.

DIVISIBILITY OF THE CENTRAL BINOMIAL COEFFICIENT $\binom{2n}{n}$

KEVIN FORD AND SERGEI KONYAGIN

ABSTRACT. We show that for every fixed $\ell \in \mathbb{N}$, the set of n with $n^\ell | \binom{2n}{n}$ has a positive asymptotic density c_ℓ, and we give an asymptotic formula for c_ℓ as $\ell \to \infty$. We also show that $\#\{n \leqslant x, (n, \binom{2n}{n}) = 1\} \sim cx/\log x$ for some constant c. We use results about the anatomy of integers and tools from Fourier analysis. One novelty is a method to capture the effect of large prime factors of integers in general sequences.

1. INTRODUCTION

That $(n+1)|\binom{2n}{n}$ for every positive integer n is a consequence of the integrality of the Catalan numbers. In [12], Pomerance raised the question of how frequently $n+k|\binom{2n}{n}$, where k is a fixed integer. Pomerance showed with a simple argument that when k is positive, almost all n have the property $n+k|\binom{2n}{n}$, and the exceptional set up to x is $O(x^{1-a_k})$ for some $a_k > 0$. When $k \leqslant 0$, he proved that the set of such n is governed by the set of such n corresponding to $k=0$; more precisely,

$$\#\left\{n \leqslant x : (n+k)\Big|\binom{2n}{n}\right\} = \#\left\{n \leqslant x : n\Big|\binom{2n}{n}\right\} + O(x^{1-a_k}).$$

Pomerance conjectured that $n|\binom{2n}{n}$ on a set of positive lower density, and showed that it has upper density at most $1-\log 2$; this is an easy consequence of the fact that if n has a prime factor larger than $\sqrt{2n}$, then $n \nmid \binom{2n}{n}$. The upper asymptotic density was later improved by Sanna [13] to $\leqslant 1 - \log 2 - 0.0551$.

Divisibility of $\binom{2n}{n}$ by higher powers of n has also been considered by several people; see the On-line Encyclopedia of Integer Sequences [11], sequences A014847, A121943, A282163, A282672. A283073, and A283074.

Our main result is the following.

Theorem 1. *Fix $\ell \in \mathbb{N}$. The set of n with $n^\ell|\binom{2n}{n}$ has a positive asymptotic density c_ℓ. The density may be computed as follows: Let $U_1, U_2, \dots$ be independent uniform-$[0,1]$ random variables, and let*

$$g_1 = \left\lfloor \frac{1}{U_1} \right\rfloor - 1,\ g_2 = \left\lfloor \frac{1}{(1-U_1)U_2} \right\rfloor - 1,\ \dots,\ g_j = \left\lfloor \frac{1}{(1-U_1)\cdots(1-U_{j-1})U_j} \right\rfloor - 1, \dots. \tag{1.1}$$

Then

$$c_\ell = \mathbb{E} \prod_{j=1}^{\infty} \left(1 - 2^{-g_j} \sum_{h=0}^{\ell-1} \binom{g_j}{h} \right).$$

Numerically, $c_1 \approx 0.114247$, which matches the accumulated data, e.g. [11, Sequence A014847]. See Section 7 for details of the calculation.

Date: September 10, 2019.

The first author was supported in part by National Science Foundation Grant DMS-1802139.

The authors are grateful to the Institute of Mathematics of the Bulgarian Academy of Sciences, which hosted their visit in June, 2018, and where the seeds of this paper were sown. The first author thanks the Institute of Mathematics at the University of Oxford for providing stimulating working conditions during a visit in March-June, 2019.

Theorem 2. *We have*

$$c_\ell \sim \rho\left(2\ell+1-\log(2\ell\log(2\ell)) - \frac{\log\log(2\ell)}{\log 2\ell}\right),$$

as $\ell\to\infty$, where ρ is the Dickman function.

The Dickman function ρ is the unique continuous solution of the differential-delay equation

$$\rho(u)=1 \quad (u\leqslant 1), \quad -u\rho'(u)=\rho(u-1) \quad (u>1). \tag{1.2}$$

Roughly, $\rho(u)$ decays like $1/\Gamma(u)$, and in fact ρ is strictly decreasing for $u>1$ and

$$\rho(u)=e^{-u(\log u+\log\log u+O(1))}. \tag{1.3}$$

Given Theorem 1, a rought heuristic for the values given in Theorem 2 is that the factor

$$1-2^{1-g_j}\sum_{h=0}^{\ell-1}\binom{g_j-1}{h}$$

is close to 1 when g_j is substantially larger than 2ℓ and is close to 0 when g_j is substantilly smaller than 2ℓ. Thus, c_ℓ should be close to the probability that $g_j\geqslant 2\ell$ for all j, which equals $\rho(2\ell)$.

In [13], Sanna considered the set $\mathcal{B}$ of positive integers n such n and $\binom{2n}{n}$ are coprime and showed that $\#(\mathcal{B}\cap[1,x]) \ll x/\sqrt{\log x}$ for all $x>1$. On the other hand, $\mathcal{B}$ contains all odd primes, and thus $\#(\mathcal{B}\cap[1,x]) \gg x/\log x$ for all $x\geqslant 2$. We sharpen these results be proving an asymptotic formula for $\#(\mathcal{B}\cap[1,x])$.

Theorem 3. *We have $\#\{n\leqslant x : (n,\binom{2n}{n})=1\}\sim cx/\log x$, where*

$$c=\sum_{k=1}^{\infty}\frac{1}{k!}\int\cdots\int_{\substack{u_i\geqslant 0\ \forall i\\ u_1+\cdots+u_k=1}} h(u_1)\cdots h(u_k)\,du_1\cdots du_k, \qquad h(x)=x^{-1}2^{1-\lfloor 1/x\rfloor}. \tag{1.4}$$

As h is bounded, the series for c converges rapidly. Numerically, $c=1.526453\ldots$ (See section 9).

1.1. **Heuristics.** For most n, the divisibility condition $n^\ell|\binom{2n}{n}$ is essentially determined by the largest prime factors of n. By Kummer's criterion (1852), if p is prime, then $p^\ell|\binom{2n}{n}$ if and only the addition of n and n in base-p has at least ℓ carries. This is equivalent to $\{n/p^s\}>\frac12$ for at least ℓ values of $s\in\mathbb{N}$. If p is large, then this means (essentially) that the base-p expansion of n has at least ℓ digits which are $\geqslant \frac{p-1}{2}$ (if a digit equals $\frac{p-1}{2}$, then it may or may not induce a carry). Supposing that $p\|n$, the final base-p digit is zero, and the leading digit is $<p/2$ with high probability. There are $k=\left\lfloor\frac{\log n}{\log p}\right\rfloor-1$ remaining base-p digits, and if these are randomly distributed (over all $n\leqslant x$ divisible by p and not by p^2) then we expect that $p^\ell|\binom{2n}{n}$ occurs with probability close to

$$1-2^{1-k}\sum_{h=0}^{\ell-1}\binom{k-1}{h}.$$

Donelly and Grimmett [3] (see also [14]) proved that the largest prime factors of a random integer have, asyptotically, the Poisson-Dirichlet distribution. A realization of this distribution is given in terms of independent uniform-$[0,1]$ random variables $U_1,U_2,\ldots$. Let $(X_1,X_2,\ldots)$ be the infinite dimensional vector formed from the decreasing rearrangement of the numbers

$$Y_1=U_1, Y_2=(1-U_1)U_2, Y_3=(1-U_1)(1-U_2)U_3,\ldots. \tag{1.5}$$

Then $(X_1, X_2, \ldots)$ has the Poisson-Dirichlet distribution. Let $p_j(n)$ denote the j-th largest prime factor of n. The paper [3] gives a simple, transparent proof that $(X_1, \ldots, X_k)$ and

$$\left(\frac{\log p_1(n)}{\log n}, \ldots, \frac{\log p_k(n)}{\log n}\right)$$

have identical distributions (asymptotically as $x \to \infty$, where n is drawn at random from $[1, x]$). For a discussion of other realizations of the Poisson-Dirichlet distribution, see Section 1 of [14]. Combining this with our heuristic above about divisibility of $\binom{2n}{n}$ by p^ℓ, we arrive at Theorem 1.

The heuristic for Theorem 3 is simpler. If n has k prime factors $p_1, \ldots, p_k$, with $p_i = x^{u_i}$, then we expect $(n, \binom{2n}{n}) = 1$ with probability $\prod_{i=1}^k 2^{1-\lfloor 1/u_i \rfloor}$. Summing over all $p_1, \ldots, p_k$ with the prime number theorem yields the result in Theorem 3.

We will make both of these heuristics precise utilizing harmonic analysis to detect the simultaneous divisibility of $\binom{2n}{n}$ by large prime factors of n. Section 3 contains the relevant estmates. In Section 2, we show that the small prime factors of n divide $\binom{2n}{n}$ with very high probability, and can safely be ignored. We prove a result about simultaneous fraction parts of quotients of primes in Section 4 that will be needed for Theorems 1 and 3. The proof of Theorem 1 occupies Section 5 and we prove Theorem 3 in Section 6. Sections 7 and 8 are devoted to the study of the constants c_ℓ, culminating in the proof of Theorem 2. Finally, we desribe how to compute c accurately in Section 9.

2. SMALL PRIME FACTORS

In this section, we will see that only the largest prime factors of n matter for Theorems 1 and 3.

Lemma 2.1. *Let p be prime, $v \in \mathbb{N}$, $\ell \in \mathbb{N}$ and $p^{\ell v} \leqslant x^{1/100}$. Then*

$$\#\left\{n \leqslant x : p^v | n,\, p^{\ell v} \nmid \binom{2n}{n}\right\} \ll \frac{x^{1-\frac{1}{3\log p}}}{p^v} e^{v/3}.$$

Proof. Suppose that $n \leqslant x$ and $p^v | n$. Write n in base-p as $n = (b_D b_{D-1} \cdots b_0)_p$, where $D = \left\lfloor \frac{\log x}{\log p} \right\rfloor$, so that $b_0 = \cdots = b_{v-1} = 0$. Also observe that the hypotheses imply that $D \geqslant 100v$ and hence that

$$\ell v \leqslant \frac{\log x}{100 \log p} \leqslant \frac{D+1}{100} < \frac{D}{99} \leqslant \frac{D-v}{98}.$$

The number of choices for b_D is at most x/p^D. By Kummer's criterion, if $p^{\ell v} \nmid \binom{2n}{n}$, then at most $\ell v - 1$ of the digits $b_v, \ldots, b_{D-1}$ are $\geqslant \frac{p}{2}$. Hence, the number of choices for $(b_v, \ldots, b_{D-1})$ is at most

$$\sum_{j=0}^{\ell v - 1} \binom{D-v}{j} \left(\frac{p-1}{2}\right)^j \left(\frac{p+1}{2}\right)^{D-v-j} \ll \left(\frac{p+1}{2}\right)^{D-v} \binom{D-v}{\ell v}$$

if $p \geqslant 3$, and $O(\binom{D-v}{\ell v})$ when $p = 2$. Recalling that $\ell v \leqslant (D-v)/98$, by Stirling's formula we have

$$\binom{D-v}{\ell v} \ll e^{0.057(D-v)}$$

and thus

$$\#\left\{n \leqslant x : p^v | n, p^{\ell v} \nmid \binom{2n}{n}\right\} \ll \frac{x}{p^v} \left(\frac{e^{0.057}(1+\frac{1}{3})}{2}\right)^{D-v} \ll \frac{x}{p^v} e^{-(D-v)/3},$$

and the claimed inequality follows. ∎

Proposition 1. *For large x, let δ satisfy $0 < \delta \leqslant 1$. For any $1 \leqslant n \leqslant x$, write $n = A_n B_n$, where $P^+(A_n) \leqslant x^\delta < P^-(B_n)$. Fix $\ell \in \mathbb{N}$. Then*

$$\#\left\{n \leqslant x : A_n^\ell \nmid \binom{2n}{n}\right\} \ll_\ell x\mathrm{e}^{-1/(300\ell\delta)}.$$

Proof. We may assume that $\frac{\log 2}{\log x} < \delta \leqslant 1/(300\ell)$, else the statement is trivial. Hence, by Lemma 1,

$$\begin{aligned}\#\left\{n \leqslant x : A_n^\ell \nmid \binom{2n}{n}\right\} &\leqslant \sum_{p \leqslant x^\delta}\left[\sum_{v \leqslant \frac{\log x}{100\ell \log p}} \#\left\{n \leqslant x : p^v | n, p^{\ell v} \nmid \binom{2n}{n}\right\} + \sum_{v > \frac{\log x}{100\ell\log p}} \frac{x}{p^v}\right] \\ &\ll \sum_{p \leqslant x^\delta}\left[x^{1-1/(3\log p)} \sum_{v \leqslant \frac{\log x}{100\ell\log p}} \frac{e^{v/3}}{p^v} + x^{1-\frac{1}{100\ell}}\right] \\ &\ll x^{1+\delta-\frac{1}{100\ell}} + x\sum_{p\leqslant x^\delta} \frac{x^{-1/(3\log p)}}{p} \\ &\ll x^{1-\frac{1}{150\ell}} + xe^{-\frac{1}{3\delta}} \\ &\ll xe^{-\frac{1}{300\ell\delta}}.\end{aligned}$$

□

Next, we prove analogous bounds for integers with a given smallest prime factor.

