Mathematical Proceedings of the Cambridge Philosophical Society

VOL. 113 MARCH 1993 PART 2

Math. Proc. Camb. Phil. Soc. (1993), **113**, 225
Printed in Great Britain

On primes not dividing binomial coefficients

BY J. W. SANDER

Institut für Mathematik, Universität Hannover, Welfengarten 1,
D-W-3000 *Hannover* 1, *Germany*

(*Received* 8 *June* 1992)

Abstract

We prove that

$$\sum_{\substack{p \leqslant n \\ p \nmid \binom{2n}{n}}} \frac{\log p}{p} \sim (1-\log 2)\log n,$$

thus dealing with open problems concerning divisors of binomial coefficients.

1. *Introduction*

In 1975, Erdös, Graham, Ruzsa and Straus[**3**], investigated the sum

$$f(n) = \sum_{\substack{p \leqslant n \\ p \nmid \binom{2n}{n}}} \frac{1}{p},$$

where p runs over the primes. They proved that

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n \leqslant x} f(n) = \sum_{k=2}^{\infty} \frac{\log k}{2^k} = c_0,$$

say, and

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n \leqslant x} (f(n))^2 = c_0^2.$$

This implies that

$$f(n) = c_0 + o(1)$$

for all n with the exception of at most $o(N)$ numbers by $n \leqslant N$. Erdös *et al.* could not decide whether $f(n)$ is bounded or not. By applying the method introduced in [**3**] to the function

$$g(n) = \sum_{\substack{p \leqslant n \\ p \nmid \binom{2n}{n}}} \frac{\log p}{p},$$

we immediately obtain

$$\lim_{x\to\infty} \frac{1}{x \log x} \sum_{n \leqslant x} g(n) = 1 - \log 2$$

and
$$\lim_{x\to\infty}\frac{1}{x(\log x)^2}\sum_{n\leqslant x}(g(n))^2=(1-\log 2)^2.$$

As above, this yields
$$g(n)=(1-\log 2)\log n+o(\log n) \tag{1}$$
for almost all n. In this paper, we will show that (1) in fact holds for all n.

THEOREM. *For* $n>1$,
$$g(n)=(1-\log 2)\log n+o(\log n).$$

We mention that the error term $o(\log n)$ could be replaced by an error term $O(\log n/s(n))$ with an explicitly given function $s(n)>0$, where $s(n)\to\infty$ for $n\to\infty$. It will become clear in the sequel (and we will comment on it in the final paragraph), why this, though easy in principle, would cause an unjustifiable amount of tedious work.

Another question raised in [**3**], also linked with prime divisors of binomial coefficients, has been treated by the author in [**5**]. For references connected with the present problem, the reader may consult [**2**].

In the remainder of this article, explicit constants $c_1, c_2, \ldots,$ may depend on k resp. K, while implicit constants occurring in $O(\,)$, $o(\,)$, or $\ll$ are absolute.

2. *Preliminaries*

Let real numbers m_i and positive integers j_i $(1\leqslant i\leqslant r)$ be given, satisfying $m_1\geqslant 1$,
$$M=\max\{|m_i|:1\leqslant i\leqslant r\},$$
and
$$1\leqslant j_1<j_2<\ldots<j_r\leqslant k.$$
Furthermore, let
$$\Lambda(x,y)=\left(\frac{\log x}{\log y}\right)^2.$$

Recently, we proved the following exponential sum estimate, which generalizes a result of Jutila [**4**]. The proof of this lemma uses exponential sum estimates of van der Corput, Vinogradov and Karacuba combined with Vaughan's identity.

LEMMA 1 (see [**6**]). *For* $2\leqslant t\leqslant n^{1/k}$, *we have*
$$\sum_{p\leqslant t}e\left(n\left(\frac{m_1}{p^{j_1}}+\ldots+\frac{m_r}{p^{j_r}}\right)\right)\leqslant c_1(t^{1-c_2\Lambda(t,Mn)}+t^{(k+2)/2}n^{-\frac{1}{2}}+t^{\frac{5}{8}}M^2)(\log Mn)^{4k},$$
where $e(x)=\exp(2\pi i x)$.

