On the least n for which $\binom{n}{k}$ has no prime divisor $\le k$

Abstract

For each integer $k \ge 2$, let $g(k) > k+1$ be the least integer n such that no prime $p \le k$ divides $\binom{n}{k}$. Equivalently, every prime divisor of $\binom{n}{k}$ is $> k$. This note records the parts of the problem that can be proved rigorously from Lucas' theorem, Kummer's theorem, the prime number theorem, and a Fourier estimate on the associated Chinese-remainder product set. In particular, we prove an exact Euler product for the density of admissible integers and show that

$$\log \Delta(k)^{-1} \asymp \frac{k}{\log k},$$

where $\Delta(k)$ is the density of admissible n. We also prove structural divisibility properties of $g(k)$ and the unconditional upper bound

$$\log g(k) \le k + O\left(\frac{k}{\log k}\right).$$

We conclude by isolating the precise remaining obstacle to the conjectural estimate $\log g(k) \asymp k/\log k$ and to any rigorous analysis of $g(k+1)/g(k)$.

1 Definitions and reformulations

For $k \ge 2$, define $g(k) > k+1$ to be the least integer n such that

$$p \nmid \binom{n}{k} \qquad \text{for every prime } p \le k.$$

Equivalently, every prime divisor of $\binom{n}{k}$ is $> k$.

It is convenient to write

$$g(k) = k + m_*(k),$$

where $m_*(k)$ is the least integer $m \ge 2$ such that

$$p \nmid \binom{k+m}{k} \qquad (\forall p \le k,\ p \in \mathbb{P}).$$

Proposition 1 (Lucas reformulation)**.** *Fix a prime $p \le k$. Write the base-p expansions*

$$k = \sum_{j\ge0} k_j^{(p)} p^j, \qquad n = \sum_{j\ge0} n_j^{(p)} p^j, \qquad 0 \le k_j^{(p)}, n_j^{(p)} \le p-1.$$

Then

$$p \nmid \binom{n}{k} \iff k_j^{(p)} \le n_j^{(p)} \quad \textit{for every } j \ge 0.$$

Proof. This is exactly Lucas' theorem. □

Proposition 2 (Kummer reformulation)**.** *Fix a prime $p \le k$. Then*

$$p \nmid \binom{k+m}{k} \iff \text{the addition of } k \text{ and } m \text{ in base } p \text{ produces no carry.}$$

Proof. By Kummer's theorem, $v_p\left(\binom{k+m}{k}\right)$ is the number of carries when adding k and m in base p. □

Thus $m_*(k)$ is the least $m \ge 2$ such that the addition of k and m is carry-free in every base $p \le k$.

2 Exact local and global densities

For fixed k, define the admissible set

$$\mathcal{A}_k := \{n \ge 0 : \ p \nmid \tbinom{n}{k} \text{ for every prime } p \le k\}.$$

We write $\Delta(k)$ for its natural density.

Local density at one prime

Fix a prime $p \le k$ and write

$$k = \sum_{j \ge 0} k_j^{(p)} p^j.$$

Let

$$L_p := \lfloor \log_p k \rfloor + 1.$$

Then $k_j^{(p)} = 0$ for $j \ge L_p$.

Proposition 3 (Local density formula)**.** *For a prime $p \le k$, the set*

$$\mathcal{A}_{k,p} := \{n \ge 0 : \ p \nmid \tbinom{n}{k}\}$$

has natural density

$$\delta_p(k) := \operatorname{dens}(\mathcal{A}_{k,p}) = \prod_{j \ge 0} \left(1 - \frac{k_j^{(p)}}{p}\right).$$

Equivalently,

$$\delta_p(k) = \frac{1}{p^{L_p}} \prod_{j=0}^{L_p - 1} (p - k_j^{(p)}).$$

Proof. By Lucas' theorem,

$$p \nmid \binom{n}{k} \iff n_j^{(p)} \in \{k_j^{(p)}, k_j^{(p)} + 1, \dots, p-1\} \quad (0 \le j < L_p).$$

