THE SUM OF BINOMIAL COEFFICIENTS AND INTEGER FACTORIZATION

Yingpu Deng
Key Laboratory of Mathematics Mechanization, NCMIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P. R. China
dengyp@amss.ac.cn

Yanbin Pan
Key Laboratory of Mathematics Mechanization, NCMIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China
panyanbin@amss.ac.cn

Received: 10/22/14, Revised: 1/18/16, Accepted: 6/3/16, Published: 6/10/16

Abstract

The combinatorial sum of binomial coefficients

$$\left[\begin{array}{c} n \\ i \end{array}\right]_r (a) := \sum_{k \equiv i \,(\mathrm{mod}\, r)} \binom{n}{k} a^{n-k}$$

has been studied widely in combinatorial number theory, especially when $a = 1$ and $a = -1$. In this paper, we connect it with integer factorization for the first time. More precisely, given a composite n, we prove that for any a coprime to n there exists a modulus r such that the combinatorial sum has a nontrivial greatest common divisor with n. Denote by $\mathrm{FAC}(n, a)$ the least r. We present some elementary upper bounds for it and believe that some bounds can be improved further since $\mathrm{FAC}(n, a)$ is usually much smaller in the experiments. We also proposed an algorithm based on the combinatorial sum to factor integers. Unfortunately, it does not work as well as the existing modern factorization methods. However, our method yields some interesting phenomena and some new ideas to factor integers, which makes it worthwhile to study further.

1. Introduction

Let n, i, r and a be integers with $n > 0$ and $r > 0$. Consider the sum of binomial coefficients

$$\left[\begin{array}{c} n \\ i \end{array}\right]_r (a) := \sum_{k \equiv i \,(\mathrm{mod}\, r)} \binom{n}{k} a^{n-k},$$

where $\binom{n}{k}$ is the binomial coefficient with the convention $\binom{n}{k} = 0$ for $k < 0$ or $k > n$. The combinatorial sum has been studied widely in combinatorial number

theory and many of its properties have been explored. For example, Weisman [18] proved that for any prime p and any positive integer α,

$$\left[\begin{array}{c} n \\ n-i \end{array} \right]_{p^\alpha} (-1) \equiv 0 \pmod{p^{\left\lfloor \frac{n}{p^{\alpha-1}(p-1)} \right\rfloor - 1}}.$$

Furthermore, Sun [13] showed that for $\alpha > 1$ and $a \equiv -1 \pmod{p}$,

$$\left[\begin{array}{c} n \\ n-i \end{array} \right]_{p^\alpha} (a) \equiv 0 \pmod{p^{\lfloor \frac{n-p^{\alpha-1}}{\varphi(p^\alpha)} \rfloor}}.$$

Other results about the combinatorial sum can be found in [12, 14, 15, 16].

However, we have to point out that the exact value of the combinatorial sum seems hard to obtain for general r, even when $a = 1$ or $a = -1$. For example, Sun [9, 10, 11] studied the values of the combinatorial sum for $a = 1$, when $r = 3, 4, 5, 6, 8, 9, 10, 12, 16$. Among them, the values are explicitly given just for $r = 3, 4, 6$, whereas the other values are implicitly given by some Lucas sequences.

Note that the former works are combinatorial in nature, which aim to obtain congruences or combinatorial identities. In this paper, we connect the combinatorial sum with integer factorization for the first time.

It is well-known that the integer factorization problem is one of the most famous computational problems, as written by Gauss (Disquisitiones Arithmeticae, 1801, art. 329):

"The problem of distinguishing prime numbers from composite numbers, and of resolving the latter into their prime factors is known to be one of the most important and useful in arithmetic. It has engaged the industry and wisdom of ancient and modern geometers to such an extent that it would be superfluous to discuss the problem at length. ... Further, *the dignity of the science itself seems to require that every possible means be explored for the solution of a problem so elegant and so celebrated* ..."

In 2004, Agrawal, Kayal and Saxena [1] gave a deterministic polynomial-time algorithm to test the primality of a number, which solved the problem of distinguishing prime numbers from composite numbers in theory. However, the problem of resolving composite numbers into their prime factors seems far from being solved, since the best known algorithm to factor integers, the number field sieve method [6], takes subexponential time.