The second tool in our proof is Vinogradov's Fourier series method, as described in [**7**], p. 32, or [**1**], lemma 2·1.

LEMMA 2. *Let* $0<\epsilon<\frac{1}{8}$. *Then there are periodic functions* $\psi(x)$ *and* $\Psi(x)$ *with period* 1 *satisfying* $0\leqslant\psi(x)\leqslant 1$, $0\leqslant\Psi(x)\leqslant 1$ *for all* $x\in\mathbb{R}$ *and*
$$\psi(x)=\begin{cases}1 & \text{for } \epsilon\leqslant x\leqslant\frac{1}{2}-\epsilon,\\ 0 & \text{for } \frac{1}{2}\leqslant x\leqslant 1,\end{cases} \tag{2}$$
and
$$\Psi(x)=\begin{cases}1 & \text{for } 0\leqslant x\leqslant\frac{1}{2},\\ 0 & \text{for } \frac{1}{2}+\epsilon\leqslant x\leqslant 1-\epsilon.\end{cases} \tag{3}$$

Moreover, $\psi(x)$ and $\Psi(x)$ have Fourier expansions of the form

$$\psi(x) = \tfrac{1}{2} - \epsilon + \sum_{0<|m|<\infty} a_m e(mx) \tag{4}$$

and
$$\Psi(x) = \tfrac{1}{2} + \epsilon + \sum_{0<|m|<\infty} A_m e(mx), \tag{5}$$

where $a_m, A_m \in \mathbb{C}$ satisfy for $m \neq 0$

$$|a_m| \ll \frac{1}{m^2\epsilon}, \quad |A_m| \ll \frac{1}{m^2\epsilon}. \tag{6}$$

Finally, we need the following easy

LEMMA 3. *Let $h_1(x)$ and $h_2(x)$ be two positive, continuous, strictly decreasing functions defined for $x > x_0$ with*
$$\lim_{x\to\infty} h_1(x) = \lim_{x\to\infty} h_2(x) = 0.$$
Then there is a positive, increasing function $s(x)$ satisfying

$$\lim_{x\to\infty} s(x) = \infty \tag{7}$$

and
$$h_1(s(x)) > h_2(x) \tag{8}$$

for sufficiently large x.

Remark. The lemma holds under much weaker conditions, but for convenience we prove it in this form.

Proof. Since h_2 is decreasing with $h_2(x) \to 0$, there is an x_1 such that for all $x > x_1$, we have $h_2(x) < \tfrac{1}{2}h_1(1)$. Since h_1 is continuous and $h_1(x) \to 0$, we have for each $x > x_1$ an $s(x) > 0$ satisfying
$$h_2(x) = \tfrac{1}{2}h_1(s(x)). \tag{9}$$
For $0 < x < y$, we have $h_2(x) > h_2(y)$, thus $h_1(s(x)) > h_1(s(y))$. Therefore $s(x) < s(y)$; in other words $s(x)$ is increasing. Since $h_2(x) \to 0$, we can see by (9) that $s(x)$ gets arbitrarily big. Being an increasing function, $s(x)$ therefore satisfies (7). In addition, (9) implies (8), which proves the lemma.

3. *Proof of the theorem*

We denote by $e(n;p)$ the exponent of p in the prime factor decomposition of n. It is well-known that
$$e\left(\binom{2n}{n};p\right) = \sum_{j=1}^{\infty}\left(\left[\frac{2n}{p^j}\right] - 2\left[\frac{n}{p^j}\right]\right). \tag{10}$$
For $x \in \mathbb{R}$, we clearly have

$$[2x] - 2[x] = \begin{cases} 1 & \text{for } \{x\} \geq \tfrac{1}{2}, \\ 0 & \text{for } \{x\} < \tfrac{1}{2}, \end{cases}$$

where $\{x\}$ denotes the fractional part of x. Thus, by (10),

$$g(n) = \sum_{\substack{p\leq n \\ e(\binom{2n}{n};p)=0}} \frac{\log p}{p} = \sum_{\substack{p\leq n \\ \{\frac{n}{p^j}\}<\frac{1}{2}\ (j>0)}} \frac{\log p}{p} = \sum_{\substack{p\leq 2n \\ \{\frac{n}{p^j}\}<\frac{1}{2}\ (j>0)}} \frac{\log p}{p}. \tag{11}$$

Before we start applying Vinogradov's Lemma 2 to the last sum, we make some preliminary considerations.

