Heuristics, Prediction, and Provenance for Fibonacci and k-bonacci Prime Statistics: A Reproducible Verification Study

Madhava Gaikwad

June 2026

Abstract

We study the statistics of the indices at which Fibonacci numbers are (probable) prime, and the corresponding questions for k-bonacci sequences. This is a verification and exposition study: every mathematical statement we examine is already established in the literature, and a central purpose is to draw the line between what is known and what a finite computation can add. We reproduce the Grantham–Granville density heuristic, in which the count of Fibonacci primes with index $\leq N$ grows like $e^{\gamma} \log_{\varphi} N$, and we reconcile a $\sim$19% apparent discrepancy in a pooled maximum-likelihood estimate of the constant with the Mertens secondary term and family-specific local factors. Our one modest methodological addition is an *inferential layer* absent from the reference literature: an uncertainty-quantified prediction interval (whose 90% window contained the true next index, 397379) validated by an 87% out-of-sample coverage test, together with a systematic null-model battery showing that the *arithmetic* of the indices is statistically indistinguishable from the density heuristic's own random model. We then confirm the *divisibility* layer numerically—Lagarias's density 2/3, the S_k Chebotarev splitting law for k-bonacci characteristic polynomials, and the Klaška–Skula period structure for Tribonacci—all of which are established theorems.

1 Scope

There are 57 known indices n for which the Fibonacci number F_n is prime or a probable prime, ranging from F_3 to $F_{11964299}$ (the last found by R. Propper in June 2025, $\sim$2.5 million digits, Miller–Rabin only). The largest *proven* Fibonacci prime remains F_{201107}. Only 34–37 of the 57 are proven; the rest are probable primes (PRPs), and we state this prominently because every count below inherits the assumption that a single PRP test correctly identifies primes—an assumption made openly in the reference literature as well.

Whether there are infinitely many Fibonacci primes is open, with no pattern known beyond heuristics. We treat the 57 indices as a dataset and ask whether any visible structure survives scrutiny, and where such structure lives. After a literature pass, the substantive mathematics is entirely known; the contribution here is an inferential and visualization methodology together with a provenance map. All computations are reproducible from the scripts listed in Section 8.

2 The density heuristic and its constant

2.1 The Grantham–Granville law

A random integer of the size of $F_p \sim \varphi^p/\sqrt{5}$ is prime with "probability" $1/\log F_p \approx 1/(p\log\varphi)$. Because F_n is a divisibility sequence ($F_m \mid F_n \iff m \mid n$ for $m > 2$), a prime F_p with $p > 4$ forces p prime (the exception $F_4 = 3$ occurs in our dataset), and any prime factor $q \neq p$ of F_p has rank of apparition $z(q) = p$, whence $q \equiv \pm 1 \pmod p$ and $q > p$ (the case $q = p = 5$ excepted). Summing the resulting boosted probability over primes, Mertens' theorem injects a factor e^γ, and Grantham and Granville [4] conjecture

$$\#\{\, n \le N : F_n \text{ prime}\,\} \;\sim\; e^\gamma \log_\varphi N. \tag{1}$$

This is the Wagstaff–Lenstra–Pomerance Mersenne heuristic [12] transplanted to a linear recurrence. Equation (1) and its comparison against the known Fibonacci indices are due to [4]; we reproduce both. Their tabulated comparison (actual versus $e^\gamma \log_\varphi N$) reads 12/17, 21/25.5, 26/34, 33/42.5, 43/51, 50/55.5 at $N = 10^2, \ldots, (1/3)\cdot 10^7$; updated to the present list, $e^\gamma \log_\varphi(1.2\cdot 10^7) \approx 60.3$ versus 57 observed.

