On the Prime Factors of $\binom{2n}{n}$

By P. Erdös, R. L. Graham, I. Z. Ruzsa and E. G. Straus

Abstract. Several quantitative results are given expressing the fact that $\binom{2n}{n}$ is usually divisible by a high power of the small primes. On the other hand, it is shown that for any two primes p and q, there exist infinitely many n for which $\binom{2n}{n}, pq = 1.$

1. Introduction. In the present paper we study the prime factors of $\binom{2n}{n}$. It is a well-known phenomenon that $\binom{2n}{n}$ is divisible by a high power of the small primes. We shall try to put this observation into a quantitative form. First of all, note that it is not known whether the smallest odd prime factor g(n) of $\binom{2n}{n}$ is bounded. A computer check shows that $g(k) \le 11$ for k < 3160, g(3160) = 13 and $g(k) \le 10$ 13 for $k \leq 10^7$. Of course, it is clear that 2 always divides $\binom{2n}{n}$. We shall show that for any two primes p and q there are infinitely many integers n such that $\binom{2n}{n}$, pq) = 1. (In fact, we shall prove a considerably sharper result.)

Set

$$f(n) = \sum_{\substack{p \neq \binom{2n}{n}; p \leq n}} \frac{1}{p}$$

where p denotes a prime. The most striking fact is that we cannot decide if f(n) is unbounded.

We are going to prove

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

Also, we shall show that

$$\lim_{x \to \infty} \frac{1}{x} \sum_{n=1}^{x} f^{2}(n) = c_{0}^{2}.$$

From these two results we immediately obtain that for all but o(n) integers $m \le n$, $f(m) = c_0 + o(1)$, and it is not hard to deduce that for all but o(n) integers $m \le n$ the number of $t \le m$ with $t \ne \binom{2m}{m}$, is $c_1m + o(m)$ for a certain absolute constant c_1 . Finally, we shall study some special questions about the divisors and prime factors of $\binom{2n}{n}$.

2. The Main Results. An elementary fact which we shall frequently use is the following:

Received June 6, 1974. AMS (MOS) subject classifications (1970). Primary 10A40, 10L10, 10H15.

Copyright © 1975, American Mathematical Society

Fact. For a prime p,

(1) $\binom{2n}{n} \neq 0 \pmod{p}$ if and only if every coefficient (or "digit") a_k in the base p expansion $n = \sum_{k \ge 0} a_k p^k$, $0 \le a_k < p$, satisfies $a_k < p/2$.

Thus the result that for any two primes p, q there are infinitely many integers n with $\binom{2n}{n}, pq = 1$ is a special case of the following:

THEOREM 1. Let A and B be positive integers satisfying $A/(p-1) + B/(q-1) \ge 1$ where p and q are integers exceeding 1. Then there exist infinitely many integers whose base p expansion has all digits $\le A$ and whose base q expansion has all digits $\le B$.

Proof. If log p and log q are commensurable, then p and q are powers of the same integer r, say, $p = r^k$, $q = r^l$. Hence, any sum $\sum_i r^{n_i k l}$ has all digits either 0 or 1 to both bases p and q.

If $\log p$ and $\log q$ are incommensurable, then there are infinitely many exponents α and β so that

(2)
$$\left| p^{\alpha} - \frac{B}{2} \frac{q^{\beta} - 1}{q - 1} \right| < \frac{B}{2} \frac{q^{\beta} - 1}{q - 1},$$

i.e., so that the base q expansion of p^{α} either has all digits $\leq B$ or has a digit $\leq B$ preceding any digit $\geq B$. For brevity we call a number (p, A)-good and (q, B)-good if their base p digits are all $\leq A$, respectively, if their base q digits are all $\leq B$. We consider the following assertion.

LEMMA. Given a number N which is (p, A)-good, say $N = a_n p^n + \cdots + a_m p^m$, $n > \cdots > m \ge 0$, $a_i \le A$, with $N = b_r q^r + \cdots + b_i q^i + \cdots + b_j q^j + \cdots$ where j is the largest index so that $b_j > B$, i is the least index > j so that $b_i < B$ (and so $b_k = B$ for i > k > j), then there exists a number N^* which is (p, A)-good satisfying $N^* \ge b_r q^r + \cdots + b_i q^i$ and so that

$$N^* = b_r q^r + \dots + b_{i+1} q^{i+1} + b_i^* q_i + \dots,$$

where

either $b_i^* = b_i$ and $N^* < N$, or $B > b_i^* > b_i$, or $B = b_i^*$ and the first digit with index less than i which is not equal to B is < B.