The constant c_2 in Lemma 1 possibly depends on k. Without loss of generality, we may assume that $0 < c_2 = c_2(k)$ is a strictly decreasing function of k. Hence

$$h_1(k) = \frac{c_2(k)}{9(4k+3)}$$

may be continued to a positive, continuous, decreasing real function $h_1(x)$ which tends to 0 for large x. Obviously $h_2(x) = (\log x)^{-\frac{1}{5}}$ also satisfies the conditions of Lemma 3. Thus, by Lemma 3, there is an increasing function $s_1(x) > 0$ with

$$\lim_{n\to\infty} s_1(n) = \infty \tag{12}$$

and

$$\frac{c_2(k)}{9(4k+3)} > (\log n)^{-\frac{1}{5}} \tag{13}$$

for $k = s_1(n)$. The same reasoning applied to the functions

$$h_1(k) = \frac{1}{c_9(k)}, \quad h_2(x) = \frac{1}{\log\log x},$$

where $c_9(k) > 0$ will be defined later (and will be increasing without loss of generality), yields an increasing function $s_2(x) > 0$ with

$$\lim_{n\to\infty} s_2(n) = \infty \tag{14}$$

and

$$\frac{c_9(k)}{\log\log n} < 1 \tag{15}$$

for $k = s_2(n)$. Now we define for sufficiently large n

$$K = s(n) = \min\{s_1(n), s_2(n), (\log\log n)^{\frac{1}{2}}\}. \tag{16}$$

Then, by (12) and (14),

$$\lim_{n\to\infty} s(n) = \infty, \tag{17}$$

and for $1 \leqslant k < K$, (13) and (15) hold.

For $1 \leqslant k < K$, let

$$n_k = (2n)^{1/k}.$$

The prime number theorem of Mertens asserts that

$$\sum_{p\leqslant n} \frac{\log p}{p} = \log n + O(1). \tag{18}$$

With this, (11) implies

$$g(n) = \sum_{k=1}^{K-1} \sum_{\substack{n_{k+1}<p\leqslant n_k \\ \{\frac{n}{p^j}\}<\frac{1}{2}\ (j>0)}} \frac{\log p}{p} + O\left(\sum_{p\leqslant n_K} \frac{\log p}{p}\right)$$

$$= \sum_{k=1}^{K-1} \sum_{\substack{n_{k+1}<p\leqslant n_k \\ \{\frac{n}{p^j}\}<\frac{1}{2}\ (1\leqslant j\leqslant k)}} \frac{\log p}{p} + O\left(\frac{1}{K}\log n\right). \tag{19}$$

For $1 \leqslant k < K$, define

$$b_k = \frac{n^{1/k}}{(\log n)^{12}}.$$

Since, by (18),

$$\sum_{b_k<p\leqslant n_k} \frac{\log p}{p} \ll \log\log n, \tag{20}$$

we have, by (19),

$$g(n) = \sum_{k=1}^{K-1} \sum_{\substack{n_{k+1} < p \leqslant b_k \\ \{\frac{n}{p^j}\} < \frac{1}{2} \quad (1 \leqslant j \leqslant k)}} \frac{\log p}{p} + O(c_3 \log\log n) + O\left(\frac{1}{K}\log n\right). \tag{21}$$