Proposition 2. *The number of integer $n \leqslant x$ for which $(n, \binom{2n}{n}) = 1$ and n has a prime factor smaller than n^δ is $O(\frac{x}{\log x} e^{-1/(3\delta)})$.*

Proof. Fix p and consider those n with smallest prime factor p and such that $p \nmid \binom{2n}{n}$. We argue as in the $\ell = 1$ case of Lemma 1, except that for fixed $b_2, \ldots, b_D$ we bound the number of possible b_1 such that $\sum_{j=1}^D p^j b_j$ has no prime factor less than p with a sieve (e.g., [7, Theorem 2.2]), obtaining

$$\#b_1 \ll \frac{p}{\log p}.$$

It follows that

$$\#\left\{n \leqslant x : n \text{ has smallest prime factor } p, p \nmid \binom{2n}{n}\right\} \ll \frac{x^{1-\frac{1}{3\log p}}.}{p \log p}.$$

Summing over $p \leqslant x^\delta$ completes the proof. □

3. Exponential sum estimates

We gather together in this section various estimates for exponential sum which we will need for the proof of Theorem 1.

The first lemma is the 'Weyl-van der Corput inequality' (see Theorems 2.2, 2.8 in [5]). It is far from the best result of its kind, but has a relatively short proof and suffices for our purposes.

Lemma 3.1. *Let $j \geqslant 2$ be an integer, let I be an interval and suppose that $f \in C^j(I)$ and that*

$$\lambda \leqslant |f^{(j)}(x)| \leqslant \alpha\lambda$$

where $\lambda > 0$, $\alpha \geqslant 1$. Then

$$\sum_{n\in I} e(f(n)) \ll |I|(\alpha^2\lambda)^{\frac{1}{4J-2}} + |I|^{1-\frac{1}{2J}}\alpha^{\frac{1}{2J}} + |I|^{1-\frac{2}{J}+\frac{1}{J^2}}\lambda^{-\frac{1}{2J}},$$

where $J = 2^{j-2}$.

We apply this lemma to bound a certain class of exponential sums.

Lemma 3.2. *Let $N \in \mathbb{N}$, and*

$$f(u) = \alpha u + \sum_{r=r_1}^{r_2} \frac{\beta_r}{u^r}, \tag{3.1}$$

where $\alpha \in \mathbb{R}$, $1 \leqslant r_1 \leqslant r_2$, and for some $A \in [1, N^{1/2}]$ we have

$$|\beta_{r_1}| \geqslant N^{r_1}A, \quad |\beta_r/\beta_{r_1}| \leqslant N^{(r-r_1)/2} \ (r_1 \leqslant r \leqslant r_2). \tag{3.2}$$

Then

$$\max_{I\subset(N,2N]} \sum_{n\in I} e(f(n)) \ll_{r_2} N\left(N^{-1/2^j} + A^{-1/4}\right),$$

where

$$j = 3 + \left\lfloor \frac{\log\left(\frac{|\beta_{r_1}|}{AN^{r_1}}\right)}{\log N}\right\rfloor. \tag{3.3}$$

Proof. We apply Lemma 3.1. Firstly, we may assume that N is sufficiently large and that

$$j \leqslant \frac{\log\log N}{\log 2}, \tag{3.4}$$

for otherwise the conclusion is trivial. Also note that $j \geqslant 3$. Denoting by $r^{(j)}$ the rising factorial $r(r+1)\cdots(r+j-1)$, and using (3.2), we have for $N < u \leqslant 2N$ the relation

$$\begin{aligned}
f^{(j)}(u) &= (-1)^j \sum_{r=r_1}^{r_2} \frac{r^{(j)}\beta_r}{u^{r+j}} \\
&= (-1)^j \frac{r_1^{(j)}\beta_{r_1}}{u^{r_1+j}}\left(1 + O\left(\sum_{r=r_1}^{r_2} \frac{(r^{(j)}/r_1^{(j)})|\beta_r/\beta_{r_1}|}{N^{r-r_1}}\right)\right) \\
&= (-1)^j \frac{r_1^{(j)}\beta_{r_1}}{u^{r_1+j}}\left(1 + O\left(\sum_{r=r_1}^{r_2} \frac{(r/r_1)^j}{N^{(r-r_1)/2}}\right)\right) \\
&= \left(1 + O_{r_2}\left(N^{-1/2}\right)\right)(-1)^j \frac{r_1^{(j)}\beta_{r_1}}{u^{r_1+j}}.
\end{aligned}$$

For large enough N it follows that

$$\lambda \leqslant |f^{(j)}(u)| \leqslant \alpha\lambda, \quad \lambda = \frac{r_1^{(j)}|\beta_{r_1}|}{2(2N)^{r_1+j}}, \quad \alpha = 2^{r_1+j+2}.$$

Inserting this bound into Lemma 3.1, we have

$$\frac{1}{N}\sum_{n\in I} e(f(n)) \ll_{r_2} \lambda^{\frac{1}{4J-2}} + N^{-\frac{1}{2J}} + N^{-\frac{2}{J}+\frac{1}{J^2}}\lambda^{-\frac{1}{2J}}, \tag{3.5}$$

where $J = 2^{j-2}$. We note that from (3.2) and the definition of j,

$$N^2 \frac{|\beta_{r_1}|}{AN^{r_1}} \leqslant N^j \leqslant N^3 \frac{|\beta_{r_1}|}{AN^{r_1}}$$

and hence that

$$\frac{A}{2^{r_1+j+1}N^3} \leqslant \lambda \leqslant r_1^{(j)} \left(\frac{A}{N^2}\right) \leqslant r_1^{(j)} N^{-3/2}.$$

When $j = 3$, therefore, the right side of (3.5) is

$$\ll_{r_2} \lambda^{1/6} + N^{-1/4} + N^{-3/4}\lambda^{-1/4} \ll N^{-1/4} + A^{-1/4}.$$

Now assume that $j \geqslant 4$ so that $J \geqslant 4$. Then the right side of (3.5) is

$$\ll_{r_2} N^{-\frac{3/2}{4J-2}} + N^{-\frac{1}{2J}} + N^{-\frac{7}{4J}}(N^3)^{\frac{1}{2J}} \ll_{r_2} N^{-\frac{1}{4J}}.$$

Combining the two cases, $j = 3$ and $j > 3$, this concludes the proof. □

We now apply Lemma 3.2 to bound analogous sums over primes.

Lemma 3.3. *Assume f satisfies* (3.1)*, where the coefficients satisfy* (3.2) *for some $A \in [1, N^{1/6}]$. Then*

$$\max_{I\subset(N,2N]} \sum_{p\in I} e(f(p)) \ll_{r_2} N(\log N)^4 \big(N^{-\frac{1}{3\cdot 2^j}} + A^{-1/10}\big),$$

where j is given by (3.3).

Proof. Our technique is standard. Throughout, constants implied by $O-$ and $\ll$- may depend on r_1, r_2. We begin by applying Vaughan's identity (see, e.g. [2, Ch. 24]) to the exponential sum in question, obtaining

$$\sum_{p\in I} e(f(p)) = O(N^{1/2}) - S_1 + S_2 + S_3,$$

where

$$S_1 = \sum_{a\leqslant N^{1/3}} \Lambda(a) \sum_{b\leqslant N^{1/3}} \mu(b) \sum_{abc\in I} e(f(abc)),$$

$$S_2 = \sum_{b\leqslant N^{1/3}} \sum_{bc\in I} \log(bc) e(f(bc)),$$

$$S_3 = \sum_{b>N^{1/3}} h(b) \sum_{\substack{bc\in I \\ c>N^{1/3}}} \Lambda(c) e(f(bc)),$$

where

$$h(b) = \sum_{\substack{d|b \\ d>N^{1/3}}} \mu(d).$$

Both S_1 and S_2 are "Type I" sums and we may apply Lemma 3.2 directly. For S_1, we fix a and b and apply Lemma 3.2 with N replaced by N/ab and β_r replaced by $\beta_r/(ab)^r$. We check that

$$A \leqslant N^{1/6} \leqslant (N/ab)^{1/2}, \qquad \left|\frac{\beta_r'}{\beta_{r_1}'}\right| = \left|\frac{\beta_r}{\beta_{r_1}}\right| (ab)^{-(r-r_1)} \leqslant \left(\frac{N}{ab}\right)^{(r-r_1)/2}.$$

Thus, for any a, b we have

$$\sum_{abc\in I} e(f(abc)) \ll \frac{N}{ab}\big((N/ab)^{-1/2^j} + A^{-1/4}\big)$$

and hence that

$$S_1 \ll N(\log^2 N)\big(N^{-\frac{1}{3\cdot 2^j}} + A^{-1/4}\big). \tag{3.6}$$

Bounding the inner sum over c in S_2 is exactly analogous, where we use partial summation to remove the logarithm factor. Since $N/b \geqslant N^{2/3}$, we obtain a stronger bound

$$S_2 \ll N(\log^2 N)\big(N^{-\frac{2}{3\cdot 2^j}} + A^{-1/4}\big). \tag{3.7}$$

For S_3, we break up the range $b \in (N^{1/3}, 2N^{2/3}]$ into $O(\log N)$ dyadic intervals of the form $(B, 2B]$ where $N^{1/3} \leqslant B \leqslant 2N^{2/3}$. Then we use Cauchy-Schwarz, followed by the trivial bound $|h(b)| \leqslant \tau(b)$ to get

$$\begin{aligned} S_3 &\ll (\log N) \max_B \Big| \sum_{B<b\leqslant 2B} h(b) \sum_{bc\in I} \Lambda(c)e(f(bc))\Big| \\ &\leqslant (\log N) \max_B \Big(\sum_{B<b\leqslant 2B} h(b)^2\Big)^{1/2} \Big(\sum_{B<b\leqslant 2B} \Big|\sum_{bc\in I} \Lambda(c)e(f(bc))\Big|^2\Big)^{1/2} \\ &\ll (\log N)^{5/2} \max_B B^{1/2} \Big(\sum_{B<b\leqslant 2B} \Big|\sum_{bc\in I} \Lambda(c)e(f(bc))\Big|^2\Big)^{1/2}. \end{aligned}$$

