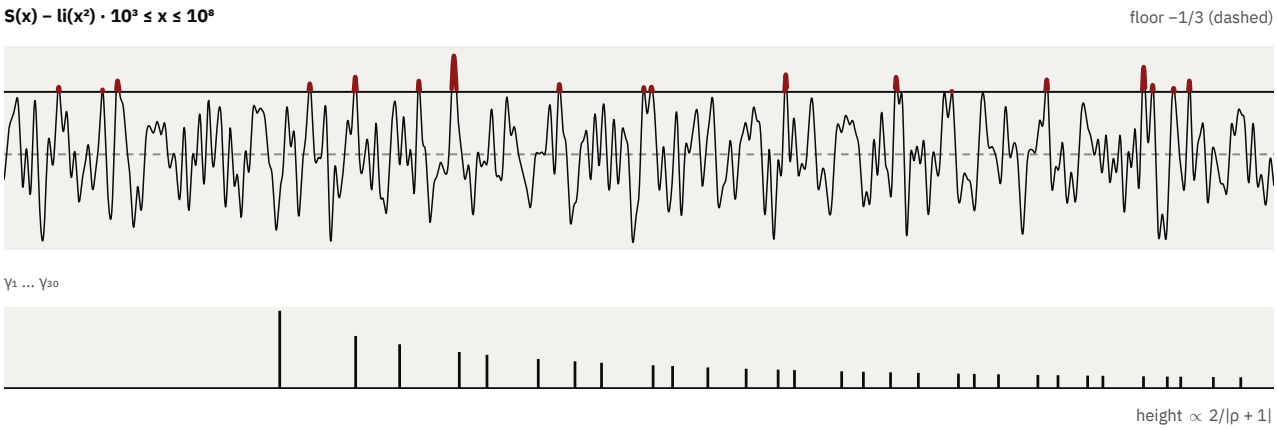



The ripple of the prime sum from the explicit formula with the first 30 zeros (top), its floor $-1/3$ left by the squares of the primes (dashed), and the comb of zero ordinates that drives it (bottom).

hexagon:2611.00130v1

# Cumulative Prime Sums and Riemann Zeros

*A fine grid, its ripple, and a family of prime races*

**Antonio Mennillo**

Naples, Italy · antoniomennillo87@gmail.com
ORCID 0009-0001-5978-6015

**NOTE ON THIS VERSION**

Version 4 revises v3 (Zenodo, October 2026, DOI 10.5281/zenodo.23269807). No result of v3 is withdrawn. New in this version:

- Theorem 4(b), stated in v3 as a heuristic, is proved under RH and LI (local limit theorem for the characteristic function $\prod J_0$).
- Corrections in the proof of Theorem 3: the lower limit of Lemma A, the error term of Step 2, the layers $k \geq 3$ of Step 3, the multiplicities in Step 5; LI is stated with multiplicity.
- All theorems are formalized in Lean 4, with the classical results they use as explicit hypotheses (Section 12).

Appendix B lists every change with respect to v3; Appendix A, the changes of v3 with respect to v2 (Zenodo, March 2026, DOI 10.5281/zenodo.19189786). Protocol, certified records and scripts are collected in the companion *Supplementary Material*.

## ABSTRACT

Summing the primes one after another and taking logarithms gives an increasing sequence of real numbers, $\sigma_m = \log S_{p_m}$ with $S_N = 1 + \sum_{p \le N} p$, which becomes exponentially dense. The ordinates $\gamma_n$ of the zeros of the Riemann zeta function fall on this grid with tiny errors: $1.4 \cdot 10^{-3}$ at the first zero, $2.4 \cdot 10^{-7}$ at the fifth, $3 \cdot 10^{-17}$ at the twentieth.

**Part I** explains why. The cell of the grid containing a height $t$ has width $(1+o(1))\sqrt{t}\, e^{-t/2}$ (Step Theorem), and every real $t > \log 3$ has a node above it within that width, by an argument going back to Bertrand's postulate (Covering Theorem). The precision observed at the zeros is therefore the resolution of the grid, and the same precision holds at every height (Genericity Principle). We compute the first-crossing primes $N_n$ with certification up to $n = 20$ and test, in a protocol fixed before the computation, whether the zeros sit closer to the nodes than a generic height. Neither the mean distance to the nearest node ($0.246$ against $0.250$, $p = 0.92$) nor the distribution of the position within the cell (Kolmogorov–Smirnov, $p = 0.79$) departs detectably from the generic case, and an integer grid used as control behaves in the same way. We also correct the half-exponent law by a factor $\sqrt{t'/t}$.

**Part II** shows where the zeros are actually visible. They are not in the position of the nodes but in the *ripple*: the deviation of the grid from its mean curve $B(x) = 1 + \mathrm{li}(x^2) - \mathrm{li}(4)$. This ripple is a sum of pure oscillations, one per zero; its spectrum shows a peak at each of the first 13 ordinates, with the amplitude predicted by the explicit formula. Under RH the ripple has size $e^{-t/4}$ on the grid, much larger than the cell $e^{-t/2}$, and RH is equivalent to this decay rate. The ripple is not centred: on $[10^3, 10^8]$ the sum of the primes stays below $\mathrm{li}(x^2)$ for about 97% of logarithmic time, and it first exceeds it at $x = 109$. We place this in a one-parameter family of prime races, $\sum_{p \le x} p^a$ against $\mathrm{li}(x^{a+1})$, and prove that under RH the normalized error has mean $-1/(2a+1)$, coming from the squares of the primes. Under RH and LI the limiting logarithmic density of the set where the race is ahead is given by a Rubinstein–Sarnak formula; it moves from $2.6 \cdot 10^{-7}$ (counting the primes, $a = 0$) to $0.06$ (summing them, $a = 1$) and tends to $1/2$ as $a \to \infty$, at the rate $\frac{1}{2} - \delta_a \sim 1/\big(2\sqrt{\pi A \log(A/2\pi)}\big)$ with $A = a + \frac{1}{2}$. At $a = -1/2$ the bias becomes the $\frac{1}{2}\log\log x$ of the weighted counting functions of Aoki and Koyama. All theorems are formalized in Lean 4.



**PART I — THE GRID**

1 Where the question comes from
2 The grid of prime sums
3 How dense the grid is
4 The zeros on the grid
5 Scaling laws
6 What the grid is good for
7 The question of the grid

**PART II — THE RIPPLE AND THE PRIME RACES**

8 Where the zeros are: the ripple of the grid
9 A one-parameter family of prime races
10 Numerical results for the family
11 Relation to the literature
12 Formalization in Lean

Reproducibility · References · Appendices A, B

PART I

# The grid

## 1 Where the question comes from

Sum the primes, $1 + 2 + 3 + 5 + 7 + \cdots$, and stop at the first prime at which the logarithm of the sum exceeds the first ordinate $\gamma_1 = 14.134725\ldots$ of a nontrivial zero of $\zeta$. This happens at $p = 4691$, where $\log S = 14.136170$: the difference is $1.4 \cdot 10^{-3}$. At the second zero the difference drops to $10^{-4}$, at the third to $10^{-5}$, at the fifth to $2.4 \cdot 10^{-7}$.

A sequence that approaches the zeros this fast looks like a message, and the previous version of this work read it as one, under the name of a "generative formula". Here we do the opposite, which is also the more useful thing: we determine exactly where the precision comes from, and only then ask what is left. The answer has two parts. The precision comes from the grid and not from the zeros (Part I). The zeros are nevertheless present in the prime sums, in a different place: the ripple of the sum around its mean curve, which carries one oscillation per zero and a bias whose size we can compute exactly (Part II).

## 2 The grid of prime sums

**Definition 1** (Cumulative sum and grid).For $N \geq 1$ let

$$S_N = 1 + \sum_{p \leq N} p,$$

and, with $p_m$ the $m$-th prime, let $\sigma_m = \log S_{p_m}$. The sequence $(\sigma_m)$ is strictly increasing: $\sigma_1 = \log 3$, $\sigma_2 = \log 6$, $\sigma_3 = \log 11, \ldots$

The initial $1$ is part of the definition and is used in the covering argument (Section 3).

**Definition 2** (First-crossing prime).For $t \geq \log 3$ let

$$N(t) = \min\{p \text{ prime} : \log S_p \geq t\}, \qquad \varepsilon(t) = \log S_{N(t)} - t \geq 0.$$

For $t = \gamma_n$ we write $N_n = N(\gamma_n)$ and $\varepsilon_n = \varepsilon(\gamma_n)$.

**Proposition 1** (The crossing is prime).*$N(t)$ is prime for every $t$.*

*Proof.* $S_N$ changes only when $N$ is prime, so the first $N$ at which $\log S_N$ exceeds $t$ is a prime. □

The proposition is true but follows from the definition, for every $t$ and not only for the zeros. (Version 1 estimated the probability that $N_1, N_2, N_3$ were prime "by chance"; the question does not arise.)

| $N$ | $S_N$ |
|---|---|
| 10 | 18 |
| 100 | 1 061 |
| 1 000 | 76 128 |
| 10 000 | 5 736 397 |

## 3 How dense the grid is

Everything depends on the step

$$\Delta_m = \sigma_{m+1} - \sigma_m = \log\Big(1 + \frac{p_{m+1}}{S_{p_m}}\Big).$$

For a height $t > \log 3$ let $m(t)$ be the index of the first-crossing prime, so that $\sigma_{m(t)-1} < t \le \sigma_{m(t)}$, and let

$$\delta(t) = \sigma_{m(t)} - \sigma_{m(t)-1} = \Delta_{m(t)-1}$$

be the width of the **cell** containing $t$. The index matters: the cell is the one *below* the arrival node, not the next step. For $t = \log 3 + 0.01$ one has $\varepsilon(t) = 0.683$, which exceeds $\Delta_2 = \log(11/6) = 0.606$ but stays below $\delta(t) = \log 2$.

**Theorem 1** (Step Theorem). *As $m \to \infty$, $\Delta_m \sim 2/m$. At height $t$ the cell has width*

$$\delta(t) = (1 + o(1))\,\sqrt{t}\,e^{-t/2}.$$

*Proof.* By the prime number theorem and partial summation, $\sum_{p \le x} p \sim x^2/(2\log x)$, and $p_{m+1} \sim p_m$. Hence $\Delta_m \sim p_m/S_{p_m} \sim 2\log p_m/p_m \sim 2/m$, since $m = \pi(p_m) \sim p_m/\log p_m$. At height $t$ we have $S \approx e^t$, so $p^2 \approx 2e^t \log p$ with $\log p = t/2 + O(\log t)$. It follows that $p = (1 + o(1))\sqrt{t}\,e^{t/2}$ and $\delta(t) = \Delta_{m(t)-1} \sim 2\log p/p = (1 + o(1))\sqrt{t}\,e^{-t/2}$, because consecutive steps are asymptotically equal. □

*Numbers.* Over all primes up to $10^8$ the ratio $\Delta_m/(\sqrt{\sigma_m}\,e^{-\sigma_m/2})$ lies between $1.038$ and $1.164$ for $t \ge 5$, and between $1.038$ and $1.072$ for $t \ge 14$. At the zeros the ratio $\delta(\gamma_n)/(\sqrt{\gamma_n}e^{-\gamma_n/2})$ decreases from $1.064$ ($n = 1$) to $1.022$ ($n = 20$), approaching $1$ slowly as expected from the $\log\log$ corrections (Figure 1).

**Theorem 2** (Covering Theorem). *For every $m \ge 1$, $S_{p_m} \ge p_{m+1}$; hence $\Delta_m \le \log 2$, with equality only for $m = 1$. Consequently $\varepsilon(\log 3) = 0$ and, for every $t > \log 3$,*

$$0 \le \varepsilon(t) < \delta(t) \le \log 2, \qquad \delta(t) = (1 + o(1))\sqrt{t}\,e^{-t/2}.$$

*Proof.* By induction. $S_2 = 3 = p_2$. If $S_{p_{m-1}} \ge p_m$, then $S_{p_m} = S_{p_{m-1}} + p_m \ge 2p_m > p_{m+1}$ by Bertrand's postulate. Hence $S_{p_{m+1}}/S_{p_m} = 1 + p_{m+1}/S_{p_m} \le 2$, with equality only in the passage $3 \to 6$. For $t > \log 3$ we have $\sigma_{m(t)-1} < t \le \sigma_{m(t)}$, so $\varepsilon(t) = \sigma_{m(t)} - t < \sigma_{m(t)} - \sigma_{m(t)-1} = \delta(t)$. The asymptotics is Theorem 1. □

*Remark (Brown's criterion).* Brown's criterion (1961) for an increasing sequence of positive integers with $a_1 = 1$ asks $a_{k+1} \le 1 + \sum_{j \le k} a_j$; for the sequence $1, 2, 3, 5, \ldots$ it reads $p_{m+1} \le S_{p_m} + 1$. The condi-

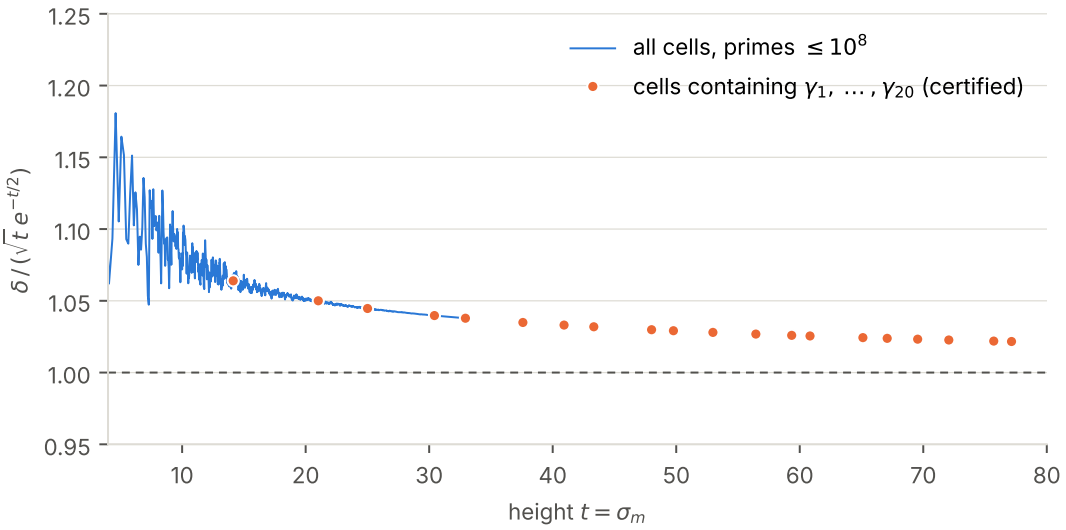



**Figure 1.** Step Theorem. Ratio of the cell width to $\sqrt{t}\,e^{-t/2}$: all cells for primes up to $10^8$ (line) and the certified cells containing $\gamma_1, \ldots, \gamma_{20}$ (dots). The zeros lie on the curve of all cells.

tion $S_{p_m} \geq p_{m+1}$ of Theorem 2 is stronger and therefore implies the criterion: every integer between 1 and $S_{p_m}$ is a sum of distinct elements of $\{1, 2, 3, 5, \ldots, p_m\}$. The grid has no gaps because the primes, together with 1, fill the integers by sums.