Applying Lemma 2, we get by (2) and (3)

$$\sum_{n_{k+1} < p \leqslant b_k} \left(\prod_{j=1}^{k} \psi\left(\frac{n}{p^j}\right)\right) \frac{\log p}{p} \leqslant \sum_{\substack{n_{k+1} < p \leqslant b_k \\ \{\frac{n}{p^j}\} < \frac{1}{2} \quad (1 \leqslant j \leqslant k)}} \frac{\log p}{p} \leqslant \sum_{n_{k+1} < p \leqslant b_k} \left(\prod_{j=1}^{k} \Psi\left(\frac{n}{p^j}\right)\right) \frac{\log p}{p}. \tag{22}$$

By (6), we have

$$\Big|\sum_{|m| \geqslant \epsilon^{-2}} a_m\, e(mx)\Big| \leqslant \sum_{|m| \geqslant \epsilon^{-2}} \frac{1}{m^2 \epsilon} \ll \epsilon, \tag{23}$$

thus by (4)

$$\Big|\sum_{0 < |m| < \epsilon^{-2}} a_m\, e(mx)\Big| \leqslant |\psi(x)| + \tfrac{1}{2} + \epsilon + \left|\sum_{|m| \geqslant \epsilon^{-2}} \frac{1}{m^2 \epsilon}\right| \ll 1.$$

Also by (23)

$$\sum_{0 < |m| < \infty} a_m\, e(mx) = \sum_{0 < |m| < \epsilon^{-2}} a_m\, e(mx) + O(\epsilon).$$

By (4), these estimates imply

$$\begin{aligned}
\prod_{j=1}^{k} \psi\left(\frac{n}{p^j}\right) &= (\tfrac{1}{2} - \epsilon)^k + \sum_{r=1}^{k} \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant k} (\tfrac{1}{2} - \epsilon)^{k-r} \prod_{l=1}^{r} \left(\sum_{0 < |m_l| < \infty} a_{m_l}\, e\left(m_l \frac{n}{p^{j_l}}\right)\right) \\
&= (\tfrac{1}{2})^k + O(c_4\, \epsilon) + \sum_{r=1}^{k} \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant k} (\tfrac{1}{2} - \epsilon)^{k-r} \\
&\quad \times \prod_{l=1}^{r} \left(\sum_{0 < |m_l| < \epsilon^{-2}} a_{m_l}\, e\left(m_l \frac{n}{p^{j_l}}\right) + O(\epsilon)\right) \\
&= (\tfrac{1}{2})^k + O(c_4\, \epsilon) + O(2^k \epsilon) + \sum_{r=1}^{k} \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant k} (\tfrac{1}{2} - \epsilon)^{k-r} \\
&\quad \times \sum_{0 < |m_1| < \epsilon^{-2}} a_{m_1} \cdots \sum_{0 < |m_r| < \epsilon^{-2}} a_{m_r}\, e\left(n\left(\frac{m_1}{p^{j_1}} + \ldots + \frac{m_r}{p^{j_r}}\right)\right) \\
&= (\tfrac{1}{2})^k + O(c_5\, \epsilon) + \sum_{r=1}^{k} \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant k} (\tfrac{1}{2} - \epsilon)^{k-r} \\
&\quad \times \sum_{0 < |m_1| < \epsilon^{-2}} a_{m_1} \cdots \sum_{0 < |m_r| < \epsilon^{-2}} a_{m_r}\, e\left(n\left(\frac{m_1}{p^{j_1}} + \ldots + \frac{m_r}{p^{j_r}}\right)\right).
\end{aligned}$$

Therefore, by (18) and (6),

$$\begin{aligned}
\sum_{n_{k+1} < p \leqslant b_k} \left(\prod_{j=1}^{k} \psi\left(\frac{n}{p^j}\right)\right) \frac{\log p}{p} &= \frac{1}{k(k+1)} (\tfrac{1}{2})^k \log n + O(c_5\, \epsilon \log n) + O(c_6 \log\log n) \\
&\quad + O\Bigg(c_7 \max_{1 \leqslant j_1 < \ldots < j_r \leqslant k} \sum_{0 < |m_1| < \epsilon^{-2}} \frac{1}{m_1^2\, \epsilon} \cdots \sum_{0 < |m_r| < \epsilon^{-2}} \frac{1}{m_r^2\, \epsilon} \\
&\quad \times \left|\sum_{n_{k+1} < p \leqslant b_k} \frac{\log p}{p} e\left(n\left(\frac{m_1}{p^{j_1}} + \ldots + \frac{m_r}{p^{j_r}}\right)\right)\right|\Bigg).
\end{aligned} \tag{24}$$

