hexagon:2611.01203v1

Parity information in a rough cofactor and exact reconstruction thresholds

Brice Pouly*

Abstract

Remove the prime powers of base at most y from an integer n sampled uniformly on $(X, 2X]$. Given the remaining cofactor and $n \bmod M$, we study the joint Liouville sign of the small parts of $n, n+1, 3n+1, 3n+2$. For $y = X^{1/(C \log\log X)}$, $C > 0$ fixed, we determine the first-order law of the signed observation cells uniformly for $1 \le M \le y$. Guessing the integer, certifying it, predicting the sign and certifying it have a common Dickman transition, with error $O_C(1/\log y)$ for all M, and distinct first corrections. The signed law also holds under von Mangoldt weighting of the cofactor. For prime cofactors counted with unit weight, the variation of prime density contributes a further term, visible when $\log M$ is of order $\sqrt{\log X}$. For products of unlabelled cofactors we deduce almost-everywhere reconstruction from a polynomial-ratio estimate of Wang and Xu and sharpen the local collision count by Pell equations. For three neighbouring small parts we determine the subpolynomial cutoff for predicting $\lambda(n)$ and construct a positive-density set on which the three observations recover the integer while each pair has many compatible origins.

1 What the quotient reveals

Factor a positive integer m as $S_y(m)R_y(m)$, where

$$S_y(m) = \prod_{p \le y} p^{v_p(m)}, \qquad R_y(m) = m/S_y(m).$$

We call these the small part and the rough cofactor. Both are one at $m = 1$. Write $\lambda(m) = (-1)^{\Omega(m)}$, counting prime factors with multiplicity. For the ordered arguments $n, n+1, 3n+1, 3n+2$, write S_i, R_i for their two factors, starting at $i = 0$. Our principal target is

$$Z(n) = \lambda(S_0 S_1 S_2 S_3).$$

The full quartic sign contains the further factor $\lambda(R_0 R_1 R_2 R_3)$.

Consider $30 < n \le 60$ and $y = 5$. On learning $R_0 = 11$, the observer can list the three candidates: the interval $30/11 < s \le 60/11$ contains $s = 3, 4, 5$, all of them smooth. Their small parts are

n	(S_0, S_1, S_2, S_3)	$Z(n)$
33	$(3, 2, 100, 1)$	$+1$
44	$(4, 45, 1, 2)$	$+1$
55	$(5, 8, 2, 1)$	-1

Returning $+1$ succeeds twice in three trials; returning a specified integer succeeds once in three. Giving the residue modulo three distinguishes all three rows. Other observations leave several possible integers with a common sign. The asymptotic cell laws below retain these differences, which the leading Dickman transition alone does not resolve.

*Research and writing with AI assistance.

1.1 The experiment and its transition

Let X be an integer tending to infinity, and put

$$L = \log X, \qquad y = X^{1/(C \log L)}, \qquad C > 0 \text{ fixed.}$$

Logarithms are natural, except in entropies, which are measured in bits. Write $\mathbb{P}_X, \mathbb{E}_X$ for the uniform law on the X integers in $(X, 2X]$. The observation is $O_M = (R_0, n \bmod M)$, where $M \geq 1$ is an integer. Its cell consists of all rows returning that value. Define

$$m_M = \mathbb{E}_X(Z \mid O_M), \qquad D_X(M) = \sup_{|W| \leq 1} |\mathbb{E}_X[ZW(O_M)]| = \mathbb{E}_X|m_M|.$$

Tests may depend on X and on the entire observed cell. The optimal probability of a correct sign decision is $(1 + D_X(M))/2$. Let $B_X(M)$ be the best probability of guessing the integer, and let $Q_X(M)$ be the probability of a singleton cell. An error-free decoder allowed to abstain certifies the integer precisely on those cells. The analogous probability for certifying Z is $\mathbb{P}_X(|m_M| = 1)$. Reconstruction means uniqueness in the dyad, with unrestricted decoding cost.

Let ρ be the Dickman function, equal to one on $[0, 1]$ and satisfying $u\rho'(u) = -\rho(u-1)$ for $u > 1$. Write

$$F(a) = e^{-\gamma} \int_0^a \rho(u)\, \mathrm{d}u, \qquad F(\infty) = 1, \qquad V(y) = \prod_{p \leq y} (1 - 1/p),$$

where γ is Euler's constant. We use $P^+(1) = 1$.

Theorem 1.1 (Prediction and reconstruction)**.** *Uniformly in all integers $M \geq 1$, each of*

$$D_X(M), \quad B_X(M), \quad Q_X(M), \quad \mathbb{P}_X(|m_M| = 1), \quad I(Z; O_M)$$

equals each of the following quantities with additive error $O_C(1/\log y)$:

$$\mathbb{P}_X(S_0 \leq M), \qquad V(y) \sum_{\substack{s \leq M \\ P^+(s) \leq y}} \frac{1}{s}, \qquad F\left(\frac{\log M}{\log y}\right). \tag{1.1}$$

If $\log M / \log y \to a \in [0, \infty]$, then $|m_M|$ converges in law to $(1 - F(a))\delta_0 + F(a)\delta_1$.

The reconstruction part has a direct geometric explanation. If $r = R_0 > X/M$ and $(r, M) = 1$, the missing multiplier s lies in an interval of length less than M, and its residue modulo M is known. This identifies the row. To control the other cells, one must prove cancellation between their signs while allowing the candidate multipliers to vary. Section 2 supplies the aggregate estimate that makes this passage possible.

At $M \asymp y$ the limiting information is $e^{-\gamma}$ bit. At $M = 1$ all five quantities tend to zero, but their first-order terms differ.

Corollary 1.2 (The cofactor alone)**.** *For $M = 1$, the following limits hold under both the uniform law and the law proportional to $\Lambda(R_0)$. The middle column omits $e^{-\gamma}$.*

Quantity	*Coefficient*	*Limit after scaling by* $\log y$
Integer certified, $Q_X(1)$	$4/3$	$0.748612644756\ldots$
Sign certified, $\mathbb{P}_X(\lvert m_1\rvert = 1)$	$3\sqrt{2}\log(1+\sqrt{2}) - 2$	$0.976575361716\ldots$
Integer guessed, $B_X(1)$	2	$1.122918967134\ldots$
Sign correlation, $D_X(1)$	$\dfrac{\sqrt{\pi}\,\Gamma(1/4)}{2\Gamma(3/4)}$	$1.472179080316\ldots$

For sign prediction, the gain in success probability over $1/2$ is half the correlation in the last row. Theorem 3.1 gives the underlying signed cell law up to $M = y$. The proof groups a residue class by its least candidate multiplier s, so its successive members are $s, s + M, \dots$. It can then sum the arithmetic error over s, even when a particular list occurs only on a short interval of cofactors.

1.2 Relation to previous work

The distribution of the small part belongs to the classical Kubilius model. Tenenbaum [18, Theorem 1, (1.7), and Corollary 2] gives a total-variation approximation under uniform sampling; subtracting the estimates up to $2X$ and X gives the dyadic version. Pinsky [10, Theorem 1] identifies the Dickman limit for the harmonic law on smooth integers, and Jaramillo and Yang [9, Lemma 5.2] give a quantitative estimate for the associated sum of geometric prime exponents. These results describe the missing factor. Our question is how that distribution enters a decision about a prescribed sign after the quotient and a residue have been observed.

The truncated Liouville function $\lambda_y(n) = \lambda(S_y(n))$ also has an established literature. Cassaigne, Ferenczi, Mauduit, Rivat and Sárközy [1, Theorems 4–5 and Lemma 4] study its progression sums and correlations, and pass from correlations to frequencies of binary words. Daboussi and Sárközy [3, Corollaries 1–2] obtain a local-product formula and binary cancellation when $y \to \infty$ and $y = X^{o(1)}$. Our dispersion estimate retains roughness, mobile multipliers and residue restrictions, with an error that remains small after summing over cells. The error $\exp(-u/8)$ in the latter reference, where $u = \log X/\log y$, does not by itself provide the first-order precision needed here for every fixed $C > 0$.

Reconstruction and sign prediction from sampled prime factors [13] uses harmonic sampling and distinguishes guessing, certification and sign prediction. Its Theorems 23.1 and 25.1 concern a rough cofactor and a message about three neighbouring small parts, with a communication law involving four Dickman variables. Its Lemma 24.2 supplies a methodological antecedent: an exact-valuation transfer with mobile slopes and summed remainders. Here the extra data are the actual residue of the origin, the target is $\lambda(S_0S_1S_2S_3)$, and the row is uniform on a dyad. The signed cells retain first-order information even when the leading Dickman limit is zero.

Two related observations give different candidate sets. *Shifted-cofactor cancellation and a multiplicity threshold for product representations* [14] samples two factors independently with harmonic weights and reveals their product. In *Observations and Shift Decoding in Divisor Incidences* [12], a list of guesses concerns a hidden displacement in a divisor incidence. Here a candidate is an integer in one dyad with a prescribed rough cofactor and residue.

The signed cell law gives four distinct decision corrections at growing modulus, with an explicit extra term for prime cofactors counted with unit weight. Section 3.3 separates the contribution of the classical smooth-number marginal from the contribution of the signs within a cell. The reconstruction conclusion for unlabelled products in Section 6 is an application of Wang and Xu [20, Lemma 2.7]; the Pell argument there gives a sharper count with the removed factor left as a parameter.

Sections 2 and 3 prove the transition and first-order laws. Sections 4 and 5 vary the sign and the residues. Sections 6 and 7 change the observations, and the last section relates these experiments to the full quartic sign under their respective sampling laws.

2 The distribution inside and outside the reconstruction range

2.1 The size of the missing factor

Write $\pi_y(s) = V(y)\mathbf{1}_{P^+(s)\le y}/s$. Under this probability law the prime exponents are independent and $\mathbb{P}(v_p(s) \ge a) = p^{-a}$ for $p \le y$. For fixed $v, c > 0$,

$$\frac{1}{X}\sum_{t\le cX} S_y(t)^{v/\log y} \ll_{v,c} 1, \qquad \mathbb{E}_{\pi_y} s^{v/\log y} \ll_v 1. \tag{2.1}$$

Indeed, the first weight has nonnegative divisor increments. Counting multiples bounds its mean by c times the Euler product for the second expectation. Its factor at p is

$$\frac{1-1/p}{1-p^{-1+v/\log y}} = 1 + O_v\left(\frac{\log p}{p\log y}\right).$$

Here $p^{v/\log y} - 1 \ll_v \log p/\log y$, and Mertens' estimate for $\sum_{p\le y}\log p/p$ proves (2.1). Markov's inequality at this exponent gives a factor $L^{-C\varepsilon v}$ on $s > X^{\varepsilon}$. Thus, for fixed $C, \varepsilon > 0$, the tail is $O_{A,C,\varepsilon}(L^{-A})$ at every prescribed order A, by choosing v afterwards.

