hexagon:2611.01049v2

Additive basis thresholds for random prime semigroups

Roy van Rijn

Abstract

Admit 2 and select each odd prime q independently with probability $\min(1, \lambda q^{-\gamma})$. We study sums from the multiplicative semigroup generated by the selected primes. At $\lambda = 1$, its least asymptotic basis order is almost surely $\lfloor (1-\gamma)^{-1} \rfloor + 1$ for $0 < \gamma < 1$. On the critical line $\gamma = 1 - 1/k$, the multiplier threshold for cofiniteness of the k-fold sumset is $1 + 1/k$. At equality the probability of cofiniteness is strictly between zero and one. Almost every realization at equality admits a finite addition of primes that makes its k-fold sumset cofinite and a finite removal of optional primes that gives infinitely many gaps. The proof combines a lower sieve and a prime-support overlap estimate for representations with a stopped martingale that compensates surviving gaps.

1 Results

Fix $\gamma > 0$ and $\lambda > 0$. The random set of primes P contains 2, and every odd prime q belongs to P with probability $p_q = \min(1, \lambda q^{-\gamma})$, independently. An *optional prime* has $0 < p_q < 1$. Let C be the multiplicative semigroup generated by P, including 1. Write kC for sums of exactly k elements of C, allowing repetitions. Let $h(C)$ be the least k for which kC is cofinite in $\mathbb{N}$, and put $h(C) = \infty$ if there is no such k. Write $A = A_{\omega,\lambda}$ for the positive random constant in the counting asymptotic proved in Lemma 2.1. All logarithms are natural.

Theorem 1 (Power-law threshold)**.** *In the model with $\lambda = 1$, for each fixed $0 < \gamma < 1$, almost surely*

$$h(C) = \left\lfloor \frac{1}{1-\gamma} \right\rfloor + 1.$$

For each fixed $k \geq 2$, kC is cofinite if $0 < \gamma < 1 - 1/k$ and has density zero if $\gamma > 1 - 1/k$, almost surely in either case. At $\gamma = 1 - 1/k$ its complement has almost surely lower density at least $\exp(-K_k A^k) > 0$, where $K_k = \Gamma(1 + 1/k)^k / k!$. For $\gamma \geq 1$, almost surely $h(C) = \infty$.

Theorem 2 (Critical multipliers)**.** *Fix $k \geq 2$ and $\gamma = 1 - 1/k$. If $0 < \lambda < 1$, kC has density zero almost surely. If $\lambda > 1$, kC has density one almost surely. Cofiniteness holds almost surely for $\lambda > 1 + 1/k$, and infinitely many gaps occur almost surely for $\lambda < 1 + 1/k$. At exact equality,*

$$0 < \mathbb{P}(kC \text{ is cofinite}) < 1.$$

More precisely, at equality let $\mathcal{E}_y$ be the event that all optional primes through y are excluded. Then

$$\lim_{y \to \infty} \mathbb{P}(kC \text{ has infinitely many gaps} \mid \mathcal{E}_y) = 1.$$

Corollary 3 (Finite changes at equality)**.** *Under the parameters $\gamma = 1 - 1/k$, $\lambda = 1 + 1/k$, almost every realization has both of the following finite modifications. Adding finitely many optional primes gives a cofinite k-fold sumset. Removing finitely many optional primes gives infinitely many gaps. If $\mathcal{I}_y$ denotes the event that every prime through y is selected, then*

$$\lim_{y\to\infty} \mathbb{P}(kC \text{ is cofinite} \mid \mathcal{I}_y) = 1, \qquad \lim_{y\to\infty} \mathbb{P}(kC \text{ is cofinite} \mid \mathcal{E}_y) = 0.$$

The modified semigroups can thus have least basis order k and $k+1$, respectively. The selected primes have the same counting asymptotic under either finite change.

At $k = 2$, the first theorem gives infinitely many gaps at $\gamma = 1/2, \lambda = 1$, with a positive lower density of gaps. Increasing the multiplier makes the gaps have density zero as soon as $\lambda > 1$. Cofiniteness has strictly intermediate probability when $\lambda = 3/2$, and holds almost surely above that value.

Erdős and Rényi [3] studied additive representations in random sequences formed by selecting individual integers. Erdős and Tetali [4] constructed an asymptotic basis with $\Theta(\log n)$ representations by k distinct summands using independent integer choices with probabilities of order $n^{1/k-1}(\log n)^{1/k}$. Our count on $\gamma = 1 - 1/k$ has order $x^{1/k}(\log x)^{\lambda-1}$, so this logarithmic power occurs at $\lambda = 1 + 1/k$. The dependence in the prime model enters through common prime divisors of summands. A change to one prime affects all semigroup elements divisible by it.

For coverage, we expose a pool of smaller primes and use their products as multipliers. A lower sieve supplies decompositions whose remaining factors have no small prime divisor. To apply Janson's inequality, we solve for two summands when counting an overlap. This gives a power saving for a proper overlap, and deleting a small family of linear relations bounds the contribution from identical supports. For gaps at equality, we expose primes one at a time. A surviving target receives a compensating weight whenever the current prime could complete a representation. Freezing the weight before it becomes too large controls variance. A change of measure turns the weight lost to freezing into an estimate for smooth semigroup elements.