In this paper, we propose a new algorithm to factor integers based on the combinatorial sum. The key observation is that for any composite n and any integer a coprime to n, there always exists a modulus r less than n such that the combinatorial sum $\left[\begin{array}{c} n \\ i \end{array} \right]_r (a)$ has a nontrivial greatest common divisor (gcd) with n. By computing the greatest common divisor, a nontrivial divisor of n can be obtained easily.

Note that the combinatorial sum can be computed efficiently for fixed a and r as in [1]. It remains to show how to find the modulus r for some fixed a. A natural way is to check every r from 1 to n by deciding whether the corresponding combinatorial sum has a nontrivial greatest common divisor with n or not. It is obvious that the time complexity of this procedure depends on the size of r. We denote by $\mathrm{FAC}(n, a)$ the least r such that the sum has a nontrivial greatest common divisor with n, and call it the *factorization number of n with respect to a*. For any even composite, which is of course easy to be factored, we find that the factorization number is at most 3 when $a = \pm 1$, which means that the even composite can also be easily factored with our algorithm. For the RSA modulus, which is considered hard to be factored, we present some elementary upper bounds for the factorization number. However, our bounds seem rather rough since the experiments show that $\mathrm{FAC}(n, a)$ is usually much smaller than the bounds. Hence, we believe that the bounds can be improved further in theory.

Due to lack of good mathematical tools to deal with the combinatorial sum, we do not know how to estimate the factorization number $\mathrm{FAC}(n, a)$ as well as possible when a is fixed. Furthermore, we do not know how to estimate $\min_a \mathrm{FAC}(n, a)$ where a runs over some specific set either. We conjecture that both of the two questions are very difficult and propose them as open problems.

We also implemented our algorithm to factor integers. Unfortunately, it did not work as well as the existing modern factorization methods, such as the number field sieve method. However, our method yields some interesting phenomena and some new ideas to factor integers, which makes it worthwhile to study further.

The paper is organized as follows. We give the definition of the factorization number of a composite in Section 2. We give some properties for the factorization number of an even composite in Section 3. In Section 4, we prove some elementary upper bounds for the factorization number of an RSA modulus. We list some experimental results in Section 5. Finally, a short conclusion and some open problems are given in Section 6.

2. The Factorization Number of a Composite

2.1. The Combinatorial Sum and its Proposition

Definition 2.1. Let n, r, i and a be integers with $n > 0$ and $r > 0$. We define the combinatorial sum of binomial coefficients as

$$\left[\begin{array}{c} n \\ i \end{array} \right]_r (a) := \sum_{\substack{0 \le k \le n \\ k \equiv i \pmod r}} \binom{n}{k} a^{n-k}.$$

For simplicity, we define

$$\left[\begin{array}{c} n \\ i \end{array} \right]_r := \left[\begin{array}{c} n \\ i \end{array} \right]_r (1).$$

The following lemma is useful to compute $\left[\begin{array}{c} n \\ i \end{array} \right]_r (a)$ when r is small.

Lemma 2.2. *Let $\zeta \in \mathbb{C}$ be a primitive r-th root of unity. Then for $0 \le i \le r-1$, we have*

$$\left[\begin{array}{c} n \\ i \end{array} \right]_r (a) = \frac{1}{r} \sum_{j=0}^{r-1} (\zeta^j + a)^n (\zeta^j)^{-i}.$$

Proof. Since $(X+a)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} X^i$, we have

$$\begin{aligned} \sum_{j=0}^{r-1} (\zeta^j + a)^n (\zeta^j)^{-i} &= \sum_{j=0}^{r-1} \left[\sum_{s=0}^{n} \binom{n}{s} \zeta^{js} a^{n-s} \right] (\zeta^j)^{-i} \\ &= \sum_{s=0}^{n} \binom{n}{s} a^{n-s} \sum_{j=0}^{r-1} \zeta^{(s-i)j} \\ &= r \cdot \sum_{\substack{0 \le s \le n \\ s \equiv i \pmod r}} \binom{n}{s} a^{n-s}. \end{aligned}$$

The lemma follows. □

2.2. Integer Factorization and the Factorization Number of a Composite

Given a composite number $n > 1$, integer factorization refers to the question of finding a nontrivial divisor d of n, i.e., $d \mid n$ and $1 < d < n$.

