappa_q,$$

  which correctly captures small-κ_q behavior but misses the saturation to 1. The full IE series *sums* to $1 - (1-\kappa_q)^{\pi(N)}$, which transitions from $\approx \pi(N)\kappa_q$ (when $\pi(N)\kappa_q \ll 1$) to ≈ 1 (when $\pi(N)\kappa_q \gg 1$).

- **Ramified or exceptional primes.** A finite set of small primes p may behave atypically (e.g. ramification). This affects at most $O(1)$ values of p and does not change the asymptotics, because $\pi(N) \to \infty$.

- **No need for the PNT.** We only used Chebyshev's inequalities to ensure $\pi(N) \gg N/\log N$, which suffices to make the exponent in (13) grow and force near certainty.

Special prime shapes

Step 0. Fix a special class of primes

Let

$$S(N) \;\subseteq\; \{\, q \le N : \; q \text{ prime} \,\}$$

be any specified family of primes up to N. Typical examples:

- *Arithmetic progressions:* $S(N) = \{q \le N : \; q \equiv a \pmod m\}$ with $(a,m) = 1$.
- *Polynomial shapes (one variable):* $S(N) = \{q \le N : \; q = f(n) \text{ prime for some } n \in \mathbb{N}\}$, e.g. $q = n^2 + 1$.
- *Two-linear forms (twin/Sophie Germain, etc.):* $S(N) = \{q \le N : \; q \text{ prime and } g(q) \text{ prime}\}$, e.g. $g(q) = 2q + 1$.
- *Mersenne primes:* $S(N) = \{q \le N : \; q = 2^r - 1 \text{ prime}\}$.

We will assume, in the spirit of (H1)–(H3), that the irreducibility model we used for all primes also applies *uniformly* to the subfamily $S(N)$: for each $q \in S(N)$ and each prime p,

$$\mathbb{P}\big(f_q \bmod p \text{ irreducible}\big) \;\approx\; \kappa_q = \frac{1}{d_q}, \qquad d_q = \deg f_q \asymp \log q,$$

and (for fixed q) the events over different p behave like independent Bernoulli trials with success probability κ_q.

Step 1. Per-q hit probability via Inclusion–Exclusion

Fix $q \in S(N)$. Define the events $E_p(q)$ as before:

$$E_p(q) = \{\, f_q(x) \bmod p \text{ is irreducible in } \mathbb{F}_p[x] \,\}.$$

By the inclusion–exclusion computation (with independence as in (H3)),

$$\mathbb{P}\Big(\exists\, p \le N :\ E_p(q)\Big) \ \approx\ 1 - (1-\kappa_q)^{\pi(N)}. \tag{15}$$

Using the elementary bounds $1 - e^{-m\kappa} \le 1 - (1-\kappa)^m \le \min\{m\kappa, 1\}$ with $m = \pi(N)$ and $\kappa = \kappa_q = 1/d_q$, we obtain

$$1 - \exp\big(-\pi(N)\kappa_q\big) \ \lesssim\ \mathbb{P}\Big(\exists\, p \le N :\ E_p(q)\Big) \ \lesssim\ \min\{\pi(N)\kappa_q,\, 1\}. \tag{16}$$

Since $d_q \asymp \log q$ and $q \le N$, there exists a fixed $C > 0$ with $\kappa_q \ge 1/(C\log q) \ge 1/(C \log N)$. Chebyshev's inequality gives $\pi(N) \ge c\,N/\log N$ for some absolute $c > 0$, so

$$\pi(N)\kappa_q \ \ge\ \frac{c\,N}{\log N}\cdot\frac{1}{C\log N} \ =\ \frac{c}{C}\cdot\frac{N}{(\log N)^2}.$$

Plugging into the *lower* bound in (16) yields

$$\mathbb{P}\Big(\exists\, p \le N :\ E_p(q)\Big) \ \gtrsim\ 1 - \exp\Big(-\frac{c}{C}\cdot\frac{N}{(\log N)^2}\Big) \ =\ 1 - o(1), \tag{17}$$

uniformly for all $q \in S(N)$.