2 Counting and moments

For $0 < \gamma < 1$, put $\alpha = 1 - \gamma$ and write $N(x) = |C \cap [1, x]|$.

Lemma 2.1 (Semigroup count)**.** *For every fixed $\lambda > 0$, almost surely*

$$N(x) \sim Ax^{\alpha}(\log x)^{\lambda-1}, \qquad A = \frac{G(\alpha)}{\alpha\Gamma(\lambda)} \in (0, \infty), \tag{1}$$

where the Dirichlet series on $\Re s > \alpha$ is $\zeta(s+\gamma)^{\lambda}G(s)$ and G is holomorphic and nonvanishing on $\Re s > \alpha/2$.

Proof. Let X_q be the selection indicator, with $X_2 = 1$. The variance of the centered prime count through u is $O(u^{\alpha})$. Kolmogorov's maximal inequality on dyadic intervals and Borel–Cantelli yield, almost surely for every $\epsilon > 0$,

$$\sum_{q\le u}(X_q - \mathbb{E}X_q) = O_{\omega,\epsilon}(u^{\alpha/2+\epsilon}).$$

Partial summation makes the corresponding centered linear terms of the logarithmic Euler product converge locally uniformly on $\Re s > \alpha/2$. Away from the finitely many capped or fixed primes these terms are $(X_q - \lambda q^{-\gamma})q^{-s}$. The terms of prime-power degree at least two converge absolutely on compact subsets almost surely. Their expected absolute majorants converge for each $\sigma > \alpha/2$ since $\gamma + 2\sigma > 1$ and $2(\gamma + \sigma) > 1$. The finitely many remaining factors have holomorphic logarithms on $\Re s > 0$. Exponentiating gives G. The logarithm is real at α. The original series converges absolutely for $\Re s > \alpha$ by taking expectations of the prime majorants.

The branch of the zeta power comes from its Euler product. For $\sigma > 1$, the inequality $3 + 4\cos v + \cos(2v) \geq 0$ applied in that product gives

$$|\zeta(\sigma)|^3 |\zeta(\sigma + it)|^4 |\zeta(\sigma + 2it)| \geq 1.$$

A zero at $1 + it$ with $t \neq 0$ would force the left side to zero as $\sigma \downarrow 1$, since the pole at 1 is simple. Hence the branch continues locally across $\Re s = \alpha$ away from α. Near α, $(s - \alpha)\zeta(s + \gamma)$ is holomorphic and equals 1 at α. Apply Delange's theorem [2, Theorem III] to the Laplace transform $\zeta(s + \gamma)^\lambda G(s)/s$ of $N(\mathrm{e}^u)$. The transform has exponent λ at $s = \alpha$ and leading coefficient $G(\alpha)/\alpha$, giving (1). □

Lemma 2.2 (Moments and cumulative representations)**.** *For $\lambda \geq 1$ and each fixed integer $l \geq 1$,*

$$\mathbb{E}N(x)^l \ll_l x^{l\alpha}(1 + \log x)^{l(\lambda-1)} \qquad (x \geq 1). \tag{2}$$

The same estimate is uniform if any one prime is forced to be selected. Let $r_k(n)$ count unordered k-term representations, with repetitions allowed. Almost surely,

$$\sum_{n \leq x} r_k(n) \sim \frac{\Gamma(1 + \alpha)^k}{k!\Gamma(1 + k\alpha)} A^k x^{k\alpha} (\log x)^{k(\lambda-1)}. \tag{3}$$

For $\lambda \geq 1$, the normalized counts in (1) and (3) also converge in every finite L^r.

Proof. The local series for simultaneous membership of l integers in C is $1 + p_q(\prod_i (1 - z_i)^{-1} - 1)$, with $p_2 = 1$. Multiply it by $\prod_i (1 - q^{-\gamma} z_i)^\lambda$. The linear coefficients cancel except at finitely many primes. At $z_i = q^{-\alpha}$, the absolute coefficient norm of the nonconstant part is $O_l(q^{-\gamma-2\alpha})$. The remaining Euler product has finite absolute coefficient norm in this weight. Forcing one prime changes just one factor to a uniformly bounded norm. The separated factors in each coordinate have coefficients $\tau_\lambda(b) b^{-\gamma}$, with $\sum_b \tau_\lambda(b) b^{-s} = \zeta(s)^\lambda$. Delange's theorem gives

$$\sum_{b \leq v} \tau_\lambda(b) b^{-\gamma} \ll v^\alpha (1 + \log v)^{\lambda-1}. \tag{4}$$

Convolution and the absolute norm give (2), using $\lambda \geq 1$.