To factor a composite n, we observe the following:

Proposition 2.3. *For any positive composite n, there exists an integer j with $1 < j < n-1$ such that $1 < \gcd(n, \binom{n}{j}) < n$.*

Proof. To prove the proposition, we consider the following two cases.

`Case 1.` n has a square divisor. Assume n has a prime divisor p such that $p^k \parallel n$ with $k > 1$, i.e., $p^k \mid n$ but $p^{k+1} \nmid n$. Notice that

$$\binom{n}{p} = \frac{n(n-1)(n-2)\cdots(n-p+1)}{p!}.$$

Since $p \nmid (n-i)$ for $1 \le i \le p-1$, we have $p^{k-1} \parallel \binom{n}{p}$. Hence $1 < \gcd(n, \binom{n}{p}) < n$.

`Case 2.` n is square-free. Let p and q be two prime divisors of n with $p < q$. As in the first case, it can be concluded that $p \nmid \binom{n}{p}$ but $q \mid \binom{n}{p}$. Hence $1 < \gcd(n, \binom{n}{p}) < n$. □

Note that every $\binom{n}{j}$ is a coefficient of the polynomial $(X+1)^n$. By Proposition 2.3, a natural way to obtain a nontrivial divisor of n is expanding $(X+1)^n$, and then computing each greatest common divisor of its coefficients and n. However, this will take exponential time since there are a total of n coefficients. To reduce the time complexity, we turn to employ another polynomial

$$(X+a)^n \equiv \sum_{i=0}^{r-1} a_i X^i \mod (X^r-1, n),$$

as considered in [1], which yields the definition of the factorization-friendly number of n.

Definition 2.4. Let n be a positive composite, a be an integer coprime to n, and r be a positive integer. Consider the polynomial with an indeterminate X:

$$(X+a)^n \equiv \sum_{i=0}^{r-1} a_i X^i \mod (X^r-1, n),$$

where a_i's are integers with $0 \le a_i \le n-1$ for $0 \le i \le r-1$. We call r a factorization-friendly number of n with respect to a if there exists an i with $0 \le i \le r-1$ such that $\gcd(n, a_i)$ is a nontrivial divisor of n.

By Definition 2.4, it is easy to conclude that for $0 \le i \le r-1$, we have

$$a_i \equiv \left[\begin{array}{c} n \\ i \end{array} \right]_r (a) \pmod{n}.$$

Next we show that the factorization-friendly number of n with respect to an arbitrary a coprime to n must exist.

Proposition 2.5. *For any positive composite n and an integer a coprime to n, $n-1$ is a factorization-friendly number of n with respect to a.*

Proof. Since

$$(X+a)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} X^i,$$

we have

$$(X+a)^n \equiv a^n + X + \sum_{i=2}^{n-2} \binom{n}{i} a^{n-i} X^i \mod (X^{n-1}-1, n).$$

The result follows by Proposition 2.3. □

By Proposition 2.5, we can always factor n by computing the corresponding $n-1$ coefficients. Unfortunately, it will take exponential time too. Generally speaking, for any factorization-friendly number r, it will take $\widetilde{O}\left(r\log^2 n\right)$ time (see [17]) to obtain all the a_i's in Definition 2.4. Therefore, we are interested in the least factorization-friendly number of n.

Definition 2.6. For any positive composite n and an integer a coprime to n, the least factorization-friendly number of n with respect to a is called the factorization number of n with respect to a and is denoted by $\mathrm{FAC}(n,a)$.

By Proposition 2.5, we immediately have

Proposition 2.7. *For any positive composite n and an integer a coprime to n, we have $FAC(n,a)\leq n-1$.*

To find the exact value of $\mathrm{FAC}(n,a)$, a natural way is to check every r from 1 to $n-1$, which yields a new algorithm to factor a composite n.