It is clear that Theorem 1 follows from this lemma, since after a finite number of modifications we must obtain an N^* which is also (q, B)-good.

Proof of Lemma. Let $T = b_{i-1}q^{i-1} + \cdots + b_0$ be the "tail" of N. If we can subtract any number $\leq T$ from N and get a (p, A)-good number, we have a modification of the first kind to an $N^* < N$. The smallest number which has to be subtracted from N in order to obtain a smaller (p, A)-good number is

$$p^m - A\left(\frac{p^m - 1}{p - 1}\right) = \left(\frac{p - A - 1}{p - 1}\right)(p^m - 1) + 1 = S.$$

So, if $T \ge S$, we set $N^* = N - S$. Thus, from now on we may assume T < S. Since

 $T \ge B((q^i - 1)/(q - 1)) + 1$, then we have

(3)
$$q^{i}-1 < \left(\frac{q-1}{B}\right)\left(\frac{p-A-1}{p-1}\right)(p^{m}-1).$$

Now, we can add any number U to N with

$$q^{i} - T \leq U \leq q^{i} - T + B((q^{j} - 1)/(q - 1)),$$

so that $N^* = N + U$ satisfies $b_i^* = b_i + 1$ and, if N^* is not (q, B)-good, then there is an index $i^* \leq i$ with $b_{i^*}^* < B$ while the first digit $b_{j^*}^* > B$ satisfies $j^* < i^*$. It therefore suffices to prove that there is a (p, A)-good number U in this range with $U < p^m$. This will follow from the following result.

Fact. For every positive integer x, the half-open interval [x, (p-1)x/A) contains a (p, A)-good integer.

Proof of Fact. The ratio between a (p, A)-good number M and the predecessor of the next (p, A)-good number N is maximal when $M = Ap^{r-1} + \cdots + Ap + A = A((p^r - 1)/(p - 1))$ and $N = p^r$. In this case the ratio is (p - 1)/A and the fact follows. \Box

Now, for $x = q^i - T$, we have

$$x \leq q^{i} - 1 - B\left(\frac{q^{i} - 1}{q - 1}\right) = \left(\frac{q - B - 1}{q - 1}\right)(q^{i} - 1)$$

$$\leq A(q^{i} - 1)/(p - 1),$$

so that

$$q^{i} - T + B\left(\frac{q^{i} - 1}{q - 1}\right) \ge \left(1 + \frac{p - 1}{A}\frac{B}{q - 1}\right)(q^{i} - T)$$
$$= \frac{p - 1}{A}\left(\frac{A}{p - 1} + \frac{B}{q - 1}\right)(q^{i} - T)$$
$$\ge \frac{p - 1}{A}(q^{i} - T) = \left(\frac{p - 1}{A}\right)x.$$

Hence, there exists a (p, A)-good number U in this interval. Finally, by (3) we have

$$\left(\frac{p-1}{A}\right) x \leq q^i - 1 < \left(\frac{q-1}{B}\right) \left(1 - \frac{A}{p-1}\right) (p^m - 1)$$
$$\leq \left(\frac{q-1}{B}\right) \left(\frac{B}{q-1}\right) (p^m - 1) = p^m - 1,$$

so that the interval [x, ((p-1)/A)x) lies below p^m . This proves the Lemma and Theorem 1 follows. \Box

One could ask whether the hypotheses of Theorem 1 can be weakened and whether similar results can be obtained for three or more bases instead of just two. At the moment we cannot decide either of these questions and perhaps a new idea will be needed.

THEOREM 2.