Modulo p^{L_p}, the first L_p base-p digits of n may be chosen independently. For the jth digit there are exactly $p - k_j^{(p)}$ admissible choices. Hence the number of admissible residues modulo p^{L_p} is

$$\prod_{j=0}^{L_p - 1} (p - k_j^{(p)}).$$

Dividing by the total number p^{L_p} of residues gives the density. Since $k_j^{(p)} = 0$ for $j \ge L_p$, the finite product is equal to the stated infinite product. □

Global density

Proposition 4 (Global density formula)**.** *The admissible set $\mathcal{A}_k$ has natural density*

$$\Delta(k) = \prod_{p \le k} \delta_p(k) = \prod_{p \le k} \prod_{j \ge 0} \left(1 - \frac{k_j^{(p)}}{p}\right).$$

Proof. For each $p \le k$, the condition $p \nmid \binom{n}{k}$ depends only on $n \bmod p^{L_p}$. Put

$$M(k) := \prod_{p \le k} p^{L_p}.$$

The moduli p^{L_p} are pairwise coprime. By the Chinese remainder theorem, the admissible residues modulo $M(k)$ are exactly the Cartesian product of the local admissible residue sets modulo p^{L_p}. Therefore the number of admissible residue classes mod $M(k)$ is the product of the local numbers, and dividing by $M(k)$ yields the product formula for $\Delta(k)$. □

3 The size of the density

Let

$$S(k) := \log \Delta(k)^{-1} = \sum_{p \le k} -\log \delta_p(k).$$

We shall prove

$$S(k) \asymp \frac{k}{\log k}.$$

We use the prime number theorem in the form $\theta(x) := \sum_{p \le x} \log p \sim x$, and also the standard interval estimate

$$\pi(x+y) - \pi(x) \ll \frac{y}{\log x} + 1 \qquad (2 \le y \le x), \tag{1}$$

which is a standard consequence of the prime number theorem (or of Brun–Titchmarsh, for the ranges needed here).

Theorem 5 (Density exponent)**.** *For $k \ge 2$,*

$$\log \Delta(k)^{-1} \asymp \frac{k}{\log k}.$$

More precisely, there exist absolute constants $c, C > 0$ such that

$$c\frac{k}{\log k} \le \log \Delta(k)^{-1} \le C\frac{k}{\log k}.$$

Proof. We first prove the lower bound. Let $p \in (k/2, 2k/3]$ be prime. Then $k = p + r$ with $r = k - p$, so the base-p expansion of k has exactly two digits,

$$k = 1 \cdot p + r.$$

Therefore

$$\delta_p(k) = \left(1 - \frac{1}{p}\right)\left(1 - \frac{r}{p}\right).$$

Since $p \le 2k/3$, we have $r = k - p \ge k/3$, and therefore

$$\frac{r}{p} \ge \frac{(k/3)}{(2k/3)} = \frac{1}{2}.$$

Thus $\delta_p(k) \le 1/2$, and so

$$-\log \delta_p(k) \ge \log 2.$$

The prime number theorem gives $\#\{p \in \mathbb{P} : k/2 < p \le 2k/3\} \gg k/\log k$, hence

$$S(k) \ge \sum_{k/2<p\le 2k/3} (-\log \delta_p(k)) \gg \frac{k}{\log k}.$$

We now prove the upper bound. Split the sum over $p \le k$ into $p \le \sqrt{k}$ and $p > \sqrt{k}$. For $p \le \sqrt{k}$ we use the trivial estimate $\delta_p(k) \ge p^{-L_p}$, so

$$-\log \delta_p(k) \le L_p \log p \le 2\log k.$$

The number of primes $p \le \sqrt{k}$ is $O(\sqrt{k}/\log k)$, so their total contribution is

$$O(\sqrt{k}).$$

Now let $p > \sqrt{k}$. Then k has only two base-p digits:

$$k = ap + b, \qquad a = \lfloor k/p \rfloor, \qquad 0 \le b < p.$$

Hence

$$\delta_p(k) = \left(1 - \frac{a}{p}\right)\left(1 - \frac{b}{p}\right).$$

Accordingly, write

$$S_>(k) := \sum_{p>\sqrt{k}} -\log \delta_p(k) = T_1 + T_2,$$

where

$$T_1 := \sum_{p>\sqrt{k}} -\log\left(1 - \frac{a}{p}\right), \qquad T_2 := \sum_{p>\sqrt{k}} -\log\left(1 - \frac{b}{p}\right).$$