By (1) the counting measure on a/x, $a \in C$, normalized by $x^\alpha (\log x)^{\lambda-1}$ and restricted to $[0, 1]$, converges weakly to $A\alpha u^{\alpha-1}\,\mathrm{d}u$. The product measures converge on $\sum_i u_i \leq 1$, whose boundary has measure zero. The simplex integral gives $\Gamma(1 + \alpha)^k/\Gamma(1 + k\alpha)$. Repeated coordinates contribute $O_k(N(x)^{k-1})$. Dividing the ordered count by $k!$ proves (3). Larger choices of l in (2) give uniform integrability of each fixed power of both normalized counts. □

We use one consequence at $\alpha = 1/k$ and $\lambda = 1 + 1/k$. For $\mathcal{E}_y$ as in Theorem 2,

$$\mathbb{E}[A^k \mid \mathcal{E}_y] \ll_k (\log y)^{-(k+1)}. \tag{5}$$

Indeed,

$$G(\alpha) = \prod_q (1 - 1/q)^{\lambda} (1 - q^{-\alpha})^{-X_q}.$$

The factors through y on $\mathcal{E}_y$ contribute $O((\log y)^{-\lambda})$ by Mertens' product estimate [6, Theorems 5–8]. The product above y is independent of the pattern. For each fixed real r, its r-th moment is bounded uniformly in y and tends to 1, since its expected local factors are $1 + O_r(q^{-1-\alpha})$. The corresponding bounds at $2r$ justify the infinite product by uniform integrability.

For $\gamma_k = 1 - 1/k < \gamma < 1$, (1) gives $N(x)^k = o(x)$. On $\gamma = \gamma_k$ the same is true if $\lambda < 1$. Both imply density zero for kC. If $\gamma \geq 1$, for every fixed $s > 0$ the expected Euler product of $\sum_{a \in C} a^{-s}$ converges, since $\sum_q p_q q^{-s} < \infty$. Taking $s < 1/k$ again gives $N(x)^k = o(x)$ almost surely.

3 Coverage from smaller selected primes

Fix $k \geq 2$ and $0 < \gamma \leq 1 - 1/k$, and write

$$u = k - 1 - k\gamma, \qquad L(n) = \frac{(\log n)^{k(\lambda-1)}}{(\log \log n)^{k\lambda}}.$$

We estimate the conditional probability that a target has no representation after exposing a fixed power of that target. All constants in this section may depend on the fixed parameters.

3.1 Choosing multipliers

Let $z = n^{\theta}$ and take level $n^{1/2}$. We choose the fixed $\theta > 0$ small enough that the ratio $\log(n^{1/2})/\log z = 1/(2\theta)$ is in the range of the lower-bound fundamental lemma [1, Theorem 3.2, specialized to $\mathbb{Q}$]. Choose $0 < \eta < \theta$ so that

$$k\eta < 1/4, \qquad (2k + k\gamma)\eta < 1, \qquad k\eta\gamma < \theta/k.$$

Fix a sufficiently small $0 < \tau < \eta$, put $U = n^{\tau}$, and let $Y = (\log n)^B$ with $B\alpha > 1$. Expose all primes through z. Take pairwise coprime squarefree multipliers $d_1, \ldots, d_k \leq n^{\eta}$ from the selected primes in $(Y, U]$ that do not divide n. Write $D = \prod_i d_i$. We will use the bound

$$\sum_{(d_i)} D^{-\alpha} \gg \frac{(\log n)^{k\lambda}}{(\log \log n)^{k\lambda}} \quad \text{almost surely for all large } n. \tag{6}$$

Without the size restrictions, the sum is $\prod_q (1 + kq^{-\alpha})$ over the allowed pool. The centered series $\sum_q (X_q - p_q) q^{-\alpha}$ and the same series with an extra $\log q$ both converge almost surely, by independence and summable variances. Also $\sum_q X_q q^{-2\alpha} < \infty$ almost surely. The omitted divisors of n contribute at most $(\log n)/(Y^{\alpha} \log Y) = o(1)$ to the first sum. Taking logarithms and using Mertens' reciprocal-prime estimate [6, Theorem 5], the unrestricted product is

$$\exp(k\lambda(\log \log U - \log \log Y) + o(1)) = (\log U / \log Y)^{k\lambda}(1 + o(1)).$$

Normalize the weights by this product. Each allowed prime is independently assigned to at most one multiplier, with probability $q^{-\alpha}/(1 + kq^{-\alpha})$ for each specified assignment. The expected

$\log D$ in this auxiliary distribution is at most $k\sum_{q\le U} X_q q^{-\alpha}\log q = O_{k,\lambda}(\log U)+O_\omega(1)$, using Mertens' weighted estimate [6, Theorem 6]. Choose τ so that this is at most $(\eta/2)\log n$ for all large n. Markov's inequality retains at least half the product weight with all $d_i \le n^\eta$, proving (6). In using this bound we can fix a deterministic lower constant and define the good exposures by that inequality. They occur eventually almost surely.