Input: a composite n.
Output: a nontrivial divisor d of n.
1. Choose some $a\in[2,n-1]$, if $1<d=\gcd(a,n)<n$, output d.
2. Otherwise, for r from 1 to $n-1$ do 3–4:
3. expand $(X+a)^n \mod (X^r-1,n)$ as $\sum_{i=0}^{r-1}a_ix^i$,
4. compute each greatest common divisor d_i of a_i and n. If $1<d_i<n$, output d_i.

We remark that the time complexity of the algorithm depends on the size of $\mathrm{FAC}(n,a)$.

3. The Factorization Number of an Even Composite Number

In this section, we show that the factorization number of an even composite number is usually very small when $a=\pm1$.

3.1. The Factorization Number of an Even Composite Number when $a=1$

Let

$$(X+1)^n\equiv\sum_{i=0}^{r-1}a_iX^i \mod (X^r-1,n).$$

Note that for $0\leq i\leq r-1$, we have

$$a_i\equiv\left[\begin{array}{c}n\\ i\end{array}\right]_r \pmod n.$$

The statements in the following proposition are well-known. The first two are easy to check, and we omit their proof. The third statement is well-known, but for completeness we provide a proof.

Proposition 3.1. *Let n be any positive integer. Then we have:*

(1) for $r = 1$, $\left[\begin{array}{c} n \\ 0 \end{array} \right]_1 = 2^n$;

(2) for $r = 2$, $\left[\begin{array}{c} n \\ 0 \end{array} \right]_2 = \left[\begin{array}{c} n \\ 1 \end{array} \right]_2 = 2^{n-1}$;

(3) for $r = 3$,

$$\begin{aligned}
\left[\begin{array}{c} n \\ 0 \end{array} \right]_3 &= \frac{1}{3}\left(2^n + 2\cos\left(\frac{n\pi}{3}\right)\right), \\
\left[\begin{array}{c} n \\ 1 \end{array} \right]_3 &= \frac{1}{3}\left(2^n + 2\cos\left(\frac{(n-2)\pi}{3}\right)\right), \\
\left[\begin{array}{c} n \\ 2 \end{array} \right]_3 &= \frac{1}{3}\left(2^n + 2\cos\left(\frac{(n+2)\pi}{3}\right)\right).
\end{aligned}$$

Proof of (3). Set $i = \sqrt{-1}$. By Lemma 2.2, let ζ be a primitive 3-rd root of unity. Then we know

$$\zeta = -\frac{1}{2} + \frac{\sqrt{3}}{2}i, \quad \zeta^2 = -\frac{1}{2} - \frac{\sqrt{3}}{2}i,$$

and

$$1 + \zeta = \frac{1}{2} + \frac{\sqrt{3}}{2}i = e^{i\frac{\pi}{3}}, \quad 1 + \zeta^2 = \frac{1}{2} - \frac{\sqrt{3}}{2}i = e^{-i\frac{\pi}{3}}.$$

Thus, we have

$$(1 + \zeta)^n = e^{i\frac{n\pi}{3}}, \quad (1 + \zeta^2)^n = e^{-i\frac{n\pi}{3}},$$

which implies

$$\left[\begin{array}{c} n \\ 0 \end{array} \right]_3 = \frac{1}{3}\sum_{j=0}^{2}(1 + \zeta^j)^n = \frac{1}{3}\left(2^n + e^{i\frac{n\pi}{3}} + e^{-i\frac{n\pi}{3}}\right) = \frac{1}{3}\left(2^n + 2\cos\left(\frac{n\pi}{3}\right)\right).$$

Similarly, we have

$$\begin{aligned}
\left[\begin{array}{c} n \\ 1 \end{array} \right]_3 &= \frac{1}{3}\sum_{j=0}^{2}(1 + \zeta^j)^n\zeta^{-j} = \frac{1}{3}\left(2^n + 2\cos\left(\frac{(n-2)\pi}{3}\right)\right), \\
\left[\begin{array}{c} n \\ 2 \end{array} \right]_3 &= \frac{1}{3}\sum_{j=0}^{2}(1 + \zeta^j)^n\zeta^{-2j} = \frac{1}{3}\left(2^n + 2\cos\left(\frac{(n+2)\pi}{3}\right)\right).
\end{aligned}$$

□

Theorem 3.2. *Let n be a positive composite number. We have:*

(1) if n is even and n is not a power of 2, then $FAC(n,1) = 1$;

(2) if n is odd, then $FAC(n,1) \geq 3$;

(3) if n is a power of 2, then $FAC(n,1) = 3$.