Next, we expand the square and then interchange the order of summation:

$$\sum_{B<b\leqslant 2B} \Big|\sum_{bc\in I} \Lambda(c)e(f(bc))\Big|^2 = \sum_{\frac{N}{2B}<c_1,c_2\leqslant \frac{2N}{B}} \Lambda(c_1)\Lambda(c_2) \sum_{b\in J} e(f(bc_1) - f(bc_2)), \tag{3.8}$$

where

$$J = \{B < n \leqslant 2B : bc_1 \in I, bc_2 \in I\}$$

is a subinterval of $(B, 2B]$. Let R be a large constant, depending on r_1, r_2. The terms above with $|c_1 - c_2| \leqslant \frac{RN}{BA^{1/5}}$ contribute at most $O(N^2(\log N)^2/(A^{1/5}B))$ to the right side of (3.8). Now suppose that $|c_1 - c_2| > \frac{RN}{BA^{1/5}}$. Write

$$f(bc_1) - f(bc_2) = \alpha b(c_1 - c_2) + \sum_{r=r_1}^{r_2} \frac{\beta_r'}{b^r}, \qquad \beta_r' = \beta_r\Big(\frac{1}{c_1^r} - \frac{1}{c_2^r}\Big).$$

We apply Lemma 3.2 with β_r replaced by β_r', N replaced by B and A replaced by

$$A' = \frac{AN^{r_1}\beta_{r_1}'}{B^{r_1}\beta_{r_1}}.$$

Since

$$|\beta_r'| \asymp |\beta_r| \frac{|c_1 - c_2|}{c_1^{r+1}},$$

we see that

$$\left|\frac{\beta_r'}{\beta_{r_1}'}\right| \ll N^{-(r-r_1)/2} c_1^{-(r-r_1)} \ll B^{-(r-r_1)/2}(N/B)^{-(r-r_1)/2},$$

so that the hypotheses (3.2) hold. Also, $A' \geqslant A^{4/5}$ if R is large enough, and therefore

$$\sum_{b\in J} e(f(bc_1) - f(bc_2)) \ll B\big(B^{-1/2^j} + A^{-1/5}\big).$$

Summing over all pairs c_1, c_2 we see that the expression in (3.8) is

$$\ll \frac{N^2}{B}(\log N)^2(N^{-1/(3\cdot 2^j)} + A^{-1/5}),$$

and we conclude that

$$S_3 \ll N(\log N)^4\big(N^{-\frac{1}{3\cdot 2^j}} + A^{-1/10}\big). \tag{3.9}$$

Combining (3.6), (3.7) and (3.9), this completes the proof. □

4. DETECTING FRACTIONAL PARTS

In this section we apply harmonic analysis to detect the simultaneous fractional parts of ratios of primes. Denote by $\{x\}$ the fractional part of x.

We begin with a result of Selberg.

Lemma 4.1. *For any $K \in \mathbb{N}$ and any non-empty interval $I \subset \mathbb{R}/\mathbb{Z}$, there is a trigonometric polynomial $S^+_{K,I}(x) = \sum_{|n|\leqslant K} a_n e(nx)$ which majorizes the indicator function of I and a trigonometric polynomial $S^-_{K,I}(x) = \sum_{|n|\leqslant K} b_n e(nx)$ which minorizes the indicator function of I, and which satisfy the following:*

- $\max(|a_n|, |b_n|) \leqslant 4/(|n|+1)$ *for all n.*
- $\int_0^1 S_{K,I}(x)^{\pm}\, dx = \textit{length}(I) \pm \frac{1}{K+1}$.

Proof. For details and explicit construction of $S^{\pm}_{K,I}$, see Chapter 1 in [10], especially formulas (16)–(22). □

Definition. A subset $\mathcal{R}$ of $\mathbb{R}^k$ is said to be *t-simple* if, for any $1 \leqslant j \leqslant k$ and any choice of $z_i \in \mathbb{R}$ $(i \neq j)$, the 1-dimensional projection $\{z_j : (z_1, \dots, z_k) \in \mathcal{R}\}$ consists of at most t disjoint intervals.

Proposition 3. *Fix ε, ρ such that $0 < \rho < \varepsilon$ and let $k \in \mathbb{N}$ with $k < 1/\varepsilon$. Suppose that $1 \leqslant m \leqslant x^{1/2}$, and $M_1, \dots, M_k$ are integers such that*

(i) *$M_i \geqslant x^{\varepsilon}$ for all i;*
(ii) *$x/2^k < M_1 \cdots M_k m \leqslant 2x$;*
(iii) *for all i, $M_i \notin \bigcup_{s\leqslant 1/\varepsilon+1} (x^{(1-\rho)/s}, 4x^{1/s}]$.*

Let $\mathcal{R}$ be any $t-$simple subset of

$$\{(x_1, \dots, x_k) : M_i < x_i \leqslant 2M_i \ (1 \leqslant i \leqslant k), x < mx_1 \cdots x_k \leqslant 2x\}.$$

and let $\mathcal{Q}$ denote the set of all k-tuples $\mathbf{q} = (q_1, \dots, q_k)$ of primes such that $\mathbf{q} \in \mathcal{R}$. For each $1 \leqslant j \leqslant k$, let $s_j = \left\lfloor \frac{\log x}{\log M_j} \right\rfloor - 1$. Then, for some $\xi > 0$, which depends only on ε,ρ and k, we have (writing $n = q_1 \cdots q_k m$)

$$\#\left\{\mathbf{q} \in \mathcal{Q} : \forall j, q_j^{\ell} \Big| \binom{2n}{n}\right\} = (1+O(\varepsilon)) \prod_{j=1}^{k} \left(1 - 2^{-s_j} \sum_{h=0}^{\ell-1} \binom{s_j}{h}\right) |\mathcal{Q}| + O_{k,\varepsilon}\left(\frac{tx^{1-\xi}}{m}\right), \tag{4.1}$$

$$\#\left\{\mathbf{q} \in \mathcal{Q} : \forall j, q_j \nmid \binom{2n}{n}\right\} = \frac{1+O(\varepsilon)}{2^{s_1+\cdots+s_k}} |\mathcal{Q}| + O_{k,\varepsilon}\left(\frac{tx^{1-\xi}}{m}\right). \tag{4.2}$$

Proof. The number of $\mathbf{q}$ such that $q_i | m$ for some i is $O((k \log x)x^{1-\varepsilon}/m)$, hence we may ignore these. For each $1 \leqslant j \leqslant k$ and $1 \leqslant s \leqslant s_j$, let $\sigma_{j,s} \in \{0, 1\}$, and denote by Σ the vector of the numbers $\sigma_{j,s}$. It is possible that $s_j = 0$ for some j, in which case terms $\sigma_{j,s}$ do not appear. For each Σ let

$$\mathcal{Q}_{\Sigma} := \left\{\mathbf{q} \in \mathcal{Q} : \left\{\frac{mq_1 \cdots q_k}{q_j^{s+1}}\right\} \in \left[\frac{\sigma_{j,s}}{2}, \frac{1+\sigma_{j,s}}{2}\right) \ (1 \leqslant j \leqslant k, 1 \leqslant s \leqslant s_j)\right\}.$$

Our main task is to prove that

$$|Q_{\Sigma}| = \frac{(1+O(\varepsilon)}{2^{s_1+\cdots+s_k}} |\mathcal{Q}| + O_{k,\varepsilon}\left(\frac{tx^{1-\xi}}{m}\right). \tag{4.3}$$

For each $\mathbf{q} \in \mathcal{Q}$, let $n = mq_1 \cdots q_k$. Since $s_k + 1 \leqslant 1/\varepsilon$, (iii) implies that $\{n/q_j^{s_j+2}\} < 1/2$. Therefore, $\binom{2n}{n}$ is divisible by q_j^ℓ if and only if for at least ℓ values of $s \in \{1, 2, \ldots, s_j\}$ we have $\{n/q_j^{s+1}\} > 1/2$. Likewise, $q_j \nmid \binom{2n}{n}$ if and only if $\{n/q_j^{s+1}\} < 1/2$ for every $1 \leqslant s \leqslant s_j$. Hence, the left side of (4.1) is the sum of $\mathcal{Q}_\Sigma$ over all Σ such that $\sum_s \sigma_{j,s} \geqslant \ell$ for all j, and the left side of (4.2) equals $\mathcal{Q}_\Sigma$ for the single Σ with $\sigma_{j,s} = 0$ for all j, s. Thus, (4.1) and (4.2) follow from (4.3).

Fix Σ. We apply Lemma 4.1 to the intervals $[0, 1/2]$ and $[1/2, 1]$ and with

$$K = \lfloor k\varepsilon^{-2} \rfloor.$$

Define

$$\psi_{0,K}^{\pm}(x) = S_{K,[0,1/2]}^{\pm}(x) = \sum_{|n| \leqslant K} c_{0,n}^{\pm} e(nx),$$
$$\psi_{1,K}^{\pm}(x) = S_{K,[1/2,1]}^{\pm}(x) = \sum_{|n| \leqslant K} c_{1,n}^{\pm} e(nx).$$

Then

$$\sum_{\mathbf{q} \in \mathcal{Q}} \prod_{j=1}^{k} \prod_{s=1}^{s_j} \psi_{\sigma_{j,s},K}^{-}(mq_1 \cdots q_k/q_j^{s+1}) \leqslant |\mathcal{Q}_\Sigma| \leqslant \sum_{\mathbf{q} \in \mathcal{Q}} \prod_{j=1}^{k} \prod_{s=1}^{s_j} \psi_{\sigma_{j,s},K}^{+}(mq_1 \cdots q_k/q_j^{s+1}). \tag{4.4}$$

Denote by $\boldsymbol{\lambda}$ the vector $(\lambda_{j,s} : 1 \leqslant j \leqslant k, 1 \leqslant s \leqslant s_j)$, where each component is bounded by K in absolute value. Focusing on the lower bound (the upper bound analysis is identical), we then have

$$|\mathcal{Q}_\Sigma| \geqslant \sum_{\mathbf{q} \in \mathcal{Q}} \sum_{\boldsymbol{\lambda}} \bigg(\prod_{j,s} c_{\sigma_{j,s},\lambda_{j,s}}^{-} \bigg) e\bigg(m \sum_{j,s} \lambda_{j,s} \frac{q_1 \cdots q_k}{q_j^{s+1}} \bigg). \tag{4.5}$$

Using Lemma 4.1, we find that the main term ($\lambda_{j,s} = 0$ for every j, s) equals

$$|\mathcal{Q}| \prod_{j,s} \bigg(\int_0^1 \psi_{\sigma_{j,s},K}^{-}(u)\, du \bigg) = \frac{|\mathcal{Q}|}{2^{s_1+\cdots+s_k}} (1 + O(1/K))^{s_1+\cdots+s_k} = \frac{1 + O(\varepsilon)}{2^{s_1+\cdots+s_k}} |\mathcal{Q}|.$$

By Lemma 4.1, $\sum_n |c_{\sigma,n}^{\pm}| \ll \log K$ and therefore we have

$$|\mathcal{Q}_\Sigma| \geqslant (1 + O(\varepsilon)) \frac{|\mathcal{Q}|}{2^{s_1+\cdots+s_k}} + E, \tag{4.6}$$

where

$$E \ll (O(\log K))^{O(k/\varepsilon)} \max_{\boldsymbol{\lambda} \neq \mathbf{0}} \bigg| \sum_{\mathbf{q} \in \mathcal{Q}} e\bigg(m \sum_{j,s} \lambda_{j,s} \frac{q_1 \cdots q_k}{q_j^{s+1}} \bigg) \bigg|.$$

