hexagon:2611.00850v1

The mean square of the Goldbach error at every large endpoint

Wren Aquila

22 November 2026

Abstract

For every sufficiently large X, the sum of squared errors in the Hardy–Littlewood prediction for the log-weighted binary Goldbach count is $\gg X^2 \log^2 X$ on $(X/1000, X]$, separately in the classes 2 and 4 modulo six. For ordinary ordered prime pairs the lower bound is $\gg X^2/\log^2 X$. Montgomery and Vaughan obtained these orders on a sequence of sharp endpoints. We localize their norm comparison by removing the unwanted output frequencies in L^1. The lower bounds persist when, on each interval of a partition with interval lengths at least $X\exp(-(\log X)^{1/3})$, the singular series $\mathfrak{S}(k)$ in the prediction is replaced by $a\,\mathfrak{S}(k)+b$, with constants a, b fitted from the data on that interval.

Use ordered representations. Put $L = \log X$, $\eta = 1/1000$, and let $\mathfrak{S}(k)$ be the usual binary singular series. Consider either the log-weighted prime count or the ordinary count, with their respective predictions:

$$\begin{aligned} R(k) &= \textstyle\sum_{p+p'=k} \log p \log p', \qquad & H(k) &= (k-1)\mathfrak{S}(k), \\ R(k) &= \#\{(p,p') : p+p' = k\}, & H(k) &= \mathfrak{S}(k) \sum_{2\le j\le k-2} \frac{1}{\log j \log(k-j)}. \end{aligned}$$

Write $K_X = XL$ for the first convention and $K_X = X/L$ for the second.

Theorem 1. *Unconditionally, for each $h = 2, 4 \bmod 6$,*

$$\liminf_{X\to\infty} \frac{1}{K_X^2} \inf_{\tau\in\mathbb{R}} \sum_{\substack{\eta X < k \le X \\ k\equiv h \ (6)}} |R(k) - \tau H(k)|^2 \ \ge \ 10^{-5}. \tag{1}$$

The two amplitudes may be chosen separately. The weighted result also holds for von Mangoldt weights and for $k\mathfrak{S}(k)$. In particular the lower order on the full sharp prefix holds for every sufficiently large endpoint X. Every interval $(X/1000, X]$ contains, in each of these two classes, an even integer with an absolute error $\gg \sqrt{X}\log X$, or $\gg \sqrt{X}/\log X$ in ordinary unit counts.

Montgomery–Vaughan (1973), Theorem 2 and Corollaries 1–3, already gave the same squared-error orders under exponential weighting and along a subsequence of sharp endpoints, and their Corollary 2 gives the pointwise sizes $\sqrt{n}\log n$ and $\sqrt{n}/\log n$ along a sequence. Their uniform sharp lower bounds (13)–(14) lose two powers of the logarithm. We have not found the every-endpoint bound at full order, or its localization to $(X/1000, X]$, in the sources we checked. A related loss occurs in the Erdős–Fuchs problem: Montgomery and Vaughan (1990, Corollary 2) prove a localized every-N bound only with a logarithmic loss and ask for a proof without it. The argument below uses the norm comparison of Montgomery and Vaughan (1973) at a finite polynomial. There it bounds the mean square over all outputs of the truncated convolution, up to $2X$; compare the finite form of their argument described by Chou, Haag, Huryn and Ledoan [1]. The step added here removes the outputs above X, and below ηX, in L^1. The output interval and both residue classes in (1) refer to the ordinary unrestricted count. No sign for the individual errors follows.