Proof. By Proposition 3.1, it is easy to prove (1) and (2). It remains to prove (3). Suppose n is a power of 2, i.e., $n = 2^m, m \geq 2$. If m is even, then $n \equiv 1 \pmod 3$ and $\frac{n+2}{3}$ is even. Thus $a_2 \equiv \left[\begin{array}{c} n \\ 2 \end{array} \right]_3 = \frac{1}{3}(2^n + 2) \pmod n$, which implies $\gcd(n, a_2) = 2$. If m is odd, we have $\gcd(n, a_1) = 2$ similarly. □

The following proposition can be used to reduce the computation of greatest common divisors.

Proposition 3.3. *Let n, r be two positive integers and i be an integer. Then we have*

$$\left[\begin{array}{c} n \\ i \end{array} \right]_r = \left[\begin{array}{c} n \\ n-i \end{array} \right]_r.$$

Proof. We have

$$\begin{aligned} \left[\begin{array}{c} n \\ n-i \end{array} \right]_r &= \sum_{\substack{0 \leq k \leq n \\ k \equiv n-i \pmod r}} \binom{n}{k} \\ &= \sum_{\substack{0 \leq n-k \leq n \\ n-k \equiv i \pmod r}} \binom{n}{n-k} \\ &= \left[\begin{array}{c} n \\ i \end{array} \right]_r. \end{aligned}$$

□

3.2. The Factorization Number of an Even Composite Number when $a = -1$

Let

$$(X-1)^n \equiv \sum_{i=0}^{r-1} a_i X^i \mod (X^r - 1, n).$$

Note that for $0 \leq i \leq r-1$, we have

$$a_i \equiv \left[\begin{array}{c} n \\ i \end{array} \right]_r (-1) = \sum_{\substack{0 \leq k \leq n \\ k \equiv i \pmod r}} \binom{n}{k} (-1)^{n-k} \pmod n.$$

Proposition 3.4. *Let n be any positive integer. Then we have:*

(1) for $r = 1$, $\left[\begin{array}{c} n \\ 0 \end{array} \right]_1 (-1) = 0$;

(2) for $r = 2$, $\left[\begin{array}{c} n \\ 0 \end{array} \right]_2 (-1) = (-1)^n 2^{n-1}$ *and* $\left[\begin{array}{c} n \\ 1 \end{array} \right]_2 (-1) = (-1)^{n-1} 2^{n-1}$;

(3) for $r = 3$,

$$\begin{aligned}
\left[\begin{array}{c} n \\ 0 \end{array} \right]_3 (-1) =& 3^{\frac{n}{2}-1} \cdot 2 \cos\left(\frac{5n\pi}{6}\right), \\
\left[\begin{array}{c} n \\ 1 \end{array} \right]_3 (-1) =& 3^{\frac{n}{2}-1} \cdot 2 \cos\left(\frac{(5n-4)\pi}{6}\right), \\
\left[\begin{array}{c} n \\ 2 \end{array} \right]_3 (-1) =& 3^{\frac{n}{2}-1} \cdot 2 \cos\left(\frac{(5n-8)\pi}{6}\right).
\end{aligned}$$

Proof. The proof is similar to the one of Proposition 3.1 and we omit the details. □

Theorem 3.5. *Let n be a positive composite number. We have:*

(1) if n is even and n is not a power of 2, then $FAC(n, -1) = 2$;

(2) if n is odd, then $FAC(n, -1) \geq 3$;

(3) if n is a power of 2, then $FAC(n, -1) = 3$.