3.2 Counting rough tuples

For fixed multipliers consider $n=\sum_i a_i$ with $a_i = d_i b_i$. Let the first $j=k-1$ coordinates range in $[n/k, n/k+n/(4k^2)]$, so all $a_i \asymp_k n$. Prescribe their residues modulo $H = 4\prod_{3\le q\le 2k} q$, where the product is over primes. At the odd primes require all summands to be nonzero. This is possible since any residue modulo an odd prime is a sum of two units. At 2, set the first j summands odd and the last odd or 2 mod 4 according to the parity of $n-k$, adjusting the residue of the j-th summand if needed. The multipliers are coprime to H.

For $i\le j$ write $b_i = r_i + Hv_i$. Integrality of $b_k = (n-\sum_{i\le j} d_i b_i)/d_k$ is equivalent to $v_j = \ell(v_1,\dots,v_{j-1}) + d_k w$, with an integral affine lift ℓ, since $(Hd_j, d_k)=1$. At each modulus r coprime to H, free residues of $v_1,\dots,v_{j-1},w$ parametrize all solutions of $\sum_i d_i b_i = n \bmod r$. Indeed, given the first $j-1$ coordinates, the remaining two range over the r solutions, in steps $(Hd_k, -Hd_j)$ of order r. In the box, each free residue vector has

$$N/r^j + O_k((n/r)^{j-1}) \quad (r\le n), \qquad N \asymp_k n^j/D.$$

This follows by counting the first $j-1$ progression coordinates and then w in an interval of fixed length $\asymp n/(d_j d_k)$ with endpoints depending on those coordinates.

At a prime $q>2k$ the proportion $g(q)$ of roots of $\prod_i b_i = 0 \bmod q$ is at most k/q. If $k\ge 3$, at least two coefficients d_i are units, so each coordinate is uniform. If $k=2$ and $q\mid d_i$, the other coordinate is forced nonzero since $q\nmid n$. The CRT and box estimate give remainder $O_k(n^{j-1}k^{\omega(r)})$ at squarefree $r\le n^{1/2}$. For the dimension condition, $g(q)\le k/q<1$ and

$$\prod_{\substack{w\le q<v\\ q>2k}} (1-g(q))^{-1} \ll_k (\log v/\log w)^k \qquad (2\le w\le v).$$

This follows from Mertens' product estimates [6, Theorems 7–8] and the summable $O_k(q^{-2})$ error in comparing $(1-k/q)^{-1}$ with $(1-1/q)^{-k}$. The same comparison gives $\prod_{2k<q<z}(1-g(q)) \gg_k (\log z)^{-k}$. The lower sieve bounds the surviving count below by a positive constant times N times this product, minus a constant times the sum of the absolute remainders to level $n^{1/2}$. That sum is $O_k(n^{j-1/2}(\log n)^{k-1})$, using $k^{\omega(r)} = \tau_k(r)$ for squarefree r. This is $o(N/(\log n)^k)$ since $D\le n^{k\eta}$. The lower sieve gives $\gg_k n^{k-1}/(D(\log n)^k)$ tuples with all odd factors of every b_i at least z.

Delete tuples with an odd square divisor in any b_i, or an odd prime dividing two b_i. These primes are all at least z and divide none of the d_i. For $q^2\mid b_i$, the modulus q^2 has root proportion $1/q^2$ and error $O_k(n^{j-1})$. Summing over $z\le q\le\sqrt{n}$ costs $O_k(N/z + n^{j-1/2})$. For $k\ge 3$, two vanishing coordinates modulo q have proportion $1/q^2$ and error $O_k(n^{j-1}/q)$. Summing costs $O_k(N/z + n^{j-1}\log n)$. For $k=2$, a shared q divides n. There are $O_\theta(1)$ such primes $q\ge z$, with count $O(N/q+1)$ each. The surviving count is still

$$\gg_k n^{k-1}/(D(\log n)^k). \tag{7}$$

In particular the odd part R of $\prod_i b_i$ is squarefree and has at most $J=\lceil k/\theta\rceil$ prime factors.

3.3 Overlaps

The probability of a successful tuple, conditional on the exposure, is $w = \lambda^{\omega(R)} R^{-\gamma} \asymp n^{-k\gamma} D^{\gamma}$. Each tuple determines the multipliers uniquely, by separation of prime ranges and the parity choice. Thus (6) and (7) give conditional mean $\gg n^u L(n)$.

Discard a tuple if its (b_i) satisfy a nonzero integral relation $\sum_i c_i b_i = 0$ with $\max_i |c_i| \le n^{\eta}$. For fixed multipliers and coefficients this and the original sum equation are inconsistent or determine two coordinates after the others are specified. Hence the total number discarded is $O_k(n^{k-2+2k\eta})$, and its conditional mean contribution is $O(n^{u-1+(2k+k\gamma)\eta}) = o(n^u L(n))$. For an identical support R there are at most k^J assignments of primes to the b_i, with the dyadic part already prescribed. For each assignment at most one multiplier vector survives. Two such vectors would give a discarded relation. Identical supports therefore contribute $O(\mu)$ to the ordered overlap sum, with μ the surviving conditional mean.