Fixing $\boldsymbol{\lambda} \neq \mathbf{0}$, let $h = \min\{j \leqslant k : \lambda_{j,s} \neq 0 \text{ for some } s\}$ and define $r = \min\{s : \lambda_{h,s} \neq 0\}$. Fixing q_i $(i \neq h)$, the t-simplicity of $\mathcal{R}$ implies that the variable q_h ranges over primes in at most t subintervals I (possibly $t = 0$) of $(M_h, 2M_h]$. We have

$$\sum_{j,s} \lambda_{j,s} \frac{q_1 \cdots q_k m}{q_j^{s+1}} = \alpha q_h + \sum_{s=r}^{s_h} \lambda_{h,s} \frac{P}{q_h^s} =: f(q_h).$$

for some real number α (depending on m and the q_i for $i \neq h$) and $P = (q_1 \cdots q_k m)/q_h$. By (ii) and (iii),

$$P \geqslant \frac{M_1 \cdots M_k m}{M_h} \geqslant \frac{x}{2^k M_h} \geqslant x^\rho 2^{-k} M_h^{s_h}. \tag{4.7}$$

We also have $|\lambda_{h,s}| \leqslant K \ll M_h^{1/10}$ for large x. Therefore, for each interval I we may apply Lemma 3.3 with

$$N = M_h, \quad r_1 = r, \quad \beta_{r_1} = P\lambda_{h,r}, \qquad A = 2^{-k}x^\rho.$$

The condition $\beta_{r_1} \geqslant N^{r_1}A$ follows from (4.7), and the lower bound $M_h \geqslant x^\varepsilon$ implies that $A \leqslant M_h$, so that (3.2) holds. We also have that

$$j \leqslant 3 + \frac{\log(KP)}{\log M_h} \leqslant 3 + \frac{\log x}{\log M_h} \leqslant 3 + 1/\varepsilon.$$

Therefore, applying Lemma 3.3, we get

$$\sum_{q_h \in I} e(f(q_h)) \ll_k M_h (\log M_h)^4 \left(M_h^{-\frac{1}{3\cdot 2^j}} + x^{-\rho/4} \right) \ll x^{-\xi} M_h.$$

Summing over all q_i $(i \neq h)$, we find that $E \ll_{k,\varepsilon} tx^{1-\xi}$. Combined with (4.6), this completes the proof of (4.3). □

5. PROOF OF THEOREM 1

Throughout this section, we will assume that k is a large integer, and that ε, δ are functions of k that tend to 0 as $k \to \infty$; precisely, we take

$$\delta = e^{-2k/3}, \qquad \varepsilon = k^{-2k}. \tag{5.1}$$

Suppose that x is a large integer. We think of k being fixed and $x \to \infty$. In this section only, we adopt the following notation for functions $f(k,x)$. The notation $f(k,x) = o(g(k,x))$ means that

$$\forall k \geqslant 1 : \lim_{x\to\infty} \frac{f(k,x)}{g(k,x)} = 0.$$

The notation $f(k,x) = \overline{o}(g(k,x))$ means that

$$\lim_{k\to\infty} \limsup_{x\to\infty} \frac{f(x,k)}{g(x,k)} = 0.$$

For example, $1/k = \overline{o}(1)$ and $e^k x^{1-1/k} = o(x)$.

5.1. **Sampling large prime factors.** Take a large integer x, and select a random integer $n \in (x, 2x]$ with uniform probability. Following Donnelly and Grimmett [3], we select at random a k-tuple $\mathbf{q}(n) = (q_1, \ldots, q_k)$ of divisors of n at random, in a size-biased fashion, together with random variables $X_1(n), \ldots, X_k(n)$. If n has fewer than k distinct prime factors, set $\mathbf{q}(n) = (1, \ldots, 1)$ and $X_1(n) = \cdots = X_k(n) = 0$. Otherwise, choose $q_1 | n$ at random with probability $\frac{\Lambda(q_1)}{\log n}$, where Λ is the von Mangoldt function. For $2 \leqslant i \leqslant k$, once $q_1, \ldots, q_{i-1}$ are chosen, select $q_i | (n/q_1 \cdots q_{i-1})$ with probability $\frac{\Lambda(q_i)}{\log n/(q_1\cdots q_{i-1})}$. Then set $X_i(n) = \frac{\Lambda(q_i)}{\log n/(q_1\cdots q_{i-1})}$ for $1 \leqslant i \leqslant k$. We observe the relation

$$q_i = n^{(1-X_1(n))\cdots(1-X_{i-1}(n))X_i(n)} \qquad (1 \leqslant i \leqslant k). \tag{5.2}$$

The following is essentially Theorem 1 of [3], although we have stated the result with a slight modification. For completeness, a proof is given in the Appendix.

Lemma 5.1. *Fix $k \in \mathbb{N}$. As $x \to \infty$, the random vector $(X_1(n), \ldots, X_k(n))$ converges weakly to the uniform distribution (that is, Lebesgue measure) on $[0,1]^k$.*

We denote $\mathbb{P}_x$, $\mathbb{E}_x$ for the probability, respectively expectation, with respect to these random n, $\mathbf{q}(n)$ and $(X_1(n), \dots, X_k(n))$, and use $\mathbb{P}$ and $\mathbb{E}$ for the uniform probability measure on $[0,1]^k$. For the latter, we work with independent, uniform-$[0,1]$ random variables $U_1, \dots, U_k$.

Definition. With x fixed, let $\mathcal{Y}_k(x)$ denote the set of k-tuples $\mathbf{y} = (y_1, \dots, y_k) \in [1, x]^k$ such that

(a) $y_i \geqslant x^\varepsilon$ for all i;
(b) $x^{1-\delta} \leqslant y_1 \cdots y_k \leqslant x^{1-\delta^2}$;
(c) for all i and all $1 \leqslant s \leqslant 1/\varepsilon + 1$, $y_i \notin [x^{(1-\varepsilon^2)/s}, 4x^{1/s}]$.

Lemma 5.2. *The set $\mathcal{Y}_k(x)$ is $(1/\varepsilon + 2)$-simple.*

Proof. Fix j and let y_i be arbitrary for $i \neq j$. Items (a) and (b) force y_j into a single interval, from which are cut at most $1/\varepsilon + 1$ intervals by (c). □

Lemma 5.3. *We have $\mathbb{P}_x(\mathbf{q}(n) \notin \mathcal{Y}_k(x)$ or some q_i not prime$) = \overline{o}(1)$.*

Proof. First, note that $\mathbb{P}_x(n \text{ has fewer than } k \text{ prime factors}) = o(1)$. Now assume that n has at least k distinct prime factors. By (5.2) and Lemma 5.1,

$$\begin{aligned}\mathbb{P}_x(\text{some } q_i < x^\varepsilon) &\leqslant \mathbb{P}_x(\text{some } q_i \leqslant n^\varepsilon)\\ &\leqslant \mathbb{P}\big((1-U_1)\cdots(1-U_{i-1})U_i \leqslant \varepsilon \text{ for some } i\big) + o(1)\\ &\leqslant \mathbb{P}\big(U_i \notin [\varepsilon^{1/k}, 1-\varepsilon^{1/k}] \text{ for some } i\big) + o(1)\\ &\leqslant 2k\varepsilon^{1/k} + o(1) = \overline{o}(1),\end{aligned}$$

upon recalling (5.1).

From (5.2), we have

$$q_1 \cdots q_k = n^{1-(1-X_1(n))\cdots(1-X_k(n))}.$$

Hence,

$$\mathbb{P}_x\big(x^{1-\delta} \leqslant q_1(n)\cdots q_k(n) \leqslant x^{1-\delta^2}\big) = \mathbb{P}_x\left(\frac{\log n}{\log x}\big(1-(1-X_1(n))\cdots(1-X_k(n))\big) \in [1-\delta, 1-\delta^2]\right).$$

By Lemma 5.1, as $k \to \infty$, the variable $1-(1-X_1(n))\cdots(1-X_k(n))$ converges in distribution to $1-(1-U_1)\cdots(1-U_k)$. Now $\mathbb{E}\log(1-U_i) = -1$ for each i, and it follows from the Law of Large Numbers that

$$\mathbb{P}\big((1-U_1)\cdots(1-U_k) \in [e^{-1.1k}, e^{-0.9k}]\big) = 1 - \overline{o}(1). \tag{5.3}$$

Recalling the definition of δ from (5.1), we conclude that

$$\mathbb{P}_x\big(q_1\cdots q_k \notin [x^{1-\delta}, x^{1-\delta^2}]\big) = \overline{o}(1).$$

The probability that (c) fails is at most the probability that n has a prime power factor in one of the intervals $[x^{(1-\varepsilon^2)/s}, 4x^{1/s}]$, which is easily bounded by Mertens' theorem by

$$\sum_{s \leqslant 1/\varepsilon+1} \sum_{x^{(1-\varepsilon^2)/s} < q \leqslant 4x^{1/s}} \frac{1}{q} \ll \frac{\varepsilon^2}{\varepsilon} = \varepsilon = \overline{o}(1).$$

Finally, if every $q_i \geqslant x^\varepsilon$ and some q_i is not prime, then n is divisible by a prime power $p^a > x^\varepsilon$ with $a \geqslant 2$. The number of such $n \in (x, 2x]$ is $O(x^{1-\varepsilon/2})$. This completes the proof. □

5.2. **Completing the proof.** From now on, the variables q_i will denote primes. Let n and $\mathbf{q}(n)$ be the random quantities described above. Our main task is to show that

$$\mathbb{P}_x\left(n^\ell \Big| \binom{2n}{n}\right) = c_\ell + \overline{o}(1). \tag{5.4}$$

Theorem 1 follows immediately upon fixing k, letting $x \to \infty$, and then letting $k \to \infty$.

We first show, using Proposition 1 and Lemma 5.3 that it suffice to consider large prime factors of n and $\mathbf{q}(n) \in \mathcal{Y}_k(x)$. Let

$$B_n = \prod_{\substack{p^a \| n \\ p > y}} p^a,$$

where y is the smallest power of two that is $> x^{2\delta}$. Applying Proposition 1, followed by an application of Lemma 5.3, we see that

$$\mathbb{P}_x\left(n^\ell \Big| \binom{2n}{n}\right) = \overline{o}(1) + \mathbb{P}_x\left(B_n^\ell \Big| \binom{2n}{n}\right) = \overline{o}(1) + \mathbb{P}_x\left(B_n^\ell \Big| \binom{2n}{n} \text{ and } \mathbf{q}(n) \in \mathcal{Y}_k(x)\right). \tag{5.5}$$

If $\mathbf{q}(n) \in \mathcal{Y}_k(x)$, then by (b), $q_1 \cdots q_k \geqslant x^{1-\delta}$. It follows that $B_n | q_1 \cdots q_k$, that is, $q_1 \cdots q_k$ contains all of the large prime factors of n. On the other hand, Proposition 1 implies that the probability that some prime factor $q < y$ of n satisfies $q^\ell \nmid \binom{2n}{n}$ is $\overline{o}(1)$. Thus

$$\mathbb{P}_x\left(B_n^\ell \Big| \binom{2n}{n} \text{ and } \mathbf{q}(n) \in \mathcal{Y}_k(x)\right) = \mathbb{P}_x\left(\mathbf{q}(n) \in \mathcal{Y}_k(x) \wedge q_j^\ell \Big| \binom{2n}{n} \ (1 \leqslant j \leqslant k)\right) + \overline{o}(1).$$