The following marginal comparison is a special case of Tenenbaum's Kubilius estimate [18, Theorem 1 and Corollary 2]. We retain a short sieve proof because its normalization is also used below. For smooth $s \le X^{1/10}$, rows with $S_0 = s$ correspond to rough integers in $(X/s, 2X/s]$. Apply the one-dimensional fundamental lemma at level $(X/s)^{1/2}$ [4, Theorem 3.6]. The local density is $1/p$, and each counting remainder is at most two. The ratio of the logarithmic level to $\log y$ is $\gg C\log L$; the fundamental-lemma error $\exp(-u\log u + O(u\log\log u))$ is therefore $O_{A,C}(L^{-A})$ for every fixed A. The rounding errors sum to $O(X^{11/20})$. Together with (2.1) and $\sum_s \pi_y(s) = 1$, this proves

$$\sum_s |\mathbb{P}_X(S_0 = s) - \pi_y(s)| \ll_{A,C} L^{-A}. \tag{2.2}$$

In particular,

$$\mathbb{P}_X(M < S_0 \le 2M) \ll_C \frac{1}{\log y} \tag{2.3}$$

uniformly in M, by summing $1/s$ over an interval of fixed multiplicative ratio.

For $M \le y$, the model mass up to M is exactly $V(y)H_M$, where $H_M = \sum_{j=1}^{M} 1/j$. Since $F(a) = e^{-\gamma}a$ on $[0,1]$ and $V(y) = (e^{-\gamma}/\log y)(1 + O(1/\log y))$, its difference from $F(\log M/\log y)$ is $O(1/\log y)$. For $M > y$, put $b = \lfloor y\rfloor$ and $a_b = \sum_{p\le b}\log p/p = \log y + O(1)$. In the notation of Jaramillo and Yang [9, Lemma 5.2], $\log s/a_b$ is their S_b, and the distribution error at argument $u > 0$ is $O((1+u^{-2})/\log b)$. Here $u = \log M/a_b \ge 1/2$ eventually. Replacing a_b by $\log y$ costs $O(1/\log y)$ even for unbounded M: the identity $u\rho(u) = \int_{u-1}^{u}\rho(v)\,\mathrm{d}v$ for $u \ge 1$ shows that $u\rho(u)$ is bounded, so $t \mapsto F(e^t)$ is Lipschitz. This proves the comparison of the final two expressions in (1.1).

2.2 An aggregate transfer for moving multipliers

The multiplier list changes with the observation. A comparison valid only for fixed affine forms would not control the total error. We first record the summed estimate, including its norm and the order in which its parameters are chosen.

Lemma 2.1 (Aggregate comparison with the local model)**.** *Let $1 \le M \le U \le X^{1/10}$ and $W_0 = X/U$. Let $\mathcal{P}$ be any subfamily of pairs*

$$(s,t) \in (U, 2U]^2 \cap \mathbb{N}^2, \qquad s \equiv t \pmod M, \qquad s, t \le X^{1/10}.$$

For every pair choose an interval $I_{s,t} \subseteq (X/s, 2X/s] \cap (X/t, 2X/t]$, and put

$$\mathbf{b}_{s,t}(r) = (r, sr+1, 3sr+1, 3sr+2, tr+1, 3tr+1, 3tr+2).$$

For arbitrary functions $F_{s,t} : \mathbb{N}^7 \to [-1,1]$, let $f_{s,t}(r) = F_{s,t}(S_y(\mathbf{b}_{s,t}(r)))$, with S_y applied coordinatewise. Let $\mu_{s,t}$ be the same expectation under independent Haar variables $r_p \in \mathbb{Z}_p$ at $p \le y$. For every fixed $A > 0$,

$$\sum_{(s,t)\in\mathcal{P}} \left| \sum_{r\in I_{s,t}\cap\mathbb{Z}} f_{s,t}(r) - |I_{s,t}|\mu_{s,t} \right| \ll_{A,C} \frac{W_0U^2}{M} L^{-A}. \tag{2.4}$$

The estimate is uniform over the tests and intervals, with no lower bound on interval length. In particular, each interval may be further restricted.

The same bound holds when the sum is over prime r, $|I_{s,t}|$ is replaced by $\int_{I_{s,t}} \mathrm{d}u/\log u$, and the local variables have normalized Haar law on $\mathbb{Z}_p^\times$.

Proof. Multivariate divisor inversion gives

$$F(\mathbf{a}) = \sum_{d_i\mid a_i} \Delta F(\mathbf{d}), \qquad \Delta F(\mathbf{d}) = \sum_{e_i\mid d_i} F(\mathbf{e}) \prod_{i=0}^{6} \mu(d_i/e_i),$$

and hence $|\Delta F(\mathbf{d})| \le \prod_{i=0}^{6} 2^{\omega(d_i)}$. This majorant is independent of the test. Retain tuples with $\prod_i d_i \le X^{1/10}$. A compatible tuple prescribes one residue class modulo the least common multiple of the reduced linear congruence moduli. Counting it in an interval costs at most two in rounding. Summing a fixed divisor function over products bounds the absolute coefficients by $O(X^{1/10}L^c)$, with c absolute. Since $\#\mathcal{P} \ll U^2/M$ and $W_0 \ge X^{9/10}$, the total rounding error is negligible compared with the right side of (2.4).

A discarded tuple has some $d_j > X^{1/70}$. For a fixed $v > 0$, its absolute contribution is bounded by the sum over $0 \le j \le 6$ of

$$L^{-Cv/70} S_y(b_j)^{v/\log y} \prod_{i=0}^{6} \tau(S_y(b_i))^2. \tag{2.5}$$

We used $\sum_{d\mid a} 2^{\omega(d)} \le \tau(a)^2$ and the Rankin factor $X^{-v/(70\log y)} = L^{-Cv/70}$.

To average this expression with mobile s, t, observe that for every nonnegative function G,

$$\frac{M}{W_0U^2} \sum_{(s,t)\in\mathcal{P}} \sum_{r\in I_{s,t}} G(rs) \ll \frac{1}{X} \sum_{n\le 2X} \tau(n)G(n). \tag{2.6}$$

For each (r,s) there are $O(U/M)$ partners t, and $n = rs$ has at most $\tau(n)$ representations. The same estimate holds for rt. An argument depending only on r instead gives the marginal $W_0^{-1}\sum_{W_0/2<r\le 2W_0} G(r)$.

Apply Hölder with fixed exponents to the seven factors in (2.5); then separate the representation weight $\tau(n)$, the divisor moments and the small-part moment. Each of the forms $n+1$, $3n+1$ and $3n+2$ is injective in n, with values at most $7X$. Thus (2.1) and $x^{-1}\sum_{b\le x} \tau(b)^j \ll_j (\log(2x))^{O_j(1)}$ give

$$\frac{M}{W_0U^2} \sum_{(s,t)\in\mathcal{P}} \sum_{r\in I_{s,t}} S_y(b_j)^{v/\log y} \prod_i \tau(S_y(b_i))^2 \ll_v L^{c_0},$$

where c_0 is independent of v. For the r marginal, the divisor increment argument for (2.1) applies with interval scale W_0 and the same cutoff y; here $y < W_0$ for large X.

In the Haar model each linear form has content at most two, so $\mathbb{P}(p^a \mid hr_p + b) \le 2p^{-a}$. Fixed divisor moments have factors $1 + O_j(1/p)$, while the small-part exponent contributes an additional bounded factor depending on v, as in (2.1). Hölder therefore gives the same bound on the model side, after multiplication by $|I_{s,t}|$ and summation. Choose the Hölder exponents first and then v so that $Cv/70 > A + c_0$. Both tails are then bounded by the right side of (2.4).

For prime r, the first small part is one and can be omitted. A retained tuple gives $O(1)$ classes modulo $q \le X^{1/10}$; its total absolute coefficient at fixed q is $\ll \tau(q)^{c_1}$. Non-unit classes contain no prime in the interval, since $r > q$. The classical Bombieri–Vinogradov theorem with maxima over reduced classes and prefix endpoints [16, Theorem 1.1], followed by partial summation, gives an error $O_B(W_0 L^{-B})$ for every fixed B. The available level is ample: $X^{1/10} \ll W_0^{1/2} L^{-B}$. Divisor weights are absorbed by Cauchy–Schwarz, the trivial bound $O(W_0/\varphi(q))$, and $\sum_{q \le X^{1/10}} \tau(q)^j/\varphi(q) \ll_j L^{O_j(1)}$. This is the usual classical theorem; see also [8, Chapter 17]. On the actual side the tails are bounded by allowing all integer r; on local units use $\mathbb{P}(p^a \mid hr_p + b) \le 2/\varphi(p^a)$. Increasing B and choosing v last proves the prime assertion. □

2.3 Two rows with the same rough part

For $n = rs$, with r rough and s smooth, write

$$Z_s(r) = \lambda(s)\lambda_y(sr+1)\lambda_y(3sr+1)\lambda_y(3sr+2).$$

The next bound supplies cancellation in cells too large to enumerate.

Lemma 2.2 (Dispersion on one band)**.** *Suppose $M \le U \le X^{1/10}$. Restrict the sum in the definition of $D_X(M)$ to $U < s \le 2U$, $s \le X^{1/10}$. Its normalized contribution is*

$$\ll_C L^{-7+o(1)} + \frac{\sqrt{M/U}}{\log y}.$$

If r is restricted to primes, the bound with the same $1/X$ normalization is $\ll_C L^{-1}(L^{-6+o(1)} + \sqrt{M/U})$. Both bounds allow a further interval restriction on r.

Proof. Put $W_0 = X/U$. Group s by its residue modulo M for each r. This refines the observation even when $(r, M) > 1$, so the triangle inequality bounds the desired sign norm by the sum of the absolute group sums. The upper sieve bounds the number N of possible groups by $O(MW_0/\log y)$. Cauchy–Schwarz reduces the second factor to signed pairs with $s \equiv t \pmod M$. There are $O(U^2/M)$ pairs, each with a common interval in $(W_0/2, 2W_0]$.

The six neighbouring forms are

$$hr+1, \quad 3hr+1, \quad 3hr+2, \qquad h = s, t, \tag{2.7}$$

and r must be rough. Two rational roots coincide only when t/s belongs to a fixed finite set. These $O(U)$ pairs each cost $O(W_0/\log y)$.