Step 2. Expected number of hits inside $S(N)$

Let

$$H_S(N) \ :=\ \#\Big\{\, q \in S(N) :\ \exists\, p \le N,\ f_q \bmod p \text{ irreducible} \Big\}.$$

By linearity of expectation and (15),

$$\mathbb{E}\, H_S(N) \ =\ \sum_{q\in S(N)} \mathbb{P}\Big(\exists\, p \le N :\ E_p(q)\Big) \ \approx\ \sum_{q\in S(N)} \Big[1 - (1-\kappa_q)^{\pi(N)}\Big].$$

Using the uniform lower bound (17), we get

$$\mathbb{E}\, H_S(N) \ \ge\ \sum_{q\in S(N)} \Big(1 - \exp\big(-\Omega(N/(\log N)^2)\big)\Big) \ =\ |S(N)| \ -\ |S(N)|\cdot\exp\big(-\Omega(N/(\log N)^2)\big).$$

Since $|S(N)| \le \pi(N)$ and the exponential factor decays faster than any power of N, the "expected misses" are negligible:

$$\boxed{\mathbb{E}\, H_S(N) \ =\ |S(N)|\cdot\big(1 - o(1)\big).} \tag{18}$$

Step 3. A note on concentration (optional, heuristic)

If we strengthen (H3) to say that *for different* q the families $\{E_p(q)\}_{p\le N}$ are weakly dependent enough (or approximately independent), then standard concentration inequalities (Chernoff/Hoeffding for sums of bounded variables) suggest that $H_S(N)$ is tightly concentrated around its mean. Heuristically,

$$H_S(N) \ =\ |S(N)|\cdot\big(1 - o(1)\big) \qquad \text{with high probability.}$$

We will not rely on this; the expectation (18) already shows that our method loses asymptotically nothing.

Step 4. How $|S(N)|$ is obtained (external number theory)

Our framework is *conditional* on an external estimate for $|S(N)|$. Some standard inputs:

- **Primes in APs.** (Dirichlet's theorem, plus effective forms.) For fixed $(a, m) = 1$,

$$|S(N)| \;=\; \#\{q \le N : \; q \equiv a \bmod m \text{ prime}\} \;\sim\; \frac{1}{\varphi(m)} \cdot \frac{N}{\log N}.$$

- **One-variable prime-producing polynomials.** (Bateman–Horn conjecture.) E.g. for $q = n^2 + 1$,

$$|S(N)| \;\sim\; C_{n^2+1} \cdot \frac{\sqrt{N}}{\log N}.$$

- **Two-linear forms (e.g. Sophie Germain).** (Bateman–Horn.) For q prime and $2q + 1$ prime,

$$|S(N)| \;\sim\; C_{\mathrm{SG}} \cdot \frac{N}{(\log N)^2}.$$

- **Mersenne primes.** (Wagstaff/Lenstra–Pomerance heuristics.) Up to bound N,

$$|S(N)| \;\approx\; \frac{e^{\gamma}}{\log 2} \log\log N.$$

Whatever the ambient asymptotic for $|S(N)|$ *is*, the expectation (18) says our detection count matches it up to a $(1 - o(1))$ factor.

Step 5. Putting it all together

Combining the "fixed p" estimate and the union estimate:

- For any *fixed* prime p,

$$\mathbb{E}\, U_p(N) \;\approx\; \frac{1}{c} \cdot \frac{\pi(N)}{\log N}, \qquad c \in \left[\tfrac{1}{\log 3}, \tfrac{1}{\log 2}\right],$$

  by the detailed summation-by-parts argument.

- For the *union over all* $p \le N$, and for any special class $S(N)$,

$$\mathbb{E}\, \#\Big\{\, q \in S(N) : \; \exists\, p \le N, \; f_q \bmod p \text{ irreducible} \Big\} \;=\; |S(N)| \cdot \big(1 - o(1)\big),$$

  by the inclusion–exclusion estimate and the uniform bound $\kappa_q \gtrsim 1/\log q$.

Summary. Under (H1)–(H3):

1. For a fixed prime p,

$$\mathbb{E}\, U_p(N) \;\approx\; \frac{1}{c} \cdot \frac{\pi(N)}{\log N}.$$

2. For any special prime class $S(N)$,

$$\mathbb{E}\, H_S(N) \;=\; |S(N)| \cdot (1 - o(1)).$$

Thus our inclusion–exclusion heuristic *loses essentially nothing*: the count of detected primes inside $S(N)$ is asymptotically the full ambient size $|S(N)|$, whatever that size is (from theorems like Dirichlet or conjectures like Bateman–Horn/Wagstaff).