1 Squared prime polynomials in the two residue classes

Fix $r = 1$ or $2 \bmod 3$, so $2r \bmod 3$ identifies the desired even class. On $[1, X]$ let $A_j = (\log j)\mathbf{1}_{\mathbb{P}}(j)$, $w_j = 1$. For unit counts use $A_j = L\mathbf{1}_{\mathbb{P}}(j)$, $w_j = L/\log j$ for $j \geq 2$, $w_1 = 0$. Then both count and model are multiplied by L^2, returning the common scale XL. Restrict both polynomials to $j \equiv r \bmod 3$, writing

$$P(x) = \sum_{j \equiv r} A_j e(jx), \qquad I(x) = \sum_{j \equiv r} w_j e(jx), \qquad e(x) = e^{2\pi i x}.$$

Put $Y = \sqrt{X}L^4$, and set

$$W(x) = \frac{9}{4} \sum_{\substack{q \leq Y \\ 3 \nmid q}} \frac{\mu^2(q)}{\varphi(q)^2} \sideset{}{^*}\sum_{a(q)} I(x - a/q)^2, \qquad T = P^2 - \tau W.$$

For $k \equiv 2r \bmod 3$, the only other prime pairs use the prime 3. Including prime powers would allow powers of 3 in these extra pairs. The contribution is $O(L^2)$ per output for either prime convention as scaled here, including at odd outputs from the prime 2. At other residues both polynomials P^2, W have zero coefficient.

On $k \equiv 2r \bmod 3$ the full series with $3 \nmid q$ is $(4/3)\mathfrak{S}(k)$ by the Euler factor at 3. Its tail after Y is bounded pointwise by the positive divisor tail used in Montgomery–Vaughan, Lemma 18.6, regardless of the restriction on q. Its squared sum over $k \leq 2X$ is $\ll XY^{-2}L^3$. The coefficients of I^2 are bounded by $\sum w_j^2 \ll X$. For $k \leq X$ in this residue class, summing $w_j w_{k-j}$ over $j \equiv r \bmod 3$ gives one third of its unrestricted sum with error $O(L^2)$, by partial summation. Indeed the nonnegative weights w_j have supremum and total variation $O(L)$ on the integer interval, so their reflected product has total variation $O(L^2)$. The residue density, Euler factor and $9/4$ multiply to one. The full series has size $O(X^{o(1)})$. Consequently the squared-root error between the coefficients of T up to X and the indicated true discrepancies is $o(XL)$ for bounded τ, including the odd outputs whose prediction is zero.

It is sufficient to work along sequences with bounded normalized square error in (1), and then $\tau = 1 + o(1)$. In fact the PNT in the two fixed progressions gives the same first mean for the count and prediction on $(\eta X, X]$. On scaled counts their sums over the desired even class are each $(1-\eta^2)X^2/8 + o(X^2)$. For the prediction the mean of $\mathfrak{S}(k)\mathbf{1}_{k \equiv 2r \ (3)}$ is $1/4$, by its even Euler product, or the positive divisor expansion and summation in the progression. The unit-count archimedean factor after scaling is $k + o(X)$ uniformly, by $\sum_{j \leq X}(w_j - 1)^2 = o(X)$. For the observed first mean use the product PNT on the triangle $\eta < u + v \leq 1$ with the two primes in class r. Cauchy bounds the error in this comparison by $O(X^{3/2}L)$.

2 Removing large output frequencies in first absolute mean

The sum $A_3(Y) = \sum_{q \leq Y, 3 \nmid q} \mu^2(q)/\varphi(q)$ is $\frac{2}{3}\log Y + O(1)$, by the Euler factor or by subtracting the terms divisible by 3 from the ordinary first mean. For fixed $s \geq 0$, damp each frequency k of T by $e^{-sk/X}$, writing T_s for the result. Damping each input polynomial by the same radial factor gives the damped squares exactly. Write $B_s = \int_0^1 e^{-2su} du$. The PNT in class r, Parseval, and the triangle inequality give

$$\limsup \frac{\|T\|_1}{XL} \leq \frac{3}{4}, \qquad \liminf \frac{\|T_s\|_1}{XL} \geq \frac{1}{2}B_s - \frac{1}{4}B_s = \frac{1}{4}B_s. \tag{2}$$

The model contribution is bounded by $\frac{9}{4}A_3(Y)\sum_{j \equiv r} w_j^2 e^{-2sj/X}$. Its leading coefficient is $1/4$. These estimates hold with $\tau = 1 + o(1)$. Restricting both input polynomials to $j \leq dX$ for fixed $0 < d < 1$, with Y, L, τ as before, gives the first upper estimate $3d/4 + o(1)$ on the original scale.