For every other pair, a prime $p > L^3$ is good unless it divides a slope or a nonzero determinant. At a good prime the seven roots, including zero, are distinct. The zero root is forbidden. At any other root the probability of an odd valuation is $(1 - 1/p)\sum_{j \ge 0} p^{-(2j+1)} = 1/(p+1)$, so the local signed mean is exactly

$$1 - \frac{1}{p} - \frac{12}{p+1}.$$

The exceptional primes above L^3 divide an integer of size $X^{O(1)}$ and have reciprocal sum $O(L^{-2})$. All remaining local factors have modulus at most one. Mertens' theorem gives a product $L^{-13+o(1)}$. Lemma 2.1, with $F(\mathbf{a}) = \mathbf{1}_{a_0=1}\prod_{i=1}^{6}\lambda(a_i)$, transfers this estimate after summation over all pairs. Consequently the two Cauchy factors satisfy

$$N \ll \frac{MW_0}{\log y}, \qquad S_2 \ll \frac{W_0U^2}{M}L^{-13+o(1)} + \frac{UW_0}{\log y}.$$

Since $X = W_0U$, their product gives

$$\frac{\sqrt{NS_2}}{X} \ll L^{-7+o(1)} + \frac{\sqrt{M/U}}{\log y}.$$

For prime r, $N \ll MW_0/L$. On local units the odd-valuation probability at a nonzero root is $p/(p^2-1)$, and the product of $1 - 12p/(p^2-1)$ is $L^{-12+o(1)}$. The prime transfer and the exceptional-pair bound $O(W_0/L)$ yield

$$S_2 \ll \frac{W_0U^2}{ML}L^{-12+o(1)} + \frac{UW_0}{L}.$$

The asserted prime bound follows from the same Cauchy calculation. □

2.4 The observable reconstruction event

Proof of Theorem 1.1. For $M < X$, take $\mathcal{A} = \{R_0 > X/M,\ (R_0, M) = 1\}$. It is observable, and each cell on $\mathcal{A}$ is a singleton. Failure of coprimality costs $O(L/y + L/X)$, by summing $1/p + 1/X$ over $p \mid M$ above y. Moreover $S_0 \le M$ implies $R_0 > X/M$, whereas $\mathcal{A}$ implies $S_0 < 2M$. Thus (2.3) gives $\mathbb{P}_X(\mathcal{A}) = \mathbb{P}_X(S_0 \le M) + O_C(1/\log y)$. For $M \ge X$, take $\mathcal{A}$ to be the whole dyad, since the residue is injective; (2.1) gives the same formula.

Rows with $s \le M$ cost at most their total mass in the sign norm. For $s > M$, sum Lemma 2.2 over $U = 2^jM$ up to $X^{1/10}$. The generic terms cost $O(L)L^{-7+o(1)}$, and the square-root terms form a convergent geometric sum. Larger s are covered by (2.1). Since $|m_M| = 1$ on $\mathcal{A}$,

$$\mathbb{E}_X[\mathbf{1}_{\mathcal{A}^c}|m_M|] = D_X(M) - \mathbb{P}_X(\mathcal{A}) \ll_C \frac{1}{\log y}. \tag{2.8}$$

This proves the correlation and sign-certification assertions and the claimed limit law. More generally, every measurable Φ with $|\Phi(x)| \le x$ on $[0,1]$ satisfies

$$\mathbb{E}_X\Phi(|m_M|) = \Phi(1)\mathbb{P}_X(\mathcal{A}) + O_C(1/\log y).$$

For mutual information use $\Phi(x) = 1 - h_2((1+x)/2)$, where h_2 is binary entropy. Concavity gives $0 \le \Phi(x) \le x$, and

$$I(Z; O_M) = \mathbb{E}_X\Phi(|m_M|) + H(Z) - 1.$$

Here $1 - H(Z) = O((\log y)^{-2})$, because $|\mathbb{E}_X Z| \le D_X(1) \ll 1/\log y$.

Conditionally on a cell the integer is uniform, so $B_X(M)$ is the number of occupied cells divided by X. Hence $\mathbb{P}_X(\mathcal{A}) \le Q_X(M) \le B_X(M)$. If $M \le X^{1/10}$, the number of cells with $r \le X/M$ is at most

$$M\#\{r \le X/M : S_y(r) = 1\} \ll X/\log y.$$

The remaining noncoprime cells cost at most the mass already bounded. If $X^{1/10} < M < X$, the moment estimate gives $\mathbb{P}_X(\mathcal{A}^c) = O_{A,C}(L^{-A}) + O(L/y + L/X)$ for every fixed A. This proves both integer-reconstruction assertions. Finally (2.2) and the quantitative Dickman comparison give all three main-term descriptions uniformly in M. □

3 The first-order law of the signed cells

The error in Theorem 1.1 is sharp in order already at $M = 1$, where $F(0) = 0$. To resolve that scale, we keep the signs on individual cells. A nonempty cell c has size k_c, contains j_c positive signs, and has common rough cofactor r_c. For a real score $h(k, j)$ and a nonnegative cofactor weight w, put

$$T_X(h, M; w) = \frac{1}{X} \sum_c w(r_c) h(k_c, j_c).$$

The score $|2j - k|$ gives the optimal sign correlation at weight one; the score 1 on nonempty cells gives the best probability of guessing the integer. We allow all scores satisfying

$$h(0, 0) = 0, \qquad |h(k, j)| \le 1 + |2j - k| \quad (0 \le j \le k,\ k \ge 1). \tag{3.1}$$

Let $J_k \sim \mathrm{Bin}(k, 1/2)$ and set $\widehat{h}(k) = \mathbb{E}h(k, J_k)$. For $m = qM + b$, $0 \le b < M$, define

$$G_{M,h}(m) = (M - b)\widehat{h}(q) + b\widehat{h}(q + 1), \qquad A_M(h) = \sum_{m \ge 1} \frac{G_{M,h}(m)}{m^2 - 1/4}. \tag{3.2}$$

For $m \le M$, at most m classes are occupied. For $m > M$, Cauchy–Schwarz gives $|G_{M,h}(m)| \le M + \sqrt{Mm}$. Thus $A_M(h) \ll 1 + \log M$ uniformly over (3.1).

Theorem 3.1 (First-order signed cell law)**.** *Put $\ell_M = \log M$, $h_1 = \widehat{h}(1)$, and*

$$E_X(M) = L \log\Big(\frac{L}{L - \ell_M}\Big) - \ell_M.$$

Uniformly for integers $1 \le M \le y$ and all scores satisfying (3.1),

$$(\log y) T_X(h, M; 1) = e^{-\gamma} A_M(h) + o_C(1), \tag{3.3}$$

$$L T_X(h, M; \mathbf{1}_{\mathrm{prime}}) = A_M(h) + h_1 E_X(M) + o_C(1), \tag{3.4}$$

$$T_X(h, M; \Lambda) = A_M(h) + o_C(1). \tag{3.5}$$

Under the law proportional to $\Lambda(R_0)$, divide (3.5) by $\mathbb{E}_X \Lambda(R_0)$ and multiply by $\log y$. The resulting main term is $e^{-\gamma} A_M(h)$ with error $o_C(1)$.

A cofactor weight changes the frequency of a cell while preserving the conditional law inside it. In (3.4) the difference between the local prime density $1/\log r$ and $1/\log X$ on singleton cells contributes $h_1 E_X(M)$. Since $E_X(M) = \ell_M^2(1 + O(\ell_M/L))/(2L)$, a fixed score with $h_1 \ne 0$ has main term $A_M(h)$ alone to additive $o(1)$ exactly when $\ell_M = o(\sqrt{L})$. If $\ell_M/\sqrt{L} \to t < \infty$, the correction tends to $h_1 t^2/2$. Weighting a prime by $\log r$ cancels this density variation in (3.5). The prime-indicator statistic in (3.4) uses the common denominator X.

3.1 Summing over the least candidate in a class

Write $\alpha = X/r$. Fix an integer $J \ge 2$ and first retain $1/2 \le \alpha < JM + 1/2$. Temporarily admit every integer multiplier in $(\alpha, 2\alpha]$. Since $M \le y$, each rough r is invertible modulo M. The observed residues therefore partition these multipliers into classes modulo M. A nonempty class has a least member s, and its full list is $s, s + M, \dots, s + (k - 1)M$, with $k \le J + 1$. The two conditions

$$s - M \le \alpha < s, \qquad s + (k - 1)M \le 2\alpha < s + kM$$

specify a single interval $I_{s,k}$ of r after the cutoff on α. A list may persist through many changes in the other residue classes. Summing by its least member counts that list only once for each r.

Lemma 3.1 (Comparison for lists with a prescribed least member)**.** *Fix J, $1 \le k \le J+1$ and a dyadic block $U \le s < 2U$, $U \ge 1$. For each s choose any subinterval $I_s \subseteq I_{s,k}$, possibly empty. Write $s_i = s + iM$ for $0 \le i < k$. Apply S_y coordinatewise to*

$$\mathbf{b}_s(r) = \bigl(r, (s_i r + 1, 3s_i r + 1, 3s_i r + 2)_{0 \le i < k}\bigr).$$

Let $f_s(r)$ be any real function of this vector of small parts with $|f_s| \le 1$. Denote its model mean by μ_s for independent Haar coordinates $r_p \in \mathbb{Z}_p$, $p \le y$, and by $\mu_s^\times$ for Haar coordinates on the local units. For every fixed $A > 0$,

$$\sum_{U \le s < 2U} \left| \sum_{r \in I_s \cap \mathbb{Z}} f_s(r) - |I_s|\mu_s \right| \ll_{A,C,J} XL^{-A}.$$

For prime r the same assertion holds with main term $\mu_s^\times \int_{I_s} du/\log u$, or with main term $\mu_s^\times |I_s|$ when the primes are weighted by $\log r$. All three estimates are uniform for $M \le y$.

Proof. For a nonempty interval, $s_i \le 2s$, $r \asymp X/U$ and $U \le 2Jy+2 = X^{o(1)}$. Apply the divisor inversion in Lemma 2.1 to the $d_0 = 1 + 3k$ coordinates, with $\prod_j d_j \le X^{1/10}$. Its coefficient is at most $\prod_j 2^{\omega(d_j)}$. A compatible tuple prescribes one class modulo the least common multiple of the reduced moduli. The rounding error over the block is $O_J(UX^{1/10}L^{O_J(1)})$.

For the discarded tuples, the analogue of (2.6) for each fixed offset i is

$$\sum_{U \le s < 2U} \sum_{r \in I_s \cap \mathbb{Z}} G((s + iM)r) \le \sum_{n \le 2X} \tau(n) G(n) \qquad (G \ge 0).$$

Indeed $s + iM$ is a divisor of n and fixes both s and r. The marginal of r is bounded by $O(U)$ copies of an interval of length $O(X/U)$. Thus Hölder's inequality, the divisor moments and (2.1) bound the total Rankin remainder by $O_{v,J,C}(XL^{c_J - Cv/(10d_0)})$, with c_J independent of v. The Haar remainder has the same bound, since $\sum_s |I_s| \ll X$ on the block and the content of each form is at most two. Choose v after A, J, C. This is where summing over s avoids a factor $X/|I_s|$ for short individual intervals.