*Check.* Up to $10^8$ the largest step is $\log 2$, attained once, at the first step.

## 4 The zeros on the grid: what the precision actually says

Table 1 lists the certified first-crossing primes for the first twenty zeros (Section 7 describes the certification). The errors $\varepsilon_n$ follow the cells, with oscillations due to the position of the zero inside its cell.

> **Corollary 1** (Genericity Principle).*The Covering Theorem holds for every real $t$, not only for the zeros. The exponential decrease of $\varepsilon_n$ does not distinguish the zeros from arbitrary heights: it is the resolution of the grid.*

This moves the question. That the zeros are close to the grid is guaranteed. The right question is whether they are **closer than chance**, and this is measured by the reduced residue $U = \varepsilon/\delta$.

*The reduced residue.* Inside a cell, $U = (\sigma_{m(t)} - t)/\delta$ is a linear function of $t$. If $t$ is uniform on a whole cell, $U$ is therefore exactly uniform on $[0, 1)$, with mean $1/2$. If $t$ is uniform on a longer interval, the same holds for every interior cell, and only the two cells cut by the endpoints introduce a deviation, negligible when the interval contains many cells. On $2 \cdot 10^5$ uniform heights in $[14, 33]$ the measured mean is $0.500$ ($0.501$ on $[14, 26]$). On the first five zeros the mean is $0.509$: no sign of a privileged closeness, but five values cannot exclude it. Section 7 takes up the question with fifteen more zeros.

*A bit count (interpretive remark, not a theorem).* Writing $N_n$ takes $\log_2 N_n \approx \gamma_n/(2 \ln 2)$ bits. The precision *guaranteed* by Theorem 2, $\varepsilon_n < \delta(\gamma_n) \approx \sqrt{\gamma_n}\, e^{-\gamma_n/2}$, amounts to about $\gamma_n/(2 \ln 2)$ bits of $\gamma_n$ (a single residue can be smaller than the cell, but the theorem promises no more). Bits in and bits out

| $n$ | $\gamma_n$ | $N_n$ | $\varepsilon_n$ | $\delta$ | $U_n$ | $D_n$ | $D_n$ (integers) |
|---|---|---|---|---|---|---|---|
| 1 | 14.134725 | 4 691 | $1.44{\cdot}10^{-3}$ | $3.41{\cdot}10^{-3}$ | 0.424 | **0.424** | 0.444 |
| 2 | 21.022040 | 176 777 | $9.86{\cdot}10^{-5}$ | $1.31{\cdot}10^{-4}$ | 0.752 | **0.248** | 0.461 |
| 3 | 25.010858 | 1 409 533 | $8.76{\cdot}10^{-6}$ | $1.94{\cdot}10^{-5}$ | 0.452 | **0.452** | 0.483 |
| 4 | 30.424876 | 23 185 159 | $4.89{\cdot}10^{-7}$ | $1.42{\cdot}10^{-6}$ | 0.345 | **0.345** | 0.177 |
| 5 | 32.935062 | 84 474 667 | $2.40{\cdot}10^{-7}$ | $4.20{\cdot}10^{-7}$ | 0.573 | **0.427** | 0.484 |
| 6 | 37.586178 | 920 765 957 | $4.11{\cdot}10^{-8}$ | $4.37{\cdot}10^{-8}$ | 0.940 | **0.060** | 0.440 |
| 7 | 40.918719 | 5 075 610 989 | $6.26{\cdot}10^{-9}$ | $8.60{\cdot}10^{-9}$ | 0.728 | **0.272** | 0.375 |
| 8 | 43.327073 | 17 393 031 977 | $2.18{\cdot}10^{-9}$ | $2.65{\cdot}10^{-9}$ | 0.823 | **0.177** | 0.196 |
| 9 | 48.005151 | 189 499 828 967 | $7.05{\cdot}10^{-11}$ | $2.69{\cdot}10^{-10}$ | 0.263 | **0.263** | 0.120 |
| 10 | 49.773832 | 466 911 439 271 | $7.81{\cdot}10^{-11}$ | $1.13{\cdot}10^{-10}$ | 0.692 | **0.308** | 0.489 |
| 11 | 52.970321 | 2 378 827 903 247 | $2.02{\cdot}10^{-11}$ | $2.35{\cdot}10^{-11}$ | 0.858 | **0.142** | 0.091 |
| 12 | 56.446248 | 13 946 348 081 951 | $2.36{\cdot}10^{-12}$ | $4.27{\cdot}10^{-12}$ | 0.552 | **0.448** | 0.229 |
| 13 | 59.347044 | 60 934 830 657 541 | $7.47{\cdot}10^{-13}$ | $1.03{\cdot}10^{-12}$ | 0.729 | **0.271** | 0.200 |
| 14 | 60.831779 | 129 555 248 013 179 | $3.99{\cdot}10^{-13}$ | $4.94{\cdot}10^{-13}$ | 0.808 | **0.192** | 0.077 |
| 15 | 65.112544 | 1 138 383 441 262 829 | $1.32{\cdot}10^{-14}$ | $6.00{\cdot}10^{-14}$ | 0.220 | **0.220** | 0.288 |
| 16 | 67.079811 | 3 088 372 227 674 287 | $1.40{\cdot}10^{-15}$ | $2.28{\cdot}10^{-14}$ | 0.061 | **0.061** | 0.393 |
| 17 | 69.546402 | 10 787 862 040 659 943 | $3.43{\cdot}10^{-15}$ | $6.75{\cdot}10^{-15}$ | 0.508 | **0.492** | 0.044 |
| 18 | 72.067158 | 38 708 026 184 211 401 | $2.57{\cdot}10^{-16}$ | $1.95{\cdot}10^{-15}$ | 0.132 | **0.132** | 0.010 |
| 19 | 75.704691 | 244 370 311 869 517 289 | $1.46{\cdot}10^{-16}$ | $3.24{\cdot}10^{-16}$ | 0.452 | **0.452** | 0.489 |
| 20 | 77.144840 | 506 690 435 563 768 739 | $3.17{\cdot}10^{-17}$ | $1.59{\cdot}10^{-16}$ | 0.200 | **0.200** | 0.396 |

**Table 1.** First-crossing primes $N_n$ for the first twenty zeros. $\varepsilon_n = \log S_{N_n} - \gamma_n$, $\delta = \delta(\gamma_n)$, $U_n = \varepsilon_n/\delta$, $D_n = \min(U_n, 1-U_n)$. Last column: $D_n$ for the integer grid $\tau_N = \log(N(N+1)/2)$. Shaded rows 1–5 are exploratory; rows 6–20 form the pre-registered test.

coincide: $N_n$ *encodes* $\gamma_n$ in a numeral system made of prime sums; it does not predict it. A formula that generates information uses fewer parameters than the values it reproduces, as Balmer's formula has one constant for all lines. Here there is one free integer per zero.

*The same with the integers.* For the grid $\tau_N = \log(N(N+1)/2)$ the step is exactly $\tau_{N+1} - \tau_N = \log(1 + 2/N)$, with asymptotic equivalent $\sqrt{2}\, e^{-\tau_N/2}$. It is finer than the prime grid by a factor $\approx \sqrt{t/2}$:

| $\gamma_n$ | step, primes | step, integers | $U$, integers |
|---|---|---|---|
| 14.13 | $3.41 \cdot 10^{-3}$ | $1.21 \cdot 10^{-3}$ | 0.56 |
| 21.02 | $1.31 \cdot 10^{-4}$ | $3.85 \cdot 10^{-5}$ | 0.54 |
| 25.01 | $1.94 \cdot 10^{-5}$ | $5.24 \cdot 10^{-6}$ | 0.52 |
| 30.42 | $1.42 \cdot 10^{-6}$ | $3.50 \cdot 10^{-7}$ | 0.18 |
| 32.94 | $4.20 \cdot 10^{-7}$ | $9.97 \cdot 10^{-8}$ | 0.48 |

For metric closeness alone the integers do better than the primes. The prime grid is not metrically privileged.

*Other weights.* The same construction with $S_N^{(a)} = 1 + \sum_{p \le N} p^a$ gives a grid whose cell at height $t$ is, by the argument of Theorem 1, of order $t^{a/(a+1)} e^{-t/(a+1)}$: a larger weight gives a coarser grid. For the squares of the primes ($a = 2$) the cell at $\gamma_1$ is $0.40\%$ of $\gamma_1$ (first-crossing prime 281), against $0.024\%$ for $a = 1$; at $\gamma_2$ it is $0.036\%$ against $0.001\%$. For approximating ordinates with prime sums the plain sum is therefore preferable to any heavier weight. (Heavier weights are the interesting ones for the races of Part II.)

## 5 Scaling laws

**Proposition 2** (Scaling law).*Along the first-crossing primes,*

$$t = \log S_{N(t)} - \varepsilon(t) = 2 \log N - \log \log N - \log 2 + o(1).$$

The leading term gives $t \approx 2 \log N$; the factor 2 comes from the quadratic growth of prime sums. At the first five zeros the right-hand side equals $14.08$, $20.98$, $24.97$, $30.39$, $32.91$, against $14.13$, $21.02$, $25.01$, $30.42$, $32.94$. The ratio $\gamma / \log N$ ($1.67$, $1.74$, $1.77$, $1.79$, $1.80$) tends to 2 only slowly because of the $-\log \log N$ term. The regression $\gamma = 1.919 \log N - 2.13$ ($R^2 = 0.99998$) absorbs that term into slope and intercept.

**Corollary 2** (Half-exponent law, corrected form).*By Theorem 1,* $N(t) = (1 + o(1)) \sqrt{t}\, e^{t/2}$*, hence*

$$\frac{N(t')}{N(t)} = e^{(t'-t)/2} \sqrt{t'/t}\, (1 + o(1)),$$

*where* $o(1) \to 0$ *as* $t \to \infty$*, uniformly in* $t' \ge t$*.*

Version 2 omitted the factor $\sqrt{t'/t}$. This, and not a "slow convergence", explains the discrepancies:

| pair | $e^{\Delta\gamma/2}$ | with $\sqrt{\gamma'/\gamma}$ | observed | agreement without → with |
|---|---|---|---|---|
| 1→2 | 31.30 | 38.17 | 37.68 | 83.1% → **98.7%** |
| 2→3 | 7.35 | 8.01 | 7.97 | 92.2% → **99.5%** |
| 3→4 | 14.98 | 16.53 | 16.45 | 91.1% → **99.5%** |
| 4→5 | 3.51 | 3.65 | 3.64 | 96.3% → **99.8%** |

This law, too, holds for any pair of heights $t < t'$.

## 6 What the grid is good for

Without the claim of predicting the zeros, the grid has two proved properties and one possible use.

- **Guaranteed approximation with explicit step.** By Theorems 1–2, every height $t > \log 3$ has a node above it at distance less than $\delta(t) \le \log 2$, and $\delta(t)$ has known asymptotics. Between two nodes there are of course intervals without nodes; what the theorem guarantees is that none is wider than the cell.
- **Exact sums.** The sums $S_{p_m}$ are integers and are computed without error (64-bit integers suffice up to the eighth zero, 128-bit integers beyond the twentieth). Logarithms and residues require controlled precision or interval arithmetic.
- **Seeds for zero finding.** Every zero has a node within one cell, and at large height the cell, $\approx \sqrt{t}\, e^{-t/2}$, is much smaller than the mean spacing $2\pi/\log(t/2\pi)$ of the ordinates. (For individual gaps we have no lower bound, so the comparison holds on average and in the verified cases.) The nodes can therefore serve as starting points for Newton's method or for a Turing-type count of sign changes, provided a function that sees the zeros, such as Hardy's $Z(t)$ or $\eta(\frac{1}{2}+it)$, guides them. A practical advantage over a uniform grid of comparable step is not established: so fine a grid can be expensive, and an explicit cost comparison would be needed.

**From the grid to the zero: an example.** "The grid does not compute the zeros" should be made precise. The grid does not contain the zeros, but it supplies starting points for an equation to be solved numerically, just as $\sqrt{2}$ is found by solving $x^2 - 2 = 0$ by successive approximations. The use of the nodes as Newton seeds, and the multiplicity formula that follows from it, are developed in the companion work *Critical Primes, Phase Jumps, and Simple Zeros* [13]; here we only illustrate it.