Proof. By Proposition 3.4, it is easy to prove (1) and (2). It remains to prove (3). Suppose n is a power of 2, i.e., $n = 2^m, m \geq 2$. Then $\frac{5n-4}{6} = \frac{5 \cdot 2^{m-1}-2}{3}$ and $\frac{5n-8}{6} = \frac{5 \cdot 2^{m-1}-4}{3}$. It is easy to conclude that either $\frac{5n-4}{6}$ or $\frac{5n-8}{6}$ must be even, thus either $\left[\begin{array}{c} n \\ 1 \end{array} \right]_3 (-1)$ or $\left[\begin{array}{c} n \\ 2 \end{array} \right]_3 (-1)$ is $2 \cdot 3^{2^{m-1}-1}$. Hence we have either $\gcd(n, a_1) = 2$ or $\gcd(n, a_2) = 2$. □

Similar to Proposition 3.3, we have

Proposition 3.6. *Let n, r be two positive integers, and i be an integer. Then we have*

$$\left[\begin{array}{c} n \\ i \end{array} \right]_r (-1) = (-1)^n \cdot \left[\begin{array}{c} n \\ n-i \end{array} \right]_r (-1).$$

4. The Factorization Number of an RSA Modulus

In this section, we consider the factorization number of an RSA modulus. An RSA modulus n is a product of two distinct odd primes, that is, $n = pq$ where $p < q$ are two distinct odd primes. The RSA modulus $n = pq$ where $p < q$ are two distinct big odd primes with the same number of bits is considered hard to be factored. To analyze the time complexity of our algorithm for an RSA modulus n, we present some upper bounds for $\mathrm{FAC}(n, a)$. First, we introduce some useful results.

4.1. Some Useful Results

Theorem 4.1 (Lucas' Theorem). *For any prime p, suppose $a = a_0 + a_1 p + \cdots + a_k p^k$, $b = b_0 + b_1 p + \cdots + b_k p^k$, where $0 \le a_i, b_i < p$ for $i = 0, 1, \ldots, k$. Then we have*

$$\binom{a}{b} \equiv \prod_{i=0}^{k} \binom{a_i}{b_i} \pmod{p}.$$

See [2] for a proof of Lucas' Theorem.

Lemma 4.2. *For any positive integer $n > 1$, let m be an integer satisfying $0 < m < n$ and $\gcd(n, m) = 1$. Then we have $n \mid \binom{n}{m}$.*

Proof. Notice that

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} = \frac{n}{m} \cdot \frac{(n-1)!}{(m-1)!(n-m)!} = \frac{n}{m} \cdot \binom{n-1}{m-1}$$

is an integer and $\binom{n-1}{m-1}$ is also an integer. We have $m \mid n\binom{n-1}{m-1}$. Since $\gcd(n, m) = 1$, we have $m \mid \binom{n-1}{m-1}$, which yields $n \mid \binom{n}{m}$. □

Lemma 4.3. *Suppose $n = pq$, where $p < q$ are two distinct odd primes. Then we have*

(1) for $0 < i < q$, $q \mid \binom{n}{pi}$;

(2) for $0 < j < p$, $p \mid \binom{n}{qj}$ but $q \nmid \binom{n}{qj}$.

Proof. Since $q \nmid pi$ for $0 < i < q$, we have $q \mid \binom{n}{pi}$ by Lucas' Theorem. Similarly, we have $p \mid \binom{n}{qj}$ for $0 < j < p$. For $0 < j < p$, we have $\binom{n}{qj} \equiv \binom{p}{j} \pmod{q}$ by Lucas' Theorem. Since $p < q$, we have $q \nmid \binom{p}{j}$. Hence $q \nmid \binom{n}{qj}$ for $0 < j < p$. □

4.2. The Factorization Number of an RSA Modulus

Now we present some upper bounds for $\mathrm{FAC}(n,1)$, where n is an RSA modulus. First we show that

Theorem 4.4. *For an RSA modulus $n = pq$ with $p < q$, p is a factorization-friendly number of n with respect to 1, which yields $FAC(n,1) \le p < \sqrt{n}$.*