Empirical justification

Setup. The following experiment was generated by running the SageMath script `counting_primes_with_polynomials.sage`. It constructs the polynomials $f_n(x)$, tests irreducibility of $f_q(x)$ mod p over $\mathbb{F}_p$, and measures both the fixed-p count $U_p(N)$ and the inclusion–exclusion union coverage over all primes $p \le P_{\max}$, restricted to various special prime classes $S(N)$.

Parameters and results (raw console output).

```
=== Parameters ===
N=50000, Pmax=50000, fixed p=101
==================

>> Fixed p baseline
Estimated c_N ~ 1.2203
U_p(N) for p=101: 448
Prediction sum_q 1/(c log q): 443.768
pi(N) ~ 5133, pi(N)/log N ~ 474.409

>> Union coverage (all primes q <= N)
All primes                      covered  5123 /  5133    ratio = 0.998

>> Special prime classes S(N) and union coverage over p <= Pmax
AP q ~=~ 1 (mod 4)                covered  2539 /  2549    ratio = 0.996
AP q ~=~ 1 (mod 3)                covered  2554 /  2556    ratio = 0.999
AP q ~=~ 1 (mod 5)                covered  1270 /  1274    ratio = 0.997
AP q ~=~ 2 (mod 5)                covered  1285 /  1289    ratio = 0.997
q = n^2 + 1                     covered    33 /    37    ratio = 0.892
q = n^2 + n + 41                covered   169 /   169    ratio = 1.000
twin primes (q, q+2)            covered   702 /   705    ratio = 0.996
Sophie Germain q                covered   669 /   670    ratio = 0.999
Mersenne primes                 covered     5 /     5    ratio = 1.000

Done.
```

Interpretation. The fixed-p count $U_p(N)$ closely matches the heuristic prediction $\sum_{q \le N} \frac{1}{c \log q} \approx \frac{1}{c} \cdot \frac{\pi(N)}{\log N}$, and the union over $p \le P_{\max}$ *almost* hits all primes $q \le N$ (ratio 0.998). Within special classes $S(N)$ (APs, polynomial shapes, twin/Sophie Germain, Mersenne), the observed coverage ratios are ≈ 1, in line with the inclusion–exclusion prediction that the detected count inside $S(N)$ is $|S(N)|(1 - o(1))$.

Remark (where we use irreducibility of f_n). Short answer: we only really use the "if n is prime, then f_n is irreducible over $\mathbb{Q}$" half. The converse ("if n is composite, then f_n is reducible") is true but not essential for our counting.

- **(1) Setting up the model for primes q.** All counting arguments restrict to q prime and study f_q. We need f_q irreducible over $\mathbb{Q}$ so that (i) it has a well-defined degree $d_q = \deg f_q$ equal to $[\mathbb{Q}[x]/(f_q) : \mathbb{Q}]$; (ii) its splitting field has a *transitive* Galois group $G_q \le S_{d_q}$, allowing the Dedekind–Frobenius dictionary (factorization mod $p \leftrightarrow$ cycle type); (iii) hence "f_q mod p is irreducible" $\iff$ "Frobenius at p is a d_q-cycle", giving the success probability $\kappa_q \approx 1/d_q$.

- **(2) Log-degree control in the probabilities.** We use $d_q \asymp \log q$ to turn κ_q into $\asymp 1/\log q$. This is applied only for prime q, i.e. to irreducible f_q.
- **(3) Inclusion–Exclusion for each fixed q.** IE needs a single success probability κ_q per p. This relies on (1): for irreducible f_q over $\mathbb{Q}$, "success" truly means "d_q-cycle" with chance $1/d_q$.

What we *do not* need:

- We never use the "$\Leftarrow$" direction for composites in the counting. Although for composite n one has $f_n = \prod_p f_p^{\nu_p(n)}$ (hence reducible), our sums run only over *prime* q.
- The IE "near full coverage" over $p \leq N$ is computed *per prime* q, so again only "prime $\Rightarrow$ irreducible over $\mathbb{Q}$" is invoked.

In one line: we use "n prime $\Rightarrow f_n$ irreducible over $\mathbb{Q}$" to justify the *d_q-cycle model* and $\kappa_q = 1/d_q$; the converse is not needed for the heuristic counts (though it explains why composites are irrelevant).