Combined with (5.5), this gives

$$\mathbb{P}_x\left(n^\ell \Big| \binom{2n}{n}\right) = \overline{o}(1) + \sum_{\mathbf{q} \in \mathcal{Y}_k(x)} \mathbb{P}_x\left(\mathbf{q}(n) = \mathbf{q} \wedge q_j^\ell \Big| \binom{2n}{n} \ (1 \leqslant j \leqslant k)\right). \tag{5.6}$$

Write $n = mq_1 \cdots q_k$. Direct computation gives

$$\mathbb{P}_x\left(\mathbf{q}(n) = \mathbf{q} \wedge q_j^\ell \Big| \binom{2n}{n} \ (1 \leqslant j \leqslant k)\right) = \frac{1}{x} \sum_{\substack{x < mq_1 \cdots q_k \leqslant 2x \\ q_j^\ell | \binom{2n}{n} \ (1 \leqslant j \leqslant k)}} \frac{(\log q_1) \cdots (\log q_k)}{\log n \log(n/q_1) \cdots \log n/(q_1 \cdots q_{k-1})}.$$

It is convenient to place each q_i into a dyadic interval. For each i, let M_i be the unique power of two such that $M_i < q_i \leqslant 2M_i$. By conditions (b) and (c) in the definition of $\mathcal{Y}_k(x)$,

$$\frac{(\log q_1) \cdots (\log q_k)}{\log n \log(n/q_1) \cdots \log n/(q_1 \cdots q_{k-1})} = (1 + o(1)) \frac{(\log M_1) \cdots (\log M_k)}{\log x \log(\frac{x}{M_1}) \cdots \log(\frac{x}{M_1 \cdots M_{k-1}})}. \tag{5.7}$$

We insert this last estimate into (5.6), obtaining

$$\begin{aligned} \mathbb{P}_x\left(n^\ell \Big| \binom{2n}{n}\right) &= \overline{o}(1) + (1 + o(1)) \sum_{\mathbf{M}} \frac{(\log M_1) \cdots (\log M_k)}{\log x \log\left(\frac{x}{M_1}\right) \cdots \log\left(\frac{x}{M_1 \cdots M_{k-1}}\right)} \\ &\quad \times \sum_{\frac{x}{2^k M_1 \cdots M_k} < m \leqslant \frac{2x}{M_1 \cdots M_k}} \sum_{\substack{\mathbf{q} \in \mathcal{R} \\ q_j^\ell | \binom{2n}{n} \ (1 \leqslant j \leqslant k)}} 1, \end{aligned} \tag{5.8}$$

where we have written $n = q_1 \cdots q_k m$ and

$$\mathcal{R} = \mathcal{R}(\mathbf{M}, m) = \{(z_1, \ldots, z_k) \in \mathcal{Y}_k(x) : M_i < z_i \leqslant 2M_i \ (1 \leqslant i \leqslant k), x < mz_1 \cdots m_k \leqslant 2x\}.$$

Now fix $\mathbf{M}$ and m. By Lemma 5.2, $\mathcal{Y}_k(x)$ is $(1/\varepsilon+2)$-simple and thus $\mathcal{R}$ is also $(1/\varepsilon+2)$-simple. We may then apply Proposition 3 to $\mathcal{R}$. Condition (iii) in that Proposition holds with $\rho=\varepsilon^2$ on account of (c). Let $s_j=\lfloor\frac{\log x}{\log M_j}\rfloor-1$ for each j, and define

$$F(b)=1-2^{-b}\sum_{h=0}^{\ell-1}\binom{b}{h},$$

By Proposition 3, we get that

$$\sum_{\substack{\mathbf{q}\in\mathcal{R}\\ q_j^\ell|\binom{2n}{n}\,(1\leqslant j\leqslant k)}}1=(1+O(\varepsilon))\prod_{j=1}^k F(s_j)\sum_{\mathbf{q}\in\mathcal{R}}1+O_{k,\varepsilon}(x^{1-\xi}),$$

for some $\xi>0$. The final error term is negligible since the number of $\mathbf{M}$ is $O((\log x)^k)$. Now sum over all m and $\mathbf{M}$, and rewrite the final result in terms of $\mathbf{q}$ using (5.7) again. By (5.8) we conclude that

$$\begin{aligned}\mathbb{P}_x\left(n^\ell\Big|\binom{2n}{n}\right)&=\overline{o}(1)+(1+O(\varepsilon))\sum_{\mathbf{q}\in\mathcal{Y}_k(x)}\mathbb{P}_x(\mathbf{q}(n)=\mathbf{q})\prod_{j=1}^k F(s_j)\\ &=\overline{o}(1)+(1+O(\varepsilon))\mathbb{E}_x\mathbf{1}_{\mathbf{q}(n)\in\mathcal{Y}_k(x)}\prod_{j=1}^k F(s_j),\end{aligned}\tag{5.9}$$

where (consistent with the earlier definition) by (c) we have

$$s_j=\left\lfloor\frac{\log x}{\log q_j}\right\rfloor-1\qquad(1\leqslant j\leqslant k).$$

Using Lemma 5.3 again, followed by Lemma 5.1, we arrive at

$$\mathbb{P}\left(n^\ell\Big|\binom{2n}{n}\right)=\overline{o}(1)+\mathbb{E}_x\prod_{j=1}^k F(s_j)=\overline{o}(1)+\mathbb{E}\prod_{j=1}^k F(g_j),$$

where g_j is defined in (1.1). Finally, by the Law of Large Numbers, cf. (5.3) we have $g_j\geqslant e^{j/2}$ for all $j\geqslant k$ with probability $1-\overline{o}(1)$ and this completes the proof of (5.4) upon recalling that

$$c_\ell=\mathbb{E}\prod_{j=1}^\infty F(g_j).$$

6. Proof of Theorem 3

The proof is similar to that of Theorem 1, but the details are simpler. In particular, we do not need the work from Section 5.1. As before, the symbols q and q_i denote primes.

For fixed $k\in\mathbb{N}$ and $\varepsilon>0$ let

$$\mathcal{N}_{k,\varepsilon}(x)=\#\left\{n=q_1\cdots q_k\in(x,2x]:\left(n,\binom{2n}{n}\right)=1,\forall i,\ q_i\geqslant x^\varepsilon\text{ and }q_i\notin\bigcup_{s\leqslant 1/\varepsilon+1}(x^{(1-\varepsilon^3)/s},4x^{1/s}]\right\}.$$

Lemma 6.1. *For any fixed $k\geqslant 2$ and $\varepsilon>0$ we have*

$$|\mathcal{N}_{k,\varepsilon}(x)|=\frac{x}{\log x}\left\{(1+O(\varepsilon))\int\cdots\int_{\substack{\varepsilon\leqslant u_1\leqslant\ldots\leqslant u_k\leqslant 1\\ u_1+\cdots+u_k=1}}h(u_1)\cdots h(u_k)\,du_1\cdots du_k+O_k(\varepsilon^2)+o(1)\right\},$$

as $x \to \infty$, where $h(v) = v^{-1}2^{1-\lfloor 1/v \rfloor}$.

Proof. Consider $n \in \mathcal{N}_{k,\varepsilon}(x)$, and write $n = q_1 \cdots q_k$ with $q_1 < \cdots < q_k$. Let

$$\mathcal{T} = \left\{ x^\varepsilon \leqslant y_1 < \cdots < y_k \leqslant x : x < y_1 \cdots y_k \leqslant 2x, \forall i : \; y_i \notin \bigcup_{s \leqslant 1/\varepsilon + 1} \left(x^{(1-\varepsilon^3)/s}, 4x^{1/s} \right] \right\},$$

so that $\mathbf{q} = (q_1, \ldots, q_k) \in \mathcal{T}$. For each i, let M_i be the unique power of two such that $M_i < q_i \leqslant 2M_i$, and for a fixed $\mathbf{M} = (M_1, \ldots, M_k)$ let $T(\mathbf{M}) = \{\mathbf{y} \in \mathcal{T} : M_i < y_i \leqslant 2M_i \; (1 \leqslant i \leqslant k)\}$.

With $\mathbf{M}$ fixed, define $s_j = \lfloor \frac{\log x}{\log M_j} \rfloor$. Then the hypotheses of Proposition 3 hold with $\rho = \varepsilon^3$. The set $\mathcal{T}$ is $(1/\varepsilon + 2)-$simple and hence by Proposition 3 with $m = 1$, we get that

$$\sum_{\substack{\mathbf{q} \in \mathcal{T}(\mathbf{M}) \\ \left(q_1 \cdots q_k, \binom{2n}{n}\right) = 1}} 1 = (1 + O(\varepsilon)) 2^{-(s_1 + \cdots + s_k)} \sum_{\mathbf{q} \in \mathcal{T}(\mathbf{M})} 1 + O_{k,\varepsilon}(x^{1-\xi}),$$

The prime number theorem implies that

$$\sum_{\mathbf{q} \in \mathcal{T}(\mathbf{M})} 1 = \frac{1 + o(1)}{\log M_1 \cdots \log M_k} \operatorname{Vol}(\mathcal{T}(\mathbf{M})).$$

Now for $\mathbf{q} \in \mathcal{T}(\mathbf{M})$, we have $s_j = \left\lfloor \frac{\log x}{\log q_j} \right\rfloor - 1$ for each j. Thus, after summing over all $\mathbf{M}$ we obtain

$$\begin{aligned}\sum_{\substack{\mathbf{q} \in \mathcal{T} \\ \left(q_1 \cdots q_k, \binom{2n}{n}\right) = 1}} 1 &= O_{k,\varepsilon}(x^{1-\xi/2}) + (1 + O(\varepsilon)) \int_{\mathcal{T}} \prod_{j=1}^k \frac{2^{1 - \lfloor \frac{\log x}{\log y_j} \rfloor}}{\log y_j} d\mathbf{y} \\ &= O_{k,\varepsilon}(x^{1-\xi/2}) + (1 + O(\varepsilon)) \frac{x}{\log x} \int_{\mathcal{U}} h(u_1) \cdots h(u_k) \, d\mathbf{u},\end{aligned}$$

where

$$\mathcal{U} = \Big\{ \varepsilon \leqslant u_1 \leqslant \cdots \leqslant u_k \leqslant 1 : u_1 + \cdots + u_k = 1, \forall i, u_i \notin \bigcup_{s \leqslant 1/\varepsilon + 1} (1 - \varepsilon^3)/s, 1/s] \Big\}.$$

Since $g()$ is bounded, the integral over the region where $(1 - \varepsilon^3)/s \leqslant u_i \leqslant 1/s$ for some i and some $s \leqslant 1/\varepsilon + 1$ contributes $O_k(\varepsilon^2)$ to the integral. This completes the proof. □

Proof of Theorem 3 from Lemma 6.1. Let $\mathcal{N}_k$ be the set of $n \in (x, 2x]$ with k distinct prime factors and with $(n, \binom{2n}{n}) = 1$. Fix $\varepsilon > 0$. Clearly

$$\mathcal{N}_1 \sim \frac{x}{\log x}.$$

Now let $k \geqslant 2$. Then one of the following is true for any $n \in \mathcal{N}_k$:

(1) $n \in \mathcal{N}_{k,\varepsilon}(x)$;
(2) n has a prime factor smaller than x^ε;
(3) n is divisible by the square of some prime larger than x^ε; or
(4) n has a prime factor in $\bigcup_{s \leqslant 1/\varepsilon + 1} (x^{(1-\varepsilon^3)/s}, 4x^{1/s}]$.