For prime r, a compatible reduced modulus q in a retained tuple satisfies $q \le X^{1/10}$. A factor removed from a congruence modulus divides the content of its form, so every d_j divides $2q$. The total coefficient at q is $\ll_J \tau(2q)^{O_J(1)}$. Nonunit classes contain no prime of I_s, since $r > q$ for large X. The endpoint form of Bombieri–Vinogradov [16, Theorem 1.1], on $r \asymp X/U$, absorbs these weights by Cauchy–Schwarz as in Lemma 2.1. It gives $O_{A,J}((X/U)L^{-A})$ per s for the retained tuples, including arbitrary endpoints. The tails use the integer majorant just proved and its local-unit analogue. Partial summation gives the logarithmically weighted prime assertion, using the estimate for arbitrary subintervals with two further powers of L to spare. □

Proof of Theorem 3.1. On the complete list put $Z_{s_i}(r) = \lambda(s_i)\lambda_y(s_i r+1)\lambda_y(3s_i r+1)\lambda_y(3s_i r+2)$. This agrees with the target for each smooth s_i and rough r. Deleting a nonsmooth multiplier from a list of length at most $J+1$ changes its score by $O_J(1)$. Any deleted multiplier satisfies $y < s_i \le 2Jy + 1$ and is pt for some $p > y$ and $t \le 2J + 1$. Their reciprocal sum is $O_J(1/\log y)$ by the prime upper bound. For a given multiplier u there are $O(X/(u \log y))$ rough cofactors and $O(X/(uL))$ prime cofactors in its dyadic interval. These bounds follow from the upper sieve and $u = X^{o(1)}$. With logarithmic prime weights the bound is $O(X/u)$. In each of the three normalizations the total cost is $o_{C,J}(1)$.

On a fixed complete list expand its score in products of its k signs. The constant coefficient is $\widehat{h}(k)$, and all other coefficients are $O_J(1)$ uniformly in h. In a nonempty product the largest selected multiplier s_* has the unique rational root $-1/(3s_*)$. Among $s_i, 3s_i, 3s_i/2$ the largest

ratio is attained only there. Away from primes dividing slopes or nonzero determinants, coincident roots give the same valuations. If $d \geq 1$ roots occur with odd multiplicity, the local signed factors for a good prime are

$$1 - \frac{1}{p} - \frac{2d}{p+1} \quad \text{with roughness,} \qquad 1 - \frac{2dp}{p^2-1} \quad \text{on units.}$$

The exceptional primes above L^3 divide an integer of size $X^{O_J(1)}$ and have reciprocal sum $O_J(L^{-2})$. Mertens' theorem bounds the two products by $L^{-3+o_J(1)}$ and $L^{-2+o_J(1)}$. Primes below L^3 cost only $L^{o_J(1)}$ in these bounds. The harmless factors $\lambda(s_i)$ may be fixed separately on each list.

Apply Lemma 3.1 to the signed products and to the constant term, with the roughness indicator included for integer r. There are $O_J(1+\log(2M))$ blocks and a bounded number of subsets per list for fixed J. The sum of interval lengths is $O_J(X(1+\log(2M)))$. The nonconstant terms therefore contribute $o_{C,J}(1)$ to each rescaled score: even using $\log M \leq L$, the integer error is $\ll_J (\log y)(1+\log(2M))L^{-3+o_J(1)}$ and the prime errors are $\ll_J (1+\log(2M))L^{-2+o_J(1)}$. The accumulated transfer errors are negligible by choosing $A > 4$.

There are $m = \lfloor 2\alpha \rfloor - \lfloor \alpha \rfloor = \lfloor \alpha + 1/2 \rfloor$ consecutive candidates. Their mean scores sum to $G_{M,h}(m)$. Substitution $r = X/\alpha$ thus gives the retained model integral

$$A^{(J)} = \int_{1/2}^{JM+1/2} G_{M,h}(\lfloor \alpha + 1/2 \rfloor) \frac{d\alpha}{\alpha^2} = \sum_{m \leq JM} \frac{G_{M,h}(m)}{m^2 - 1/4}.$$

Its multiplier for integer r is $(\log y)V(y)$, and for logarithmically weighted primes it is one. The prime number theorem and partial summation give $(\log y)V(y) = e^{-\gamma}(1 + O((\log y)^{-2}))$, with enough precision for $|A^{(J)}| \ll_J 1 + \log(2M)$.

For primes counted with weight one and rescaled by L, the integrand acquires the factor $L/(L - \log \alpha)$. Write $v(\alpha) = \log \alpha/(L - \log \alpha)$ for its excess over one. Up to $\alpha = M + 1/2$, apart from the endpoint, $G_{M,h}(m) = mh_1$. On $[1, M]$ replacing m by α in the excess integral costs $O(L^{-1}\int_1^\infty (\log u)u^{-2}du)$. The endpoint pieces $[1/2, 1]$ and $[M, M + 1/2]$ cost $O(L^{-1})$. Above $M + 1/2$ the integral of $|G|/\alpha^2$ is $O(1)$ by (3.2), and $\sup |v| \ll_J \log(2M)/L = o(1)$ on the retained range. The resulting extra term is

$$h_1 \int_1^M \frac{\log u}{L - \log u} \frac{du}{u} + o_J(1) = h_1 E_X(M) + o_J(1).$$

For $\alpha \geq JM + 1/2$, (3.1) bounds the score by the number of occupied cells plus the sum of the absolute signed sums. There are at most M cells per $r \leq X/(JM)$. The upper sieve bounds the rescaled contribution from their number by $O_C(1/J)$ for each weight. In the signed term all multipliers exceed JM. Sum Lemma 2.2 over the dyadic ranges starting at JM, with the same interval restriction on r. Its square-root sum costs $O_C(J^{-1/2})$, and (2.1) treats $s > X^{1/10}$. For a prime weighted by $\log r$, use $\log r \leq L + O(1)$. The omitted part of $A_M(h)$ is also $O(J^{-1/2})$ uniformly in M. Take $X \to \infty$ at fixed J and then let $J \to \infty$.

Finally, proper prime powers $r \geq X^{9/10}$ contribute $O(X^{-1}X^{1/10}X^{1/2}L^2)$ to the normalized score since $|h(k, j)| \leq 2k$. For smaller r, every row has $s > X^{1/10}$ and its weighted mass is $O_{A,C}(L^{1-A})$ by (2.1). Summing the prime number theorem over smooth $s \leq X^{1/10}$ in the dyads $(X/s, 2X/s]$ gives, with the same tail estimate,

$$\mathbb{E}_X \Lambda(R_0) = V(y)^{-1}(1 + O_{A,C}(L^{-A})). \tag{3.6}$$

This proves the assertions for Λ, including the normalized law, since $(\log y)V(y)$ has the precision established above. □

3.2 Which finite cells contribute

For $h(k,j) = |2j-k|$, let $g_k = \mathbb{E}|\varepsilon_1 + \cdots + \varepsilon_k|$, where the ε_i are independent fair signs, and set $g_0 = 0$. Write $A_M = A_M(h)$. Adding a candidate adds one sign to one residue class. Since $g_{q+1} - g_q = \mathbb{P}(\sum_{i=1}^{q} \varepsilon_i = 0)$, summation by parts gives

$$A_M = \sum_{q\ge 0} \mathbb{P}\left(\sum_{i=1}^{q} \varepsilon_i = 0\right) \sum_{j=qM+1}^{(q+1)M} \frac{1}{j-1/2}. \tag{3.7}$$

Here $1/(m^2-1/4) = 1/(m-1/2) - 1/(m+1/2)$. For $M=1$, the generating function of return probabilities yields

$$A_1 = \int_0^1 \frac{t^{-1/2}}{\sqrt{1-t^2}}\,\mathrm{d}t = \frac{\sqrt{\pi}\,\Gamma(1/4)}{2\Gamma(3/4)}.$$

Write $b_M = \sum_{j=1}^{M} (j-1/2)^{-1} = 2H_{2M} - H_M$. This is the $q=0$ term in (3.7).

For guessing the integer, the score is one per occupied cell, so $G_{M,h}(m) = \min(m, M)$. For certifying it, the score is $\mathbf{1}_{k=1}$, giving $G_{M,h}(m) = \max(0, \min(m, 2M-m))$. Summing first differences gives, uniformly for $M \le y$,

$$(\log y) B_X(M) = e^{-\gamma} b_M + o_C(1), \tag{3.8}$$

$$(\log y) Q_X(M) = e^{-\gamma}(2b_M - b_{2M}) + o_C(1). \tag{3.9}$$

Along any growing modulus in this range, their rescaled difference tends to $e^{-\gamma}\log 2$.

For any Φ with $|\Phi(x)| \le x$ on $[0,1]$, take $h(k,j) = k\Phi(|2j-k|/k)$ on nonempty cells. Theorem 3.1 gives the first-order term of $\mathbb{E}_X \Phi(|m_M|)$ uniformly over these functions. For $\Phi(x) = x^2$, $\widehat{h}(k) = 1$ for $k \ge 1$, so

$$\mathbb{E}_X m_M^2 = B_X(M) + o_C(1/\log y).$$

For $\Phi(x) = 1 - h_2((1+x)/2)$, Theorem 3.1 gives the corresponding first-order series for mutual information, with its own score $A_M(h)$ and the $O((\log y)^{-2})$ entropy correction. For $\Phi(x) = \mathbf{1}_{x=1}$ it gives sign certification: the score is k on monochromatic cells, rather than one on singleton cells.

For $M=1$ and each fixed $k \ge 1$, the rescaled row probability of a cell of size k with j positive signs is

$$\lim_{X\to\infty} (\log y)\mathbb{P}_X(k_{O_1} = k, j_{O_1} = j) = e^{-\gamma} \frac{k}{k^2-1/4}\, 2^{-k} \binom{k}{j}. \tag{3.10}$$

Apply the theorem to the score supported at (k,j), and multiply by k. To sum over all monochromatic cells, apply the theorem directly to $\Phi(x) = \mathbf{1}_{x=1}$; this also controls the tail. The resulting coefficient is

$$\sum_{k\ge 1} \frac{k 2^{1-k}}{k^2-1/4} = 3\sqrt{2}\log(1+\sqrt{2}) - 2.$$

For the evaluation, split $k/(k^2-1/4)$ into half the sum of $(k \pm 1/2)^{-1}$ and sum the odd-power series for $\operatorname{arctanh}(1/\sqrt{2})$. Since $b_1 = 2$ and $2b_1 - b_2 = 4/3$, Corollary 1.2 follows, including its weighted version.