The equation is that of the zeros on the critical line. We use Hardy's function $Z(t) = e^{i\theta(t)}\zeta(\frac{1}{2}+it)$, real for real $t$, whose real zeros are exactly the ordinates of the zeros of $\zeta$ on the line (equivalently one can use $\eta(\frac{1}{2}+it)$, which has the same zeros in the strip). $Z$ is evaluated by a finite procedure with controlled error (Riemann–Siegel, or an accelerated series for $\eta$). Newton's method updates the candidate by

$$t_{j+1} = t_j - \frac{Z(t_j)}{Z'(t_j)},$$

and near a simple zero it converges quadratically: each error is of the order of the square of the previous one. Error $|t_j - \gamma_1|$ at 40 digits (Figure 2):

| starting point $t_0$ | $\lvert t_0 - \gamma_1\rvert$ | after 1 | after 2 | after 3 |
|---|---|---|---|---|
| prime-grid node, $\log 1\,378\,036 = 14.13617$ | $1.4 \cdot 10^{-3}$ | $1.7 \cdot 10^{-7}$ | $2.4 \cdot 10^{-15}$ | $4.8 \cdot 10^{-31}$ |
| integer-grid node, $\log(1659 \cdot 1660/2) = 14.13540$ | $6.7 \cdot 10^{-4}$ | $3.7 \cdot 10^{-8}$ | $1.1 \cdot 10^{-16}$ | $1.0 \cdot 10^{-33}$ |
| an arbitrary height, $14.2$ | $6.5 \cdot 10^{-2}$ | $3.3 \cdot 10^{-4}$ | $9.1 \cdot 10^{-9}$ | $6.7 \cdot 10^{-18}$ |

The two roles are distinct: the grid proposes heights to examine; $Z$ (or $\eta$) recognizes and refines the candidates. The initial node and the zero found are different numbers, and every sufficiently close starting point converges: from $14.2$ one more step suffices. A tiny residual is a numerical verification; certifying a zero and its uniqueness in an interval requires rigorous error bounds (interval arithmetic, Turing's method). A specific contribution of the prime grid would be to show that its starting points make the search more efficient or more controllable than a control grid: this is a case of the question of Section 7.

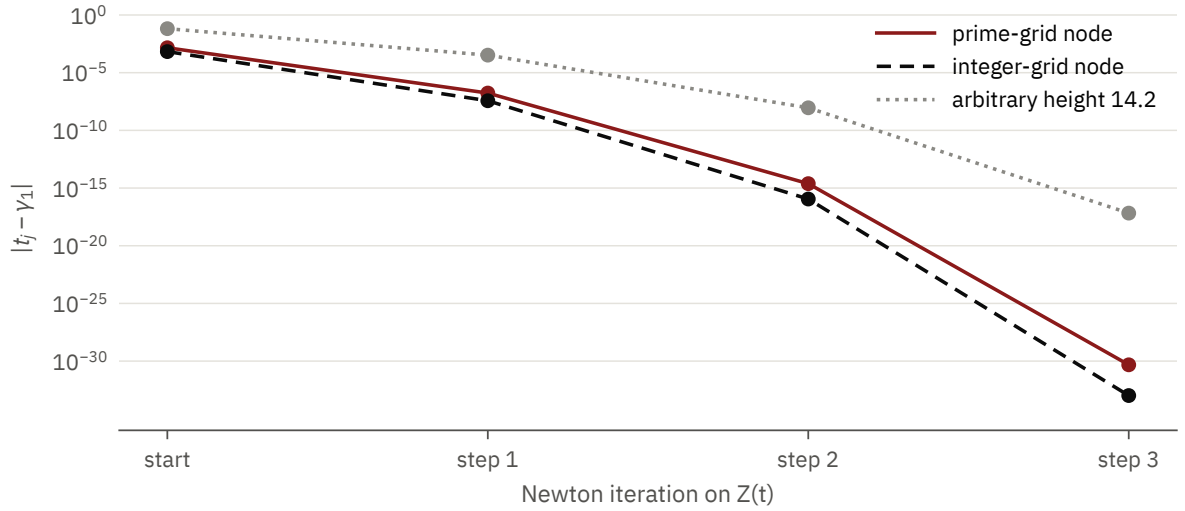



**Figure 2.** Newton iteration on Hardy's $Z(t)$ towards $\gamma_1$, from the three starting points of the table above. Each step roughly doubles the number of correct digits whatever the seed: the grid proposes, $Z$ refines.

*Remark (Zeros off the line).* The grid uses only the ordinate $t$. It would assign the same node to a hypothetical zero $\beta + i\gamma$ with $\beta \neq \frac{1}{2}$ and to a zero on the line with the same ordinate. It does not distinguish the two situations, and therefore the position of the nodes says nothing about the Riemann Hypothesis. (The ripple of Part II does: see Proposition 3.)

*Remark (Diagonal operator).* The operator $\hat{H} = \sum_m \sigma_m |m\rangle\langle m|$, with natural domain $\{a \in \ell^2 : \sum \sigma_m^2 |a_m|^2 < \infty\}$, is self-adjoint with discrete and simple spectrum: since $\sigma_m \to \infty$, the spectrum has no finite accumulation point. By Theorem 2, however, the distance of a height $t$ from the spectrum tends to 0 as $t \to \infty$. Every large height has a very close eigenvalue, and the closeness of a zero to the spectrum selects nothing. A Hilbert–Pólya candidate would need a selection rule coming from $\zeta$, not from the grid.

## 7 The question of the grid

By the Genericity Principle the prime grid is **canonical**: it is made of the same atoms as $\zeta$, it covers by a theorem, and its step is explicit. It is not shown to be **privileged**: in the cases examined it behaves like an arbitrary height or an integer grid. What remains is whether its arithmetic origin leaves a trace in some observable.

> **Question** (the grid question).*Is there a statistic $\Phi$, built from the primes only, such that $\Phi$ evaluated on the nodes of the prime grid separates the zeros of $\zeta$ from generic heights, while the same $\Phi$ on a control grid with the same density and the same covering (the integers, or a Cramér model satisfying Brown's criterion) does not?*

Two concrete forms.

### *7.1 Metric form*

Are the reduced residues $U_n = \varepsilon_n/\delta(\gamma_n)$ equidistributed in $[0,1)$? A mean of $U$ close to $1/2$ can hide a symmetric attraction towards both ends of the cell, so we use the distance to the nearest node in cell units,

$$D_n = \min(U_n, 1 - U_n),$$

uniform on $[0, \frac{1}{2})$ under the null model, with mean $1/4$. Small values of $D$ would measure attraction towards the nodes.

*Exploratory data.* On the first five zeros $D = 0.42, 0.25, 0.45, 0.35, 0.43$, with mean $0.38$: if anything the zeros sit **farther** from the nodes than chance (probability under the null of a mean $\geq 0.38$: about $0.02$). The integer grid gives $0.44, 0.46, 0.48, 0.18, 0.48$, mean $0.41$ (about $0.005$). The effect appears on two grids with different nodes, so it is not specific to the primes. With five points, a statistic chosen after seeing the data and two statistics examined, it has no evidential value. This is why the protocol below was fixed before computing further terms.

*Protocol (fixed in writing before the computation, 9 October 2026).* Zeros 1–5 are exploratory and do not enter the test; test data are $n = 6, \ldots, 20$. The primary statistic is the mean of $D_n$, against $D_n$ i.i.d. uniform on $[0, \frac{1}{2})$, two-sided at level $0.05$ with the $p$-value by simulation ($10^6$ samples, fixed seed); the secondary statistic is Kolmogorov–Smirnov on $U_n$. The same statistics are computed in parallel on the integer grid $\tau_N = \log(N(N+1)/2)$. The declared expectation was no significant outcome. The full protocol, the certification pipeline and the certified records are given in the Supplementary Material.

*Computing the terms.* A sieve reaches only the fifth zero, because it needs all primes up to $x \approx \sqrt{\gamma}\, e^{\gamma/2}$. The sum of the primes up to $x$ can, however, be computed without a sieve, by combinatorial algorithms in sublinear time (Legendre-type in about $x^{3/4}$ operations and $\sqrt{x}$ memory; Meissel–Lehmer variants in about $x^{2/3}$), and then completed with a segmented sieve in a short interval around $N_n$. The program *primesum* by K. Walisch [19] computes the sum of the primes up to $x \leq 10^{31}$ with 128-bit integers, in seconds up to $10^{18}$ on a recent desktop processor.

*Outcome (9 October 2026).* On the fifteen test zeros the mean of $D$ is **0.246** against 0.250 ($p = 0.92$, two-sided), and Kolmogorov–Smirnov on $U$ gives $p = 0.79$. The integer grid gives $0.256$ ($p = 0.88$) and $p = 0.96$. The weak "repulsion" of the first five zeros (mean 0.38) does not recur (Figures 3 and 4).

*What this outcome says, and what it does not.*

- **It says** that on the fifteen test zeros there is no significant departure of the mean of $D$ from $1/4$, nor of the distribution of $U$ from the uniform law. The 95% confidence interval for the mean of $D$ is $[0.17, 0.32]$: a strong attraction, bringing the mean below $0.17$, is excluded at this level.
- **It does not say** that the zeros sit in their cells "like an arbitrary number". A mean of $D$ near $1/4$ is compatible with very regular structures: for instance all $D$ equal, or points systematically avoiding half of the cell (in that case $D$ has the same law as in the uniform case, because the folding $\min(U, 1-U)$ erases the difference). The individual data exclude these two extreme examples: the $D_n$ range from $0.06$ to $0.49$, and the $U_n$ fall in both halves of the cell (nine above $1/2$, six below). With fifteen points, however, Kolmogorov–Smirnov detects only large departures (threshold $0.34$ at the 5% level).

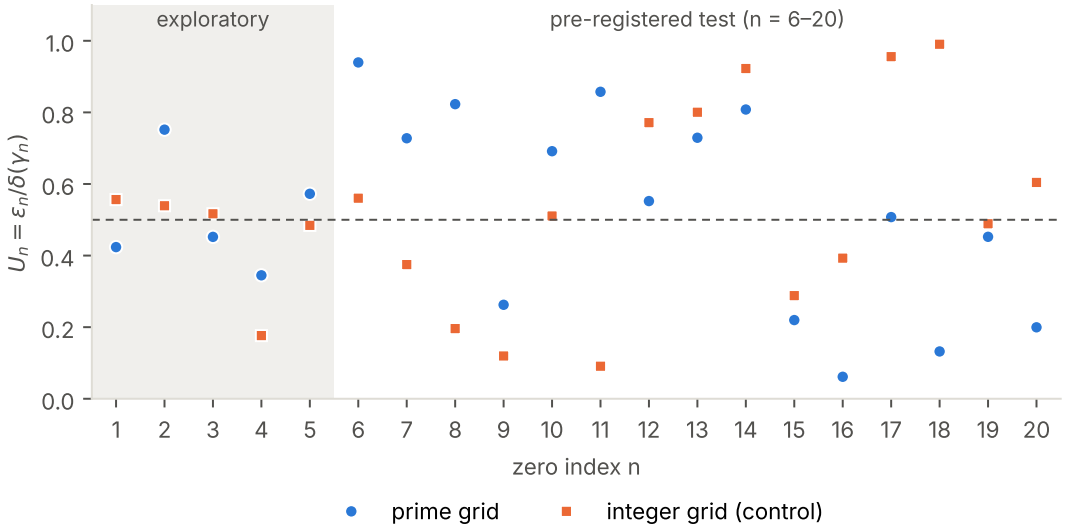



**Figure 3.** Reduced residues $U_n$ for the prime grid and for the integer grid. Shaded: the exploratory zeros $n = 1$–$5$. Under the null model, $U$ is uniform on $[0, 1)$.

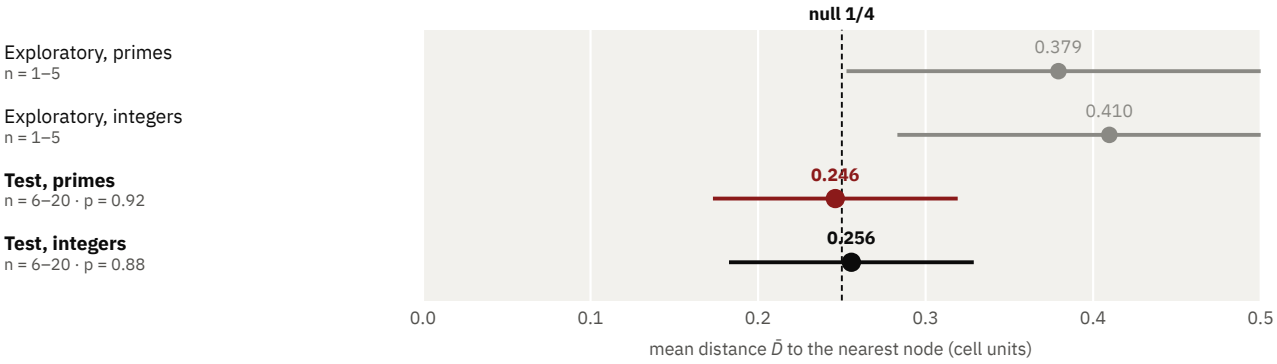



**Figure 4.** The pre-registered test at a glance. Mean distance $\bar{D}$ to the nearest node, with 95% intervals, for the exploratory zeros ($n = 1$–$5$) and the test zeros ($n = 6$–$20$), on the prime grid and on the integer control. The null value is $1/4$.

Moderate effects, regularities not captured by these two statistics and dependencies between terms remain open.

- **To close the question** one would need a relevance threshold fixed in advance (an equivalence test, not a difference test) and many more terms. Any further analysis chosen now would be exploratory.

### *7.2 Structural form*

The grid comes from the *addition* of the primes; $\zeta$ comes from their *multiplication*, the Euler product, and the zeros are seen by evaluating, on any dense grid, a function built multiplicatively from the same primes. Does choosing the additive grid of the same primes, instead of a control grid, change anything measurable beyond resolution?

For the metric form, up to the twentieth zero, the test detects no effect, which is not the same as excluding one. The structural form is open as well. Part II shows that the zeros *are* present in the prime sums, though not where Part I looked for them: not in the position of the nodes relative to the ordinates, but in

the deviation of the whole grid from its mean curve.

---

# PART II The ripple and the prime races

## 8 Where the zeros are: the ripple of the grid

### *8.1 The mean curve and the ripple*

By the prime number theorem, $\sum_{p\le x} p \sim \mathrm{li}(x^2)$, where $\mathrm{li}(y) = \mathrm{PV}\int_0^y du/\log u$. With the conventions of Definition 1 we compare

$$S(x) = 1 + \sum_{p\le x} p \qquad \text{with} \qquad B(x) = 1 + \mathrm{li}(x^2) - \mathrm{li}(4),$$

and call **ripple** the normalized deviation

$$f(u) = \big(S(x) - B(x)\big)\,\frac{\log x}{x^{3/2}}, \qquad u = \log x.$$

(The constants $1$ and $\mathrm{li}(4) = 2.97$ are irrelevant at the scale of $x^{3/2}/\log x$; in Sections 9–10 we drop them and compare $\sum_{p\le x} p$ with $\mathrm{li}(x^2)$ directly.)