Lemma 6.1 gives the size of $\mathcal{N}_{k,\varepsilon}(x)$. By Proposition 2, the number of n satisfying (2) is $O(e^{-1/(3\varepsilon)}x/\log x)$. The number of n satisfying (3) is evidently $\ll x^{1-\varepsilon/2}$. Fixing s, the number of $n \in \mathcal{N}_k$, with all prime factors $\geqslant x^\varepsilon$ and with a prime factor in $I = (x^{(1-\varepsilon^3)/s}, 4x^{1/s}]$ is zero for $s = 1$, and when $s \geqslant 2$ it is at most

$$\sum_{q_1 \in I} \sum_{\substack{q_2,\cdots,q_{k-1} \\ \forall i:\, q_i \geqslant x^\varepsilon \\ q_1\cdots q_{k-1} \leqslant 2x^{1-\varepsilon}}} \pi\left(\frac{x}{q_1\cdots q_{k-1}}\right) \ll \sum_{q_1\in I} \sum_{q_2,\ldots,q_k \in (x^\varepsilon, x]} \frac{x}{\varepsilon q_1 \cdots q_{k-1} \log x}$$

$$\ll \frac{x}{\log x} \frac{(\log 2/\varepsilon)^{k-1}\varepsilon^3}{\varepsilon}.$$

After summing the above over $s \leqslant 1/\varepsilon + 1$, we conclude that

$$|\mathcal{N}_k| = \frac{x}{\log x}\left\{ \frac{1}{k!} \int\cdots\int_{\substack{\varepsilon \leqslant u_1,\ldots,u_k \leqslant 1 \\ u_1+\cdots+u_k=1}} g(1/u_1)\cdots g(1/u_k)\, d\mathbf{u} + O\left(e^{-1/(3\varepsilon)} + \varepsilon(\log 2/\varepsilon)^{k-1} + o(1)\right)\right\}.$$

The function $g()$ is bounded above by 2, thus upon letting $\varepsilon \to 0$ we find that

$$|\mathcal{N}_k| \sim \frac{x}{k! \log x} \int\cdots\int_{\substack{0 \leqslant u_1,\ldots,u_k \leqslant 1 \\ u_1+\cdots+u_k=1}} h(u_1)\cdots h(u_k)\, d\mathbf{u} \qquad (x \to \infty) \tag{6.1}$$

for each fixed k. On the other hand, if n has more than K prime factors, then n has a prime factor $< x^{1/K}$, and by Proposition 2, there are $O(e^{-K/3}x/\log x)$ such integers. That is, for any fixed K,

$$\#\{\mathcal{B} \cap [1, x]\} = \sum_{k=1}^{K} \mathcal{N}_k + O\left(e^{-K/3} \frac{x}{\log x}\right).$$

Summing (6.1) and then letting $K \to \infty$, Theorem 3 follows. □

7. NUMERICAL ESTIMATES OF THE DENSITY

It is convenient here to go back to the variables Y_i given in (1.5). Moreover, in order for the product in the definition to be nonzero, we need $Y_i \leqslant \frac{1}{\ell+1}$ for all i. In particular, this shows that

$$c_\ell \leqslant \rho(\ell+1) = e^{-(1+o(1))\ell \log \ell} \tag{7.1}$$

as $\ell \to \infty$, where ρ is the Dickman function. We have

$$c_\ell = \mathbb{E} \prod_{j=1}^\infty g(y_j), \quad g(y) = \begin{cases} 1 - 2^{1-\lfloor 1/y \rfloor} \sum_{h=0}^{\ell-1} \binom{\lfloor 1/y \rfloor - 1}{h} & \text{if } 0 < y \leqslant \frac{1}{\ell+1} \\ 0 & \text{if } y > \frac{1}{\ell+1}. \end{cases} \tag{7.2}$$

We estimate c_ℓ using Laplace transforms. By Theorem 3.2 of [9], we have that

$$F(s) := \int_0^\infty e^{-st}\Big(\mathbb{E}\prod_{j=1}^\infty g(ty_j)\Big)\, dt = \frac{1}{s}\exp\left(\int_0^\infty \frac{g(z)-1}{z} e^{-sz}\, dz\right) \qquad (\Re s > 0). \tag{7.3}$$

Theorem 3.2 of [9] is only stated for real $s > 0$, but the proof gives the result in the full half-plane $\Re s > 0$. The left side of (7.3) is an entire function of $s \in \mathbb{C}$, since

$$\mathbb{E}\prod_{j=1}^\infty g(ty_j) \leqslant \rho(t(\ell+1))$$

decays faster than exponentially in t; however the right side is only well defined for $\Re s > 0$. We massage the right side using the standard function

$$E_1(z) = \int_z^\infty \frac{e^{-t}}{t}\, dt. \tag{7.4}$$

Since $g(z) = 0$ for $z > \frac{1}{\ell+1}$ we may decompose

$$\int_0^\infty \frac{g(z)-1}{z} e^{-sz}\, dz = \int_0^{\frac{1}{\ell+1}} \frac{g(z)-1}{z} e^{-sz}\, dz - E_1\left(\frac{s}{\ell+1}\right).$$

We next use the fact that $g(z)$ is a step-function with jumps at the points $1/k$, where k is an integer satisfying $k \geqslant \ell+1$. Using the Pascal relation, adn in the notation of Stieltjes integration, we have

$$\begin{aligned} dg\left(\frac{1}{k}\right) = g\left(\frac{1}{k-1}\right) - g\left(\frac{1}{k}\right) &= -2^{2-k}\sum_{h=0}^{\ell-1}\binom{k-2}{h} + 2^{1-k}\sum_{h=0}^{\ell-1}\left(\binom{k-2}{h-1}+\binom{k-2}{h}\right) \\ &= -2^{1-k}\binom{k-2}{\ell-1}. \end{aligned}$$

Thus, applying (Stieltjes) integration by parts we find that

$$\begin{aligned} \int_0^{(1/(\ell+1))^+} (g(z)-1)\frac{e^{-sz}}{z}\, dz &= E_1\left(\frac{s}{\ell+1}\right) + \int_0^{(1/(\ell+1))^+} E_1(sz) dg(z) \\ &= E_1\left(\frac{s}{\ell+1}\right) - \sum_{k\geqslant \ell+1} 2^{1-k}\binom{k-2}{\ell-1} E_1\left(\frac{s}{k}\right). \end{aligned}$$

Here we used that $\lim_{y\to 0^+} g(y) = 1$ and $\lim_{z\to 0} E_1(sz)(g(z)-1) = 0$. Inserting this into (7.3) and inverting, we conclude the following:

Proposition 4. *For any $\sigma > 0$, we have*

$$c_\ell = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} \frac{e^s}{s} \exp\Big\{ - \sum_{k\geqslant \ell+1} 2^{1-k}\binom{k-2}{\ell-1} E_1\left(\frac{s}{k}\right)\Big\}\, ds.$$

Computing c_ℓ was accomplished with the Python scripts `mpmath`, which have a built-in function for numerically inverting the Laplace transform, and which can can be computed to arbitrary precision [1].

```
from mpmath import *
mp.dps=100 # digit accuracy of internal computations
def F(s,l):
    x=mpf('0.0')
    for k in range(l+1,200):x=x+2**(1-k)*binomial(k-2,l-1)*mp.e1(s/k)
    return(mp.exp(-x)/s)
c = lambda l : mp.invertlaplace(lambda z: F(z,l),mpf(1.0))
```

TABLE 1. Python code to compute c_ℓ

[1] We are not completely confident in these numerical values. They are the result of comparing the mpmath numbers with different precision, and the displayed digits are those that are stable when increasing the working precision.

We get

$$\begin{aligned} c_1 &= 0.11424743\ldots \\ c_2 &= 0.003227778\ldots \\ c_3 &= 0.000031511777490\ldots \\ c_4 &= 1.33012994669\ldots \times 10^{-7} \\ c_5 &= 2.83248121476\ldots \times 10^{-10} \\ c_6 &= 3.40390904801\ldots \times 10^{-13}. \end{aligned}$$

When $\ell = 1$, this is in fairly good aggreement with accumulated numerical data, e.g. [11, Sequence A014847].

8. Proof of Theorem 2

We use Proposition 4 and invert using the saddle-point method, as in §III.5 of [15]. By the shape of the binomial distribution, $g(z)$ transitions from being close to 1 to being very small in the vicinity of $z = \frac{1}{2\ell}$. Recall the definition (7.4) of $E_1(z)$ and define

$$\mathrm{Ein}(s) := \gamma + \log s + E_1(s) = \int_0^s \frac{1 - e^{-t}}{t}\, dt, \tag{8.1}$$

which is an entire function of s. By [15, Theorem 5.10, §III.5], we have

$$\hat{\rho}(s) := \int_0^\infty \rho(t) e^{-ts}\, dt = e^{\gamma - \mathrm{Ein}(s)}. \tag{8.2}$$

To bound the integral in Proposition 4, we define

$$J(w,u) := \sum_{k=\ell+1}^{\infty} 2^{1-k}\binom{k-2}{\ell-1}\left(E_1(w) - E_1\left(\frac{wu}{k}\right)\right) = E_1(w) - \sum_{k=\ell+1}^{\infty} 2^{1-k}\binom{k-2}{\ell-1} E_1\left(\frac{wu}{k}\right). \tag{8.3}$$

In this notation, plus (8.1), Proposition 4 implies that

$$\begin{aligned} c_\ell &= \frac{1}{2\pi i u}\int_{\sigma - i\infty}^{\sigma + i\infty} e^s \exp\left\{\gamma - \mathrm{Ein}(s/u) + J(s/u, u)\right\} ds \\ &= \frac{1}{2\pi i}\int_{\sigma - i\infty}^{\sigma + i\infty} e^{uw} \exp\left\{\gamma - \mathrm{Ein}(w) + J(w, u)\right\} dw, \end{aligned} \tag{8.4}$$

where $u \geqslant 1$ is an arbitrary parameter, to be chosen later to make $J(s/u, u)$ small when $s \approx \sigma$.

Comparing (8.4) with (8.2), we will see that the optimal choise of u is very close to the optimal value needed to compute $\rho(u)$ by inverting $\hat{\rho}$, namely

$$\sigma = -\xi_0 := -\xi(u), \tag{8.5}$$

where $\xi = \xi(u)$ satisfies $e^\xi = 1 + u\xi$. We note that

$$\xi(u) = \log(u \log u) + \frac{\log\log u}{\log u} + O\left(\frac{(\log\log u)^2}{\log^2 u}\right). \tag{8.6}$$

We record estimates for $\hat{\rho}(s)$ on vertical segments from [15, Lemma 5.12, Ch. III].

Lemma 8.1. *Let $u \geqslant 2$ and $\xi = \xi(u)$. For $w = -\xi + i\tau$, we have*

$$\hat{\rho}(w) = e^{\gamma - \mathrm{Ein}(w)} = \begin{cases} O\bigg(\exp\Big\{ -\mathrm{Ein}(-\xi) - \frac{\tau^2 u}{2\pi^2} \Big\} \bigg) & \text{if } |\tau| \leqslant \pi \\ O\bigg(\exp\Big\{ -\mathrm{Ein}(-\xi) - \frac{u}{\pi^2+\xi^2} \Big\} \bigg) & \text{if } |\tau| > \pi \\ \frac{1}{w}\Big(1 + O\Big(\frac{1+u\xi}{|w|} \Big) \Big) & \text{if } |\tau| > 1 + u\xi. \end{cases}$$

We also use a standard bound for the binomial distribution which follows quickly, for example, from Hoeffding's inequality applied to Bernouilli random variables X_i with $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2$.