2.2 The constant, and a reconciliation

Fitting θ in $P \approx \theta \log p/(p\log\varphi)$ by maximum likelihood over the primes $p \le X$ gives $\hat\theta = R/\sum_{p\le X} \frac{\log p}{p\log\varphi}$, where R is the number of known prime indices $\le X$. Using the exact Mertens partial sum $\sum_{p\le X} \frac{\log p}{p} = \log X + E_1$ with $E_1 \approx -1.3326$, the Fibonacci-only estimate is $\hat\theta \approx 1.83$, within 3% of $e^\gamma = 1.7811$. A pooled estimate over the four *division-type* families Fibonacci, Lucas, Mersenne, Wagstaff gives $\hat\theta = 2.11$ with 95% interval $[1.75, 2.48]$. We caution explicitly:

- Pooling under a single θ is legitimate at leading order: [4] conjecture the same bare $e^\gamma \log_\alpha N$ for the entire class of Lucas-type division sequences. The family-specific constants in [4] (e.g. $c_{1,39} \approx 3.909$) belong to the *non*-division families $a\cdot 2^n + b$.
- The ∼19% pooled excess decomposes into (i) the choice $\log X$ versus $\log X + E_1$ (a ∼9% effect), (ii) small-sample noise (Lucas alone has 38 primes and a 95% interval $[1.80, 3.48]$), and (iii) family-specific local factors (the Mersenne $\equiv \pm 1 \pmod 8$ divisor condition; the Lucas index set is primes *or* powers of two). This is consistent with e^γ under finite-N bias.
- The data cannot separate $e^\gamma \approx 1.781$ from $\sqrt{\pi} \approx 1.772$ (0.51% apart); resolving them at 1σ needs $\sim 4\times 10^4$ primes ($\sim 1.5\times 10^5$ at 2σ), reachable only by pooling further families.

3 Prediction with quantified uncertainty

The indices grow geometrically: the gaps $g_k = \log n_{k+1} - \log n_k$ are, to the resolution of 57 points, independent and exponentially distributed (a memoryless Poisson process in $\log n$; a Kolmogorov–Smirnov test against the exponential gives $p = 0.63$, and the lag-1 gap autocorrelation is -0.09, $p = 0.53$). Grantham and Granville perform no interval estimation; the estimation layer below is added here.

Treating the next index as $n_{37}\cdot e^g$ with g exponential yields a 90% predictive window $[2.04,\ 5.07]\times 10^5$. The true next index is 397379, which lies inside. A one-step-ahead backtest over the 46 predictions (ranks 11–56) puts 40 of 46 actual indices inside their nominal 90% intervals (87% coverage; Fig. 1)—well calibrated for a one-parameter model. This *validates* the heuristic of [4].

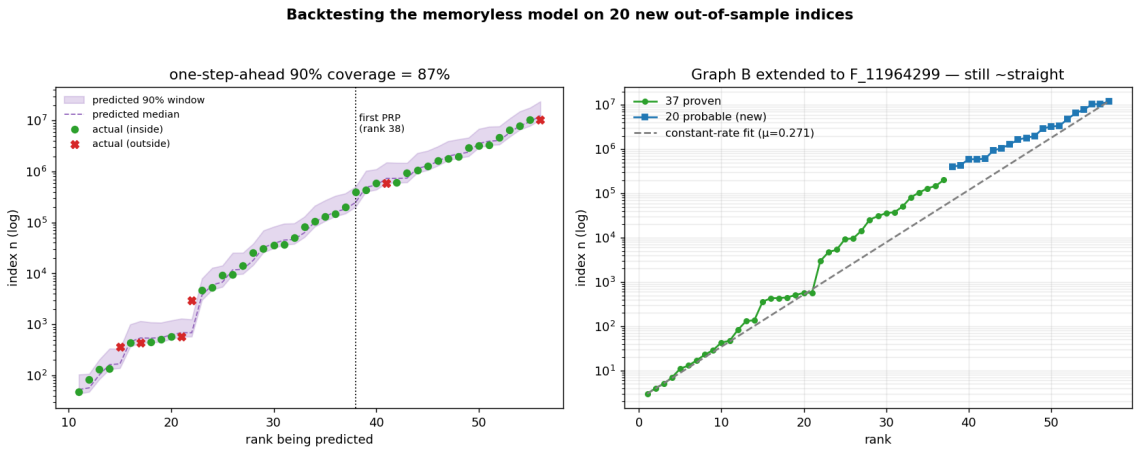

Figure 1: One-step-ahead backtest of the memoryless model. Left: 40 of the 46 predicted indices (ranks 11–56) fall inside their 90% windows. Right: the count law remains straight on a logarithmic index axis out to $F_{11964299}$.