We define $$A = A(n) = \exp((\log n)^{\frac{3}{4}}).$$

Then, for $t \geqslant A$, small $\delta > 0$ and sufficiently large n, we have by (13) for $k < K$, thus satisfying (13),

$$\log n(\log\log n)^{\frac{1}{2}} \leqslant (\log n)^{1+\delta} = (\log n)^{-\frac{1}{10}}(\log n)^{\frac{11}{10}+\delta}$$
$$< \left(\frac{c_2}{9(4k+3)}\right)^{\frac{1}{2}}(\log t)^{\frac{4}{3}(\frac{11}{10}+\delta)} < \left(\frac{c_2}{9(4k+3)}\right)^{\frac{1}{2}}(\log t)^{\frac{3}{2}}.$$

This implies

$$(4k+3)\log\log n < \tfrac{1}{9}c_2\frac{(\log t)^3}{(\log n)^2} = \tfrac{1}{9}c_2\,\Lambda(t,n)\log t,$$

hence for $t \geqslant A$,

$$t^{1-\frac{1}{9}c_2\Lambda(t,n)}(\log n)^{4k} < \frac{t}{(\log t)^3}. \tag{25}$$

We choose $\epsilon = (\log\log n)^{-2}$. Clearly, for $0 < m < n^2$,

$$\Lambda(t, mn) > \Lambda(t, n^3) = \tfrac{1}{9}\Lambda(t,n).$$

Thus we have by (25), for $0 < m < \epsilon^{-2}$, $t \geqslant A$ and sufficiently large n,

$$t^{1-c_2\Lambda(t,mn)}(\log n)^{4k} < \frac{t}{(\log t)^3}. \tag{26}$$

For $k < K$, we have $b_k > n_{k+1} \geqslant A$ by (16). Hence, by partial summation, Lemma 1 and (26), we get for $0 < |m_i| < \epsilon^{-2}$ and $k < K$ with K as in (16)

$$\sum_{n_{k+1}<p\leqslant b_k}\frac{\log p}{p}e\left(n\left(\frac{m_1}{p^{j_1}}+\dots+\frac{m_r}{p^{j_r}}\right)\right)$$
$$= \frac{\log b_k}{b_k}\left(\sum_{n_{k+1}<p\leqslant b_k}e\left(n\left(\frac{m_1}{p^{j_1}}+\dots+\frac{m_r}{p^{j_r}}\right)\right)\right)$$
$$+\int_{n_{k+1}}^{b_k}\left(\sum_{n_{k+1}<p\leqslant t}e\left(n\left(\frac{m_1}{p^{j_1}}+\dots+\frac{m_r}{p^{j_r}}\right)\right)\right)\frac{\log t-1}{t^2}\,dt$$
$$\ll c_1\frac{\log b_k}{b_k}\left(\frac{b_k}{(\log b_k)^3}+(b_k^{(k+2)/2}n^{-\frac{1}{2}}+b_k^{\frac{5}{6}}\epsilon^{-4})(\log n)^{4k}\right)$$
$$+c_1\int_{n_{k+1}}^{b_k}\left(\frac{t}{(\log t)^3}+(t^{(k+2)/2}n^{-\frac{1}{2}}+t^{\frac{5}{6}}\epsilon^{-4})(\log n)^{4k}\right)\frac{\log t}{t^2}\,dt$$
$$\ll c_1\left(\frac{1}{(\log b_k)^2}+b_k^{k/2}n^{-\frac{1}{2}}(\log n)^{4k+1}+b_k^{-\frac{1}{6}}(\log n)^{4k+2}\right)$$
$$+c_1\left(\frac{1}{\log n_{k+1}}+n^{-\frac{1}{2}}(\log n)^{4k+1}b_k^{k/2}+n_{k+1}^{-\frac{1}{6}}(\log n)^{4k+2}\right)$$
$$\ll c_1((\log n_{k+1})^{-1}+n^{-1/(6(k+1))}(\log n)^{4k+4}) \ll c_1\frac{1}{\log n}.$$