The explicit formula of Riemann and von Mangoldt, transported to the weight $p$ (Section 9, Step 4 with $a = 1$), gives under RH

$$f(u) = -\frac{1}{3} - 2\,\mathrm{Re}\sum_{\gamma>0} \frac{e^{i\gamma u}}{\frac{3}{2} + i\gamma} + O\Big(\frac{1}{u}\Big).$$

Each zero contributes one pure oscillation in the variable $u = \log x$: frequency $\gamma$, amplitude $2/|\frac{3}{2} + i\gamma| \approx 2/\gamma$. The ripple is the sum of these oscillations around a constant level $-1/3$, which Section 9 explains.

### *8.2 The spectrum*

We sampled $f$ on $2^{16}$ points with $u \in [\log 10^3, \log 10^8]$, applied a Hann taper, and computed $|\hat f(\omega)|$ for $\omega \in [0.5, 60]$. As a control we did the same with random integers of density $1/\log n$ (a Cramér model), summed in the same way.

- Every one of the 13 ordinates below 60 appears as a peak (local maxima above 5 times the median), with no spurious peak. The offsets are at most $0.02$.
- The heights agree with the prediction $L/(2|\rho + 1|)$, $L = \log(10^8/10^3)$, of the explicit formula, within a few per cent.
- The control has no peaks at the ordinates; its spectrum is a broad floor about 20 times higher (median $0.196$ against $0.0097$).
- In the time domain, the explicit formula truncated at 100 zeros already has correlation $0.92$ with the data; with 20 000 zeros and the exact prime-power terms, $0.998$ (Section 10).

| zero $\gamma_n$ | peak $\omega$ | offset | measured height | predicted $L/(2\lvert\rho+1\rvert)$ |
|---|---|---|---|---|
| 14.135 | 14.14 | +0.00 | 0.4072 | 0.4050 |
| 21.022 | 21.03 | +0.01 | 0.2672 | 0.2731 |
| 25.011 | 25.01 | −0.00 | 0.2312 | 0.2297 |
| 30.425 | 30.41 | −0.01 | 0.1854 | 0.1890 |
| 32.935 | 32.94 | +0.01 | 0.1734 | 0.1746 |
| 37.586 | 37.58 | −0.00 | 0.1561 | 0.1530 |
| 40.919 | 40.92 | −0.00 | 0.1369 | 0.1406 |
| 43.327 | 43.32 | −0.01 | 0.1340 | 0.1328 |
| 48.005 | 48.01 | +0.00 | 0.1176 | 0.1199 |
| 49.774 | 49.77 | +0.00 | 0.1166 | 0.1156 |
| 52.970 | 52.97 | −0.00 | 0.1096 | 0.1086 |
| 56.446 | 56.46 | +0.01 | 0.1010 | 0.1019 |
| 59.347 | 59.33 | −0.02 | 0.0931 | 0.0970 |

**Table 2.** Peaks of the spectrum of the ripple against the zeros.

This is a verification of the explicit formula, not a new property of the zeros: any function of the primes with a known mean behaviour carries the same oscillations. Its interest here is to locate precisely what the grid does and does not see. The comparison with other functions of the primes on the same window and with the same procedure is in Table 3 and Figure 6.

Weighting by $p$ instead of counting changes little: the amplitudes $2/|\rho+a|$ are almost independent of $a$. The large gain comes from including the prime powers, that is, from removing the layer of the squares of the primes, which is not a flat level but a staircase with its own fluctuations. The same layer is the source of the bias studied in Section 9.

### *8.3 Two scales on the grid, and RH*

**Proposition 3** (Ripple and RH).
*(a) Under RH,* $S(x)-B(x)=O(x^{3/2}\log x)$*, hence* $|\log S(x)-\log B(x)|=O\big((\log x)^2/\sqrt{x}\big)$.
*(b) Conversely, if* $S(x)-B(x)=O(x^{3/2+\epsilon})$ *for every* $\epsilon>0$*, then RH holds.*

*Proof.* With $\vartheta(t)=\sum_{p\le t}\log p$ we have $\sum_{p\le x}p=\int_{2^-}^{x}\frac{t}{\log t}\,d\vartheta(t)$.

(a) Under RH, $\vartheta(t)=t+E(t)$ with $E(t)=O(\sqrt{t}\log^2 t)$ (von Koch, together with $\psi-\vartheta=O(\sqrt{t})$). The main term gives $\int_2^x t\,dt/\log t=\mathrm{li}(x^2)-\mathrm{li}(4)$. Integrating the error by parts, $\int\frac{t}{\log t}\,dE=\big[\frac{tE}{\log t}\big]-\int E\,d(\frac{t}{\log t})=O(x^{3/2}\log x)$. Dividing by $B(x)\sim x^2/(2\log x)$ gives the bound on the logarithms.

(b) Conversely, $\vartheta(x)=\int_{2^-}^{x}\frac{\log t}{t}\,d(S(t)-1)$. The main term $B$ contributes $\int\frac{\log t}{t}\cdot\frac{t}{\log t}\,dt=x+O(1)$; the remainder $F=S-B=O(t^{3/2+\epsilon})$ contributes $\big[\frac{\log t}{t}F\big]-\int F\,d\big(\frac{\log t}{t}\big)=O(x^{1/2+2\epsilon})$. So $\vartheta(x)=$

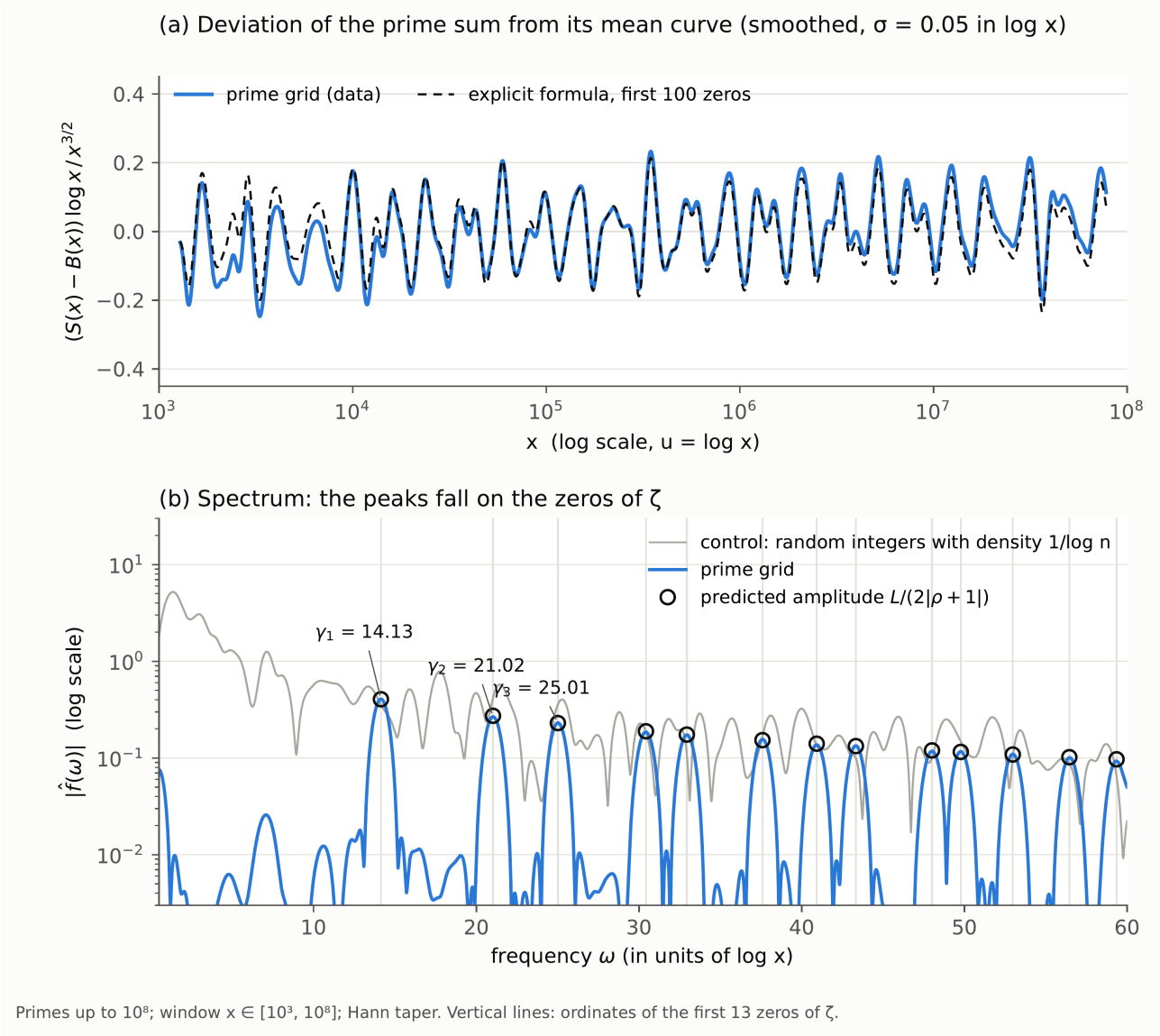



**Figure 5.** (a) Ripple of the prime grid (smoothed, $\sigma = 0.05$ in $\log x$) against the explicit formula with the first 100 zeros. (b) Spectrum of the ripple: the peaks fall on the ordinates of the zeros (vertical lines), with the heights predicted by the explicit formula (circles); the Cramér control (grey) shows no peaks there. Primes up to $10^8$, window $x \in [10^3, 10^8]$ — the same window used for the bias (Figure 8, Table 5); script `07_spectrum.py`.

$x + O(x^{1/2+2\epsilon})$ for every $\epsilon$, which is equivalent to RH. □

On the grid, height $t = \log S \approx 2 \log x$, so $x \approx e^{t/2}$ up to powers of $t$. The two scales are:

- **the cell**, $\delta(t) \asymp \sqrt{t}\, e^{-t/2}$: how finely the grid resolves heights (Part I);
- **the ripple**, $|\log S - \log B| \asymp e^{-t/4}$ up to powers of $t$: how far the grid deviates from its mean curve, and where the zeros are.

If the zeros had real parts up to $\Theta > \frac{1}{2}$, the ripple would be of order $e^{-(1-\Theta)t/2}$. RH is equivalent to the ripple decaying at rate $1/4$. The ripple is much larger than the cell, and this explains Part I: the position of a node with respect to $\gamma_n$ is a phenomenon at the scale of the cell, while the zeros act at the scale of the ripple.

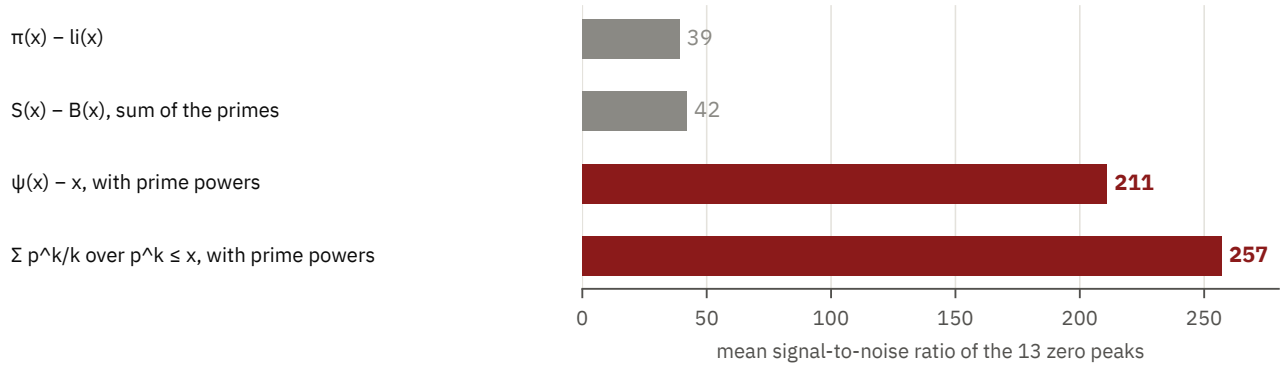



**Table 3** · **Figure 6.** Mean signal-to-noise ratio of the 13 zero peaks for four functions of the primes, same window and procedure. Weighting by $p$ changes little; including the prime powers gives the large gain.

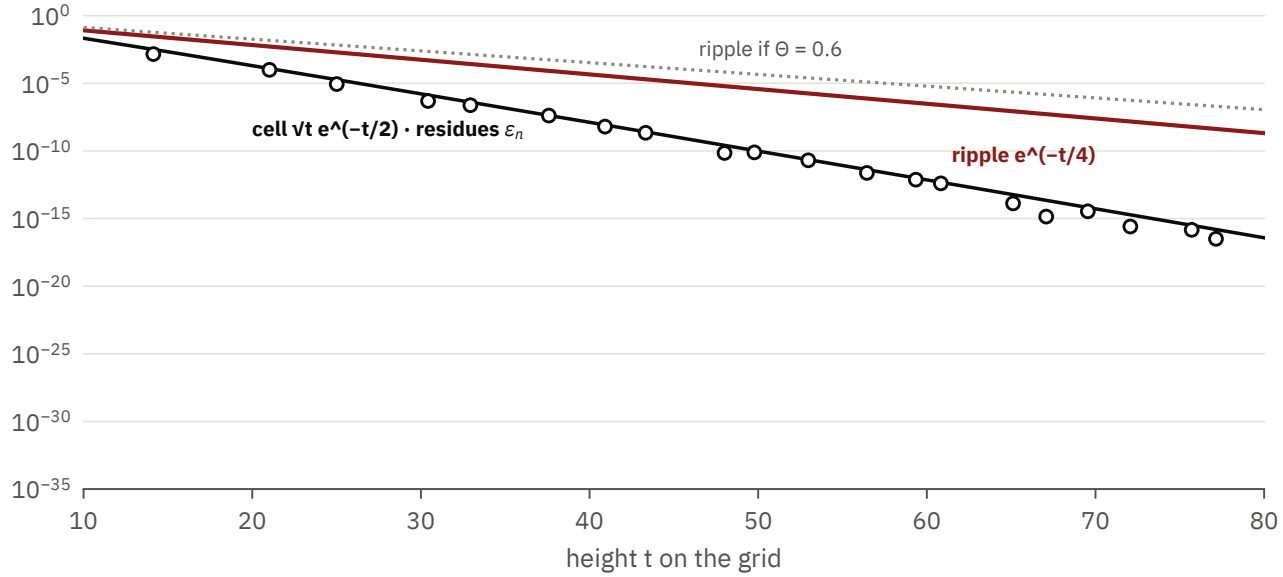



**Figure 7.** The two scales on the grid, on a logarithmic axis. The cell $\sqrt{t}\,e^{-t/2}$ (Part I) with the certified residues $\varepsilon_n$ at the first twenty zeros (dots, always below it), and the ripple scale $e^{-t/4}$ under RH (Part II). At $t = 77$ the ripple is about $10^8$ times larger than the cell; a zero off the line, with real part $\Theta > \frac{1}{2}$, would lift the ripple to $e^{-(1-\Theta)t/2}$ (dotted, $\Theta = 0.6$).