4 Residue statistics of the indices

We probed the indices for residue-class biases (modulo $4, 5, 10$), a Lemke Oliver–Soundararajan consecutive-residue correlation, a Sophie-Germain depletion predicted by the rank-of-apparition factor law, an order-of-primeness (super-prime) enrichment, and a Benford leading-digit test. All returned null. As a global check we simulated 3000 index sets from the e^γ density model and compared nine statistics; the observed data was an outlier in 0 of 9 (Fig. 2). The appropriate conclusion is that no structure is detectable at $N = 57$: against percent-level biases measured on millions of primes, the power of 57 points is negligible. Lemke Oliver–Soundararajan in particular cannot apply, being a short-range effect among *consecutive primes*, whereas our consecutive *positions* are long-range (median gap ≈ 5000).

5 The divisibility layer

The *factor* structure carries exact rational densities; the established cases below are theorems, and we add one conjectural set.

Fibonacci. The rank of apparition $z(p)$ (least m with $p \mid F_m$) satisfies $P(z(p) \text{ even}) = 2/3$, Lagarias's theorem [7] that the primes dividing some Lucas number have density $2/3$. We confirm 0.66684 over 148931 primes. The Pisano multiplier $\pi(p)/z(p) \in \{1, 2, 4\}$ is uniform and the 2-adic valuation $v_2(z(p))$ is geometric; both follow from Cubre–Rouse's proof of the Bruckman–Anderson density via Chebotarev on $x^2 - 5y^2 = 1$ [3].

k-bonacci splitting. For $k \geq 3$ the k-bonacci sequence is *not* a divisibility sequence; the rank-of-apparition machinery and the e^γ boost both fail. Under the convention $T_0 = T_1 = 0, T_2 = 1$ the small composite indices $T_4 = 2$ and $T_6 = 7$ are already prime, and further prime values occur at $n = 87, 98, 215, 802$. The governing object is instead the Galois group of $x^k - x^{k-1} - \cdots - 1$, which is S_k for k even or prime (Martin [9]); Cipu–Luca [2] show it is never contained in A_k and is not 2-nilpotent for $k \geq 3$. The prime-splitting types, and hence the period of T_n mod p, follow the