3.3 The four decision corrections

At a growing modulus the four decisions share the Dickman main term. Their first corrections can still be separated throughout $M \le y$. Write

$$c_{\mathrm{sign}} = \sum_{j\ge 1} 4^{-j}\binom{2j}{j}\log\Big(1+\frac{1}{2j}\Big), \qquad c_{\mathrm{mono}} = -\sum_{j\ge 2}\frac{j-1}{2^j}\log\Big(1+\frac{1}{j}\Big).$$

Theorem 3.2 (Decision corrections)**.** *Let $M \to \infty$, $M \le y$, and $a = \log M/\log y$. For each quantity T in the table,*

$$(\log y)(T - F(a)) \longrightarrow e^{-\gamma}(\gamma + 2\log 2 + c).$$

The same limits hold for the corresponding decisions under the probability law proportional to $\Lambda(R_0)$.

Decision quantity T	c
Integer certified, $Q_X(M)$	$-\log 2$
Sign certified, $\mathbb{P}_X(\lvert m_M\rvert = 1)$	c_{mono}
Integer guessed, $B_X(M)$	0
Optimal sign correlation, $D_X(M)$	c_{sign}

In particular the advantage in sign success probability over a fair guess is $D_X(M)/2$.

Proof. Counting occupied classes gives (3.8). Since $b_M = 2H_{2M} - H_M = \log M + \gamma + 2\log 2 + o(1)$ and $F(a) = e^{-\gamma}a$ for $a \le 1$, this gives the third row. Equation (3.9) gives the first. For the last row subtract b_M in (3.7). At a fixed $q \ge 1$ the inner sum tends to $\log(1+1/q)$ and is $O(1/q)$. The return probability is $O(q^{-1/2})$, so dominated convergence gives c_{sign}.

For sign certification the expected score on a class of size q is $w_q = q2^{1-q}$ for $q \ge 1$, with $w_0 = 0$. Each new multiplier increases $G_{M,h}$ by $w_{q+1} - w_q$ when it enters a class of size q. The summation by parts used in (3.7) now has first block b_M. At $q \ge 1$, the increment is $(1-q)2^{-q}$. The remaining blocks thus converge absolutely to c_{mono}. All four applications of Theorem 3.1 also hold with the normalized von Mangoldt weight. □

Numerically $c_{\mathrm{mono}} \simeq -0.2628$ and $c_{\mathrm{sign}} \simeq 0.6253$. To see that $-\log 2 < c_{\mathrm{mono}} < 0 < c_{\mathrm{sign}}$, observe that the positive weights $(j-1)2^{-j}$ for $j \ge 2$ sum to one, and $\log(1+1/j) < \log 2$ there. Thus nonsingleton cells with a common sign contribute at the same $1/\log y$ scale as the extra success of an integer guess. For comparison, (2.2) gives $\mathbb{P}_X(S_0 \le M) = V(y)H_M + O_{A,C}(L^{-A})$ on this range. Its correction to $F(a)$ is $e^{-\gamma}\gamma/\log y + o(1/\log y)$. The scalar friable-component expansion is classical [19, Lemma 2.1, (2.2)]. The other constants above come from the sizes and sign compositions of the observation cells.

4 Other multiplicative signs

The arithmetic reason for a fair sign on a bounded cell was the unique root belonging to its largest multiplier. This permits changes at all four sites. Replace Z by

$$Z_f = a_X(S_0)f_1(S_1)f_2(S_2)f_3(S_3),$$

where a_X is any sign-valued function and the f_i take values in $\{-1, 1\}$ and are multiplicative on coprime arguments. They may all depend on X.

Proposition 4.1. *If, for a fixed $\eta > 1/2$,*

$$\sum_{\substack{p\le y\\ f_2(p)=-1}} \frac{1}{p} \ge \eta \log\log y, \tag{4.1}$$

then the unweighted conclusions of Theorems 1.1 and 3.1 hold uniformly for these functions, with the same main terms. At $\eta = 1/2$ there are such families for which

$$\liminf_{X\to\infty} (\log y) D_{X,f}(1) \ge e^{-\gamma} A_1 + 1,$$

where $D_{X,f}$ denotes the sign correlation for Z_f.

Proof. For a generic pair, the two distinct roots of $3sr+1, 3tr+1$ give negative sign whenever the corresponding valuation is exactly one at a prime counted in (4.1). At a good $p > L^3$, the local signed mean is positive and at most

$$1 - \frac{1}{p} - 4\mathbf{1}_{f_2(p)=-1}(p^{-1} - p^{-2}).$$

Prime-power divisor increments still have modulus at most two. After excluding small and exceptional primes, the product is $L^{-1-4\eta+o(1)}$. In Lemma 2.2, the generic contribution becomes $L^{-1-2\eta+o(1)}$ per band; summing $O(L)$ bands costs $o(1/\log y)$ when $\eta > 1/2$.

For any nonempty product on a bounded candidate list, the largest s_* supplies a unique root with the prescribed sign. Its local product is $L^{-1-2\eta+o(1)}$ before conditioning on roughness. Lemma 3.1 applies to these signs with the same divisor majorant, uniformly in their values. After summing the retained list lengths and rescaling, the error is $\ll_{C,J} (\log y)(1+\log(2M))L^{-1-2\eta+o(1)} \le L^{1-2\eta+o(1)} = o(1)$ throughout $M \le y$. The integer-reconstruction counts are independent of the signs, and all claimed unweighted conclusions follow.

For the endpoint, choose a terminal segment $\mathcal{P}$ of primes up to y whose reciprocal sum is $(\log\log y)/2 + o(1)$, stopping as soon as the required lower bound is reached. Its smallest prime tends to infinity by Mertens' theorem. Set $a_X = f_1 = f_3 = 1$, set $f_2(p^a) = -1$ for $p \in \mathcal{P}$ and $a \ge 1$, and set all other prime-power values to one. One-form divisor truncation gives

$$\mathbb{E}_X Z_f = \prod_{p\in\mathcal{P}} (1 - 2/p) + O_C(L^{-3}) = \frac{1+o(1)}{\log y}.$$

On a fixed list of multipliers the roots $-1/(3s)$ are distinct at all sufficiently large primes. Their reciprocal mass diverges, and the local factors on rough units are $1 - 2d/(p-1)$ for $d \ge 1$ selected rows. Thus nonempty products vanish after conditioning on roughness, by the same truncation.

For fixed J, cells with $X/r < J + 1/2$ have total signed mass $o_J(1/\log y)$, whereas the sum of their absolute signed masses is

$$\frac{e^{-\gamma} \sum_{m\le J} g_m/(m^2 - 1/4) + o_J(1)}{\log y}.$$

The remaining cells contribute at least the absolute value of their combined signed mass, which is $(1 + o_J(1))/\log y$. Letting $J \to \infty$ proves the endpoint assertion. □

The endpoint example concerns universality of the first-order signed coefficient for this family of functions that varies with X. It does not contradict the additive $O(1/\log y)$ transition formula, and it does not give a necessary condition for cancellation of every fixed multiplicative function.

5 Residues of neighbouring cofactors

Suppose the observer sees O_M and also $R_1, R_2, R_3 \bmod M$. Let $V_X(M)$ be the optimal correlation with the original target Z. These extra residues constrain candidates inside each cell. We retain them in the local signed measure, so the comparison remains valid for a test chosen after observing the residues.

Proposition 5.1. *For each fixed M, $V_X(M) \asymp_{M,C} 1/\log y$. Uniformly up to $M \le L^{4/3+o(1)}$, one has $V_X(M) \le L^{-1+o(1)}$. For every fixed $\varepsilon > 0$, $V_X(M) = o(1)$ uniformly up to $M \le L^{2-\varepsilon}$.*

Proof. The lower bound follows by ignoring the extra observations and using Theorem 3.1. In a band $s \asymp U$, put $W_0 = X/U$. The first Cauchy factor counts $O(M^4 W_0/\log y)$ groups. The square sum retains equality of the origin residue and all three neighbouring cofactor residues. Pairs with coincident rational roots still cost $O(W_0/\log y)$ each. For the other pairs we need the signed Haar measure retaining all six neighbour residues and roughness. We shall bound its full ℓ^1 norm by $L^{-7+o(1)}$.

The residue-preserving transfer. Fix a desired saving $B > 0$ and assume $M \le L^2$. First separate the primes dividing M. If $M = 1$ this step is empty; otherwise put

$$Q = M^b, \qquad b = 3 + \left\lceil (B+4)\frac{\log L}{\log 2} \right\rceil, \qquad Q = X^{o(1)}.$$

Discard a pair at r when a neighbour has p-valuation at least $(b-1)v_p(M)$ for some $p \mid M$. Per pair the actual and Haar costs are $O_B(W_0 L^{-B-1} + \log(2M))$, using at most two solutions for each divisibility condition. For every remaining class of $r \bmod Q$, the complete M-supported part a of each neighbour is known, and $aM \mid Q$. Its quotient after division by a is therefore known modulo M. Dividing by the remaining invertible small part determines the rough residue. Roughness at primes of M is fixed on the class as well.

With this class held fixed, any bounded test is a function of the seven small parts supported outside M. Multivariate inversion has coefficients bounded by $\prod_i 2^{\omega(d_i)}$, as in Lemma 2.1. Truncating at $\prod_i d_i \le X^{1/10}$ costs $O(QX^{1/10}L^{O(1)})$ in rounding per pair. For the tails, reunite the disjoint classes before applying the moment bounds. Each s has $O(U)$ partners if congruence restrictions are temporarily dropped. The argument of (2.6) then gives aggregate error $O_{B,C}(W_0 U^2 L^{-B})$, uniformly over all bounded tests and all pairs.

On a finite residue space Ω, the dual identity for a signed measure is

$$\|\sigma\|_1 = \sum_{\omega\in\Omega} |\sigma(\omega)| = \sup_{F:\Omega\to[-1,1]} \left| \sum_{\omega\in\Omega} F(\omega)\sigma(\omega) \right|. \tag{5.1}$$

We use this full norm, not half of it. The preceding coefficient and tail bounds hold uniformly for a separate test on every pair. We may therefore take the suprema in (5.1) after transferring. No additional sum over residue vectors, and no factor M^6, is incurred.

Cancellation with the residues retained. Work on $((\mathbb{Z}/M\mathbb{Z})^\times)^6$. Let e be point mass at the identity, $e_{i,p}$ point mass at p on axis i, and u_i uniform mass on that axis. Retaining the small parts in this group is equivalent to retaining the rough residues: inversion and multiplication by fixed residues merely relabel the observations.