We first estimate T_1. For $\sqrt{k} < p \le 2\sqrt{k}$ there are $O(\sqrt{k}/\log k)$ such primes and each summand is $O(1)$, hence their contribution is $O(\sqrt{k}/\log k)$. For $p > 2\sqrt{k}$ we have $a/p \le k/p^2 \le 1/2$, so

$$-\log\left(1 - \frac{a}{p}\right) \le 2\frac{a}{p} \le 2\frac{k}{p^2}.$$

Therefore

$$T_1 \ll \frac{\sqrt{k}}{\log k} + k \sum_{p>2\sqrt{k}} \frac{1}{p^2} \ll \sqrt{k}.$$

It remains to estimate T_2. For each integer $a \in [1, \lfloor \sqrt{k} \rfloor]$, let

$$I_a := \left(\frac{k}{a+1}, \frac{k}{a}\right].$$

If $p \in I_a \cap \mathbb{P}$, then $\lfloor k/p \rfloor = a$, so again $k = ap + b$ and

$$1 - \frac{b}{p} = a + 1 - \frac{k}{p} =: u_p \in (0, 1].$$

Thus

$$T_2 = \sum_{1 \le a \le \sqrt{k}} \sum_{p \in I_a \cap \mathbb{P}} - \log u_p.$$

For $m \ge 1$ define

$$E_{a,m} := \{p \in I_a \cap \mathbb{P} : \ e^{-m} < u_p \le e^{-(m-1)}\}.$$

Since $u_p > 0$ and $u_p \ge 1/p \ge 1/k$, only $m \le 1 + \log k$ occur. Hence

$$\sum_{p \in I_a \cap \mathbb{P}} - \log u_p \le \sum_{1 \le m \le 1 + \log k} m \, \# E_{a,m}.$$

We claim that $E_{a,m}$ lies in an interval of length

$$\ll e^{-m} \frac{k}{a^2}.$$

Indeed, the function

$$f_a(x) := a + 1 - \frac{k}{x}$$

is increasing on I_a, and if $f_a(x_1) = u_1$, $f_a(x_2) = u_2$, then

$$x_2 - x_1 = \frac{k}{a + 1 - u_2} - \frac{k}{a + 1 - u_1} = \frac{k(u_2 - u_1)}{(a + 1 - u_2)(a + 1 - u_1)} \ll \frac{k(u_2 - u_1)}{a^2}.$$

Applying this with $u_2 - u_1 \asymp e^{-m}$ proves the claim.

By the interval estimate (1),

$$\# E_{a,m} \ll 1 + \frac{e^{-m} k}{a^2 \log(k/a + 2)}.$$

Therefore

$$\sum_{p \in I_a \cap \mathbb{P}} - \log u_p \ll \sum_{1 \le m \le 1 + \log k} m + \frac{k}{a^2 \log(k/a + 2)} \sum_{m \ge 1} m e^{-m}.$$

Since $\sum_{m \ge 1} m e^{-m} < \infty$ and $\sum_{m \le 1 + \log k} m \ll \log^2 k$, this yields

$$\sum_{p \in I_a \cap \mathbb{P}} - \log u_p \ll \log^2 k + \frac{k}{a^2 \log(k/a + 2)}.$$

Summing over $1 \le a \le \sqrt{k}$ gives

$$T_2 \ll \sqrt{k} \log^2 k + \sum_{1 \le a \le \sqrt{k}} \frac{k}{a^2 \log(k/a + 2)} \ll \sqrt{k} \log^2 k + \frac{k}{\log k} \ll \frac{k}{\log k}.$$

Combining the estimates for T_1, T_2, and the small primes $p \le \sqrt{k}$, we obtain

$$S(k) \ll \frac{k}{\log k}.$$

This completes the proof. □

Corollary 6. *There exist absolute constants $c, C > 0$ such that*

$$\exp\left(-C \frac{k}{\log k} \right) \le \Delta(k) \le \exp\left(-c \frac{k}{\log k} \right).$$