Since $K \ll (\log\log n)^{\frac{1}{2}}$ by (16), (24) yields for $k < K$

$$\sum_{n_{k+1} < p \leqslant b_k} \left(\prod_{j=1}^{k} \psi\left(\frac{n}{p^j}\right)\right) \frac{\log p}{p} = \frac{1}{k(k+1)} (\tfrac{1}{2})^k \log n + O\left(c_6 \frac{\log n}{(\log\log n)^2}\right) + O\left(c_7 \frac{(\log\log n)^{2k}}{\log n}\right)$$

$$= \frac{1}{k(k+1)} (\tfrac{1}{2})^k \log n + O\left(c_8 \frac{\log n}{(\log\log n)^2}\right). \tag{27}$$

An analogous argument with (5) and (6) yields

$$\sum_{n_{k+1} < p \leqslant b_k} \left(\prod_{j=1}^{k} \Psi\left(\frac{n}{p^j}\right)\right) \frac{\log p}{p} = \frac{1}{k(k+1)} (\tfrac{1}{2})^k \log n + O\left(c_8 \frac{\log n}{(\log\log n)^2}\right). \tag{28}$$

By (20),

$$\sum_{k=1}^{K-1} \sum_{b_k < p \leqslant n_k} \frac{\log p}{p} \ll K \log\log n.$$

Hence, by (21), (22), (27), (28), (15) and (17), we have

$$g(n) = \sum_{k=1}^{K-1} \frac{1}{k(k+1)} (\tfrac{1}{2})^k \log n + O\left(c_9 \frac{\log n}{(\log\log n)^2}\right) + O\left(\frac{1}{K} \log n\right)$$

$$= \sum_{k=1}^{K-1} \frac{1}{k(k+1)} (\tfrac{1}{2})^k \log n + O\left(\frac{\log n}{\log\log n}\right) + O\left(\frac{\log n}{s(n)}\right)$$

$$= \sum_{k=1}^{K-1} \frac{1}{k(k+1)} (\tfrac{1}{2})^k \log n + o(\log n). \tag{29}$$

Obviously
$$\sum_{k=K}^{\infty} (\tfrac{1}{2})^k = O((\tfrac{1}{2})^K) = O\left(\frac{1}{K}\right),$$

and thus
$$g(n) = \sum_{k=1}^{\infty} \frac{1}{k(k+1)} (\tfrac{1}{2})^k \log n + o(\log n). \tag{30}$$

Integrating the geometric series twice, we get for $|x| < 1$

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} x^{k+1} = x + (1-x)\log(1-x).$$

Therefore (30) implies the theorem.

4. *Final remarks*

In order to be able to prove our theorem with an error term $O(\log n/s(n))$, the dependence of the constants c_1 and c_2 of k in Lemma 1 must be given explicitly (see (29)). Such a version of Lemma 1, however, seems not to be worth the effort for the present purpose.

REFERENCES

[1] R. C. BAKER. *Diophantine Inequalities* (Clarendon Press, 1986).
[2] P. ERDÖS and R. L. GRAHAM. *Old and New Problems and Results in Combinatorial Number Theory* (Enseignement Math., 1980).
[3] P. ERDÖS, R. L. GRAHAM, I. Z. RUZSA and E. G. STRAUS. On the prime factors of $\binom{2n}{n}$. *Math. Comp.* **29** (1975), 83–92.
[4] M. JUTILA. On the numbers with a large prime factor, II. *J. Indian Math. Soc.* **38** (1974), 125–130.
[5] J. W. SANDER. On a sum over primes, to appear.
[6] J. W. SANDER. Prime power divisors of binomial coefficients. *J. Reine Angew. Math.*, **430** (1992), 1–20.
[7] I. M. VINOGRADOV. *The Method of Trigonometrical Sums in the Theory of Numbers* (Interscience Publishers).