Proof. It is sufficient to prove that p is a factorization-friendly number of n with respect to 1. Write $q = ap + k$, where $a > 0$ and $0 < k < p$. We will prove the theorem by showing that

$$\gcd\left(n, \left[\begin{array}{c} n \\ k \end{array}\right]_p\right) = p.$$

Set $I = \{k + ps \mid s \in \mathbb{Z}, 0 \le k + ps \le n\}$. Obviously, $0, n$ and $pi (0 < i < q)$ are not in I but q is in I. If qj is in I for some j with $0 < j < p$, then $k \equiv qj \equiv kj \pmod p$, which yields $j \equiv 1 \pmod p$, that is, $j = 1$. Hence $q \in I$ and the other elements in I are coprime to n. By Lemma 4.2, we have

$$\left[\begin{array}{c} n \\ k \end{array}\right]_p = \sum_{\substack{0 \le t \le n \\ t \equiv k \pmod p}} \binom{n}{t} \equiv \binom{n}{q} \pmod n.$$

By Lemma 4.3, we have $p \mid \binom{n}{q}$ but $q \nmid \binom{n}{q}$. Hence $\gcd\left(n, \left[\begin{array}{c} n \\ k \end{array}\right]_p\right) = p$. □

Together with Theorem 3.2, we have

Corollary 4.5. *For an integer $n = 3q$ where $q > 3$ is a prime, we have $FAC(n,1) = 3$.*

In fact, we can do a little better.

Theorem 4.6. *Let $n = pq$ be an RSA modulus with $p < q < 2p$, and c be a positive integer such that $3c \le 2p - q - 1$. Then $p - c$ is a factorization-friendly number of n with respect to 1, which yields $FAC(n,1) \le p - c$.*

Proof. Write $q = p + k$, where $0 < k < p$. Since k is even, we have $p \ge 3c + 3$, which yields $c + 1 < p - c$ and $k + c < p - c$. It is easy to show that $\gcd(p - c, c) = 1$ since $(p - c) + c = p$ is a prime. Hence the set $\{-ci \pmod{p-c} \mid i = 1, 2, \ldots, c + 1\}$ has exactly $c + 1$ elements. Similarly, it can be shown that $\gcd(p - c, k + c) = 1$ and the set $\{(k + c)j \pmod{p-c} \mid j = 1, 2, \ldots, c\}$ has c elements. Thus there must exist an element a in the first set but not in the second set. Let $a \equiv -ci_1 \pmod{p-c}$ with $1 \le i_1 \le c + 1$ and $0 < a < p - c$.

Set $I = \{a + (p - c)s \mid s \in \mathbb{Z}, 0 \le a + (p - c)s \le n\}$. Obviously, $0 \notin I$. Since $n = pq \equiv c(k + c) \pmod{p-c}$, we have $n \notin I$ by the choice of a. For

$0 < i < q$, if $pi \in I$, then $pi \equiv a \pmod{p-c}$. We have $ci \equiv -ci_1 \pmod{p-c}$, which implies $i \equiv -i_1 \pmod{p-c}$. Hence $i = p-c-i_1+(p-c)s, s \geq 0$. Since $p-c-i_1+p-c = 2p-2c-i_1 \geq 2p-3c-1 \geq p+k = q$, we have $i = p-c-i_1 \triangleq i_0$, which means that for $0 < i < q$ only $pi_0 \in I$. For $0 < j < p$, if $qj \in I$, then $qj \equiv a \pmod{p-c}$. We have $(k+c)j \equiv a \pmod{p-c}$. The equation has a unique solution j_0 with $0 \leq j_0 < p-c$. We know $j_0 \geq c+1$ by the choice of a. Since $j_0+p-c > p$, we have $j = j_0$, which means that for $0 < j < p$ only $qj_0 \in I$. Thus, we get

$$\left[\begin{array}{c} n \\ a \end{array} \right]_{p-c} \equiv \binom{n}{pi_0} + \binom{n}{qj_0} \pmod{n}.$$

By Lemma 4.3, we have

$$q \nmid \left[\begin{array}{c} n \\ a \end{array} \right]_{p-c}.$$

Since $q - i_0 + 1 \leq p$, we have $p \mid \binom{n}{pi_0}$, which implies

$$p \mid \left[\begin{array}{c} n \\ a \end{array} \right]_{p-c}.$$

Hence, we have

$$\gcd\left(n, \left[\begin{array}{c} n \\ a \end{array} \right]_{p-c}\right) = p.$$