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

Proof. We have by definition

$$\sum_{n=1}^{x} f(n) = \sum_{n=1}^{x} \sum_{p \neq \binom{2n}{n}; p \leq n} \frac{1}{p} = \sum_{p \leq x} \frac{A(p;x)}{p}$$

where A(p; x) denotes $|\{k: p \le k < x, p \ne \binom{2k}{k}\}|$. Let $\eta = 1/s > 0$ be small. We first show that the contribution from the primes $p \le x^{\eta}$ is negligible. (The reason for this is that $\binom{2n}{n}$ almost surely is a multiple of a small prime.) Choose r and t so that $1 \le t < p$ and $tp^r \le x < (t+1)p^r$. By (1) it is clear that

$$A(p;x) \leq A(p;(t+1)p')$$

$$\leq (t+1)\left(\frac{p+1}{2}\right)^r \leq \left(\frac{t+1}{t}\right)\left(\frac{p+1}{2p}\right)^r x \leq 2\left(\frac{2}{3}\right)^r x$$

Thus,

$$\frac{1}{x} \sum_{p < x^{\eta}} \frac{A(p; x)}{p} \leq \frac{1}{x} \sum_{r \geq s} \sum_{x^{1/(r+1)} < p < x^{1/r}} \frac{A(p; x)}{p}$$
$$\leq \frac{c_2}{x} \sum_{r \geq s} 2\left(\frac{2}{3}\right)^r x \log\left(1 + \frac{1}{r}\right) = o(c_3^{-s})$$

for some $c_3 > 1$, which is negligible for s sufficiently large.

Therefore, it suffices to consider only those p exceeding $x^{1/s}$. Note that for any $0 \le \epsilon \le \alpha \le 1 - \epsilon$,

(4)
$$\sum_{x^{\alpha} \leqslant p < x^{\alpha + \epsilon}} \frac{A(p; x)}{p} \leqslant x \sum_{x^{\alpha} \leqslant p < x^{\alpha + \epsilon}} \frac{1}{p} < c_4 \ \epsilon x$$

where $c_4 = c_4(\alpha)$. But for each p with $x^{1/(r+1)+\epsilon} ,$

$$A(p;x) = x/2^r + o(x).$$

Thus,

$$\frac{1}{x} \sum_{n=1}^{x} f(n) = \frac{1}{x} \sum_{p \le x} \frac{A(p; x)}{p}$$

$$= \frac{1}{x} \left(\sum_{p \le x^{1/s}} \frac{A(p; x)}{p} + \sum_{p > x^{1/s}} \frac{A(p; x)}{p} \right)$$

$$= \frac{1}{x} \sum_{r=1}^{s-1} \sum_{x^{1/(r+1)}
$$= \frac{1}{x} \sum_{r=1}^{s-1} \left(\sum_{x^{1/(r+1)}$$$$

$$= \sum_{r=1}^{s-1} \frac{1}{2^r} \log\left(1 + \frac{1}{r}\right) + \epsilon_s + o(1)$$

where $\epsilon_s \rightarrow 0$ as $s \rightarrow \infty$. Hence

$$\lim_{x \to \infty} \frac{1}{x} \sum_{n=1}^{x} f(n) = \sum_{r=1}^{\infty} \frac{1}{2^r} \log\left(1 + \frac{1}{r}\right) = \sum_{r=2}^{\infty} \frac{1}{2^r} \log r = c_0. \quad \Box$$

THEOREM 3.

$$\lim_{x \to \infty} \frac{1}{x} \sum_{n=1}^{x} f^{2}(n) = c_{0}^{2}.$$

Proof. We can write

$$\frac{1}{x}\sum_{n=1}^{x}f^{2}(n) = \frac{1}{x}\sum_{n=1}^{x}\sum_{p,q\neq\binom{2n}{n}; p,q\leqslant n}\frac{1}{pq} = \frac{1}{x}\sum_{p,q\leqslant x}\frac{A(p,q;x)}{pq}$$

where

$$A(p, q; x) = \left| \left\{ k: p, q \leq k \leq x, \left(\begin{pmatrix} 2k \\ k \end{pmatrix}, pq \right) = 1 \right\} \right|.$$

As before we first choose a large s and then a small ϵ depending on s. The exact dependence will be clear shortly. We partition the pairs of primes (p, q) into three classes:

I. $p, q \le x^{1/s}$; II. $p \le x^{1/s} < q$ or $q \le x^{1/s} < p$; III. $p, q > x^{1/s}$.

For class I it can be seen by the same argument as in the proof of Theorem 2 that

$$\frac{1}{x} \sum_{p,q \le x^{1/s}} \frac{A(p,q;x)}{pq} = o(c_5^{-s}), \quad c_5 > 1.$$