For real u, let $g = 0$ up to $1/2$, $g = 2u - 1$ on $[1/2, 1]$, and $g = 1$ thereafter. The coefficient multiplier $h(u) = e^{-su}g(u)$ acting at $u = k/X$ has $L^1(\mathbb{T})$ operator norm at most

$$\|\widehat{h}\|_1 \le \sqrt{\pi}\,(\|h\|_2^2 + \|h'\|_2^2/(4\pi^2))^{1/2} \le C_s := \sqrt{\pi/(2s)}e^{-s/2}\sqrt{1 + (s+2)^2/(4\pi^2)}. \tag{3}$$

This follows by Fourier inversion, averaging translated copies of the polynomial, and Cauchy–Schwarz on the real Fourier variable. Only $s > 0$ is used in (3). Removing this part of T_s leaves the multiplier $e^{-su}(1 - g(u))$ supported, on the available nonnegative frequencies, in $[0, 1]$.

To also remove $u \le \eta$ use $v_\eta(u) = 2(1 - |u|/(2\eta))_+ - (1 - |u|/\eta)_+$. It equals 1 through η, vanishes from 2η onwards, and lies in $[0, 1]$. Its multiplier norm is at most three since each tent has a nonnegative real Fourier transform of integral one. On its support $g = 0$, and the coefficients up to $2\eta X$ are exactly those of the polynomial formed by restricting both input slots to $j \le 2\eta X$. Radial damping is an L^1 contraction for analytic polynomials by the Poisson kernel. Thus the removed small-end contribution has norm at most $(9\eta/2 + o(1))XL$. The remaining multiplier has modulus at most one and support $\eta X < k \le X$. From (2)–(3) its output has L^1 norm, divided by XL, at least

$$B_{12}/4 - \tfrac{3}{4}C_{12} - 9/2000 - o(1) > 0.004 - o(1).$$

The L^2 norm is larger. Returning from the truncated model and discarding odd outputs costs $o(XL)$, proving (1). Omitting or inserting proper prime powers in the two unrestricted slots changes the full squared-output polynomial in L^2 by $O(\sqrt{X})\,O(\sqrt{XL}) = o(XL)$. Indeed Chebyshev's bound gives $\sum_{p^j \le X,\ j \ge 2} \log p \ll \sqrt{X}$, which bounds the supremum norm of the difference of the two input polynomials. Their sum has L^2 norm $O(\sqrt{XL})$ by Parseval. Replacing $k - 1$ by k costs $X^{1/2+o(1)}$.

3 A fitted local mean and singular-series amplitude

The same constant 10^{-5} in (1) survives the following larger fit. Write $H_0(k) = H(k)/\mathfrak{S}(k)$ on the chosen even class. Choose *any* partition of $(\eta X, X]$ into intervals, each of length at least $X\exp(-(\log X)^{1/3})$. In each interval B fit two arbitrary constants, replacing the prediction by

$$(a_B\mathfrak{S}(k) + b_B)H_0(k).$$

The partition and constants can depend on the observed counts. Thus arbitrarily many fitted local main terms are permitted at the indicated resolution. We give details since a small relative change in the main need not be small in squared error. Montgomery and Vaughan (1973, p. 209) already raised further main terms and judged an essential improvement unlikely; Vaughan (1972, Theorem 2) proves a lim sup version, with finitely many power-type main terms, for representation functions of general sequences.