Put $w = \exp(L^\alpha)$ for fixed $0 < \alpha < 1$. At a good prime $w < p \le y$, the root zero is excluded and the six simple roots have negative sign at valuation one. The signed local measure ν_p satisfies

$$\nu_p = (1 - 7/p)e - p^{-1}\sum_{i=1}^{6} e_{i,p} + O_{\ell^1}(p^{-2}),$$

and its convolution logarithm is

$$\log_* \nu_p = -p^{-1}\left(7e + \sum_{i=1}^{6} e_{i,p}\right) + O_{\ell^1}(p^{-2}).$$

The logarithm converges absolutely since $\|\nu_p - e\|_1 = O(1/p)$. Let $H = \sum_{w<p\le y} 1/p$. Siegel–Walfisz and partial summation give

$$\left\| \sum_{\substack{w<p\le y \\ p \text{ good}}} \frac{e_{i,p}}{p} - Hu_i \right\|_1 = o(1).$$

Indeed, for $v \ge w$ the moduli satisfy $M \le (\log v)^{2/\alpha}$; a sufficiently large logarithmic saving absorbs the number of reduced classes. Omitting slope and determinant primes costs $O(L/w)$.

The group is abelian and $u_i * u_i = u_i$. Therefore

$$\exp_*(-Hu_i) = e + (e^{-H} - 1)u_i, \qquad \|\exp_*(-Hu_i)\|_1 \le 2.$$

Every other local measure has norm at most one. Exponentiating the sum of convolution logarithms bounds the full norm by $2^6 e^{-7H+o(1)}$. Since $H = (1 - \alpha + o(1)) \log L$, taking arbitrarily small fixed $\alpha > 0$ gives $L^{-7+o(1)}$.

Summing over multiplier bands. For fixed M, the resulting per-band estimate is $O_{M,C}(L^{-4+o(1)} + U^{-1/2}/\log y)$. Summing the bands, with the small-part tail as before, gives $O_{M,C}(1/\log y)$. For growing M, first bound $s \le 2L^{12}$ by the upper sieve, at cost $O(\log L/\log y)$. On larger bands rough r is coprime to M, and equality of origin residues forces $s \equiv t \pmod M$. There are $O(U^2/M)$ generic pairs and $O(U)$ exceptional pairs. Choosing B large absorbs the transfer error after these restrictions. The Cauchy factors and summation give

$$V_X(M) \ll_C \frac{\log L}{\log y} + M^{3/2} L^{-3+o(1)} + \frac{M^2 L^{-6}}{\log y}. \tag{5.2}$$

The two uniform ranges follow by substitution. □

The exponents 4/3 and 2 describe sufficient ranges established by this estimate. No converse or optimality assertion for these exponents is made.

6 Reconstruction from unlabelled products

The labelled pair (R_0, R_1) fixes n modulo R_0R_1, since the two cofactors divide consecutive integers. Large labelled quotients therefore identify the row by the Chinese remainder theorem. Their product erases the assignment of each prime factor to a site.

Almost-everywhere recovery from this unlabelled product follows from an earlier polynomial-ratio estimate of Wang and Xu [20, Lemma 2.7]. We give the deduction, then retain the special square identities to obtain a more precise collision count with the size of the removed factor left as a parameter.

Theorem 6.1 (Reconstruction from products)**.** *Let $A(n) = an + b$ and $B(n) = cn + d$ be fixed affine forms with positive integer slopes, integer intercepts and $ad - bc \ne 0$. The product $R_y(A(n))R_y(B(n))$ identifies n within the entire dyad, except on $O_{A,B,H,C}(XL^{-H})$ rows for every fixed $H > 0$. The same conclusion holds for $\prod_{i=0}^{3} R_i$.*

Deduction from Wang–Xu. Let $P = AB$, or let $P(n) = n(n+1)(3n+1)(3n+2)$, and put $S(n) = S_y(P(n))$, $T(n) = R_y(P(n))$. Complete multiplicativity of S_y and R_y makes $T(n)$ the observed product. If $S(n) \le Y$ and $T(n') = T(n)$ for rows in the dyad, write

$$\frac{P(n')}{P(n)} = \frac{S(n')}{S(n)} = \frac{u}{v}, \qquad (u, v) = 1. \tag{6.1}$$

Then $v \le Y$ and $u \ll_P Y$, because P varies by a bounded factor on a large dyad. There is no restriction on $S(n')$. The ratio $u = v$ gives $n = n'$ by eventual strict monotonicity.

Wang–Xu's lemma states, for a polynomial P of degree d_P with at least two distinct complex roots, that

$$\sum_{1 \le a < b \le T} \#\{x, z \le N : aP(z) = bP(x)\} \ll_{d_P, \varepsilon} T^2 N^{1/2+\varepsilon}.$$

Both polynomials here meet the root condition. If necessary, a fixed translation makes their values positive on all positive integers. Interchanging the two rows when needed and taking $T \asymp_P Y$, $N \asymp X$ bounds the distinct-ratio pairs in (6.1) by

$$O_{P,\varepsilon}(Y^2 X^{1/2+\varepsilon}). \tag{6.2}$$

The small-part moment bound, applied to each injective affine form and combined by Hölder, gives

$$\#\{n \in (X, 2X] : S(n) > X^{1/10}\} \ll_{P,C,H} XL^{-H}$$

for every fixed $H > 0$. Take $Y = X^{1/10}$ and any fixed $0 < \varepsilon < 3/10$ in (6.2). The power-saving collision term is absorbed by XL^{-H} for every fixed H. □

6.1 The sharper count supplied by a square identity

For the two-form product,

$$4acP(n) + (ad - bc)^2 = (2acn + ad + bc)^2.$$

For the four-form product, put $v_0 = n(n+1)$. Then $P = v_0(9v_0 + 2)$ and

$$9P(n) + 1 = (9n(n+1) + 1)^2. \tag{6.3}$$

Identity (6.3) also underlies *Small roots of split quadratic congruences and fundamental units in a quartic family* [15, (1.3) and Lemma 3.1]. There the observation is the complete squarefree kernel, and the argument concerns units and separation of indices. Here the observation removes small-prime valuations; the ratio in (6.1) is the additional datum needed to count its collisions.

Lemma 6.1 (Collision count with a bound on the removed factor)**.** *For either fixed polynomial in Theorem 6.1, and $1 \le Y \le X$,*

$$\#\left\{ n \in (X, 2X] : \begin{array}{l} S(n) \le Y,\ \exists n' \in (X, 2X],\ n' \ne n, \\ T(n') = T(n) \end{array} \right\} \ll_P Y^2 \log^2(2Y) \log(2X). \tag{6.4}$$

No bound on the removed factor $S(n')$ of a competing row is required.

Proof. Write $U(n)^2 = kP(n) + e$ using the corresponding identity, where $e > 0$ is fixed and U is positive and strictly increasing on large dyads. The ratio in (6.1) has $v \le Y$, $u \ll_P Y$ and $u \ne v$. Write $uv = Dc_0^2$ with D squarefree. Eliminating P gives the exact norm equation

$$(uU(n))^2 - D(c_0 U(n'))^2 = u(u - v)e =: K \ne 0. \tag{6.5}$$

If $D = 1$, factoring the difference of squares gives $O(\tau(|K|))$ possibilities.

If $D > 1$, the positive algebraic integer $\alpha = uU(n) + c_0U(n')\sqrt{D}$ generates an integral ideal of norm $|K|$ in the maximal order of $\mathbb{Q}(\sqrt{D})$. A split rational prime of exponent a contributes at most $a + 1$ choices of ideal exponents; an inert or ramified prime contributes at most one. There are therefore at most $\tau(|K|)$ such ideals.

Two positive generators of the same ideal and the same norm have a norm-one unit as their ratio. If this ratio ε exceeds one, its integral trace satisfies $\varepsilon + \varepsilon^{-1} \geq 3$, and hence $\varepsilon \geq (3 + \sqrt{5})/2$. All relevant generators lie in $[1, X^{O_P(1)}]$ for $Y \leq X$, so each ideal has $O_P(\log(2X))$ relevant generators, uniformly in D. Their two coefficients determine the ordered pair (n, n'). This is the ideal-and-unit method for generalized Pell equations [2, Theorem 1.1 and its proof]; the lower bound on a norm-one unit makes its required uniformity explicit.

Finally,

$$\sum_{u \ll_P Y} \sum_{\substack{v \leq Y \\ v \neq u}} \tau(|u(u-v)e|) \leq \tau(e) \sum_{u \ll_P Y} \tau(u) \sum_{\substack{v \leq Y \\ v \neq u}} \tau(|u-v|)$$
$$\ll_P Y^2 \log^2(2Y),$$

using $\sum_{n \leq t} \tau(n) \ll t \log(2t)$. This proves (6.4). □

The bound (6.4) improves the input (6.2) for these specific families. With $Y = X^{1/10}$ and the same smooth-part tail, it gives a second proof of Theorem 6.1.

6.2 The reflected product

For integers N with $3X \leq N \leq 4X$, the observation $R_y(n)R_y(N-n)$ recovers the unordered pair $\{n, N-n\}$, except on $O_{H,C}(XL^{-H})$ rows, uniformly in N and for every fixed $H > 0$. Use

$$U_N(n) = |2n - N|, \qquad U_N(n)^2 = N^2 - 4n(N-n).$$

The products $n(N-n)$ are comparable to X^2 on the dyad, so the same bounded-ratio reduction applies. Ratio one gives exactly the reflection symmetry. For a distinct ratio, the norm count gains a factor $\tau(N^2) = X^{o(1)}$, uniformly in this range. Each absolute square root has at most two origins, and a zero root gives at most a bounded number of pairs for a fixed ratio. The resulting collision count with $S(n) \leq X^{1/10}$ is $X^{1/5+o(1)}$; the moment bound for the two injective forms $n, N-n$ supplies the stated logarithmic exceptional-set estimate.

7 Three neighbouring small parts

We now change both the observation and the target. For a cutoff $2 \leq z = z(X) \leq 7X$, let

$$D_1 = S_z(n+1), \qquad D_2 = S_z(3n+1), \qquad D_3 = S_z(3n+2).$$

The observation keeps the site labels and all prime exponents. The target is the full sign $\lambda(n)$.

Theorem 7.1 (The subpolynomial cutoff)**.** *Writing*

$$P(X, z) = \sup_{|H| \leq 1} |\mathbb{E}_X[\lambda(n)H(D_1, D_2, D_3)]|,$$

one has, for every sequence of cutoffs in the stated range,

$$P(X, z) \longrightarrow 0 \quad \Longleftrightarrow \quad \frac{\log z}{\log X} \longrightarrow 0.$$

The necessity follows from complete revelation of one neighbour. The sufficiency requires comparison of whole observation cells, with their exact valuations and site labels. We first state the distribution estimate used to sum the signed sieve remainders.

Lemma 7.1 (Liouville in progressions with divisor weights)**.** *For every fixed $A > 0$ and nonnegative integer J,*

$$\sum_{q\le X^{0.31}} \tau(q)^J \max_{\substack{a \bmod q\\(a,q)=1}} \left| \sum_{\substack{X<n\le 2X\\ n\equiv a \pmod q}} \lambda(n) \right| \ll_{A,J} X(\log X)^{-A}. \tag{7.1}$$

The same holds on $(X/g, 2X/g]$ for each fixed positive integer g, with the constant also depending on g.