# 9 A one-parameter family of prime races

## *9.1 The bias of the prime sum*

The ripple of Section 8 is not centred at zero. On $[10^3, 10^8]$ the sum of the primes stays below its mean curve for about 97% of logarithmic time. On $x \geq 2$ it first exceeds $\mathrm{li}(x^2)$ at $x = 109$ (just above $x = 1$, $\mathrm{li}(x^2)$ is negative, so the comparison starts at $x = 2$):

| $x$ | $\sum_{p\le x} p$ | $\mathrm{li}(x^2)$ | difference |
|---|---|---|---|
| 103 | 1 264 | 1 312.0 | $-48.0$ |
| 107 | 1 371 | 1 402.3 | $-31.3$ |
| **109** | **1 480** | **1 448.4** | **+31.6** |
| 112 | 1 480 | 1 518.9 | $-38.9$ |
| 113 | 1 593 | 1 542.7 | $+50.3$ |
| 127 | 1 720 | 1 893.6 | $-173.6$ |

The same phenomenon for $\pi(x)$ against $\mathrm{li}(x)$ is Chebyshev's bias. There the first crossing is not known: it is below $1.4 \cdot 10^{316}$ and believed to lie near that value. For the sum of the primes the race is visible by hand. The reason is a single term of the explicit formula, and it is natural to study it for every weight.

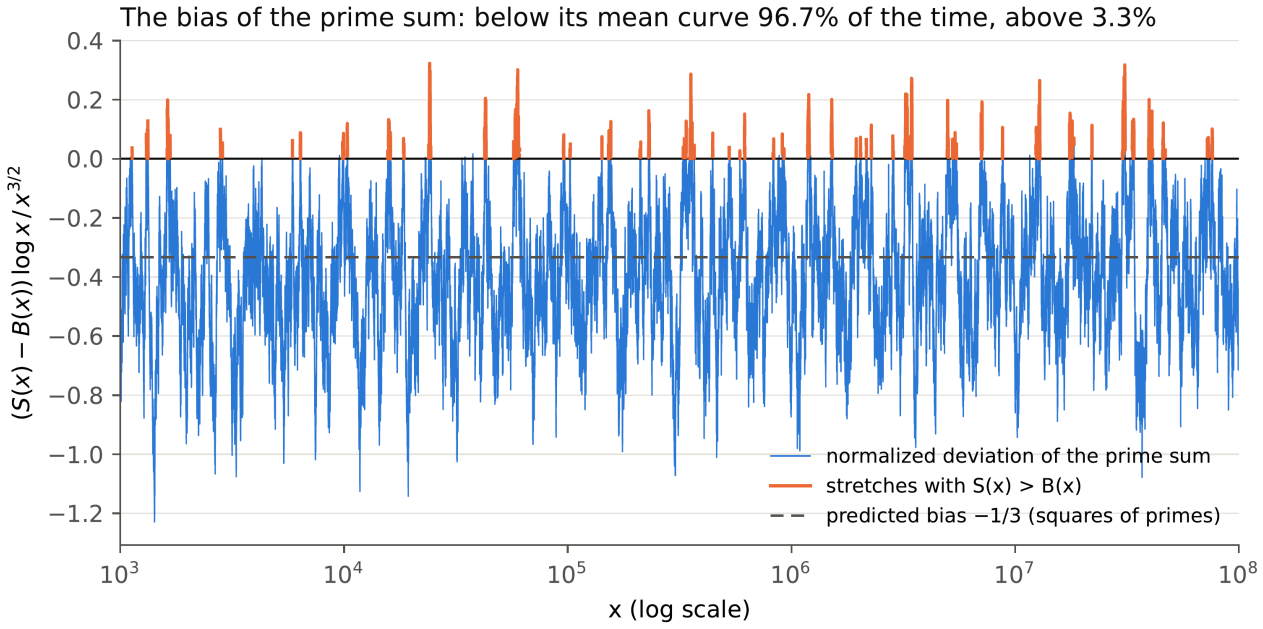



**Figure 8.** The normalized deviation of the prime sum on $[10^3, 10^8]$, its predicted bias $-1/3$ (dashed), and the stretches where $S(x) > B(x)$ (orange).

### *9.2 Notation*

Fix a real $a > -\frac{1}{2}$. For $x \ge 2$ put

$$\theta_a(x) = \sum_{p \le x} p^a, \qquad \Pi_a(x) = \sum_{p^k \le x} \frac{p^{ka}}{k} = \sum_{k \ge 1} \frac{1}{k}\, \theta_{ka}(x^{1/k}),$$

and

$$E_a(x) = \left(\theta_a(x) - \mathrm{li}(x^{a+1})\right) \frac{\log x}{x^{a+1/2}}, \qquad y = \log x.$$

The substitution $u = t^{a+1}$ gives $\mathrm{li}(x^{a+1}) = \mathrm{PV}\int_0^x t^a\, dt/\log t$, which is why $\mathrm{li}(x^{a+1})$ is the natural comparison function for the weight $p^a$.

- For $a = 0$, $E_0$ is the classical normalized error $(\pi(x) - \mathrm{li}(x))\log x/\sqrt{x}$ of Rubinstein and Sarnak.

- For $a = 1$, $E_1$ is the ripple of Section 8 (up to the irrelevant constants).
- For $a = -w$ with $0 \le w < \frac{1}{2}$, $\theta_a$ is the weighted counting function $\pi_w(x) = \sum_{p \le x} p^{-w}$ of Aoki–Koyama and Shimada–Koyama, compared with $\mathrm{li}(x^{1-w})$.

### *9.3 The bias law*

**Theorem 3** (Bias law).*Assume RH.*
*(i) $E_a(e^y)$ has a limiting logarithmic distribution $\mu_a$, and its mean is*

$$\int t \, d\mu_a(t) = -\frac{1}{2a+1}.$$

*(ii) Assume in addition LI, that is, that the positive ordinates $\gamma$ of the zeros of $\zeta$, counted with multiplicity, are linearly independent over $\mathbb{Q}$ (in particular these zeros are simple). Then $\mu_a$ is absolutely continuous and symmetric about $-1/(2a+1)$, and the logarithmic density of $\{x : \theta_a(x) > \mathrm{li}(x^{a+1})\}$ exists and equals*

$$\delta_a = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin(\xi/(2a+1))}{\xi} \prod_{\gamma > 0} J_0\Big(\frac{2\xi}{|\frac{1}{2} + a + i\gamma|}\Big) \, d\xi.$$

*For $a = 0$ this is the classical bias $-1$ and the density $2.6 \cdot 10^{-7}$ of Rubinstein and Sarnak.*

### *9.4 Proof*

**Step 1: the layer $k = 1$.** By Stieltjes integration against $t^a/\log t$,

$$\Pi_a(x) = 2^a + \int_{5/2}^{x} \frac{t^a}{\log t} \, d\psi(t), \qquad \psi(t) = \sum_{n \le t} \Lambda(n),$$

because the jump of $\psi$ at $p^k$ is $\log p$ and $(\log p)\, p^{ka}/\log(p^k) = p^{ka}/k$; the only prime power below $5/2$ is $2$, which contributes $2^a$. The lower limit $5/2$ lies between prime powers, which keeps the boundary terms away from the jumps, and it lies in the range where the explicit formula below holds.

We use the truncated explicit formula (Montgomery–Vaughan [22], Theorem 12.5, p. 400; see also Davenport [6], ch. 17): for a fixed $c > 1$, $t \ge c$ and $T \ge 2$,

$$\psi_0(t) = t - \sum_{|\gamma| \le T} \frac{t^\rho}{\rho} - \log 2\pi - \tfrac{1}{2}\log(1 - t^{-2}) + R_0(t,T), \qquad |R_0(t,T)| \ll_c \frac{t \log^2(tT)}{T} + \log t \cdot \min\Big(1, \frac{t}{T\langle t\rangle}\Big),$$

where $\psi_0$ is $\psi$ with the value at prime powers replaced by the mean of the two one-sided limits, $\langle t \rangle$ is the distance from $t$ to the nearest prime power other than $t$, and zeros are counted with multiplicity. We take $c = 2$. Define $R(t,T)$ by

$$\psi(t) = t - \sum_{|\gamma| \le T} \frac{t^\rho}{\rho} + R(t,T).$$

For $t \ge 5/2$ and $T \ge 2$ we have $|R(t,T)| \ll t\log^2(tT)/T + \log t$: the constant terms are bounded, $|\psi - \psi_0| \le \frac{1}{2}\log t$, and the minimum is at most $1$.

Put $\kappa = a + \frac{1}{2} > 0$ and, for $\operatorname{Re} s > 0$, $L(s;x) = \int_{5/2}^{x} t^{s-1}\,dt/\log t$. Defining it directly as an integral avoids any choice of branch of li. Then

$$\Pi_a(x) = 2^a + L(a+1;x) - \sum_{|\gamma|\le T} L(\rho + a; x) + \int_{5/2}^{x} \frac{t^a}{\log t}\, dR(t,T),$$

where the zero terms come from $d(t^\rho/\rho) = t^{\rho-1}dt$.

**Lemma A** (pointwise form of the layer $k = 1$)*.Assume RH. Let $Y \ge 2$, $T = e^Y$ and $3 \le x \le e^Y$. Then*

$$\Pi_a(x) = \operatorname{li}(x^{a+1}) - \frac{x^\kappa}{\log x} \sum_{|\gamma|\le T} \frac{x^{i\gamma}}{\rho + a} + O_a\Big(\frac{x^\kappa}{\log^2 x} + Y^2 e^{-Y/2}\frac{x^\kappa}{\log x} + 1\Big).$$

*Proof.*

1. *Main term.* $2^a + L(a+1;x) = \operatorname{li}(x^{a+1}) - \operatorname{li}((5/2)^{a+1}) + 2^a = \operatorname{li}(x^{a+1}) + O(1)$.

2. *Zero terms.* Put $s = \rho + a$, so $\operatorname{Re} s = \kappa$ under RH, and substitute $t = e^v$, $u = \log x$, $v_0 = \log(5/2)$. Two integrations by parts give

$$L(s;x) = \int_{v_0}^{u} \frac{e^{sv}}{v}\, dv = \Big[\frac{e^{sv}}{sv} + \frac{e^{sv}}{s^2v^2}\Big]_{v_0}^{u} + \frac{2}{s^2}\int_{v_0}^{u} \frac{e^{sv}}{v^3}\, dv.$$

   The last integral is $\ll_\kappa |s|^{-2}(1 + e^{\kappa u}/u^3)$. The upper boundary gives $x^s/(s\log x) + O(x^\kappa/(|s|^2\log^2 x))$; the lower boundary gives $(5/2)^s(1/(sv_0) + 1/(s^2v_0^2))$. Now sum over $|\gamma| \le T$. Every term with $|s|^{-2}$ is absolutely summable, because $\sum_\gamma |\rho + a|^{-2} < \infty$; together they give $O(x^\kappa/\log^2 x + 1)$ uniformly in $T$. The lower-boundary term with $1/s$ is $(5/2)^a v_0^{-1}\sum_{|\gamma|\le T}(5/2)^\rho/(\rho + a)$. Writing $1/(\rho + a) = 1/\rho - a/(\rho(\rho + a))$, the second piece is absolutely summable, and the first is $\sum_{|\gamma|\le T}(5/2)^\rho/\rho = 5/2 - \psi(5/2) + R(5/2,T) = O(1)$ uniformly in $T$, by the truncated formula at $t = 5/2$ (here $\psi(5/2) = \log 2$ and $\langle 5/2\rangle = 1/2$). What remains is the main zero term $\sum_{|\gamma|\le T} x^{\rho+a}/((\rho + a)\log x) = (x^\kappa/\log x)\sum_{|\gamma|\le T} x^{i\gamma}/(\rho + a)$.