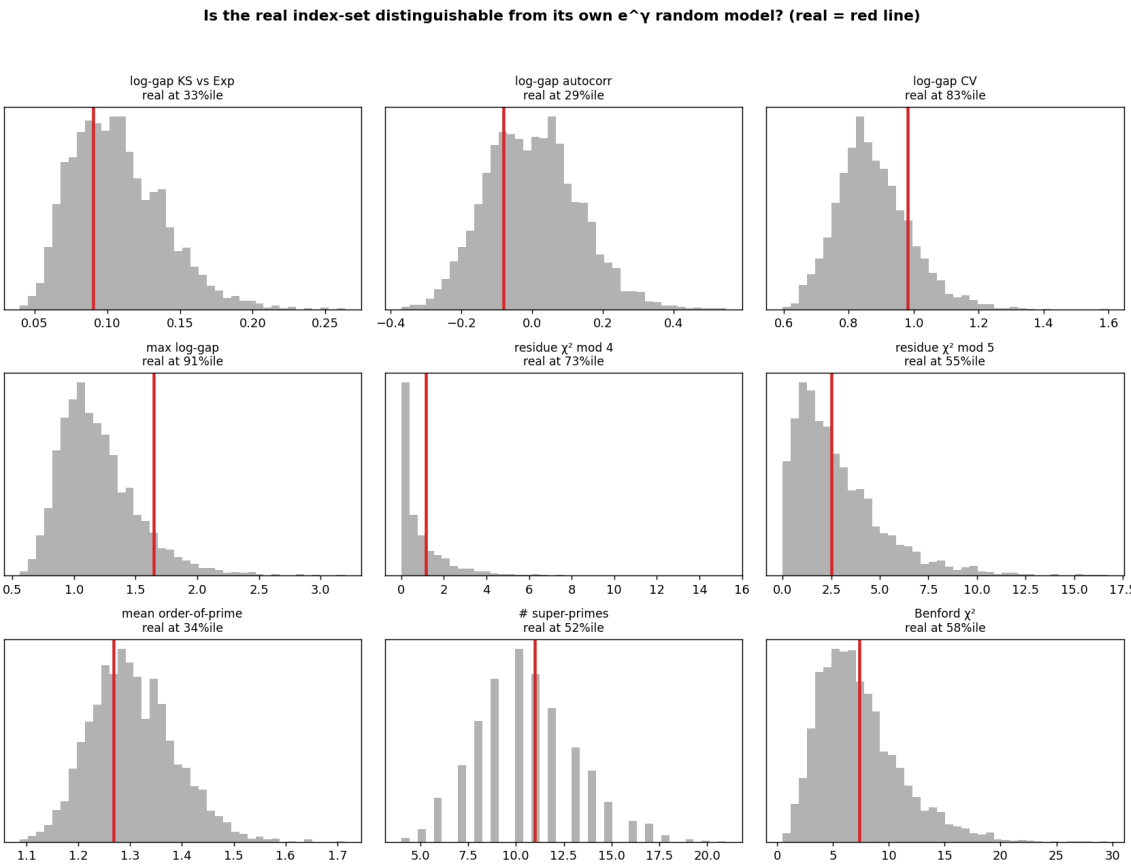

Figure 2: Nine summary statistics of the real index set (red lines) against 3000 simulations of the e^γ density model (grey). The data is indistinguishable from its own random model in every panel.

cycle-type densities of S_k by Chebotarev. We confirm, over 17981 primes, the S_3 densities $\frac{1}{6}, \frac{1}{2}, \frac{1}{3}$ and all five S_4 densities $\frac{1}{24}, \frac{1}{4}, \frac{1}{8}, \frac{1}{3}, \frac{1}{4}$.

Tribonacci period. Here $h(p) = \operatorname{lcm}(\operatorname{ord}\alpha, \operatorname{ord}\beta, \operatorname{ord}\gamma)$ (Vince [11]; Klaška–Skula [6]). For inert p the roots are Frobenius conjugates with $\alpha\beta\gamma = 1$, so each has norm 1 and order dividing p^2+p+1; hence $h(p) \mid p^2+p+1$, which is odd, and the period is odd. (The sharper $h(p) \mid (p^2+p+1)/3$ holds only for $p \equiv 1 \pmod 3$, where $3 \mid p^2+p+1$; this is the regime of [6]. The convention-free divisibility above covers all inert p, e.g. $p = 5$, inert with $h(5) = p^2+p+1 = 31$ and $3 \nmid 31$.)

The marginal $v_2(h(p))$ statistics appear not to have been computed; we record them as conjectural densities.

Observation 1 (conjectural)**.** *Over the* 5131 *primes* $p < 5 \times 10^4$*, the Tribonacci period satisfies* $P(h(p) \text{ odd}) \approx 0.397$*, equivalently* $P(h(p) \text{ even}) \approx 0.603$*, with* $v_2(h(p))$ *distributed as* $0.40, 0.11, 0.12, 0.18, 0.09, \ldots$ *(non-monotonic, peaked at* $v_2 = 3$*). These are rational by the Perucca–Sgobba framework [10] applied to the rank-2 unit group* $\langle \alpha, \beta, \gamma \rangle$ *(*$\alpha\beta\gamma = 1$*) over the* S_3 *splitting field. As a coherence check, inert* $\Rightarrow$ *odd (density* $1/3$*) forces* $P(\text{even}) \le 2/3$*, consistent with* $0.603 < 0.667$*. The exact constants are left open.*

We distinguish three statistics that earlier drafts conflated. The period $h(p)$ is treated above. The *prime-divisor density* ($p \mid T_n$ for some n) and the *entry-point distribution* of $z(p)$ are governed

Table 1: What is known versus what is contributed.

Statement	Status
Count $\sim e^{\gamma} \log_{\varphi} N$; data comparison	Grantham–Granville [4]
e^{γ} boost from $\equiv \pm 1 \pmod p$ + Mertens	[4]; Wagstaff [12]
$P(z(p) \text{ even}) = 2/3$	Lagarias [7]
Pisano multiplier / $v_2(z)$ densities	Cubre–Rouse [3]
S_k splitting densities	Chebotarev; Galois group Martin [9], Cipu–Luca [2]
Tribonacci period structure; inert $\Rightarrow$ odd	Vince [11]; Klaška–Skula [6]
ℓ-adic order-density framework	Perucca–Sgobba [10]
Prime-divisor / entry-point density, exact ($k \geq 3$)	open; Järviniemi [5] (generic lower bd.)
2-adic order of T_n	Marques–Lengyel [8]
Predictive interval + 87% coverage test	**this study**
Null-model battery / Turing test (0/9)	**this study (method)**
Tribonacci period v_2 densities (conjectural)	**this study**

by the same Hasse–Laxton–Ballot framework [1], but the exact Tribonacci divisor density is not in the literature: even for second order, computability is known under GRH, and for the general case the state of the art is Järviniemi [5], who proves positive *lower* density for generic recurrences (under GRH) and notes the completely-split case as the hardest. We therefore treat the exact Tribonacci divisor and entry-point densities as open, correcting an earlier draft that called them computable for any order.