Use the scaled coefficient convention from the proof, so H_0 is of order X on this interval. Write $K = \exp(L^{1/3})$. The first-mean comparisons over any subinterval B, refined to either of the classes $k \equiv 0, 1 \bmod 5$ as well as the chosen class modulo six, are

$$\begin{aligned} \sum_{k\in B, c(30)} R(k)/H_0(k) &= \mu_c\,|B|/30 + O_\eta(Xe^{-c_0\sqrt{L}}), \\ \sum_{k\in B, c(30)} \mathfrak{S}(k) &= \mu_c\,|B|/30 + O(X^{o(1)}). \end{aligned} \tag{4}$$

Here $\mu_c = 15/8$ or $45/32$ according as $k = 0$ or $1 \bmod 5$. For the second line use the positive divisor expansion at odd primes on each even class. The local factor at 5 has the two values $5/4$ and $15/16$, multiplying the common conditional mean $3/2$ on $k \equiv h \bmod 6$. The total floor error in summing the

divisor weights through X is $X^{o(1)}$; the mean tail beyond X contributes $X^{o(1)}$ as well by the divisor bound.

For the first line, in each fixed unit class modulo 30 replace the prime weights in each input slot by $30/\varphi(30)$ times the weights on the integers in that class. The PNT in these fixed progressions, also after partial summation for unit counts, gives prefix error $O(X\exp(-2c_0\sqrt{L}))$ for sufficiently small $c_0 > 0$. In a two-slot triangular prefix $m + n \le t$, apply this first in the inner variable, then to the other variable against a monotone partial sum. The resulting error is $O(X^2\exp(-2c_0\sqrt{L}))$. The normalized weight $X/H_0(k)$ has supremum $O_\eta(1)$ and total variation $O_\eta(L^2)$. Indeed each first difference of the integer two-slot convolution is $O(L^2)$ by the variation and supremum of one-slot weights. Partial summation therefore gives the first stated error. The integer comparison at each k has one thirtieth of the input weight per permissible class with error $O(L^2)$. There are four or three permissible input classes, respectively. This gives the same μ_c, and the finitely many excluded small primes are harmless.

Suppose the normalized fitted error stays bounded along a sequence. Testing the discrepancies on either subprogression in B after division by $H_0(k)$ costs $O(\sqrt{|B|}\,L)$ by Cauchy. Equations (4) give two linear equations in a_B, b_B, whose coefficient rows after normalization are $(\mu_c + O(X^{o(1)}/|B|), 1 + O(1/|B|))$. Their limiting matrix is nonsingular. The right-hand sides are $\mu_c + O_\eta(Ke^{-c_0\sqrt{L}} + L\sqrt{K/X})$. Consequently $\max_B(|a_B - 1| + |b_B|) = O_\eta(e^{-c_1\sqrt{L}})$ with fixed $c_1 > 0$.

Projection of a trigonometrical polynomial onto one interval of integer frequencies has L^1 operator norm $O(L)$ up to degree $2X$, by the Dirichlet kernel. Projection onto the chosen fixed congruence class is a contraction by finite character averaging. The original model has $\|W\|_1 \ll XL$. The integer prediction $H_0(k)$ before the residue projection is furnished by the square of the full integer polynomial, of L^1 norm $O(X)$. The total L^1 cost of these fitted corrections in the truncated model is therefore $O(KLe^{-c_1\sqrt{L}}XL) = o(XL)$. The localized multiplier in the proof has bounded L^1 norm: on analytic polynomials use the Poisson contraction, (3), and the low-frequency tents. The full-series coefficient approximation still costs $o(XL)$ in L^2, since the intervals are disjoint and the amplitudes bounded. Hence the preceding lower bound survives.

References

[1] L. Chou, S. Haag, J. Huryn and A. Ledoan, The error term in counting prime pairs, J. Number Theory **278** (2026), 422–450.

[2] H. L. Montgomery and R. C. Vaughan, Error terms in additive prime number theory, Quart. J. Math. Oxford (2) **24** (1973), 207–216.

[3] H. L. Montgomery and R. C. Vaughan, On the Erdős–Fuchs theorems, in *A Tribute to Paul Erdős*, Cambridge University Press, 1990, 331–338.

[4] R. C. Vaughan, On the addition of sequences of integers, J. Number Theory **4** (1972), 1–16.