3. *The remainder $R$.* Integrate by parts:

$$\int_{5/2}^{x} \frac{t^a}{\log t}\, dR = \Big[\frac{t^a R}{\log t}\Big]_{5/2}^{x} - \int_{5/2}^{x} R(t,T)\Big(\frac{t^a}{\log t}\Big)'\, dt,$$

   with $|(t^a/\log t)'| \ll_a t^{a-1}/\log t$ for $t \ge 5/2$. Using the bound on $R$ and $x \le T$: the boundary term at $x$ is $\ll \frac{x^{a+1}\log^2(xT)}{T\log x} + x^a \le x^\kappa \cdot \frac{x^{1/2}(2Y)^2}{T\log x} + x^a \ll Y^2 e^{-Y/2}\frac{x^\kappa}{\log x} + x^a$; the integral of the first part of $R$ is $\ll \frac{\log^2(xT)}{T}\int t^a\, dt/\log t$, which is the same quantity; the integral of the $\log t$ part is $\ll \int_{5/2}^{x} t^{a-1}dt \ll 1 + x^a + \log x$. Finally $x^a + \log x = O(x^\kappa/\log^2 x)$, because $\kappa - a = \frac{1}{2}$. □

**Step 2: the layer $k = 2$, the source of the bias.** The term $k = 2$ of $\Pi_a$ is

$$\tfrac{1}{2}\theta_{2a}(\sqrt{x}) = \tfrac{1}{2}\int_2^{\sqrt{x}} \frac{t^{2a}}{\log t}\, dt\,\Big(1 + O\big(e^{-c\sqrt{\log x}}\big)\Big) = \tfrac{1}{2}\big(\operatorname{li}(x^{a+1/2}) - \operatorname{li}(2^{2a+1})\big)\Big(1 + O\big(e^{-c\sqrt{\log x}}\big)\Big),$$

by partial summation from the prime number theorem with the error term of de la Vallée Poussin, $\pi(y) = \operatorname{li}(y) + O(y\,e^{-c\sqrt{\log y}})$ (Davenport [6], ch. 18), valid for the weight $t^{2a}$ because $2a > -1$. (Under RH, which Theorem 3 assumes, von Koch's bound gives the same with a better error; the plain asymptotic

$\pi(y) \sim y/\log y$ would give only $1 + o(1)$.) Since $\mathrm{li}(x^{a+1/2}) = \frac{x^{a+1/2}}{(a+\frac12)\log x}(1 + O(1/\log x))$,

$$\boxed{\tfrac12\theta_{2a}(\sqrt{x}) = \frac{x^{a+1/2}}{(2a+1)\log x}\Big(1 + O(\tfrac{1}{\log x})\Big).}$$

Two factors of 2 cancel here: the layer weight $1/k = 1/2$, and $\log\sqrt{x} = \frac12\log x$. What is left is the $1/(2a+1)$ coming from the exponent of $(\sqrt{x})^{2a+1}$. In Shimada–Koyama (their Lemma 2.3, notation $w = -a$) this same prime-square layer appears as an error of order $x^{1/2-w}\log x/(1-2w)$, bounded in absolute value. Here it is kept with its sign and its constant, and it is the bias.

**Step 3: the layers $k \geq 3$.** For $a \geq 0$, $\theta_{ka}(x^{1/k}) \leq x^a \cdot x^{1/k}$, and $k$ runs only up to $\log x/\log 2$; hence $\sum_{k\geq 3}\frac1k\theta_{ka}(x^{1/k}) \ll x^{a+1/3}\log x$. For $-\frac12 < a < 0$ put $\delta = a + \frac12 \in (0, \frac12)$ and $b_k = \min(k|a|, 1-\delta)$. Since $p^{ka} \leq p^{-b_k}$, $\theta_{ka}(y) \leq \sum_{n\leq y} n^{-b_k} \leq y^{1-b_k}/\delta$, so the $k$-th term is at most $x^{(1-b_k)/k}/\delta$. If $b_k = k|a|$ the exponent is $1/k - |a| \leq a + \frac13$; otherwise it is $\delta/k \leq \delta/3$. Hence the total is $\ll_a x^{\max(a+1/3,\,\delta/3)}\log x$, with no transition cases. Since $a + \frac13 < a + \frac12$ and $\delta/3 < a + \frac12$, in both ranges of $a$ the total is $O(x^{a+1/2-\eta}\log x)$ for some $\eta = \eta(a) > 0$, in particular $o(x^{a+1/2}/\log x)$.

**Step 4: assembling.** We have $\theta_a = \Pi_a - \frac12\theta_{2a}(\sqrt{x}) - \sum_{k\geq 3}(\cdots)$. Normalizing by $\log x/x^{a+1/2}$ and pairing conjugate zeros, with $y = \log x$:

$$E_a(e^y) = -\frac{1}{2a+1} - 2\,\mathrm{Re}\sum_{0<\gamma\leq T}\frac{e^{i\gamma y}}{\frac12 + a + i\gamma} + \varepsilon_a(y, T),$$

where, for $3 \leq e^y \leq e^Y$ and $T = e^Y$, the error satisfies the **pointwise** bound

$$|\varepsilon_a(y, T)| \ll_a \frac{1}{y} + Y^2 e^{-Y/2}.$$

It collects Lemma A multiplied by $\log x/x^\kappa$, which gives $O(1/y + Y^2e^{-Y/2} + ye^{-\kappa y})$; the relative error $O(1/\log x)$ of Step 2; and Step 3, of order $O(x^{a+1/2-\eta}\log x)$, which after normalization is $O(1/y)$. Consequently

$$\frac{1}{Y}\int_{y_0}^{Y}|\varepsilon_a(y, e^Y)|^2\,dy \ll \frac{1}{Y}\int_{y_0}^{Y}\frac{dy}{y^2} + Y^4e^{-Y} \ll \frac{1}{Y} \to 0.$$

No mean-value estimate for the tail $\sum_{|\gamma|>T}$ is needed: taking $T = e^Y \geq x$, which is exactly the truncation $X = e^Y$ in the condition of Akbary–Ng–Shahabi, makes the truncation error pointwise small. This replaces Lemma 2.1 of Rubinstein–Sarnak. The layer $k = 2$ is the only new deterministic term, and it was computed in Step 2.

**Step 5: limiting distribution and mean.** We apply Akbary–Ng–Shahabi [1], Theorem 1.2, in the simplified form of their Corollary 1.3(a). Its hypotheses are an expansion $\varphi(y) = c + \mathrm{Re}\sum_{\lambda_n\leq X} r_n e^{i\lambda_n y} + \mathcal{E}(y, X)$ with $r_n \ll \lambda_n^{-\beta}$ for some $\beta > \frac12$, $\sum_{T<\lambda_n\leq T+1} 1 \ll \log T$, and $\lim_{Y\to\infty}\frac1Y\int_{y_0}^{Y}|\mathcal{E}(y, e^Y)|^2dy = 0$. Here $\varphi = E_a$, $c = -1/(2a+1)$, the $\lambda_n$ are the distinct positive ordinates $\gamma$, and $r_\gamma = -2m_\gamma/(\frac12 + a + i\gamma)$, where $m_\gamma$ is the multiplicity. Since $m_\gamma \ll \log\gamma$, $r_\gamma \ll \lambda^{-3/4}$, so $\beta = \frac34 > \frac12$ (under LI all $m_\gamma = 1$ and $\beta = 1$); Step 4 gives the mean-square condition. The additive constant $c$ is part of their expansion, so no reduction is needed. Hence $E_a$ is $B^2$-almost periodic and has a limiting distribution $\mu_a$.

*Mean.* A $B^2$-almost periodic function has a mean value equal to its zero-frequency Fourier coefficient (Besicovitch). Here that coefficient is $c$: every $e^{i\gamma y}$ with $\gamma \neq 0$ has mean zero, and $E_a - c$ is the $B^2$-limit of trigonometric polynomials without constant term, while the mean is continuous in the $B^2$ seminorm.

The mean square is finite (ANS, Theorem 1.14: $c^2 + \frac{1}{2}\sum |r_\gamma|^2$), so the distributions $\nu_Y$ of $E_a(e^y)$ on $[y_0, Y]$ (normalized Lebesgue measure) have uniformly bounded second moments. Since $\nu_Y \to \mu_a$ weakly, the first moments converge:

$$\int t\, d\mu_a(t) = \lim_{Y\to\infty} \frac{1}{Y}\int_0^Y E_a(e^y)\, dy = -\frac{1}{2a+1}.$$

(ANS state the mean only in passing, as "a similar argument would also establish"; this is why the lines above are written out.) This proves (i).

**Step 6: the density under LI.** Under LI the vector $(\gamma y \bmod 2\pi)_{\gamma>0}$ is equidistributed on the infinite torus (Kronecker–Weyl; ANS, Theorem 1.9). So $\mu_a$ is the law of

$$X_a = -\frac{1}{2a+1} + \sum_{\gamma>0} \rho_\gamma \cos U_\gamma, \qquad \rho_\gamma = \frac{2}{|\frac{1}{2} + a + i\gamma|},$$

with $U_\gamma$ i.i.d. uniform on $[0, 2\pi)$. Since $\mathbb{E}[e^{i\xi\rho\cos U}] = J_0(\rho\xi)$,

$$\hat{\mu}_a(\xi) = e^{-i\xi/(2a+1)} \prod_{\gamma>0} J_0(\rho_\gamma \xi).$$

The product decays faster than any power of $\xi$, because infinitely many factors are each $\le 1$ and eventually $\ll |\xi|^{-1/2}$. So $\mu_a$ is absolutely continuous, in particular $\mu_a(\{0\}) = 0$, and it is symmetric about $-1/(2a+1)$. For the symmetric part $Z_a = X_a + 1/(2a+1)$ and $b = 1/(2a+1) > 0$, the inversion formula of Rubinstein–Sarnak (§3) gives

$$\mathbb{P}(Z_a > b) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin(b\xi)}{\xi}\, \hat{\mu}_{Z_a}(\xi)\, d\xi.$$

Since $\delta_a = \mu_a((0,\infty)) = \mathbb{P}(Z_a > b)$, this is (ii). □

### *9.5 The critical point* $a = -\frac{1}{2}$

**Remark 1.** The deterministic term of Step 2 is, uniformly in $a \ge -\frac{1}{2}$,

$$B_a(x) = \tfrac{1}{2}\int_2^{\sqrt{x}} \frac{t^{2a}}{\log t}\, dt.$$

For $a > -\frac{1}{2}$, $B_a(x) \sim x^{a+1/2}/((2a+1)\log x)$, which is the bias of Theorem 3. At $a = -\frac{1}{2}$, $B_{-1/2}(x) = \frac{1}{2}(\log\log\sqrt{x} - \log\log 2) = \frac{1}{2}\log\log x + O(1)$. Unconditionally, by Mertens, $\frac{1}{2}\sum_{p\le\sqrt{x}} p^{-1} = \frac{1}{2}\log\log x + O(1)$.

So the normalized bias $-1/(2a+1)$ diverges as $a \downarrow -\frac{1}{2}$, and at $a = -\frac{1}{2}$ the deterministic term becomes $-\frac{1}{2}\log\log x$. This is the $q = 1$ counterpart of the $\frac{1}{2}\log\log x$ in Aoki–Koyama (their (1.3), a race between residue classes mod 4 under the Deep Riemann Hypothesis, where the squares of the odd primes all fall in the class $1 \bmod 4$). In the notation $w = -a$ of the weighted counting functions, the bias $-1/(1-2w)$ diverges exactly at their critical point $w = \frac{1}{2}$.

*Status.* The deterministic part is proved. At $a = -\frac{1}{2}$ the zero contribution is $\sum \mathrm{li}(x^{i\gamma}) = (1/\log x)\sum x^{i\gamma}/(i\gamma)\,(1 + O(1/(\gamma\log x)))$, which under RH is bounded in mean square by $O(1/\log x)$. A complete statement, "$\theta_{-1/2}(x) - \mathrm{li}(\sqrt{x}) = -\frac{1}{2}\log\log x + O(1)$ in logarithmic mean square", requires writing out the boundary terms and the constant; we leave it at the level of a remark.

### *9.6 The bias fades as the weight grows*

**Theorem 4.** *Assume RH and LI, and put* $A = a + \frac{1}{2}$.
*(a)* $\delta_a \to \frac{1}{2}$ *as* $a \to \infty$.
*(b)* $\frac{1}{2} - \delta_a \sim b_a/(\sigma_a\sqrt{2\pi})$, *with* $\sigma_a^2 \sim \log(A/2\pi)/(2A)$, *hence* $\frac{1}{2} - \delta_a \sim 1/\big(2\sqrt{\pi A \log(A/2\pi)}\big)$.

*Proof of (a).* The variance of $Z_a$ is $\sigma_a^2 = \frac{1}{2}\sum_{\gamma>0}\rho_\gamma^2 = \sum_{\gamma>0} 2/(A^2+\gamma^2)$. With $N'(\gamma) \sim \frac{1}{2\pi}\log\frac{\gamma}{2\pi}$ and $\int_0^\infty \log(\gamma/2\pi)\,d\gamma/(A^2+\gamma^2) = \frac{\pi}{2A}\log\frac{A}{2\pi}$, this gives $\sigma_a^2 \sim \log(A/2\pi)/(2A)$. (The constant $2\pi$ matters numerically: at $a = 100$ it gives $\sigma = 0.1174$ against $0.1175$ from the zeros, while $\log A/(2A)$ would give $0.151$.) The summands $\rho_\gamma \cos U_\gamma$ are independent, centred and bounded by $\max\rho_\gamma \le 2/A$. Since $\max\rho_\gamma/\sigma_a \to 0$, Lindeberg's condition holds, and $Z_a/\sigma_a$ converges in law to $N(0,1)$, as in the limit $q \to \infty$ of Rubinstein–Sarnak (their Theorem 1.5). The bias is $b_a = 1/(2A)$, so $b_a/\sigma_a \sim 1/\sqrt{2A\log(A/2\pi)} \to 0$. The limit law is continuous, so the convergence of distribution functions is uniform (Pólya). Hence $\delta_a = \mathbb{P}(Z_a/\sigma_a > b_a/\sigma_a) \to \mathbb{P}(N > 0) = \frac{1}{2}$. □

*Proof of (b).* Write $W = Z_a/\sigma_a$ and $h_a = b_a/\sigma_a \to 0$. By Step 6, $W$ has characteristic function $\varphi_W(\xi) = \prod_{\gamma>0} J_0(\rho_\gamma\xi/\sigma_a)$. The obstacle noted in v3 is real: Berry–Esseen does not suffice, since $\sum\rho_\gamma^3/\sigma_a^3$ is of the same order as $h_a$. We work instead with the density of $W$.

*Lemma J.* For $|x| \le 2$, $0 \le J_0(x) \le e^{-x^2/4}$. Indeed, with $y = x^2/4 \le 1$, $e^{-y} - J_0(x) = \sum_{k\ge2}(-1)^k c_k$ with $c_k = y^k(1/k! - 1/(k!)^2)$; for $k \ge 2$, $c_{k+1}/c_k = \frac{y}{k+1}\cdot\frac{1-1/(k+1)!}{1-1/k!} \le \frac{5y}{9} < 1$, so the series alternates with decreasing terms and positive first term, and the difference is $\ge 0$. And $J_0 \ge 0$ on $[0,2]$, since its first zero is $2.404\ldots$

*A uniform bound.* $J_0(u) = \int_0^1 \cos(u\cos 2\pi t)\,dt$, so $|J_0(u)| < 1$ for $u \ne 0$; since $J_0$ is continuous and $J_0(u) \to 0$, $\theta := \sup_{|u|\ge\sqrt{2}}|J_0(u)| < 1$. We also use $|J_0(u)| \le C|u|^{-1/2}$ for $|u| \ge 1$.