For a squarefree common divisor d of two supports, let $D_n(d)$ count tuples whose support contains d. Assign its primes to the summands, obtaining pairwise coprime $e_i \mid a_i$ with product d. There are at most k^J assignments. Fix all summands except the two with smallest e_i, e_ℓ. The sum equation and the CRT leave $O(n/(e_i e_\ell) + 1)$ choices for them. Since $e_i e_\ell \le d^{2/k}$,

$$D_n(d) \ll n^{k-1} d^{-1} (1 + d^{2/k}/n).$$

A nontrivial gcd proper in at least one support has $z \le d \le n^k/z$. Thus, on that interval,

$$d^{\gamma} D_n(d) \ll n^{k-1} (d^{-1/k} + d^{1/k}/n) \ll n^{k-1} z^{-1/k}.$$

The joint probability of successes with this gcd is the product of their success probabilities times $d^{\gamma} \lambda^{-\omega(d)}$. Sum over divisors of the first support, at most 2^J, and use the maximum success weight $O(n^{-k\gamma+k\eta\gamma})$ on the other tuple. Together with the identical-support estimate, this bounds the ordered overlap sum by

$$\Delta \ll \mu(1 + n^{u+k\eta\gamma} z^{-1/k}).$$

By Janson's inequality [5, formula (2)], on the good exposures the conditional probability of a gap is at most

$$\exp\Big(-cL(n) \min(n^u, n^{\theta/k - k\eta\gamma})\Big). \tag{8}$$

If $u > 0$, the right side is summable in n, so Borel–Cantelli and eventual occurrence of good exposures give cofiniteness. If $u = 0$, (8) is summable when $k(\lambda - 1) > 1$. For any $\lambda > 1$, the expected fraction of gaps on good exposures in the dyadic block $(2^{j-1}, 2^j]$ is at most

$$O\Big(\exp\{-cj^{k(\lambda-1)}/(\log j)^{k\lambda}\}\Big).$$

This is summable in j. Markov and Borel–Cantelli give density zero for gaps on good exposures. Since almost surely there are only finitely many other targets, this proves density one for kC.

3.4 Positive probability of coverage at equality

Set $\gamma = 1 - 1/k$ and $\lambda = 1 + 1/k$. In the multiplier construction allow primes from a fixed lower cutoff $Y_0 > 2k$, beyond the capped primes, through $U = n^{\tau}$. Call n suitable if

$$\sum_{\substack{Y_0 < q \le U \\ q \mid n}} X_q q^{-\alpha} \le 1.$$

The full pool product before removing prime divisors of n satisfies

$$\prod_{Y_0<q\le U}(1+kq^{-\alpha})^{X_q}\sim D_\omega(\log U)^{k\lambda}.$$

For its logarithm the centered terms have summable variances, and their expected terms differ from $k\lambda/q$ by an absolutely summable series. The constant D_ω has unbounded upper support. Selecting all primes of a long finite initial interval raises the centered partial sum arbitrarily high with positive probability. Its independent remainder has variance tending to zero as the interval increases.

For a suitable target the exclusion of prime divisors costs a factor at most e^k in the pool product. The size truncation of the multipliers retains at least half its weight, by the same bound on the expected $\log D$ as before. It follows that for any large fixed K a positive-probability event gives multiplier weight at least $K(\log n)^{k\lambda}$ for all sufficiently large suitable n. On each corresponding good exposure, (7) and the overlap estimates give $\mu\gg K\log n$ and $\Delta\ll\mu$, with constants independent of K. For large K the conditional gap probability is $O(n^{-2})$. Borel–Cantelli shows that with positive probability every sufficiently large suitable n is represented.

Choose fixed $L>Y_0$ for which covering every suitable $n>L$ has positive probability. This event is increasing in the prime choices, since adding primes increases the sets of represented targets and can only shrink the set of suitable targets. It has positive probability together with selection of all primes through L, by conditioning those finitely many bits to one. On this event, from $n>L$ divide out a selected pool prime whenever the current target exceeds L and is unsuitable. If the process stops at a suitable target above L, multiply its representation by the removed factors. If it stops at an integer at most L, the remaining integer and all removed factors belong to C. At least one removed prime $q>Y_0$ itself belongs to kC. For $q>L$ it is suitable because $q^\tau<q$. For $q\le L$ use $k-1$ copies of 1 and $q-k+1$. Multiplying its representation by $n/q\in C$ covers n in this case as well. This proves positive probability of cofiniteness.