For class II we observe that

$$\frac{1}{x} \sum_{p \leqslant x^{1/s}} \sum_{x^{1/s} \leqslant q \leqslant s} \frac{A(p, q; x)}{pq}$$
$$\leqslant \frac{1}{x} \sum_{p \leqslant x^{1/s}} \frac{A(p; x)}{p} \sum_{x^{1/s} \leqslant q \leqslant x} \frac{1}{q} < c_6^{-s} \log s = o(1)$$

as $s \rightarrow \infty$.

Finally, we consider class III. Suppose $x^{1/(r+1)} , <math>x^{1/(t+1)} < q \le x^{1/r}$, $1 \le r$, $t \le s$. Let $w_1 < w_2 < \cdots < w_{r+t}$ be the sequence of numbers p^i and q^i , $1 \le i \le r$, $1 \le j \le t$, arranged in increasing order. The numbers k which contribute to A(p, q; x) satisfy the condition

$$k \equiv z_k \pmod{w_k}, \quad 0 \le z_k \le w_k/2, \ k = 1, 2, \cdots, r + t.$$

Hence, if we now assume

(5)
$$w_{k+1}/w_k > x^{\epsilon}, \quad k = 1, \cdots, r+t-1, \ x/w_{r+t} > x^{\epsilon},$$

then we see by (1) that

$$A(p, q; x) = \frac{x}{2^{r+t}} + o(x).$$

Summing this over p and q, we get the main contribution of $c_0^2 x + o(x)$. It is easily seen that the contributions of the pairs (p, q) not satisfying (5) is negligible. For if (5) does not hold, then either

$$p^r \ge x^{1-\epsilon}$$
, i.e., $x^{(1-\epsilon)/r} \le p \le x^{1/r}$

or

$$q^t \ge x^{1-\epsilon}$$
, i.e., $x^{(1-\epsilon)/t} \le q \le x^{1/t}$

or

$$x^{-\epsilon} \leq p^u q^v \leq x^{\epsilon}$$
 for some $1 \leq u \leq r, \ 1 \leq v \leq t$.

Summing 1/pq in these cases, we get o(1) as $\epsilon \rightarrow 0$. Strictly speaking, we only proved

$$\sum_{n \leq x} f^2(n) \leq c_0^2 x + o(x).$$

However, Theorem 3 follows at once by Theorem 2 and the arithmetic-quadratic mean inequality. \Box

COROLLARY. For all $\epsilon > 0$,

$$\lim_{x\to\infty}\frac{1}{x}|\{n\leqslant x\colon |f(n)-c_0|>\epsilon\}|=0.$$

Observe that, in fact, the proofs of Theorems 2 and 3 show that for $0 < \alpha, \beta < 1$ with $\beta - \alpha > \eta > 0$ we have for almost all n,

(6)
$$\sum_{p \neq \binom{2n}{n}; n^{\alpha}$$

uniformly in η . From this it now follows by the sieve method that:

Theorem 4. For $\alpha < 1$,

$$\left|\left\{m: 1 \leq m \leq n^{\alpha}, m \neq \binom{2n}{n}\right\}\right| = c(\alpha)n^{\alpha} + o(n^{\alpha}),$$

ı,

where $c(\alpha) \rightarrow 1$ as $\alpha \rightarrow 0$. (In fact, $c(\alpha)$ can be explicitly calculated.)

A well-known averaging argument gives: If $\epsilon > 0$ and $r \ge r(\epsilon)$, then for any p satisfying $x^{1/r} , then with the exception of at most <math>x/c_{\epsilon}^{r}$ $n's \le x$, we have $p^{\alpha} || \binom{2n}{n}$ with $n^{\frac{1}{2}-\epsilon} < p^{\alpha} < n^{\frac{1}{2}+\epsilon}$, where $c_{\epsilon} > 1$. This leads to the following:

THEOREM 5. Suppose for some $\epsilon > 0$ and $m \leq x$, $p^{\alpha} | m \Rightarrow p^{\alpha} < x^{\epsilon}$. Then

$$\left|\left\{n \leq x : m \not \binom{2n}{n}\right\}\right| < \frac{x}{c_7^{1/\epsilon}},$$

where $c_7 > 1$.