Proof. The classical Bombieri–Vinogradov estimate is stated explicitly for λ as well as μ in Shao [17, p. 961]. To pass from that exceptional-modulus formulation to a sum, split the moduli into bands $(Q/2, Q]$. At either endpoint $x = X, 2X$, the nonexceptional moduli contribute $O_B(XL^{-B})$ on a band, for any prescribed $B > 0$. There are $O_B(QL^{-B})$ exceptional moduli, whose uncentred sums are bounded trivially by $O(X/Q + 1)$ each. They give the same total bound. The centring term is absorbed by the prime number theorem for Liouville, $\sum_{n\le x} \lambda(n) \ll_D XL^{-D}$ for every fixed D, and $\sum_{q\le X^{0.31}} 1/\varphi(q) \ll L$. After summing the $O(L)$ bands and increasing the input saving,

$$\sum_{q\le X^{0.31}} E_q \ll_B XL^{-B}, \qquad E_q = \max_{(a,q)=1} \left| \sum_{\substack{X<n\le 2X\\ n\equiv a \pmod q}} \lambda(n) \right|.$$

The strict power margin below $X^{1/2}$ absorbs every fixed logarithmic loss in the distribution level.

Since $E_q \le X/q + 1$, divisor moments give, for a fixed K_J,

$$\sum_{q\le X^{0.31}} \tau(q)^{2J} E_q \le X \sum_{q\le X^{0.31}} \frac{\tau(q)^{2J}}{q} + \sum_{q\le X^{0.31}} \tau(q)^{2J} \ll_J XL^{K_J}.$$

Consequently Cauchy–Schwarz gives

$$\sum_q \tau(q)^J E_q \le \left(\sum_q E_q\right)^{1/2} \left(\sum_q \tau(q)^{2J} E_q\right)^{1/2} \ll_{B,J} XL^{(K_J-B)/2}.$$

Choose $B > 2A + K_J$. For fixed g, the same argument applies at $X/g, 2X/g$ with the same positive power margin. □

Proof of Theorem 7.1. If $D_1 > X$, then $D_1 = n + 1$, because every proper divisor of $n + 1 \le 2X + 1$ is at most X. Take $H = \lambda(D_1 - 1)$ on that event and zero elsewhere. Along any subsequence with $z \ge X^\theta$, where $0 < \theta \le 1$ is fixed, the smooth-number theorem on the shifted dyad gives lower limiting mass at least $\rho(1/\theta) > 0$ [7, Theorem 1.1]. This proves necessity, including for oscillating cutoffs.

Now assume $z = X^{o(1)}$. For an upper bound we may include primes 2 and 3 even if the original cutoff is smaller. Discard valuations at these primes exceeding a fixed depth k, at cost $O(2^{-k}) + O(1/X)$, and split into classes modulo $Q_0 = 6^{k+2}$. The Haar cost obeys the same

bound. For $p \geq 5$, the three neighbouring roots are distinct units. Let d be the product of all prime powers with $5 \leq p \leq z$ marked at the three sites. Under the integer and Haar laws,

$$\mathbb{E} \log d \ll \log(2z). \tag{7.2}$$

Indeed, sum $\log p$ times the divisibility probability over p^a. On integers the total rounding error is $O(z \log X/X)$, and the main term is bounded by a constant multiple of $\sum_{p \leq z} \log p/(p-1)$. Markov's inequality permits discarding $d > X^{1/10}$ with loss $o(1)$, for fixed k.

For a retained d, assigning each $p^a \| d$ to its site prescribes a single residue class modulo d. After substitution, one forbidden lift at each $p \mid d$ excludes the next power; at an unmarked prime there are three forbidden residues. Apply upper and lower fundamental-lemma sieves at level $X^{1/5}$ separately to the nonnegative populations $(1+\lambda(n))/2$ and $(1-\lambda(n))/2$. Their dimension is at most three, uniformly in d, and the ratio of the logarithmic level to $\log z$ tends to infinity. The main term for either population is half the Haar mass of the cell times $X(1+o(1))$ [4, Theorem 3.6]. For a single retained class modulo Q_0, that Haar mass is

$$\frac{1}{Q_0 d} \prod_{\substack{5 \leq p \leq z \\ p \mid d}} (1-1/p) \prod_{\substack{5 \leq p \leq z \\ p \nmid d}} (1-3/p).$$

The total mass of all retained cells is at most one.

We now verify that the arithmetic remainders may be summed over all cells. A sieve divisor $e \leq X^{1/5}$ gives classes modulo $q = Q_0 de \ll_k X^{3/10}$. If p divides both d and e, its extra occurrence enforces the next valuation: the modulus is de, rather than $[d,e]$. For fixed q, all divisor choices, site assignments and roots have multiplicity $O_k(\tau(q)^J)$ for one fixed J. Integer rounding therefore costs $O_k(X^{3/10} L^{O(1)}) = o_k(X)$.

The classes are reduced outside $2, 3$. Removing their common divisor $g \mid Q_0$ gives a reduced class for n/g, and $\lambda(n) = \lambda(g)\lambda(n/g)$ by complete multiplicativity. For fixed k the finitely many such g are covered by Lemma 7.1, since the moduli are eventually below $X^{0.31}$. It bounds all signed remainders by $O_{A,k}(XL^{-A})$. Summing the two nonnegative sieve bounds over the retained cells yields total absolute signed discrepancy $o_k(X)$. The discarded integer and Haar masses are $O(2^{-k}) + o(1)$. Taking first $X \to \infty$ with k fixed, then $k \to \infty$, proves $P(X,z) \to 0$.

Gorodetsky [5, Theorem 2.3 and Corollary 2.5] gives total-variation models for general sieve sequences with positive distribution level. The root assignments and the two Liouville populations are the extra information retained in the cell comparison just proved. □

7.1 Recovery when every neighbouring factorization is incomplete

The preceding necessity argument uses a completely smooth neighbour. There is also a positive-density mechanism where all three readings omit a prime, while every pair is shared by many rows. A finite example is $X = 4000$, $z = 11$, $n = 4034$:

$$(D_1, D_2, D_3) = (15, 49, 8), \qquad \operatorname{lcm}(15, 49, 8) = 5880 > X.$$

Every reading is a proper divisor of its argument. In $(4000, 8000]$, the pairs $(15, 49)$, $(15, 8)$ and $(49, 8)$ occur respectively two, five and three times, whereas the triple occurs only at 4034.

Theorem 7.2 (Three incomplete readings)**.** *Fix $0 < \theta \leq 1/120$, put $z = X^\theta$, and let $0 < \kappa < 59/5$. There is a set E of $\gg_{\theta,\kappa} X$ rows recognizable from (D_1, D_2, D_3). On E, the triple determines n in the entire dyad and every neighbour has a prime factor above z. Outside $o_{\theta,\kappa}(X)$ rows of E, each of the three labelled pairs is shared by at least $X^{\kappa\theta}$ rows of E.*

The proof uses the ineffective form of Harper's distribution theorem. Accordingly, an effective threshold for X is not asserted.

Proof. Choose $\kappa < \sigma < 59/5$ and $2 < \nu < 13.8 - \sigma$, with $\nu \le 8$. Define E by requiring $n \equiv 1 \pmod{24}$, all readings incomplete,

$$\operatorname{lcm}(D_1, D_2, D_3) > X, \qquad D_i D_j \le X/z^{\sigma} \quad (i \ne j). \tag{7.3}$$

Divisibility by the three readings prescribes a unique class modulo their least common multiple, because $3 \nmid D_2 D_3$. On an actual observation the congruences are compatible. A modulus exceeding X leaves at most one row in the dyad; after recovering it, all other requirements can be checked. Thus E is recognizable. It remains to construct many such rows.

Prescribed prime factors. Write $n + 1 = 2pd$, with p prime in $(z^{32}, z^{33}]$, d z-smooth, and $pd \equiv 1 \pmod{12}$. Set

$$U_2 = (3pd - 1)/2, \qquad U_3 = 6pd - 1.$$

Both are coprime to six, and $U_3 - 4U_2 = 1$. Require $B_i r_i \mid U_i$, using the following disjoint bands:

Object	Construction	Exponent of z
p	one prime in $(z^{32}, z^{33}]$	$(32, 33]$
B_2	twenty distinct primes in $(z^{0.87}, z^{0.88}]$	$(17.4, 17.6]$
B_3	twenty distinct primes in $(z^{0.90}, z^{0.91}]$	$(18, 18.2]$
r_2	one prime in $(z^{1.01}, z^{1.02}]$	$(1.01, 1.02]$
r_3	one prime in $(z^{1.03}, z^{1.04}]$	$(1.03, 1.04]$
m	$12B_2B_3r_2r_3$	≤ 37.86, apart from 12
$m\ell^a$	$\ell^a \le z^2$	≤ 39.86, apart from 12

At fixed p, each choice of the unordered marked primes prescribes one reduced class of d modulo $m = 12B_2B_3r_2r_3$. The bands recover the choices from m, and p is invertible modulo m.

Put $Y = X/(2p)$. The number N_p of smooth d coprime to six in the exact interval $((X+1)/(2p), (2X+1)/(2p)]$ is

$$N_p = (1 + o_\theta(1))\frac{Y}{3}\,\rho\left(\frac{\log Y}{\log z}\right) \asymp_\theta Y.$$

This follows by inclusion–exclusion at $2, 3$ and the compact-uniform smooth-number asymptotic [7, Theorem 1.1]. The forty primes in B_2B_3 each remove at most $O(Y/z^{0.87} + 1)$ rows from this population. The primes in r_2r_3 exceed z and cannot divide d. Thus the coprime population in the progression main term is $N_p(1 + o_\theta(1))$, uniformly in the marked moduli.

The distribution margin. Use the ineffective form of Harper's Theorem 1 [6], with its maximum over reduced classes, at both interval endpoints. The largest needed modulus is $\mathcal{Q} \le 12z^{39.86}$, allowing one further prime power $\ell^a \le z^2$. Since $Y \ge \frac{1}{2}X^{1-33\theta}$,

$$\frac{1 - 33\theta}{2} - 39.86\theta = \frac{1}{2} - 56.36\theta \ge \frac{91}{3000} > 0. \tag{7.4}$$

The smooth population is comparable to Y for fixed θ. Consequently Harper's term $\sqrt{\Psi(Y, z)}\,\mathcal{Q}(\log Y)^{7/2}$ is smaller than Y by a fixed power of X, up to logarithms. The

theorem's remaining terms, in its ineffective form, give arbitrary logarithmic saving. Hence the summed maximal progression discrepancy is $O_{B,\theta}(YL^{-B})$ for every fixed B.