*Lemma W (local limit theorem).* $W$ has a continuous density $f_W$, and $f_W(z) \to (2\pi)^{-1/2}e^{-z^2/2}$ uniformly in $z$. We have $f_W(z) - (2\pi)^{-1/2}e^{-z^2/2} = (2\pi)^{-1}\int e^{-iz\xi}(\varphi_W(\xi) - e^{-\xi^2/2})d\xi$, and we split the integral.

- $|\xi| \le A\sigma_a$. Since $\rho_\gamma \le 2/A$, every argument $\rho_\gamma|\xi|/\sigma_a$ is at most 2, so by Lemma J $|\varphi_W(\xi)| \le \exp(-\frac{\xi^2}{4\sigma_a^2}\sum\rho_\gamma^2) = e^{-\xi^2/2}$. The pointwise convergence $\varphi_W(\xi) \to e^{-\xi^2/2}$ is the central limit theorem of part (a); dominated convergence applies.
- $|\xi| > A\sigma_a$. The $K$ ordinates $0 < \gamma \le A$ have $\rho_\gamma \ge \sqrt{2}/A$, so their arguments are at least $\sqrt{2}$, and $K \to \infty$ by the Riemann–von Mangoldt formula. Bounding four of these factors by $C|u|^{-1/2}$, the other $K-4$ by $\theta$ and the rest by 1,

$$\int_{|\xi|>A\sigma_a} |\varphi_W| \le C^4\theta^{K-4}\,\frac{\sigma_a^2A^2}{2}\int_{|\xi|>A\sigma_a}\frac{d\xi}{\xi^2} = C^4\theta^{K-4}A\sigma_a \to 0,$$

  because $K \gg A\log A$ while $A\sigma_a \ll \sqrt{A\log A}$. The Gaussian tail beyond $A\sigma_a$ also tends to 0.

*Conclusion.* With $b_a$ as above, $\delta_a = \mathbb{P}(Z_a > b_a) = \mathbb{P}(W > h_a)$. $W$ is symmetric and absolutely continuous, so $\mathbb{P}(W > 0) = \frac{1}{2}$ and

$$\tfrac{1}{2} - \delta_a = \mathbb{P}(0 < W \le h_a) = \int_0^{h_a} f_W = h_a\big((2\pi)^{-1/2} + o(1)\big)$$

by Lemma W. With $b_a = 1/(2A)$ and $\sigma_a^2 \sim \log(A/2\pi)/(2A)$ this is $1/(2\sqrt{\pi A \log(A/2\pi)})$. □

Numerically, the Gaussian law with the exact $\sigma_a$ is already accurate at $a = 1$ ($0.0598$ against $0.0605$) and agrees to three decimals from $a = 10$ on; the closed form in $A$ becomes accurate only for $A \gg 2\pi, \gamma_1$ (Table 6).

The family therefore interpolates between the extreme bias of Chebyshev ($a = 0$) and the absence of bias ($a \to \infty$), and at the other end ($a \to -\frac{1}{2}$) it joins the logarithmic bias of the Deep Riemann Hypothesis school.

## 10 Numerical results for the family

All densities are computed with the $J_0$ formula using 20 000 zeros from Arb, plus a Gaussian factor for the tail beyond the last zero. The same code gives $\delta_0 = 2.61 \cdot 10^{-7}$ for $\pi$ against li, the value of Rubinstein–Sarnak: this validates the pipeline.

| $a$ | bias $b_a$ | $\sigma_a$ | $b_a/\sigma_a$ | $\delta_a$ (RH + LI) | first crossing |
|---|---|---|---|---|---|
| 0 | 1 | 0.2149 | 4.65 | $\mathbf{2.61{\cdot}10^{-7}}$ | none up to $10^8$ (known: below $1.4{\cdot}10^{316}$) |
| 0.25 | 0.667 | 0.2145 | 3.11 | $\mathbf{6.8{\cdot}10^{-4}}$ | none up to $10^8$ |
| 0.5 | 0.5 | 0.2144 | 2.33 | $\mathbf{9.2{\cdot}10^{-3}}$ | 24 109 |
| 0.75 | 0.4 | 0.2143 | 1.87 | **0.0308** | 113 |
| 1 | 0.333 | 0.2142 | 1.56 | **0.0605** | 109 |
| 1.5 | 0.25 | 0.2139 | 1.17 | **0.123** | 7 |
| 2 | 0.2 | 0.2135 | 0.94 | **0.177** | 3 |
| 3 | 0.143 | 0.2126 | 0.67 | **0.254** | 3 |

**Table 4.** The family: bias $b_a = 1/(2a+1)$, spread, limiting density, first crossing. $\theta_a(x) = \sum_{p \le x} p^a$ against $\mathrm{li}(x^{a+1})$. "First crossing" is the first prime $p$ with $\theta_a(p) > \mathrm{li}(p^{a+1})$, checked exactly with mpmath.

For $a \ge 1.5$ the first crossing happens among the first primes, where small-$x$ effects dominate, and the first crossing stops being an informative statistic. In the range $0 < a \le 1$ it moves from astronomically large ($a = 0$) to 24 109 ($a = \frac{1}{2}$) and 109 ($a = 1$). For small $a$ the spread is essentially constant, $\sigma_a^2 \approx 2\sum_{\gamma>0} \gamma^{-2} \approx 0.0462$: the knob moves the level, not the wave.

| $a$ | measured fraction | predicted fraction | sign agreement | correlation | mean of $E_a$ | intervals | limit $\delta_a$ |
|---|---|---|---|---|---|---|---|
| 1 | **0.0331** | **0.0329** | 99.66% | 0.9984 | $-0.398$ | 8 199 | 0.0605 |
| 2 | **0.1482** | **0.1476** | 98.99% | 0.9983 | $-0.231$ | 22 187 | 0.177 |

**Table 5.** Measured against predicted on $[10^3, 10^8]$. "Measured" is the exact logarithmic measure of $\{\theta_a > \mathrm{li}(x^{a+1})\}$, computed gap by gap between consecutive primes. "Predicted" uses the explicit formula with 20 000 zeros plus the exact prime-power layers $k \ge 2$, on $2^{16}$ points.

On this range the measured fraction is below the limit $\delta_a$: the mean of $E_a$ is still more negative than $-1/(2a+1)$ because of the layers $k \geq 3$ and the $O(1/\log x)$ corrections. The explicit formula with these terms reproduces the measured fraction to the third decimal. For $a = 1$, adding back the exact prime-power terms $K$ to the data gives a mean of $f + K = -0.0001$ (against $-0.398$ for $f$): the whole bias is accounted for by the powers of the primes. The rms error of the prediction is $0.0125$, against an rms of $0.219$ for the signal.

| $a$ | $\delta_a$ ($J_0$) | Gaussian, exact $\sigma_a$ | closed form | $\sigma_a$ | $\sigma_a^2 \cdot 2A/\log(A/2\pi)$ |
|---|---|---|---|---|---|
| 1 | **0.0605** | 0.0598 | — | 0.2142 | — |
| 5 | **0.3346** | 0.3324 | — | 0.2098 | — |
| 10 | **0.4072** | 0.4061 | 0.379 | 0.2005 | 1.64 |
| 20 | **0.4469** | 0.4465 | 0.443 | 0.1813 | 1.14 |
| 50 | **0.4730** | 0.4729 | 0.473 | 0.1455 | 1.03 |
| 100 | **0.4832** | 0.4831 | 0.483 | 0.1175 | 1.00 |

**Table 6.** Large $a$ (Theorem 4). Closed form: $\frac{1}{2} - 1/\big(2\sqrt{\pi A \log(A/2\pi)}\big)$. $\delta_a \to \frac{1}{2}$ as in Theorem 4(a); the Gaussian law with the exact $\sigma_a$ is accurate throughout, and the closed-form variance takes over from $A \approx 50$.

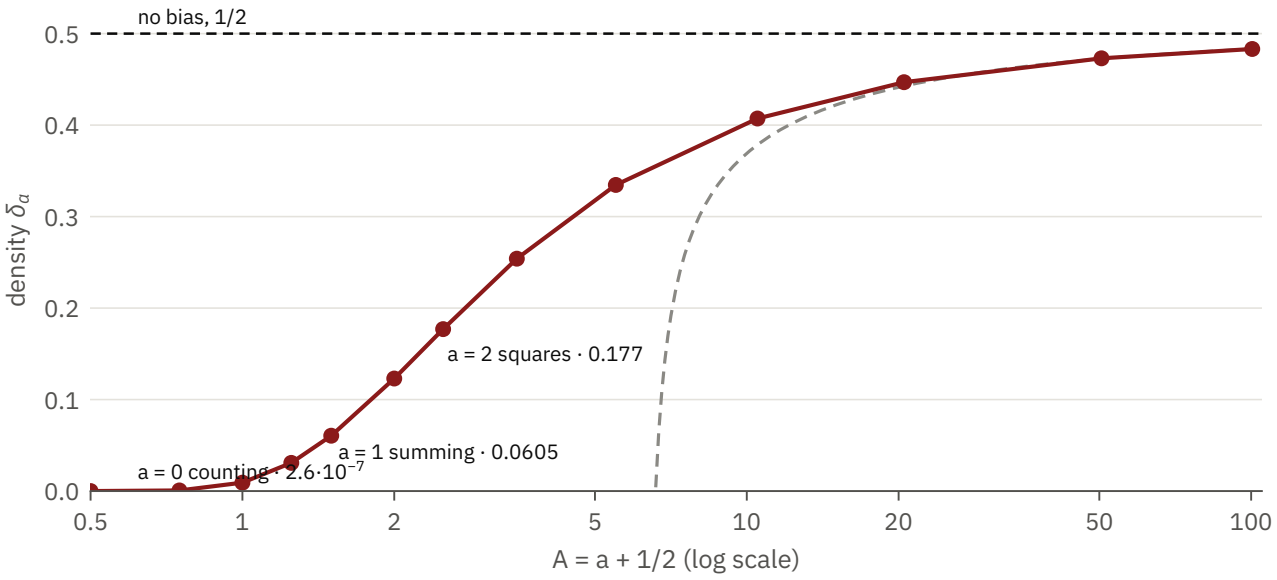



**Figure 9.** The whole family in one curve: limiting density $\delta_a$ of the set where the race $\theta_a$ is ahead of $\mathrm{li}(x^{a+1})$, against $A = a + \frac{1}{2}$ on a logarithmic axis (Tables 4 and 6). It climbs from Chebyshev's $2.6 \cdot 10^{-7}$ at $a = 0$ towards the unbiased $\frac{1}{2}$; the dashed curve is the closed form of Theorem 4(b), drawn where $A > 2\pi$.

## 11 Relation to the literature

- **The method** is that of Rubinstein–Sarnak (1994) [15]: limiting logarithmic distributions under RH, densities under LI. The general existence theorem used in Step 5 is that of Akbary–Ng–Shahabi (2014) [1]. For $a = 0$ Theorem 3 is their classical case.
- **Weighted counting functions.** Aoki–Koyama (2023) [2] and Shimada–Koyama (2025) [16] study

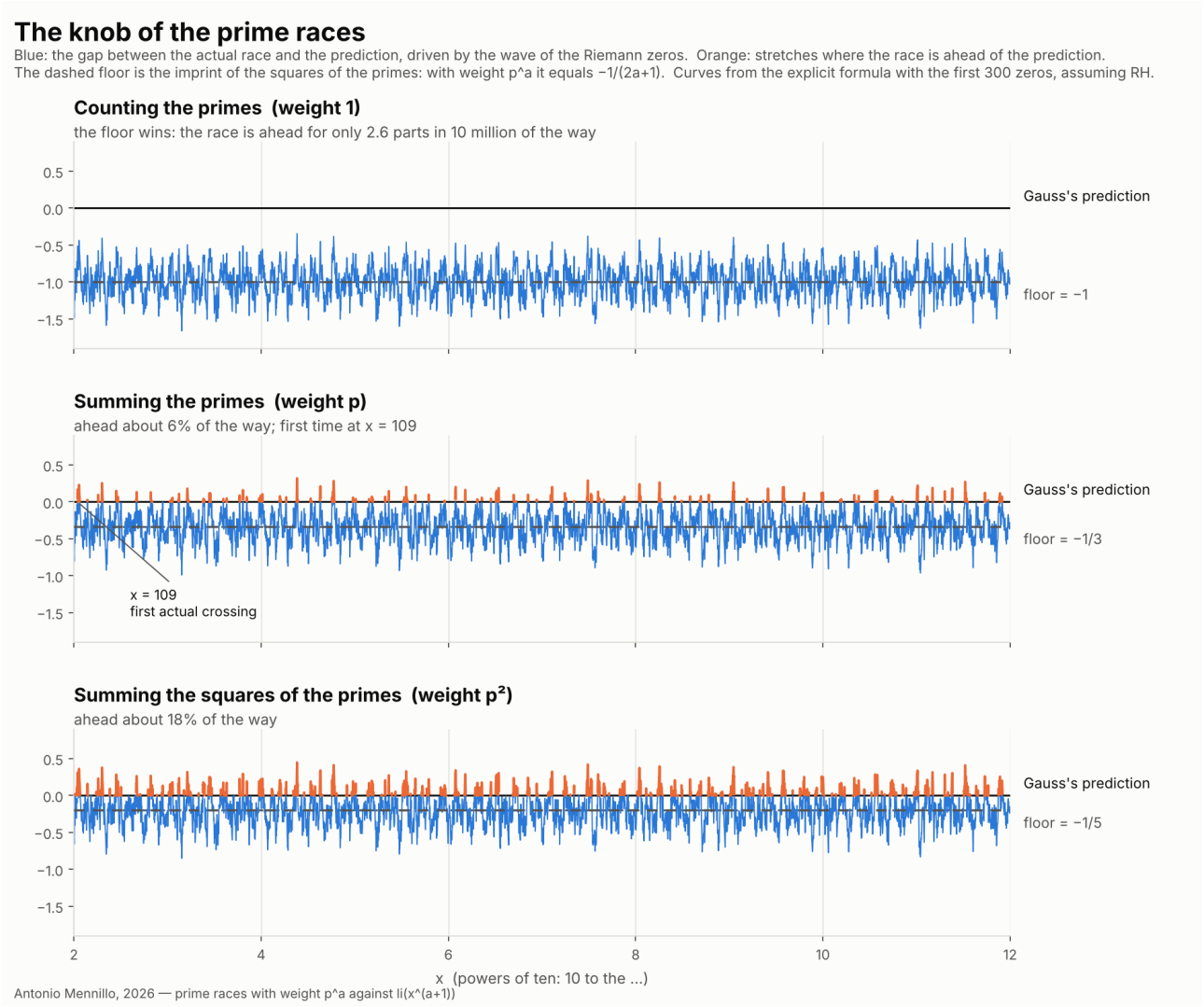



**Figure 10.** The knob of the prime races. Normalized error $E_a$ for $a = 0$ (counting the primes), $a = 1$ (summing them) and $a = 2$ (summing their squares), from the explicit formula with the first 300 zeros, assuming RH. The dashed floor is the bias $-1/(2a + 1)$ left by the squares of the primes; orange marks the stretches where the race is ahead of $\mathrm{li}(x^{a+1})$. The wave is almost the same in the three panels; the knob moves the floor.