PROOF. Suppose $x^{1/r+1} < p^{\alpha} < x^{1/r}$. By the above remark, it is certainly true that at most $x/c_{\epsilon}^{r} n$'s $\leq x$ have $p^{\beta} \| \binom{2n}{n}$ with $p^{\beta} < n^{\frac{1}{2}-\epsilon}$. Since at most r different prime powers dividing m can lie between $x^{1/r+1}$ and $x^{1/r}$, then these prime powers can knock out at most $rx/c_{\epsilon}^{r} n$'s less than x. On the other hand, if $p^{\beta} \| \binom{2n}{n}$ with $p^{\beta} \geq n^{\frac{1}{2}-\epsilon}$, then the prime p can cause no trouble provided $n^{\frac{1}{2}-\epsilon} \geq x^{\epsilon}$, i.e., $n \geq x^{2\epsilon/(1-2\epsilon)}$. Therefore

$$\left| \left\{ n \le x \colon m \neq \binom{2n}{n} \right\} \right| < \sum_{r>1/\epsilon} \frac{rx}{c_{\epsilon}'} + x^{2\epsilon/(1-2\epsilon)}$$
$$< \frac{x}{c_{7}^{1/\epsilon}} \text{ for some } c_{7} > 1. \quad \Box$$

By the preceding methods, we can prove the following result.

THEOREM. Let p be fixed. Then

$$\left|\left\{n \leq x : p^{\alpha} \| \binom{2n}{n}, p^{\alpha} \notin (n^{\frac{1}{2}-\epsilon}, n^{\frac{1}{2}+\epsilon})\right\}\right| = o(x).$$

 ≈ -3

In fact, as we have already observed, this result holds for $p = o(x^{\eta})$.

We would like to be able to turn this result around, i.e., to show that for fixed n,

$$\left|\left\{p \leq n : p^{\alpha} \parallel \binom{2n}{n}, p^{\alpha} \notin (n^{\frac{1}{2}-\epsilon}, n^{\frac{1}{2}+\epsilon})\right\}\right|$$

is in some sense small. For example, put $f_{\epsilon}(n) = \Sigma' 1/p$, where the dash in the summation indicates that the sum is extended over all primes $p \leq n$ for which $p^{\alpha} || \binom{2n}{n}$ and $p^{\alpha} \notin (n^{\frac{1}{2}-\epsilon}, n^{\frac{1}{2}+\epsilon})$. It seems possible that for every $\epsilon > 0$, $f_{\epsilon}(n) < c(\epsilon)$. By the methods of Theorems 2 and 3, we can prove that the limits

$$\lim_{x \to \infty} \frac{1}{x} \sum_{n \le x} f_{\epsilon}(n) = c_{\epsilon} \quad \text{and} \quad \lim_{x \to \infty} \frac{1}{x} \sum_{n \le x} f_{\epsilon}^{2}(n) = c_{\epsilon}^{2}$$

exist.

By methods similar to those we have employed earlier, it is not difficult to prove the following

(7)
$$\sum_{\substack{p \mid \binom{2n}{n}; p \leq n}} \frac{1}{p} > c \log \log n.$$

There is no doubt that (7) holds for any $c > 1 - \epsilon$, and this would follow, of course, from the boundedness of $\sum_{\substack{p \neq \binom{2n}{n}; p \leq n}} 1/p$. In this connection we would like to state the following conjecture:

$$\sum_{p \le n} {* \frac{1}{p}} = (\frac{1}{2} + o(1)) \log \log n$$

where the * indicates that the summation is extended over all primes p such that n = kp + r, where p/2 < r < p and k is integral.

Before closing the paper, we make a few random remarks about divisibility properties of binomial coefficients. It is well known that $\binom{2n}{n}/(n+1)$ is always an integer. Balakran [1] proved that $(n+1)^2 |\binom{2n}{n}$ for infinitely many n and by his method one can prove that for every k there are infinitely many n, so that $(n+1)^k |\binom{2n}{n}$, and also for every k there are infinitely many n for which (2n)!/(n+1)!(n+k)!is an integer. (In fact, this even holds if $k < c \log n$ if c is a sufficiently small absolute constant.) It seems certain that for every k there are infinitely many integers n for which (2n)!/(n+k)!(n+k)! is an integer, but we cannot prove this even for k = 2.