Lemma 8.2. *We have*

$$2^{1-k}\binom{k-2}{\ell-1} \ll \exp\left\{ -\frac{(k-2\ell)^2}{2k} \right\}.$$

Lemma 8.3. *Let A_ℓ be the random variable with*

$$\mathbb{P}(A_\ell = k) = a_{k,\ell} := 2^{1-k}\binom{k-2}{\ell-1} \qquad (k \geqslant \ell+1).$$

Then, for $\ell \geqslant 4$ we have

(a) $\mathbb{E}A_\ell = 2\ell + 1$*;*
(b) $\mathbb{E}|A_\ell - 2\ell|^B \ll_B \ell^{B/2}$ *for all $B \geqslant 0$;*
(c) $\mathbb{E}A_\ell^{-1} = \dfrac{1}{2\ell} + O\left(\dfrac{1}{\ell^3}\right)$*;*
(d) $\mathbb{E}A_\ell^{-2} = \dfrac{1}{4\ell^2} + \dfrac{1}{8\ell^3} + O\left(\dfrac{1}{\ell^4}\right)$.
(e) $\mathbb{E}A_l e^{z/A_\ell} \ll \ell e^{z/(2\ell)}$ *uniformly for $0 \leqslant z \leqslant \ell^{4/3}$.*

Remark. *The random variables are well-defined since $\sum_k \mathbb{P}(A_\ell = k) = g(0^+) - g(1/\ell) = 1$.*

Proof. Identity (a) follows from

$$\mathbb{E}A_\ell = 1 + \mathbb{E}(A_\ell - 1) = 1 + \sum_k (k-1)a_{k,\ell} = 1 + 2\ell \sum_k a_{k,\ell+1} = 2\ell + 1.$$

The estimate (b) follows from Lemma 8.2:

$$\mathbb{E}|A_\ell - 2\ell|^B \ll \sum_{k>\ell} |k-2\ell|^B e^{-\frac{1}{2k}(k-2\ell)^2} \ll \ell^{B/2}.$$

We prove (c) and (d) in a manner similar to that of the proof of (a). First, for $k \geqslant 4$ we have

$$\frac{1}{k} = \frac{1}{k-2} - \frac{2}{(k-2)(k-3)} + O\left(\frac{1}{k^3}\right)$$

and thus

$$\begin{aligned} \mathbb{E}A_\ell^{-1} &= O\left(\frac{1}{\ell^3}\right) + \sum_k \left(\frac{1}{k-2} - \frac{2}{(k-2)(k-3)} \right) a_{k,\ell} \\ &= O\left(\frac{1}{\ell^3}\right) + \frac{1}{2(\ell-1)} \sum_k a_{k,\ell-1} - \frac{2}{4(\ell-1)(\ell-2)} \sum_k a_{k,\ell-2} \\ &= \frac{\ell-3}{2(\ell-1)(\ell-2)} + O\left(\frac{1}{\ell^3}\right) = \frac{1}{2\ell} + O\left(\frac{1}{\ell^3}\right). \end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}A_\ell^{-2} &= \sum_{k\geqslant \ell+1} a_{k,\ell}\left(\frac{1}{(k-2)(k-3)} - \frac{5}{(k-2)(k-3)(k-4)} + O\left(\frac{1}{k^4}\right)\right)\\
&= O\left(\frac{1}{\ell^4}\right) + \frac{1}{4(\ell-1)(\ell-2)}\sum_k a_{k,\ell-2} - \frac{5}{8(\ell-1)(\ell-2)(\ell-3)}\sum_k a_{k,\ell-3}\\
&= \frac{2\ell-11}{8(\ell-1)(\ell-2)(\ell-3)} + O\left(\frac{1}{\ell^4}\right)\\
&= \frac{1}{4\ell^2} + \frac{1}{8\ell^3} + O\left(\frac{1}{\ell^4}\right).
\end{aligned}$$

Finally we prove part (e) using Lemma 8.2. Let $k_0 = \lfloor 2\ell - 10\ell^{2/3}\rfloor$ and $k_1 = 4\ell$. We have

$$\begin{aligned}
\mathbb{E}A_\ell e^{z/A_\ell} &\ll \ell\, e^{z/k_0} + \ell \sum_{k=k_0+1}^{2\ell} \exp\left\{-\frac{(2\ell-k)^2}{2k} + \frac{z}{k}\right\} + \ell \sum_{k>10\ell} \exp\left\{-\frac{(k-2\ell)^2}{2k} + \frac{z}{k}\right\}\\
&\ll \ell\, e^{z/(2\ell)} + \ell \sum_{k=k_0+1}^{2\ell} e^{-\ell^{1/3}} + \ell \sum_{k=k_1}^{\infty} e^{-k/8+z/k_1}\\
&\ll \ell\, e^{z/(2\ell)},
\end{aligned}$$

as required. □

We use the previous two lemmas to estimate $J(w,u)$, as defined in (8.3).

Proposition 5. *Suppose that $u = 2\ell + O(\log \ell)$ and $\xi = \xi(u)$. Then, on the vertical line $\Re w = -\xi$ we have the crude bound*

$$J(w,u) \ll \frac{e^{\xi}}{|w|} \ll \frac{\ell \log \ell}{|w|}. \tag{8.7}$$

Furthermore, if $|w| \leqslant \ell^{1/4}$ then we have the asymptotic

$$J(w,u) = e^{-w}\left[\frac{u-w-1}{2\ell} - 1 + O(|w|^2\ell^{-3/2})\right]. \tag{8.8}$$

Proof. Using integration by parts, we see that

$$\begin{aligned}
E_1(w) - E_1\left(\frac{wu}{k}\right) &= \int_1^{u/k} \frac{e^{-wz}}{z}\,dz\\
&= \frac{e^{-w} - e^{-wu/k}(k/u)}{w} - \frac{1}{w}\int_{\frac{u}{k}}^{1} \frac{e^{-wz}}{z^2}\,dz\\
&\ll \frac{e^{\xi} + e^{\xi u/k}(k/u)}{|w|} + \frac{(k/u)\max(e^{\xi}, e^{\xi u/k})}{|w|}\\
&\ll \frac{(e^{\xi} + e^{\xi u/k})(1 + k/u)}{|w|}.
\end{aligned} \tag{8.9}$$

Apply (8.3), followed by an application of Lemma 8.3 (a) and (e). We have

$$\begin{aligned} J(w,u) &\ll \frac{1}{|w|}\sum_{k=\ell+1}^{\infty} 2^{1-k}\binom{k-2}{\ell-1}(e^{\xi}+e^{\xi u/k})(1+k/u) \\ &= \frac{1}{|w|}\mathbb{E}\big(1+A_\ell/u\big)\big(e^{\xi}+e^{\xi u/A_\ell}\big) \\ &\ll \frac{\mathbb{E}A_\ell\big(e^{\xi}+e^{\xi u/A_\ell}\big)}{u|w|} \\ &\ll \frac{\ell e^{\xi}+\ell e^{\xi u/(2\ell)}}{\ell|w|}, \end{aligned}$$

and (8.7) follows from the bounds on u.

Now suppose that $|w| \leqslant \ell^{1/4}$. By (8.6), (8.9) and Lemma 8.2, the terms in the definition (8.3) of $J(w,u)$ corresponding to $|k-2\ell| > 100(\ell\log\ell)^{1/2}$ have total sum

$$\ll \frac{e^{2\xi}}{|w|}\sum_{|k-2\ell|>100(\ell\log\ell)^{1/2}} (1+k/u)a_{k,\ell} \ll \frac{1}{\ell^{100}}. \tag{8.10}$$

When $|k-2\ell| < 100(\ell\log\ell)^{1/2}$, the fraction $u/k = 1+O(\sqrt{\frac{\log\ell}{\ell}})$. Hence

$$\begin{aligned} E_1(w) - E_1\left(\frac{wu}{k}\right) &= e^{-w}\int_0^{\frac{u}{k}-1} \frac{e^{-wv}}{1+v}\,dv \\ &= e^{-w}\int_0^{\frac{u}{k}-1} \big(1-(w+1)v+O(|w|^2v^2)\big)\,dv \\ &= -e^{-w}\left[1-\frac{u}{k}+(w+1)\left(1-\frac{u}{k}\right)^2+O\left(|w|^2\frac{|k-u|^3}{\ell^3}\right)\right]. \end{aligned}$$

By Lemma 8.3 (b),

$$\mathbb{E}|k-u|^3 \ll \mathbb{E}|k-2\ell|^3 + |2\ell-u|^3 \ll \ell^{3/2}$$

and thus the big-O term above is $\ll |w|^2\ell^{-3/2}$. Reintroducing the summands $|k-2\ell| \geqslant 100(\ell\log\ell)^{1/2}$, which are negligible by (8.10), we find using Lemma 8.3 (c) and (d) that

$$\begin{aligned} J(w,u) &= O\left(\frac{1}{\ell^{100}}\right) - e^{-w}\left[1-u\mathbb{E}A_\ell^{-1}+(w+1)\mathbb{E}\left(1-\frac{u}{A_\ell}\right)^2+O(|w|^2\ell^{-3/2})\right] \\ &= O\left(\frac{1}{\ell^{100}}\right) - e^{-w}\left[1-\frac{u}{2\ell}+(w+1)\left(\left(1-\frac{u}{2\ell}\right)^2+\frac{u^2}{8\ell^3}\right)+O(|w|^2\ell^{-3/2})\right] \\ &= e^{w}\left[\frac{u-w-1}{2\ell}-1+O(|w|^2\ell^{-3/2})\right]. \end{aligned}$$

Here we used repeatedly the bounds $|w| \geqslant 1$ and $|u-2\ell| \ll \log\ell$. This completes the proof of (8.8). □

We now complete the proof of Theorem 2. Begin with the w-integral on the right side of (8.4) and define

$$u = 2\ell+1-\xi(2\ell), \qquad \sigma = u\xi(u). \tag{8.11}$$

Since

$$\xi'(u) = \frac{\xi+1}{u(\xi-1)+1} \ll \frac{1}{u}$$

and $\xi(2\ell) \ll \log \ell$, it follows that

$$\xi(2\ell) = \xi(u) + O\left(\frac{\log \ell}{\ell}\right)$$

and hence that

$$u = 2\ell + 1 - \xi(u) + O\left(\frac{\log \ell}{\ell}\right).$$

Plugging this into (8.8), we see that when $w = -\xi + i\tau$ and $|\tau| < \ell^{1/4}$, we have the bound

$$J(-\xi + i\tau, u) = e^{-w}\left(\frac{-i\tau}{2\ell} + O(|w|^2\ell^{-3/2})\right) \ll |\tau|\log \ell + \frac{\log^3 \ell + |\tau|^2 \log \ell}{\ell^{1/2}} \quad (|\tau| < \ell^{1/4}). \tag{8.12}$$

We now insert the estimates (8.12), (8.7) and the bounds from Lemma 8.1 into the right side of (8.4). Let

$$\tau_1 = 100\sqrt{\frac{\log u}{u}}, \quad \tau_2 = \pi, \quad \tau_3 = 1 + u\xi(u).$$

Write $w = -\xi + i\tau$, $\xi = \xi(u)$.