4 Structural facts about $g(k)$

A divisibility theorem

Theorem 7 (Forced divisibility)**.** *Let*

$$Q(k) := \prod_{p \le k} p^{v_p(k+1)}.$$

Then

$$Q(k) \mid (g(k) - k).$$

In particular, if $k+1$ *is composite, then* $Q(k) = k+1$ *and therefore*

$$g(k) \ge 2k+1.$$

Proof. Write $g(k) = k + m_*(k)$. Fix a prime $p \le k$, and set $t_p := v_p(k+1)$. Then the last t_p base-p digits of k are all equal to $p-1$. By the Kummer reformulation, the addition of k and $m_*(k)$ in base p is carry-free. Therefore the last t_p base-p digits of $m_*(k)$ must all be 0, that is,

$$p^{t_p} \mid m_*(k).$$

Multiplying over all primes $p \le k$ gives

$$Q(k) \mid m_*(k) = g(k) - k.$$

If $k+1$ is composite, then every prime factor of $k+1$ is at most k, so

$$Q(k) = \prod_{p \le k} p^{v_p(k+1)} = k+1.$$

Hence $g(k) - k$ is a positive multiple of $k+1$, and since $g(k) > k+1$, necessarily $g(k) \ge 2k+1$. □

A Fourier upper bound for $g(k)$

We now prove the best unconditional upper bound obtained in the discussion above.

For each prime $p \le k$, let

$$B_p(k) := \left\{ b \bmod p^{L_p} : \text{ the addition of } k \text{ and } b \text{ in base } p \text{ is carry-free} \right\}.$$

Equivalently, if

$$k = \sum_{j=0}^{L_p-1} k_j^{(p)} p^j,$$

then

$$B_p(k) = \left\{ \sum_{j=0}^{L_p-1} u_j p^j : \ 0 \le u_j \le p - 1 - k_j^{(p)} \right\} \subset \mathbb{Z}/p^{L_p}\mathbb{Z}.$$

By the Chinese remainder theorem, the admissible shifts modulo

$$M(k) := \prod_{p \le k} p^{L_p}$$

form a product set

$$B(k) \subset \mathbb{Z}/M(k)\mathbb{Z}, \qquad |B(k)| = \Delta(k)M(k),$$

and

$$m_*(k) = \min\{m \geq 2 : \ m \bmod M(k) \in B(k)\}.$$

For a finite cyclic group $\mathbb{Z}/M\mathbb{Z}$ and a subset $A \subset \mathbb{Z}/M\mathbb{Z}$, let

$$\widehat{1_A}(r) := \sum_{x \bmod M} 1_A(x)\, e^{2\pi i r x/M} \qquad (r \in \mathbb{Z}/M\mathbb{Z})$$

be the Fourier transform of its indicator function.

Lemma 8 (A counting lemma)**.** *Let $A \subset \mathbb{Z}/M\mathbb{Z}$, put $K := |A|$, and suppose that for some $\Gamma \geq 0$ one has*

$$\sum_{1 \leq r \leq X} |\widehat{1_A}(r)| \leq K\Gamma X \qquad (1 \leq X \leq M/2).$$

Then for every integer $H \geq 1$,

$$\#\{1 \leq n \leq H : \ n \bmod M \in A\} = \frac{KH}{M} + O\big(K\Gamma(1 + \log H)\big),$$

with an absolute implied constant.

Proof. Let

$$N_A(H) := \sum_{n=1}^{H} 1_A(n).$$

By Fourier inversion,

$$N_A(H) = \frac{KH}{M} + \frac{1}{M}\sum_{r \neq 0} \widehat{1_A}(r) S_H(-r),$$

where

$$S_H(r) := \sum_{n=1}^{H} e^{2\pi i r n/M}.$$

For $1 \leq r \leq M/2$,

$$|S_H(r)| \ll \min\left(H, \frac{M}{r}\right).$$

Using the symmetry $|\widehat{1_A}(M - r)| = |\widehat{1_A}(r)|$, we obtain

$$\left|N_A(H) - \frac{KH}{M}\right| \ll \frac{H}{M} \sum_{1 \leq r \leq M/H} |\widehat{1_A}(r)| + \sum_{M/H < r \leq M/2} \frac{|\widehat{1_A}(r)|}{r}.$$