$\pi_w(x; q, a) = \sum p^{-w}$ over primes in a residue class, with *decreasing* weights $0 \le w < 1$ (in their words, regarding "smaller primes as heavier elements"). They prove $\pi_w \sim \mathrm{Li}(x^{1-w})/\varphi(q)$, sign changes between classes for $w < \frac{1}{2}$, and, under the Deep Riemann Hypothesis, a $\frac{1}{2}\log\log x$ bias at $w = \frac{1}{2}$. The prime-square layer appears in Shimada–Koyama's Lemma 2.3 as an error term. Our family is the same object with $a = -w$, continued to *increasing* weights $a > 0$, compared with $\mathrm{li}(x^{a+1})$ rather than across residue classes, and with the prime-square layer read as a bias. The law $-1/(2a + 1)$ connects their two regimes: it diverges exactly at their critical point.

- **Highly biased races.** Fiorilli (2014) [8] shows that races between residue classes can be made arbitrarily biased; Hayani (2026) [9] identifies the square roots, the order of vanishing at $\frac{1}{2}$ and, in new examples, higher roots as sources of bias in prime ideal races. In our family the bias comes from the square roots alone, and its strength is controlled by a single real parameter.

- **Generalized Skewes numbers.** Bailleul–Hayani–Untrau (2026) [3] bound the first sign change in races between residue classes. For our family the first crossing is computed directly (Table 4).

- **Weighted error terms.** Recent work on correlations of error terms for weighted prime counting functions (arXiv:2507.13504) treats the weights $1$, $\log p$, $1/p$, $\log p/p$, not $p^a$.
- **Status of novelty.** To our knowledge the family $\sum_{p\le x} p^a$ against $\mathrm{li}(x^{a+1})$ with $a > 0$, its bias law and its limiting densities have not been studied before. A search of arXiv, zbMATH Open and the papers above found no prior treatment; a preprint of K. Shimada on the refined weighted prime number theorem, cited by Shimada–Koyama, could not be accessed. We would be grateful to be informed of any earlier occurrence.

## 12 Formalization in Lean

Every theorem, proposition and corollary of this paper, and the certified facts of Table 1, are formalized in Lean 4 [20] on top of mathlib [21], in the public repository github.com/anthem87/cps (release tag `cps-v4`, AGPL-3.0-or-later), with the shared library github.com/anthem87/lean-base. The development has no `sorry`; the only axioms are `propext`, `Classical.choice` and `Quot.sound`. Classical theorems not yet in mathlib enter as explicit hypotheses of the statements that use them, never as axioms: the prime number theorem; von Koch's bound under RH; the truncated explicit formula; the Riemann–von Mangoldt formula; Akbary–Ng–Shahabi, Corollary 1.3(a), Theorem 1.9 and (1.22); the inversion formula of Rubinstein–Sarnak; the decay $J_0(u) \ll |u|^{-1/2}$; Mertens' second theorem. RH and LI enter only as antecedents. RH uses mathlib's `RiemannHypothesis`.

- **Proved without hypotheses:** Definition 1, Proposition 1, Theorem 2 (Covering), Brown's criterion, Proposition 3(b) (with Landau's theorem proved), Step 3 of Theorem 3, Remark 1, the integer values of Table 1 and the primality of $N_1, \dots, N_{20}$, the first crossing at $x = 109$ (by rational bounds on li).
- **Proved from the classical hypotheses above:** Theorem 1, Proposition 2, Corollary 2, Proposition 3(a), Lemma A, Theorem 3 (i) and (ii), Theorem 4 (a) and (b).
- **Certificate:** the crossings $\sigma_{m-1} < \gamma_n \le \sigma_m$ of Table 1 depend on the rational enclosures of $\gamma_1, \dots, \gamma_{20}$ computed with Arb and on the prime sums computed with primesum; they enter as one stated certificate.
- **Not formalized:** the spectrum (Section 8, Tables 2–3), the densities of Tables 4–6 including $\delta_0 = 2.6 \cdot 10^{-7}$, the pre-registered test of Section 7 and the remarks without a formal statement.

Three proofs in Lean take a different route from the text, with the same statements: Theorem 4(a) proves the convergence of the characteristic functions $\prod J_0(\rho_\gamma \xi/\sigma_a) \to e^{-\xi^2/2}$ directly and concludes with Lévy's continuity theorem, instead of citing Lindeberg; Theorem 4(b) uses the integrated form $\frac{1}{2} - \delta_a = (h_a/\pi) \int_0^\infty \mathrm{sinc}(h_a s)\, \varphi_W(s)\, ds$ and an $L^1$ form of Lemma W; the absolute continuity in Theorem 3(ii) is obtained from the inversion formula.

### Reproducibility

All numbers were recomputed on 9 October 2026. Primes up to $10^8$ by a sieve ($5\,761\,455$ primes); sums in 64-bit integers; logarithms and residues with mpmath at 30 digits. Zeros: the first 1 000 with mpmath

at 33 digits, checked against an independent table; the first 20 000 as certified balls from Arb (python-flint 0.9). Null model of Section 4: $2 \cdot 10^5$ uniform heights (numpy `default_rng(1)`); the probability for $D$ in the exploratory data uses $10^6$ simulations (`default_rng(7)`); the pre-registered test uses `default_rng(20261009)`. Newton iterations use `mpmath.siegelz` and its derivative. The protocol, the certified records `run.jsonl`, the list of terms `b_file.txt` and the sixteen scripts are described in the Supplementary Material.


## Acknowledgments

AI tools were used to assist with the numerical computations, the certification pipeline, the Lean formalization and parts of the drafting. The author takes full responsibility for the content.


## References


[1] A. Akbary, N. Ng, M. Shahabi, *Limiting distributions of the classical error terms of prime number theory*, Q. J. Math. 65 (2014), 743–780; arXiv:1306.1657.

[2] M. Aoki, S. Koyama, *Chebyshev's bias against splitting and principal primes in global fields*, J. Number Theory (2023); arXiv:2203.12266.

[3] A. Bailleul, M. Hayani, T. Untrau, *A Wasserstein metric approach to generalized Skewes' numbers. I. Prime number races*, arXiv:2603.20093 (2026).

[4] J. L. Brown Jr., *Note on complete sequences of integers*, Amer. Math. Monthly 68 (1961), 557–560.

[5] P. L. Chebyshev, *Mémoire sur les nombres premiers*, J. Math. Pures Appl. 17 (1852), 366–390 (Bertrand's postulate).

[6] H. Davenport, *Multiplicative Number Theory*, 3rd ed., Graduate Texts in Mathematics 74, Springer, ch. 17–18.

[7] M. Deléglise, J. Rivat, *Computing $\pi(x)$: the Meissel, Lehmer, Lagarias, Miller, Odlyzko method*, Math. Comp. 65 (1996), 235–245.

[8] D. Fiorilli, *Highly biased prime number races*, Algebra Number Theory 8 (2014); arXiv:1210.6946.

[9] M. Hayani, *On the role of higher roots in prime ideal races*, arXiv:2607.23150 (2026).

[10] F. Johansson, *Arb: efficient arbitrary-precision midpoint-radius interval arithmetic*, IEEE Trans. Comput. 66 (2017), 1281–1292.

[11] J. C. Lagarias, V. S. Miller, A. M. Odlyzko, *Computing $\pi(x)$: the Meissel–Lehmer method*, Math. Comp. 44 (1985), 537–560.

[12] D. H. Lehmer, *On the exact number of primes less than a given limit*, Illinois J. Math. 3 (1959), 381–388.

[13] A. Mennillo, *Critical Primes, Phase Jumps, and Simple Zeros: Additive Sieving and the Riemann Spectrum*, Zenodo (2026), DOI 10.5281/zenodo.19190058.

[14] A. M. Odlyzko, *On the distribution of spacings between zeros of the zeta function*, Math. Comp. 48 (1987), 273–308.

[15] M. Rubinstein, P. Sarnak, *Chebyshev's bias*, Experiment. Math. 3 (1994), 173–197.

[16] K. Shimada, S. Koyama, *Weighted prime number theorem on arithmetic progressions with refinements*, Mathematics 13 (2025), 2564.

[17] J. Sorenson, J. Webster, *Strong pseudoprimes to twelve prime bases*, Math. Comp. 86 (2017), 985–1003.

[18] A. M. Turing, *Some calculations of the Riemann zeta-function*, Proc. London Math. Soc. (3) 3 (1953), 99–117.

[19] K. Walisch, *primesum*, fast prime summation, github.com/kimwalisch/primesum.

[20] L. de Moura, S. Ullrich, *The Lean 4 theorem prover and programming language*, CADE-28, LNCS 12699 (2021), 625–635.

[21] The mathlib Community, *The Lean mathematical library*, CPP 2020, 367–381.

[22] H. L. Montgomery, R. C. Vaughan, *Multiplicative Number Theory I. Classical Theory*, Cambridge Studies in Advanced Mathematics 97, Cambridge University Press, 2007, ch. 12.

## Appendix A. Changes with respect to version 2

| VERSION 2 (MARCH 2026) | THIS VERSION |
|---|---|
| "Generative formula" $\gamma_n = \log S_{N_n} + O(e^{-\gamma_n/2})$ | approximation by a grid; $N_n$ is defined from $\gamma_n$ (Definition 2), so it does not generate $\gamma_n$ |
| (5): $\varepsilon_n = O(e^{-\gamma/2}/\sqrt{\gamma})$, with $m(n) \sim e^{\gamma/2}\sqrt{\gamma}$ | $e^{\gamma/2}\sqrt{\gamma}$ is the asymptotics of $N_n$, not of the index $m = \pi(N_n) \sim 2e^{\gamma/2}/\sqrt{\gamma}$. Correct form: $\varepsilon < (1+o(1))\sqrt{\gamma}\, e^{-\gamma/2}$ (Theorem 2). Figure 1 of v2 already used this form. |
| §2.5: "a random spectrum would show no such trend" | the trend holds for every $t$; $U = 0.509$ on the zeros against $0.500$ on random heights; pre-registered test on $n = 6$–$20$ (Section 7) |
| "the error drops by a factor $\sim$300 at each step" | factors 15, 11, 18, 2 |
| Half-exponent law without $\sqrt{t'/t}$ | with the factor, agreement 98.7–99.8% |
| Tab. 3.2 and §5.2: $N_2 = 176\,753$ | $N_2 = 176\,777$ |
| Tab. 2.3, $n = 4$: $3 \cdot 10^{-6}\,\%$ | $1.6 \cdot 10^{-6}\,\%$ |
| §6.1: $N_4 \sim 10^7$, $N_5 \sim 2 \cdot 10^7$ | $2.3 \cdot 10^7$ and $8.4 \cdot 10^7$ |
| §4 and §6.3: mass–bit correspondence and $\psi(2S)$ | removed. $I_n = \gamma_n/\ln 2$ rewrites $\gamma_n$ in bits and carries no evidence of its own; in the mass–bit figure, $\chi_{c0}$ ($\gamma = 23.170$) and $\Upsilon(1S)$ ($\gamma = 64.30$) did not use zeros of $\zeta$. |
| Abstract: "verified against CMS open data" | the data were PDG masses; the physics part is no longer in this work |
| §7.2, question 4 (zeros off the line): open | settled: the position of the nodes is blind to the real part (Section 6); the ripple is not (Proposition 3) |
| §7.2, question 3 (Hilbert–Pólya): open | settled negatively for $\hat{H}$ alone (Section 6) |
| Data availability: `zeros.txt`, `someprimes.txt`, verification script | the v2 record contained only the PDF and the figure script; this version deposits data and scripts (Supplementary Material) |
| — | new: Part II (ripple, Proposition 3, Theorems 3–4, Tables 2–6, Figures 5–10) |

## Appendix B. Changes with respect to version 3

| VERSION 3 (OCTOBER 2026) | THIS VERSION |
|---|---|
| Theorem 4(b): "Heuristic, with numerical support" | proved under RH and LI (Lemma J, Lemma W) |
| Lemma A: lower limit $3/2$; the truncated explicit formula used at $t = 3/2$ | lower limit $5/2$ with the term $2^a$; the formula cited from Montgomery–Vaughan, Theorem 12.5, with its range $t \geq c > 1$ |
| Step 2: $(1 + o(1))$ from the prime number theorem, then boxed with $1 + O(1/\log x)$ | prime number theorem with the error term of de la Vallée Poussin |
| Step 3: transition values $ka \approx -1$ contribute $O(\log\log x)$ | a uniform exponent $b_k = \min(k\lvert a\rvert, 1 - \delta)$, no transition cases |
| Step 5: $r_\gamma = -2/(\frac{1}{2} + a + i\gamma)$, $\beta = 1$ | multiplicities $m_\gamma$ included, $\beta = \frac{3}{4}$ without LI |
| Theorem 3(ii): LI for the ordinates | LI for the ordinates counted with multiplicity |
| Corollary 2: $1 + o(1)$ | $o(1)$ as $t \to \infty$, uniformly in $t' \geq t$ |
| Section 9.1: first crossing at $x = 109$ | on $x \geq 2$ |
| — | new: Section 12, formalization in Lean 4 |