Our fist task is to show that the part of the integral with $|\tau| > \tau_1$ is negligible. When $\tau_1 \leqslant |\tau| \leqslant \tau_2$, Lemma 8.1 and (8.12) imply that

$$\begin{aligned} e^{\gamma - \mathrm{Ein}(w) + J(w,u)} &\ll e^{-\mathrm{Ein}(-\xi) - \tau^2 u/(2\pi^2) + O(|\tau|\log \ell)} \\ &\ll e^{-\mathrm{Ein}(-\xi) - 1000 \log u}. \end{aligned}$$

When $\tau_2 \leqslant |\tau| \leqslant \tau_3$, Lemma 8.1, (8.7) and (8.12) together imply

$$\begin{aligned} e^{\gamma - \mathrm{Ein}(w) + J(w,u)} &\ll e^{-\mathrm{Ein}(-\xi) - \frac{u}{\pi^2 + \xi^2} + O(\ell^{3/4}\log \ell)} \\ &\ll e^{-\mathrm{Ein}(-\xi) - \frac{u}{2\xi^2}}, \end{aligned}$$

and when $|\tau| > \tau_3$, Lemma 8.1 and (8.7) give

$$e^{\gamma - \mathrm{Ein}(w) + J(w,u)} = \frac{1}{w}\left(1 + O\left(\frac{\ell \log \ell}{|w|}\right)\right).$$

We find that the portion of the w-integral in (8.4) corresponding to $|\tau| \geqslant \tau_1$ is

$$\begin{aligned} &\ll \frac{e^{-u\xi - \mathrm{Ein}(-\xi)}}{\ell^{500}} + e^{-u\xi}\left|\int_{\tau_3}^{\infty} \frac{e^{i\tau u}}{\tau}\left(1 + O\left(\frac{\ell \log \ell}{\tau}\right)\right) d\tau\right| \\ &\ll \frac{e^{-u\xi - \mathrm{Ein}(-\xi)}}{\ell^{500}} + e^{-u\xi} \ll \frac{e^{-u\xi - \mathrm{Ein}(-\xi)}}{\ell^{500}}, \end{aligned}$$

upon appealing to the easy bound $-\mathrm{Ein}(-\xi) \gg \xi^{-1}e^{\xi} \gg \ell$.

Finally, we consider $|\tau| \leqslant \tau_1$. By Lemma 8.1 and (8.7) it follows that

$$\frac{1}{2\pi i}\int_{-\xi - i\tau_1}^{-\xi + i\tau_1} e^{uw} e^{\gamma - \mathrm{Ein}(w) + J(w,u)}\, dw = K(u) + O\left(e^{-u\xi - \mathrm{Ein}(-\xi)} \frac{\log^2 \ell}{\ell}\right),$$

where

$$K(u) = \frac{1}{2\pi i}\int_{-\xi - i\tau_1}^{-\xi + i\tau_1} e^{uw} e^{\gamma - \mathrm{Ein}(w)}\, dw.$$

Extending the limits to $-\xi \pm i\infty$ produces a small error term by Lemma 8.1 and it follows from (8.2) that

$$\rho(u) - K(u) \ll \int_{|\tau| > \tau_1} |e^{uw - \mathrm{Ein}(w)}|\, dw \ll \frac{e^{-\xi - \mathrm{Ein}(-\xi)}}{\ell^{100}}.$$

Gathering these estimates together, we deduce that

$$c_l = \rho(u) + O\left(\frac{\log^2 \ell}{\ell} e^{-u\xi - \mathrm{Ein}(-\xi)}\right).$$

By Theorem 5.13 of [15, Ch. III], we have

$$\rho(u) = \left(1 + O\left(\frac{1}{u}\right)\right)\left(\frac{\xi}{2\pi(u(\xi-1)+1)}\right)^{1/2} e^{\gamma - u\xi - \mathrm{Ein}(-\xi)} \gg \frac{1}{u^{1/2}} e^{-u\xi - \mathrm{Ein}(-\xi)} \tag{8.13}$$

and thus

$$c_l = \rho(u)\left(1 + O\left(\frac{\log^2 \ell}{\ell^{1/2}}\right)\right). \tag{8.14}$$

Finally, we estimate the error made by replacing u by

$$u^* = 2\ell + 1 - \log(2\ell \log(2\ell)) - \frac{\log\log(2\ell)}{\log 2\ell}$$

in (8.14). By (8.6),

$$|u - u^*| \ll \frac{(\log\log \ell)^2}{\log^2 \ell}.$$

Hence, using (8.13), (8.6), the bound $\xi'(u) \ll 1/u$ and the bounds

$$\mathrm{Ein}(-\xi(u)) - \mathrm{Ein}(-\xi(u^*)) \ll \frac{e^{\xi(u)}}{\xi(u)} |\xi(u^*) - \xi(u)| \ll |u - u^*|,$$
$$u\xi(u) - u^*\xi(u^*) \ll |u - u^*| \log u,$$

we see that

$$\rho(u) \sim \rho(u^*) \quad (u \to \infty).$$

Combining this with (8.14), this completes the proof of Theorem 2.

9. Numerical computation of c

The terms with $k = 1$ and $k = 2$ in (1.4) contribute 1, respectively, $\sum_{m=2}^{\infty} 2^{1-m} \log\left(\frac{m}{m-1}\right) = 0.507833922868438392189041\ldots$. Define

$$f(t) = \sum_{k=3}^{\infty} \frac{1}{k!} \underset{\substack{u_i \geqslant 0\ \forall i \\ u_1 + \cdots + u_k = t}}{\int \cdots \int} h(u_1) \cdots h(u_k)\, du_1 \cdots du_k,$$

so that $c = f(1) + 1.507833922868438392189041\ldots$. Extend the definition of h to $(0, \infty)$ by defining $h(u) = 1/u$ for $u \geqslant 1$. In this way, $h(u) = 1/u$ for $u > 1/2$, and thus h is C^∞ near $t = 1$. As in previous sections, define the Laplace transform

$$F(s) = \int_0^\infty f(t) e^{-st}\, dt = e^J - 1 - J^2/2. \quad J = \int_0^\infty h(u) e^{-su}\, du.$$

Using that $h(u) = u^{-1}2^{1-m}$ for $\frac{1}{m+1} < u \leqslant \frac{1}{m}$, $m \geqslant 1$, and recalling the definition (7.4) of $E_1(z)$, we quickly derive

$$\int_0^\infty h(u)e^{-su}\,du = \sum_{m=1}^\infty 2^{1-m}\int_{1/(m+1)}^{1/m} \frac{e^{-su}}{u}\,du + \int_1^\infty \frac{e^{-su}}{u}\,du$$
$$= \sum_{m=2}^\infty 2^{1-m}E_1(s/m).$$

Again, we use the Python package `mpmath` to numerically invert the Laplace transform $F(s)$, and this gives $c = f(1) = 1.526453\ldots$.

Appendix A. Proof of Lemma 5.1

Recall that for random $\mathbf{q} = \mathbf{q}(n) = (q_1, \ldots, q_k)$ we defined

$$X_i(n) = \frac{\Lambda(q_i)}{\log(\frac{n}{q_1\cdots q_{i-1}})}. \tag{A.1}$$

It suffices to show that for any real numbers $0 < a_i < b_i < 1$ $(1 \leqslant i \leqslant k)$,

$$\mathbb{P}_x(a_i \leqslant X_i(n) \leqslant b_i\ (1 \leqslant i \leqslant k)) \to \prod_{i=1}^k (b_i - a_i) \qquad (x \to \infty). \tag{A.2}$$

Below, constants implied by $O-$ an $\ll -$ may depend on k and the a_i, b_i. From (5.2), if $X_i \leqslant b_i$ for all i then

$$\frac{n}{q_1 \cdots q_{i-1}} \geqslant n^{(1-b_1)\cdots(1-b_{i-1})}. \tag{A.3}$$

Hence, writing $c = (1-b_1)\cdots(1-b_k)\min_i a_i$, we have $q_i > n^c$ for all i under the assumption that $a_i \leqslant X_i(n) \leqslant b_i$ for every i. If some q_i is not prime, then n is divisible by a prime power $p^a > x^{c/2}/\log x$ with $a \geqslant 2$ and the number of such $n \in (x, 2x]$ is $O(x^{1-c/2})$. Thus, we may assume that the q_i are all prime. We calculate, using (A.1),

$$\mathbb{P}_x(a_i \leqslant X_i(n) \leqslant b_i\ (1 \leqslant i \leqslant k)) = \frac{1}{x}\sum_{x<n\leqslant 2x}\ \sum_{\substack{q_1|n\\ a_1\leqslant X_1(n)\leqslant b_1}} X_1(n)\cdots \sum_{\substack{q_k|n\\ a_1\leqslant X_k(n)\leqslant b_1}} X_k(n).$$

On the right side, the variables q_i are no longer random, but we still define $X_i(n)$ by (A.1). Since $\log x \leqslant \log n \leqslant \log(2x)$, the above expression is bounded below by

$$(1 + O(1/\log x)) \sum_{a_1\log(2x)\leqslant \log q_1\leqslant b_1\log x} \frac{\log q_1}{q_1}\cdots \sum_{a_k\log(\frac{2x}{q_1\cdots q_{k-1}})\leqslant \log q_k\leqslant b_k\log(\frac{x}{q_1\cdots q_{k-1}})} \frac{\log q_k}{\log\frac{x}{q_1\cdots q_{k-1}}},$$

and bounded above by the same expression with "x" and "$2x$" interchanged in the logarithms.

For each fixed $q_1, \ldots, q_{i-1}$, Mertens' estimate gives

$$\sum_{a_i\log(\frac{x}{q_1\cdots q_{i-1}})+O(1)\leqslant \log q_i\leqslant b_i\log(\frac{x}{q_1\cdots q_{i-1}})+O(1)} \frac{\log q_i}{\log\frac{x}{q_1\cdots q_{i-1}}} = b_i - a_i + O\left(\frac{1}{\log x}\right),$$

and the desired result (A.2) follows.

REFERENCES

[1] P. Billingsly, *On the distribution of large prime divisors*, Collection of articles dedicated to the memory of Alfréd Rényi, I. Period. Math. Hungar. **2** (1972), 283–289.

[2] H. Davenport, *Multiplicative number theory*, 3rd ed., Graduate Texts in Mathematics vol. 74, Springer-Verlag, New York, 2000.

[3] P. Donnelly and G. Grimmett, *On the asymptotic distribution of large prime factors*, J. London Math. Soc. (2) **47** (1993), 395–404.

[4] P. Erdős, R. L. Graham, I. Z. Ruzsa, E. G. Straus, *On the prime factors of* $\binom{2n}{n}$, Collection of articles in honor of Derrick Henry Lehmer on the occasion of his seventieth birthday, Math. Comp. **29** (1975) 83–92.

[5] S. W. Graham and G. Kolesnik, *Van der Corput's method of exponential sums*, London Math. Soc. Lecture Note, vol. 126, Cambridge University Press, 1991.

[6] A. Granville, *Arithmetic properties of binomial coefficients. I. Binomial coefficients modulo prime powers*, Organic Mathematics (Burnaby, BC, 1995), 253–276, CMS Conf. Proc., 20, Amer. Math. Soc., Providence, RI, 1997.

[7] H. Halberstam and H.-E. Richert, *Sieve Methods*, Academic Press, London, 1974.

[8] R. R. Hall, G. Tenenbaum, *Divisors,* Cambridge Tracts in mathematics vol. **90**, 1988.

[9] K. Handa, *The two-parameter Poisson-Dirichlet process*, Bernoulli **15** (2009), 1082–1116.

[10] H. L. Montgomery, *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*, CBMS Regional Conference Series in Mathematics vol. 84, Amer. Math. Soc., 1994.

[11] N. A. Sloane, *The on-line encyclopedia of integer sequences.* `http://oeis.org`

[12] C. Pomerance, *Divisors of the middle binomial coefficient*, Amer. Math. Monthly **122** (2015), 636–644.

[13] C. Sanna, *Central binomial coefficients divisible by or coprime to their indices.* Int. J. Number Theory **14** (2018), no. 4, 1135–1141.

[14] G. Tenenbaum, *A rate estimate in Billingsley's theorem for the size distribution of large prime factors*, Quart. J. Math. Oxford **51** (2000), no. 3, 385–403.

[15] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory, 3rd ed.*, Amer. Math Soc.

DEPARTMENT OF MATHEMATICS, 1409 WEST GREEN STREET, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL 61801, USA

E-mail address: ford126@illinois.edu

STEKLOV MATHEMATICAL INSTITUTE, 8 GUBKIN STREET, MOSCOW, 119991, RUSSIA

E-mail address: konyagin23@gmail.com