An old result of P. Erdös (see [3]) states that there is an absolute constant c so that if n!/a!b! is an integer then $a + b < n + c \log n$, but for infinitely many values of n and some c > 0, n!/a!b! is an integer with $a + b = n + c \log n$. In fact, it is not hard to show that for all n with the exception of a sequence of density 0, $(2n)!/n! [n + c \log n]!$ is an integer. We do not give the details of any of these results (the proofs are fairly simple).

There is one curious problem here. As stated before, n!/a!b! cannot be an integer for $a + b \ge n + c \log n$. It is possible that this is due only to the small primes.

90

More precisely, is the following result true: To every c there is a k so that for infinitely many n (all $n > n_0(k, c)$?) there are suitable a and b such that $a + b > n + c \log n$ and n!/a!b! has no prime factor > k in its denominator? Also, suppose $a > \epsilon n$, $b > \epsilon n$, and $a + b > n + c \log n$. Can it happen that n!(a + b - n)!/a!b! is an integer?

Finally, while there is no doubt that there exist infinitely many pairs $\binom{2m}{m}$, $\binom{2n}{n}$ which have the same set of prime divisors, e.g., $\binom{174}{87}$, $\binom{176}{88}$ or $\binom{1214}{607}$, $\binom{1216}{608}$, we are not at present able to prove this.

Let us denote by A(n) the least integer which does not divide $\binom{2n}{n}$. Of course, A(n) is always a prime power. It is not hard to show that except for a set of density 0,

(8)
$$\exp((\log n)^{\frac{1}{2}-\epsilon}) < A(n) < \exp((\log n)^{\frac{1}{2}+\epsilon}).$$

It would not be difficult to obtain sharper results than (8), but an asymptotic formula seems hard. Below we tabulate the first 100 values of A(n).

8	1		2	17	ABLE I				
n	A(n)	n	A(n)	n	A(n)	n	A(n)	<u>n</u>	A(n)
1	3	21	7	41	13	61	5	81	3
2	2 ²	2	7	2	13	2	5	2	3
3	3	3	7	3	13.	3	7	3	32
4	3	4	7	4	11	4	2 ²	4	3
5	5	5	5	5	11	5	7	5	3
6	5	6	5	6	11	6	7	6	17
7	5	7	3	7	11	7	3 ²	7	3 ²
8	2 ²	8	3	8	2 ³	8	2 ³	8	32
9	3	9	3 ²	9	7	9	13	9	17
0	3	30	3	50	5	70	7	90	3
1	5	1	3	1	5	1	7	1	3
2	3	2	2 ²	2	5	2	7	2	32
3	3	3	2 ³	3	13	3	7	3	3
4	7	4	2 ³	4	3 ²	4	24	4	3
5	7	5	5	5	5	5	17	5	19
6	2 ²	6	3	6	5	6	24	6	2 ³
7	7	7	3	7		7	19	7	24
8	2 ³	8	3 ²	8	5 7	8	13	8	7
9	3 ²	9	3	9	7	9	13	9	7
0	2 ³	40	3	60	5	80	2 ³	100	777

TABLE 1

We wish to thank N. J. A. Sloane for obtaining the numerical evidence we cite at the beginning of the paper. Bell Laboratories Murray Hill, New Jersey 07974 and Mathematics Institute of the Hungarian Academy of Science Budapest, Hungary

Bell Laboratories Murray Hill, New Jersey 07974

Eötvös Lorand University Budapest, Hungary

Mathematics Department University of California Los Angeles, California 90024

1. H. BALAKRAN, "On the values of n which make (2n)!/(n + 1)! (n + 1)! an integer," J. Indian Math. Soc., v. 18, 1929, pp. 97-100.

P. ERDÖS, "On some divisibility properties of (²ⁿ_n)," Canad. Math. Bull., v. 7, 1964, pp. 513-518. MR 30 #52.

3. P. ERDÖS, "Aufgabe 557," Elemente Math., v. 23, 1968, pp. 111-113.