The first term is $O(K\Gamma)$ by hypothesis. For the second, put

$$F(X) := \sum_{1 \leq r \leq X} |\widehat{1_A}(r)|,$$

so that $F(X) \leq K\Gamma X$. By partial summation,

$$\sum_{Y < r \leq M/2} \frac{|\widehat{1_A}(r)|}{r} \ll \frac{F(M/2)}{M} + \int_Y^{M/2} \frac{F(t)}{t^2}\, dt \ll K\Gamma + K\Gamma \log \frac{M}{Y}.$$

Taking $Y = M/H$ gives the stated bound. □

We now specialize to $A = B(k)$. Let

$$\mathcal{P}_k := \{p \in \mathbb{P} : 2\sqrt{k} < p \le k\}.$$

For $p \in \mathcal{P}_k$, one has $L_p = 2$. Write

$$k = ap + b, \qquad a = \lfloor k/p \rfloor, \qquad 0 \le b < p.$$

Then the local admissible set modulo p^2 is the rectangle

$$B_p(k) = \{u + vp : \ 0 \le u \le p - b - 1, \ 0 \le v \le p - a - 1\}.$$

Hence

$$|B_p(k)| = (p - b)(p - a),$$

and for every $r \in \mathbb{Z}$,

$$\widehat{1_{B_p(k)}}(r) = \left(\sum_{u=0}^{p-b-1} e^{2\pi i r u/p^2} \right) \left(\sum_{v=0}^{p-a-1} e^{2\pi i r v/p} \right).$$

If $p \nmid r$, then the second factor is the negative of the sum of the a omitted pth roots of unity, so its absolute value is at most a. Therefore

$$\frac{|\widehat{1_{B_p(k)}}(r)|}{|B_p(k)|} \le \frac{a}{p - a} =: \rho_p \qquad (p \nmid r).$$

If $p \mid r$, then trivially

$$\frac{|\widehat{1_{B_p(k)}}(r)|}{|B_p(k)|} \le 1.$$

Thus for every $p \in \mathcal{P}_k$,

$$\frac{|\widehat{1_{B_p(k)}}(r)|}{|B_p(k)|} \le \rho_p \, 1_{p \nmid r} + 1_{p \mid r}, \qquad \rho_p = \frac{a}{p - a}. \tag{2}$$

Lemma 9 (Global Fourier majorant)**.** *Define*

$$\beta_p := \rho_p + \frac{1 - \rho_p}{p} \qquad (p \in \mathcal{P}_k), \qquad \Gamma_k := \prod_{p \in \mathcal{P}_k} \beta_p.$$

Then for every $X \ge 1$,

$$\sum_{1 \le r \le X} |\widehat{1_{B(k)}}(r)| \le |B(k)| \, \Gamma_k \, X.$$

Proof. For $p \notin \mathcal{P}_k$ we use the trivial bound

$$|\widehat{1_{B_p(k)}}(r)| \le |B_p(k)|.$$

For $p \in \mathcal{P}_k$, apply (2). Multiplying over primes gives

$$|\widehat{1_{B(k)}}(r)| \le |B(k)| \prod_{p \in \mathcal{P}_k} \big(\rho_p 1_{p \nmid r} + 1_{p \mid r} \big).$$

Now

$$\rho_p 1_{p \nmid r} + 1_{p \mid r} = \rho_p \Big(1 + (\rho_p^{-1} - 1) 1_{p \mid r} \Big).$$

Hence

$$|\widehat{1_{B(k)}}(r)| \le |B(k)| \Big(\prod_{p\in\mathcal{P}_k} \rho_p \Big) \prod_{p\in\mathcal{P}_k} \Big(1 + (\rho_p^{-1} - 1) 1_{p|r} \Big).$$

Expanding the product and summing over $1 \le r \le X$ yields

$$\sum_{1\le r\le X} |\widehat{1_{B(k)}}(r)| \le |B(k)| \Big(\prod_{p\in\mathcal{P}_k} \rho_p \Big) \sum_d \lambda(d) \left\lfloor \frac{X}{d} \right\rfloor,$$

where the sum runs over squarefree integers d all of whose prime factors lie in $\mathcal{P}_k$, and

$$\lambda(d) := \prod_{p|d} (\rho_p^{-1} - 1).$$

Therefore

$$\sum_{1\le r\le X} |\widehat{1_{B(k)}}(r)| \le |B(k)|\, X \left(\prod_{p\in\mathcal{P}_k} \rho_p \right) \prod_{p\in\mathcal{P}_k} \left(1 + \frac{\rho_p^{-1} - 1}{p} \right).$$