It follows that the total number of marked configurations is $(1+o(1))\mathcal{A}$, where

$$\mathcal{A} = \sum_p N_p \sum_m \frac{1}{\varphi(m)} \asymp_\theta X. \tag{7.5}$$

Every band has positive limiting reciprocal prime mass. For twenty distinct unordered primes, its contribution tends to the twentieth power of that mass divided by 20!; repeated choices have vanishing weight. The sum of progression errors over p is $O_{B,\theta}(XL^{-B})$, because $\sum_p 1/p \ll 1$ on the prescribed band. A row has $O_\theta(1)$ markings: each marked prime is at least $z^{0.87}$ and each affected integer is $O(X)$.

Unprescribed small factors. Discard rows on which a prime of B_i divides U_i twice. Their number is bounded, even after summing over integers q, by

$$O\left(\sum_{z^{0.87}<q\ll\sqrt{X}} (X/q^2+1)\right) = o(X).$$

Also discard rows where some prime $\ell \le z$ has a power greater than z^2 dividing a U_i. For each ℓ, use the first such power; its cost is $O(X/z^2+1)$, and the total is $o(X)$. On the fixed residue class, these divisibility conditions are affine congruences of bounded content. Bounded marking multiplicity transfers both losses to $o_\theta(X)$ configurations.

For surviving configurations put $E_i' = S_z(U_i)/B_i$. Every contribution $\log \ell$ to $\log E_i'$ comes from $\ell^a \mid U_i$ with $5 \le \ell \le z$, $\ell^a \le z^2$ and $\ell \nmid m$. A prime marked at the other site is excluded by $(U_2, U_3) = 1$; the square exclusions remove further occurrences of primes marked at the same site. For fixed p, m, each remaining condition adds one reduced class of d modulo ℓ^a. Its main term is at most $N_p/(\varphi(m)\varphi(\ell^a))$.

A combined modulus has at most forty-three prime bases outside $2, 3$. Choosing the added base determines its exponent and the original marked modulus, so regrouping the progression errors has multiplicity at most forty-three. The factor $\log \ell \le \log z$ is absorbed by an additional logarithmic saving in Harper's theorem. Thus at each site the sum of $\log E_i'$ is at most

$$\mathcal{A} \sum_{\substack{5\le\ell\le z\\ \ell \text{ prime}}} \frac{\ell \log \ell}{(\ell-1)^2} + o_\theta(X) = (1+o_\theta(1))\mathcal{A} \log z.$$

Markov's inequality leaves at least $(1-2/\nu-o(1))\mathcal{A}$ configurations with $E_2'E_3' \le z^\nu$, a positive fraction since $\nu > 2$.

From the marks to the readings. For these configurations,

$$D_1 = 2d, \qquad D_2 = 4B_2E_2', \qquad D_3 = B_3E_3'.$$

The B_i are coprime to each other and to D_1. Consequently

$$\operatorname{lcm}(D_1, D_2, D_3) > Xz^{17.4+18-33} = Xz^{2.4} > X.$$

The primes p, r_2, r_3 make the readings incomplete. Using $D_1 \le (2X+1)/z^{32}$ gives

$$\begin{aligned} D_1D_2 &\ll Xz^{17.6+\nu-32},\\ D_1D_3 &\ll Xz^{18.2+\nu-32},\\ D_2D_3 &\le 4z^{35.8+\nu}. \end{aligned}$$

The first two are at most X/z^σ for large X because $\nu+\sigma<13.8$; the third is so because $(35.8+\nu+\sigma)\theta<1$. Removing bounded marking multiplicity proves $|E|\gg_{\theta,\kappa} X$.

For each labelled pair, the number of possible values with product at most X/z^σ is $O(X\log X/z^\sigma)$, by summing $\lfloor X/(az^\sigma)\rfloor$ over a. Fibres containing fewer than z^κ rows of E cover only $O(X\log X/z^{\sigma-\kappa})=o(X)$ rows. Sum this bound over the three pairs to obtain the conclusion. □

For any prescribed sign on the dyad, evaluating it at the recovered row and returning zero off E gives correlation $|E|/X$. The multiplicity assertion for pairs establishes ambiguity of the integer; it gives no sign balance inside a pair cell. The stated range of θ,κ comes from the displayed prime bands, the distribution margin and the extra-factor moment bound.

8 Returning to the full sign and changing the sampling law

Small-part parity and the full quartic sign are different targets. Write $C_4(n)=Z(n)\prod_{i=0}^3\lambda(R_i)$. For any observation O, any O-measurable sign d, and any underlying probability measure,

$$|\mathbb{E}C_4(n)|\le\sup_{|W|\le1}|\mathbb{E}[ZW(O)]|+2\mathbb{P}\left(\prod_{i=0}^3\lambda(R_i)\ne d(O)\right). \tag{8.1}$$

This follows by adding and subtracting $Zd(O)$. The first term is controlled under the dyadic sampling laws treated above, for the observations specified in the statements: O_M, or its enrichment by the neighbouring cofactor residues in the proved ranges. The second term requires prediction of the remaining signs under that same observation and measure.

For example, observing the entire product $R_0R_1R_2R_3$ makes its sign known. But Theorem 6.1 also makes the row recoverable with probability $1-o(1)$, so the optimal correlation for predicting Z from that product is $1-o(1)$. The decomposition (8.1) therefore supplies no small upper bound from this observation. The fixed-form quantity in *An explicit bound below one half for a quartic Liouville correlation* [11] is moreover averaged with a different, logarithmic measure.

For a sampled-prime experiment, let I be a finite interval of integers in $[3,\infty)$ with harmonic law $\mathbb{P}(v)=1/(vH_I)$, where $H_I=\sum_{v\in I}1/v$. Use the four arguments $(v,v-1,3v-1,3v-2)$ and their full Liouville signs. At two chosen sites a prime label $p\mid m$ may be drawn with probability $v_p(m)\log p/\log m$, including repeated draws with replacement. These probabilities sum to one, so marginalizing the draws preserves the starting law of v. If U,D are the products of the two observed-site signs and the two remaining signs, then for a transcript O and any sign predictor $d(O)$,

$$|\mathbb{E}_I(UD)|\le\sup_{|h|\le1}|\mathbb{E}_I[Uh(O)]|+2\mathbb{P}_I(D\ne d(O)).$$

The expectations include the draw randomness. This follows from $UD=Ud(O)+U(D-d(O))$; the general covariance comparison appears in [13, Section 22]. The observation, target and sampling law must all be kept fixed when applying either inequality.

Finite calculations. The source archive includes `verify_finite.py` for the worked examples, cell decisions, four cofactor constants and square identities. The program `verify_lists.py` enumerates complete residue classes by their least candidate, tests the local signs at good primes, and checks the prime-density correction against the cell integrals.

AI assistance. OpenAI's ChatGPT assisted with the arithmetic models and the original proofs. OpenAI's Codex assisted with the uniform Dickman estimate, the decision and multiplicity bounds, the bibliography and the English exposition. ChatGPT also assisted with the aggregate transfer, the weighted progression arguments and the finite calculations. Codex assisted in extending the signed cell law, deriving the prime-weight correction and preparing its reproducible checks.

References

[1] J. Cassaigne, S. Ferenczi, C. Mauduit, J. Rivat and A. Sárközy, *On finite pseudorandom binary sequences III: The Liouville function, I*, Acta Arith. 87 (1999), no. 4, 367–390. doi:10.4064/aa-87-4-367-390.

[2] B. Conrad, *Generalized Pell equation*, Math 154, Stanford University, Theorem 1.1.

[3] H. Daboussi and A. Sárközy, *On the correlation of the truncated Liouville function*, Acta Arith. 108 (2003), no. 1, 61–76. doi:10.4064/aa108-1-6.

[4] K. Ford, *Sieve methods*, lecture notes, University of Illinois, Spring 2023, Theorem 3.6. `https://ford126.web.illinois.edu/sieve2023.pdf`.

[5] O. Gorodetsky, *A Kubilius model for sieve-theoretic sequences*, Anal. Math. 52 (2026), no. 3, 999–1022. doi:10.1007/s10476-026-00178-w.

[6] A. J. Harper, *Bombieri–Vinogradov and Barban–Davenport–Halberstam type theorems for smooth numbers*, 2012, Theorem 1. doi:10.48550/arXiv.1208.5992.

[7] A. Hildebrand and G. Tenenbaum, *Integers without large prime factors*, J. Théor. Nombres Bordeaux 5 (1993), no. 2, 411–484. doi:10.5802/jtnb.101.

[8] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, AMS Colloquium Publications 53, American Mathematical Society, 2004, Chapter 17. doi:10.1090/coll/053.

[9] A. Jaramillo and X. Yang, *Approximation of smooth numbers for harmonic samples: a Stein method approach*, 2025, Lemma 5.2. doi:10.48550/arXiv.2409.16527.

[10] R. G. Pinsky, *A natural probabilistic model on the integers and its relation to Dickman-type distributions and Buchstab's function*, in *Probability and Analysis in Interacting Physical Systems*, Springer Proceedings in Mathematics & Statistics 283, Springer, 2019, 267–294. doi:10.1007/978-3-030-15338-0_10.

[11] B. Pouly, *An explicit bound below one half for a quartic Liouville correlation*, 2026. doi:10.5281/zenodo.24443184.

[12] B. Pouly, *Observations and Shift Decoding in Divisor Incidences*, Zenodo, 2026. doi:10.5281/zenodo.24578881.

[13] B. Pouly, *Reconstruction and sign prediction from sampled prime factors*, 2026. doi:10.5281/zenodo.24443266.

[14] B. Pouly, *Shifted-cofactor cancellation and a multiplicity threshold for product representations*, Zenodo, 2026. doi:10.5281/zenodo.24578832.

[15] B. Pouly, *Small roots of split quadratic congruences and fundamental units in a quartic family*, 2026. doi:10.5281/zenodo.24443299.

[16] A. Sedunova, *A partial Bombieri–Vinogradov theorem with explicit constants*, Publ. Math. Besançon Algèbre Théorie Nr. (2018), 101–110, Theorem 1.1. doi:10.5802/pmb.24.

[17] X. Shao, *Gowers norms of multiplicative functions in progressions on average*, Algebra Number Theory 11 (2017), no. 4, 961–982. doi:10.2140/ant.2017.11.961.

[18] G. Tenenbaum, *Crible d'Ératosthène et modèle de Kubilius*, in K. Győry, H. Iwaniec and J. Urbanowicz (eds.), *Number Theory in Progress*, vol. 2, de Gruyter, 1999, 1099–1130. doi:10.1515/9783110285581.1099. Author's corrected text.

[19] G. Tenenbaum, *Integers with a large friable component*, Acta Arith. 124 (2006), no. 3, 287–291. doi:10.4064/aa124-3-6.

[20] V. Y. Wang and M. W. Xu, *Paucity phenomena for polynomial products*, Bull. Lond. Math. Soc. 56 (2024), no. 8, 2718–2726. doi:10.1112/blms.13095. Lemma 2.7.