4 Gaps below the critical multiplier

Take $\alpha=1/k$ and $1\le\lambda<1+1/k$. Write $a=k(\lambda-1)<1$, choose $a<b<1$, and expose all primes through $Y=\exp((\log x)^b)$. Let $\mathcal{F}_Y$ denote that exposure and C_Y the semigroup using only those selected primes. Take $I_x=(x,2x]$ or $I_x=[1,x]$, with integer x, and put

$$Z_x=\frac{|I_x\setminus kC|}{x},\qquad V_x=\frac{\sum_{n\in I_x}r_k(n)}{x(\log x)^a}.$$

Along each fixed geometric sequence of x's, $N_{C_Y}(2x)=o(x^\alpha)$ almost surely. For this, apply Rankin's inequality at $s=\alpha-1/\log Y$. The expected Euler product through Y is $O((\log Y)^K)$ for some fixed K: its factors differ from 1 by $O(q^{-1+1/\log Y})$, and $q^{1/\log Y}\le\mathrm{e}$. Thus $\mathbb{E}N_{C_Y}(2x)\ll(2x)^s(\log Y)^K$. Markov bounds the probability that the count exceeds $x^\alpha\exp(-(\log x)^{1-b}/2)$ by $O(\exp(-c(\log x)^{1-b}))$, a summable bound on that sequence.

Switching on an unexposed prime $q\le 2x$ affects only sums with at least one summand divisible by q. Write C^+ for the semigroup with q included. The number of its multiples of q

through $2x$ is $N_{C^+}(2x/q)$. The changes in Z_x and V_x are bounded respectively by

$$\frac{k}{x}N_{C^+}(2x/q)N_{C^+}(2x)^{k-1}, \qquad \frac{k}{x(\log x)^a}N_{C^+}(2x/q)N_{C^+}(2x)^{k-1}.$$

The forced-prime bound (2) and Hölder's inequality give expected squares $O(q^{-2\alpha}(\log x)^{2a})$ for the change in Z_x and $O(q^{-2\alpha})$ for the change in V_x. Reveal the remaining primes after $\mathcal{F}_Y$. Orthogonality of the successive conditional expectation differences, with Jensen's inequality over the future bits, bounds each conditional variance by the sum of conditional squared changes times $p_q(1-p_q)$. Taking expectation over the exposure gives

$$\mathbb{E}\operatorname{Var}(Z_x \mid \mathcal{F}_Y) + \mathbb{E}\operatorname{Var}(V_x \mid \mathcal{F}_Y) \ll (\log x)^{2a}\sum_{q>Y} p_q q^{-2\alpha} \ll (\log x)^{2a}\exp(-\alpha(\log x)^b).$$

By Chebyshev and Borel–Cantelli, both variables differ from their conditional means by $O(\exp(-\alpha(\log x)^b/4))$ almost surely along each fixed geometric sequence. Equation (3) shows $V_x \to K_k A^k$ for either interval.

Discard the $o(x)$ targets represented using only exposed primes. Every remaining possible representation has conditional success probability at most $\lambda Y^{-\gamma}$. Harris' inequality on the decreasing events that those representations fail [5, formula (3)] gives the lower bound $\exp(-\mathbb{E}[r_k(n) \mid \mathcal{F}_Y]/(1-\lambda Y^{-\gamma}))$ for the conditional gap probability. Averaging over the retained targets by Jensen gives

$$\mathbb{E}[Z_x \mid \mathcal{F}_Y] \geq (1-o(1))\exp\left(-\frac{(\log x)^a\,\mathbb{E}[V_x \mid \mathcal{F}_Y]}{(1-o(1))(1-\lambda Y^{-\gamma})}\right).$$

Since $b > a$, the concentration error is negligible against this lower bound. If $\lambda > 1$, taking dyadic blocks gives infinitely many gaps almost surely. At $\lambda = 1$, use $I_x = [1, x]$ on a countable sequence of geometric grids whose ratios decrease to one. The numerator of Z_x is nondecreasing between grid points. The resulting lower density is at least $\exp(-K_k A^k)$.

This finishes the strict sides of the critical-multiplier threshold. It also proves Theorem 1. Indeed, when $\lambda = 1$, cofinite orders $k \geq 2$ are exactly those with $k > 1/(1-\gamma)$ for $0 < \gamma < 1$, by the count, the upper bound and the critical gap result. The count excludes order one. Intersecting the probability-one statements over integer k gives the formula for $h(C)$.

5 Stopped martingales at equality

In this section $\alpha = 1/k$ and $\lambda = 1 + 1/k$. Write $N_Q(X) = |C_Q \cap [1, X]|$, for $Q \geq 2$.

Lemma 5.1 (Uniform smooth moment)**.** *For all $X, Q \geq 2$,*

$$\mathbb{E}N_Q(X)^k \ll_k X\log(2X)\exp\left(-c_k\frac{\log X}{\log Q}\right). \tag{9}$$

Proof. Set $\eta = \min(\alpha/4, 1/\log Q)$ and $\sigma = \alpha - \eta$. The local multivariate membership series at $q \leq Q$ has constant term 1 and coefficient p_q at each nonzero exponent vector, with $p_2 = 1$. At $q > Q$ it equals 1. Majorize its product coefficientwise by $\prod_i \zeta(s_i + \gamma + \eta)^\lambda H(\mathbf{s})$. For $q \leq Q$, the local factor of H has constant 1, linear coefficients $(p_q - \lambda q^{-\gamma-\eta})_+$, and all coefficients of

degree at least two equal to p_q. At $q > Q$ take $H_q = 1$. This is a coefficientwise majorant because the zeta factors are nonnegative and supply any missing linear mass.