But

$$\rho_p \left(1 + \frac{\rho_p^{-1} - 1}{p} \right) = \rho_p + \frac{1 - \rho_p}{p} = \beta_p,$$

so the right-hand side is $|B(k)|\Gamma_k X$, as claimed. □

Lemma 10 (Size of Γ_k)**.** *One has*

$$\log \Gamma_k \le -k + O\Big(\frac{k}{\log k} \Big).$$

Proof. Fix $p \in \mathcal{P}_k$ and write $k = ap + b$ as above. Since $p > 2\sqrt{k}$, we have $a < p/2$, so $p - a > p/2$. Also

$$\beta_p = \rho_p + \frac{1 - \rho_p}{p} \le \rho_p + \frac{1}{p} \le \frac{a}{p-a} + \frac{1}{p} \le \frac{a+1}{p-a} \le \frac{2(a+1)}{p}.$$

For each integer a with $1 \le a < \sqrt{k}/2$, let

$$I_a := \left(\frac{k}{a+1}, \frac{k}{a} \right].$$

Then the primes in I_a are exactly those for which $\lfloor k/p \rfloor = a$. Therefore

$$\log \Gamma_k \le \sum_{a<\sqrt{k}/2} \sum_{p\in I_a\cap\mathbb{P}} \big(\log(2(a+1)) - \log p \big).$$

Summing the negative terms over all primes $p \in \mathcal{P}_k$ gives

$$-\sum_{p\in\mathcal{P}_k} \log p = -\theta(k) + \theta(2\sqrt{k}) = -k + O(\sqrt{k}).$$

For the positive terms we use the interval estimate (1). Since $|I_a| \asymp k/a^2$ and $I_a \subset [k/(a+1), k/a]$ with comparable endpoints,

$$\#(I_a \cap \mathbb{P}) \ll \frac{k}{a(a+1)\log(k/a+2)} + 1.$$

Hence

$$\sum_{a<\sqrt{k}/2} \#(I_a \cap \mathbb{P}) \log(2(a+1)) \ll \frac{k}{\log k} \sum_{a\geq 1} \frac{\log(a+1)}{a(a+1)} + \sum_{a<\sqrt{k}/2} \log(a+1).$$

The first sum converges, and the second is $O(\sqrt{k}\log k) = O(k/\log k)$. Therefore

$$\sum_{a<\sqrt{k}/2} \#(I_a \cap \mathbb{P}) \log(2(a+1)) \ll \frac{k}{\log k}.$$

Combining the two parts yields

$$\log \Gamma_k \leq -k + O\left(\frac{k}{\log k}\right),$$

as claimed. □

Theorem 11 (Best unconditional upper bound proved here)**.** *For $k \geq 2$,*

$$\log g(k) \leq k + O\left(\frac{k}{\log k}\right).$$

More precisely,

$$\log(g(k) - k) \leq k + O\left(\frac{k}{\log k}\right).$$

Proof. Apply the counting lemma to $A = B(k)$, $M = M(k)$, $K = |B(k)|$, and $\Gamma = \Gamma_k$. We obtain

$$\#\{1 \leq n \leq H : \ n \bmod M(k) \in B(k)\} = \frac{|B(k)|H}{M(k)} + O\big(|B(k)|\Gamma_k(1 + \log H)\big).$$

Choose

$$H := CM(k)\Gamma_k \log\big(2M(k)\Gamma_k\big),$$

with C a sufficiently large absolute constant. Then the main term is

$$\frac{|B(k)|H}{M(k)} = C|B(k)|\Gamma_k \log\big(2M(k)\Gamma_k\big),$$

which dominates the error term for C large enough. Hence there is at least one admissible residue in $[1, H]$, so

$$m_*(k) \ll M(k)\Gamma_k \log\big(2M(k)\Gamma_k\big).$$

It remains to estimate $M(k)$. Since

$$L_p = 1 + \lfloor \log_p k \rfloor,$$

we have

$$\log M(k) = \sum_{p\leq k} L_p \log p = \sum_{p\leq k} \log p + \sum_{p\leq k} \lfloor \log_p k \rfloor \log p = \theta(k) + \psi(k),$$