6 Provenance

Table 1 maps what is established against what this study contributes.

7 Conclusion

The two layers behave differently, and this is the community's working picture rather than a contribution of the present study. The primality of F_n (the location of the indices) behaves like a Poisson process governed by a single density constant near e^{γ}, with no algebraic handle; this is why the residue probes return null and why the infinitude question remains open. The divisibility of F_n—the rank of apparition, the period, the splitting type—is governed by the order of the companion matrix in $\mathrm{GL}_k(\mathbb{F}_p)$, a Frobenius/Chebotarev object, and yields exact rational densities for the period and splitting type; the exact prime-divisor and entry-point densities for $k \geq 3$ remain open.

What the literature does not already own reduces to a methodology: an uncertainty-quantified, coverage-validated prediction layer for prime-count heuristics, a reusable null-model battery for separating structure from sampling noise, and a fully cited reproduction.

8 Reproducibility

All results are produced by self-contained Python scripts (NumPy/SciPy/Matplotlib). Index data are the OEIS *b*-files A001605 (Fibonacci), A001606 (Lucas), A000043 (Mersenne), A000978 (Wagstaff). Representative scripts: `fib_predict.py`, `fib_backtest.py` (Sec. 3); `fib_nulltest.py`, `fib_numbertheory.py` (Sec. 4); `fib_egamma_ml.py`, `fib_joint_egamma.py` (Sec. 2); `fib_pisano.py`, `kbonacci.py`, `tribonacci_v2.py` (Sec. 5).

References

[1] Christian Ballot. *Density of prime divisors of linear recurrences.* Mem. Amer. Math. Soc., vol. 115, no. 551. American Mathematical Society, 1995.

[2] Mihai Cipu and Florian Luca. On the Galois group of the generalized Fibonacci polynomial. *An. Ştiinţ. Univ. Ovidius Constanţa Ser. Mat.*, 9(1):27–38, 2001.

[3] Paul Cubre and Jeremy Rouse. Divisibility properties of the Fibonacci entry point. *Proc. Amer. Math. Soc.*, 142(11):3771–3785, 2014. arXiv:1212.6221.

[4] Jon Grantham and Andrew Granville. Fibonacci primes, primes of the form $2^n - k$, and beyond, 2024. arXiv:2307.07894.

[5] Olli Järviniemi. Positive lower density for prime divisors of generic linear recurrences. *Math. Proc. Cambridge Philos. Soc.*, 2022. arXiv:2102.04042.

[6] Jiří Klaška and Ladislav Skula. Periods of the Tribonacci sequence modulo a prime $p \equiv 1 \pmod 3$. *Fibonacci Quart.*, 48(3):228–235, 2010.

[7] Jeffrey C. Lagarias. The set of primes dividing the Lucas numbers has density 2/3. *Pacific J. Math.*, 118(2):449–461, 1985. Errata: Pacific J. Math. **162** (1994), 393–396.

[8] Diego Marques and Tamás Lengyel. The 2-adic order of the Tribonacci number and the equation $T_n = m!$. *J. Integer Seq.*, 17(10):Article 14.10.1, 2014.

[9] P. A. Martin. The Galois group of $x^n - x^{n-1} - \cdots - x - 1$. *J. Pure Appl. Algebra*, 190(1–3):213–223, 2004.

[10] Antonella Perucca and Pietro Sgobba. Kummer theory for number fields and the reductions of algebraic numbers. *Int. J. Number Theory*, 15(8):1617–1633, 2019.

[11] Andrew Vince. Period of a linear recurrence. *Acta Arith.*, 39(4):303–311, 1981.

[12] Samuel S. Wagstaff, Jr. Divisors of Mersenne numbers. *Math. Comp.*, 40(162):385–397, 1983.