At the weight $z_i = q^{-\sigma}$, the sum of odd-prime linear corrections is

$$\ll_k \sum_{q \le Q} \frac{q^\eta - 1}{q} \ll_k \eta \sum_{q \le Q} \frac{\log q}{q} \ll_k 1$$

by $\eta \log Q \le 1$ and Mertens' estimate [6]. Higher degrees contribute $O_k(q^{-\gamma-2\sigma})$ per odd prime, which is summable since $\gamma + 2\sigma \ge 1 + \alpha/2$. The factor at 2 is bounded. Hence the weighted absolute coefficient mass of H is uniformly bounded. Partial summation from (4) gives, uniformly in η,

$$\sum_{b \le v} \tau_\lambda(b) b^{-\gamma-\eta} \ll v^{\alpha-\eta}(1 + \log v)^{\lambda-1},$$

since $\alpha - \eta \ge 3\alpha/4$. Convolution gives $O_k(X^{k(\alpha-\eta)}(\log(2X))^{k(\lambda-1)})$. Here $k(\lambda - 1) = 1$, and $k\eta \gg_k 1/\log Q$, proving (9). □

5.1 Compensating surviving targets

Let $\mathbb{P}_y$ and $\mathbb{E}_y$ denote probability and expectation conditional on $\mathcal{E}_y$, with y beyond the mandatory primes and $p_q \le 1/2$ above y. Take integer x, set $X = 2x$, and expose the remaining primes $y < q \le X$ in increasing order. For $x < n \le 2x$, say q is dangerous for n if n has a representation using q and otherwise only primes already selected. At least one summand must be divisible by q. This predicate is defined for all past configurations, even when n has an earlier representation, and is increasing in the previous prime bits.

Give every initially unrepresented n weight 1 and every other n weight zero. Choose $0 < t < \alpha/2$. At a dangerous prime, if the current weight w is positive and unfrozen, first test whether $w/(1 - p_q) > q^t$. In that case freeze w before exposing the prime. Otherwise set the weight to zero upon selection and to $w/(1 - p_q)$ upon exclusion. At other steps the weight stays fixed. The decision to freeze is predictable and each update has conditional mean w, so the weights are martingales. A positive terminal weight on an unfrozen target certifies a gap. Its first representation would have killed the weight at its largest prime divisor.

Let Z be the terminal total weight divided by x, and S the normalized part that is frozen. Only polylogarithmically many targets can be represented from the mandatory primes alone, so $\mathbb{E}_y Z = 1 - o(1)$. Each unfrozen terminal weight is at most X^t.

5.2 Controlling the variance and the frozen part

At q, all updated targets use the same Bernoulli variable. If B_q is the sum of their weights just before the update divided by x, the conditional variance of the increment is $p_q B_q^2/(1 - p_q)$. Each active weight is at most q^t. Let $F_q(\alpha)$ be the Euler product through q, with the actual preceding choices and q provisionally selected. The number of dangerous targets is at most the number of ordered tuples with some coordinate divisible by q from this provisional semigroup. Rankin bounds that number by $kXq^{-\alpha}F_q(\alpha)^k$, using $k\alpha = 1$. The unconditioned local moment factors at large primes $\ell < q$ are $1 + 2k\lambda/\ell + O_k(\ell^{-1-\alpha})$. The factor at q is uniformly bounded, and the exclusions in $\mathcal{E}_y$ can only decrease the product. Mertens' estimate and orthogonality of increments give

$$\operatorname{Var}_y Z \ll_k \sum_{q > y} q^{-1-\alpha+2t}(\log q)^{2k\lambda}. \tag{10}$$

This tends to zero as y increases, uniformly in x.

For an initially unrepresented target n, use its terminal weight as the density of a new probability measure relative to $\mathbb{P}_y$. Its mean is 1, and the density at an earlier exposure step is the current weight. The tilt forces exclusion when an update takes place. At all other steps, including the decision that freezes a weight, the prime retains its probability p_q. Under coupling by common uniforms, the tilted prime set is therefore contained in the original prime set.

Put $H_n(Q) = \sum_{y<q\le Q} \mathbf{1}_{\{q \text{ dangerous for } n\}}[-\log(1-p_q)]$. This is increasing in the exposed bits. Before a freeze on a tilted path, every dangerous prime is excluded and multiplies the weight by $1/(1-p_q)$. Thus at the proposed freeze at q the current log-weight is the hazard just before q, and the proposed update gives $H_n(q) > t\log q$. Split the primes into intervals $(Q_{j+1}, Q_j]$, where $Q_j = X^{\theta_j}$ and $\theta_j = 2^{-j}$. In the original process

$$\sum_{x<n\le 2x} \mathbb{E}_y H_n(Q) \le 2\mathbb{E}_y N_Q(X)^k.$$

Indeed, danger is determined by the past. Its probability times p_q is at most the expected count of representations with largest prime divisor q. Also $-\log(1-p_q) \le 2p_q$. On the j-th interval a freeze requires $H_n(Q_j) > (t/2)\theta_j \log X$. The expected stopped mass is the sum of the tilted freeze probabilities over initially unrepresented targets, divided by x. Monotone coupling and Markov's inequality bound it by

$$\mathbb{E}_y S \ll_t \sum_{j:Q_j>y} \frac{\mathbb{E}_y N_{Q_j}(X)^k}{X\theta_j \log X}. \tag{11}$$