where

$$\psi(k) := \sum_{p^m \leq k} \log p.$$

By the prime number theorem, $\theta(k) = k + o(k)$ and $\psi(k) = k + o(k)$, hence

$$\log M(k) = 2k + O\left(\frac{k}{\log k}\right).$$

Combining this with the bound for $\log \Gamma_k$ gives

$$\log\bigl(M(k)\Gamma_k\bigr) \le k + O\left(\frac{k}{\log k}\right).$$

Therefore

$$\log m_*(k) \le k + O\left(\frac{k}{\log k}\right),$$

and since $g(k) = k + m_*(k)$, the same bound holds for $\log g(k)$. □

5 What is not yet proved

The density theorem suggests the conjectural estimate

$$\log g(k) \asymp \frac{k}{\log k},$$

but the arguments above do not reach that scale. The obstacle can be stated cleanly.

After dividing out the forced factor

$$Q(k) = \prod_{p \le k} p^{v_p(k+1)},$$

one may write

$$g(k) = k + Q(k)h_*(k).$$

The integer $h_*(k)$ is the least positive element of a reduced product set

$$C(k) \subset \mathbb{Z}/N(k)\mathbb{Z}$$

of density

$$\widetilde{\Delta}(k) = \frac{|C(k)|}{N(k)} = Q(k)\Delta(k),$$

whose local factors are still anchored digit boxes coming from base-p expansions. If one could prove

$$h_*(k) \ll \widetilde{\Delta}(k)^{-1}(\log N(k))^A$$

for some absolute A, then the conjectural upper bound for $g(k)$ would follow from the density theorem.

A Fourier formulation of the same issue is as follows. For $H \asymp N(k)/|C(k)| = \widetilde{\Delta}(k)^{-1}$ one has

$$\#\{1 \le n \le H : \ n \bmod N(k) \in C(k)\} = \frac{|C(k)|H}{N(k)} + \frac{1}{N(k)} \sum_{r \ne 0} \widehat{1_{C(k)}}(r) S_H(-r).$$

The main term is of order 1. Thus one would need a genuine cancellation estimate of the form

$$\sum_{r \ne 0} \widehat{1_{C(k)}}(r) S_H(-r) = o\bigl(N(k)\bigr) \qquad \text{for } H \asymp \widetilde{\Delta}(k)^{-1}.$$

The local Fourier factors are explicit, but the arguments in this note only control absolute values and therefore lose too much phase information. This is the precise remaining obstacle.

6 A remark on the ratio $g(k+1)/g(k)$

No rigorous nontrivial theorem about

$$\liminf_{k\to\infty} \frac{g(k+1)}{g(k)} \qquad \text{or} \qquad \limsup_{k\to\infty} \frac{g(k+1)}{g(k)}$$

is proved in this note.

What can be proved exactly is that the local densities can jump very sharply. If $k = p^a - 1$, then the base-p digits of k are all $p-1$ up to place $a-1$, so

$$\delta_p(k) = p^{-a}.$$

But for $k+1 = p^a$, the base-p digits are $1, 0, \dots, 0$, so

$$\delta_p(k+1) = 1 - \frac{1}{p}.$$

Thus the single-prime local density changes by a factor

$$\frac{\delta_p(k+1)}{\delta_p(k)} = p^a \left(1 - \frac{1}{p}\right) \asymp k.$$

This strongly suggests that $g(k+1)/g(k)$ should fluctuate wildly, but a rigorous proof would again require control of the first admissible residue, not only of the density.

Summary of proved results. The arguments above establish the following rigorously:

- the exact density formulas

$$\delta_p(k) = \prod_{j\ge 0} \left(1 - \frac{k_j^{(p)}}{p}\right), \qquad \Delta(k) = \prod_{p\le k} \delta_p(k);$$

- the density asymptotic

$$\log \Delta(k)^{-1} \asymp \frac{k}{\log k};$$

- the divisibility property

$$Q(k) = \prod_{p\le k} p^{v_p(k+1)} \mid (g(k) - k),$$

  and in particular $g(k) \ge 2k+1$ whenever $k+1$ is composite;

- the unconditional upper bound

$$\log g(k) \le k + O\left(\frac{k}{\log k}\right).$$