The conditional count in (11) is dominated by the unconditioned count. By Lemma 5.1, the summands are at most $O(2^j \exp(-c_k 2^j))$. For each fixed j, y, the limsup of the summand as $x \to \infty$ is at most $2^j \mathbb{E}_y A^k$, from the L^k count limit and $N_{Q_j}(X) \le N(X)$. Equation (5) and dominated convergence give $\lim_{y\to\infty} \limsup_{x\to\infty} \mathbb{E}_y S = 0$.

By (10) and Chebyshev, the probability that $Z < 1/2$ is small for large y and large x. Markov bounds the probability that $S > 1/4$. Otherwise $Z - S \ge 1/4$, certifying at least $x/(4X^t)$ genuine gaps. Taking the limsup along dyadic x gives conditional probability tending to one, as $y \to \infty$, of infinitely many blocks with at least $x/(4X^t)$ gaps. Every $\mathcal{E}_y$ has positive probability, so infinitely many gaps occur with positive probability unconditionally. Section 3 gives positive probability of cofiniteness, finishing Theorem 2.

6 Finite changes of selected primes

Let B denote the cofinite event at equality. It is measurable, since representability for each target depends on finitely many prime bits. The saturation of B under finite changes of optional bits is the countable union of its finite-coordinate translates. It is measurable and belongs to the tail sigma field. For each prefix, membership can be tested by replacing that prefix with one fixed choice and then testing the same saturation. The same argument applies to the saturation of B^c. Both saturations have positive probability by Theorem 2, so both have probability one by Kolmogorov's zero-one law.

Take a realization in both saturations. A finite change that gives B may be replaced by adding its newly selected primes, by monotonicity. A change that gives B^c may be replaced by

removing the primes it deletes, for the same reason. In particular, forcing all optional bits up to y to one eventually gives B, and forcing them to zero eventually gives B^c. Under the original product law the resulting realizations have the conditional laws given $\mathcal{I}_y$ and $\mathcal{E}_y$, respectively. Dominated convergence gives both conditional limits in Corollary 3. Theorem 2 also implies that B is outside the tail sigma field, even modulo null events.

For the selected primes themselves, the prime number theorem and partial summation give

$$|P \cap [1, x]| \sim \frac{\lambda}{\alpha} \frac{x^\alpha}{\log x} \qquad \text{almost surely.}$$

The variance of this prime count is at most its mean. Chebyshev on geometric grids, followed by Borel–Cantelli and a squeeze as the grid ratios approach one, proves the stated almost-sure asymptotic. Finite changes affect the count by $O(1)$. For the generated semigroup, a finite change preserves the growth exponents in (1) and multiplies the leading constant by

$$\prod_{q \text{ added}} (1 - q^{-\alpha})^{-1} \prod_{q \text{ removed}} (1 - q^{-\alpha}).$$

This follows by multiplying the corresponding finite Euler factors and reapplying the counting argument.

In fact, at equality the two outcomes have least basis orders k and $k + 1$, respectively, for almost every realization and for all its finite optional modifications. Orders below k are excluded by (1), and order $k + 1$ is cofinite by the coverage proof, since $\gamma < 1 - 1/(k + 1)$. The latter assertion holds simultaneously under all finite optional changes: any prescribed change of finitely many optional bits is absolutely continuous with respect to the original law, and the possible changes form a countable family.

A quantitative description of the endpoint probability in terms of the counting constant and local representation densities remains open.

AI assistance and review. ChatGPT (OpenAI) assisted with mathematical arguments, literature checks, exposition, proof review and verification code.

References

[1] Y. Cao, Z. Huang and R. Zhang, *Effective equidistribution, arithmetic purity of strong approximation, and geometric sieve for affine quadrics*, arXiv:2509.23320v1, 2025.

[2] H. Delange, *Généralisation du théorème de Ikehara*, Ann. Sci. École Norm. Sup. (3) **71** (1954), 213–242. doi:10.24033/asens.1023.

[3] P. Erdős and A. Rényi, *Additive properties of random sequences of positive integers*, Acta Arith. **6** (1960), 83–110. doi:10.4064/aa-6-1-83-110.

[4] P. Erdős and P. Tetali, *Representations of integers as the sum of k terms*, Random Structures Algorithms **1** (1990), 245–261. doi:10.1002/rsa.3240010302.

[5] S. Janson and L. Warnke, *The lower tail: Poisson approximation revisited*, Random Structures Algorithms **48** (2016), 219–246. arXiv:1406.1248v1.

[6] J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), 64–94. doi:10.1215/ijm/1255631807.