Therefore $p-c$ is a factorization-friendly number of n with respect to 1. □

Corollary 4.7. *Let $n = pq$ be an RSA modulus with $p < q < 2p$. Write $q = p+k$ with $0 < k < p$. If $k < \varepsilon p$ for some $0 < \varepsilon < 1$ and $p \geq \frac{3}{1-\varepsilon}$, then $FAC(n,1) \leq p - \lfloor \frac{1-\varepsilon}{3} p \rfloor \approx \frac{2+\varepsilon}{3} p$.*

Proof. Put $c = \lfloor \frac{1-\varepsilon}{3} p \rfloor$ and then the result follows by Theorem 4.6. □

Theorem 4.8. *Let $n = pq$ be an RSA modulus with $p < q < 2p$. Write $q = p+k$ with $0 < k < p$. If $n \equiv -1 \pmod 4$ and $2 < k < \frac{2}{3}p$, then $r := \frac{p}{2} + \frac{3}{4}k$ is a factorization-friendly number of n with respect to 1, which yields $FAC(n,1) \leq \frac{p}{2} + \frac{3}{4}k$.*

Proof. Since $n \equiv -1 \pmod 4$, it can be concluded that one of p and q is congruent to 1 modulo 4, whereas the other is congruent to -1 modulo 4. Hence we have $k \equiv 2 \pmod 4$, which yields that r is a positive integer and $r < p$.

Let $a \in \mathbb{Z}$ satisfy $0 \leq a < r$ and $a \equiv (r-1)p \pmod r$. We have $a > 0$ since $r > 1$. Set $I = \{a + rs \mid s \in \mathbb{Z}, 0 \leq a + rs \leq n\}$. By a similar analysis as in the proof of Theorem 4.6, we have $p(r-1) \in I$, $q(r-3) \in I$, and the other elements in I are coprime to n. Thus

$$\left[\begin{array}{c} n \\ a \end{array} \right]_{r} \equiv \binom{n}{p(r-1)} + \binom{n}{q(r-3)} \pmod{n}.$$

Similarly, we have

$$\gcd\left(n, \left[\begin{array}{c} n \\ a \end{array}\right]_r\right) = p,$$

which implies that r is a factorization-friendly number of n with respect to 1. □

Corollary 4.9. *With notation as in Theorem 4.8, if we further suppose $k < \varepsilon p$ with $0 < \varepsilon \leq \frac{2}{3}$, then we have FAC$(n, 1) \leq (\frac{1}{2} + \frac{3}{4}\varepsilon)p$.*

Remark 4.10. Comparing the bounds in Corollary 4.7 and Corollary 4.9, it is easy to see that $\frac{2+\varepsilon}{3} = \frac{1}{2} + \frac{3}{4}\varepsilon$ when $\varepsilon = \frac{2}{5}$, $\frac{2+\varepsilon}{3} < \frac{1}{2} + \frac{3}{4}\varepsilon$ when $\varepsilon > \frac{2}{5}$, and $\frac{2+\varepsilon}{3} > \frac{1}{2} + \frac{3}{4}\varepsilon$ when $\varepsilon < \frac{2}{5}$.

Moreover, from the proofs of Theorems 4.4, 4.6 and 4.8 we know that there is only one binomial coefficient left when the combinatorial sum is reduced modulo p (or modulo q). Therefore, all the bounds for FAC$(n, 1)$ above hold also for FAC(n, a) with $a \in \mathbb{Z}$ coprime to n.

Finally, we have to point out that the upper bounds for FAC$(n, 1)$ above are rather rough, since our experiments show that FAC$(n, 1)$ is usually much smaller than p. See Section 5 for more details.

4.3. The Factorization Number of an RSA Modulus with Twin Primes

We next present two interesting results, which provide partial evidence of our conjecture that a number that can be easily factored by other methods can also be easily factored by our method.

Theorem 4.11. *Let $n = pq$ be an RSA modulus. If $q = p + 2$, i.e., p and q are twin primes, then we have FAC$(n, 1) \leq 6$.*

Proof. Since FAC$(15, 1) = 3$ by Corollary 4.5, we can assume $p \geq 5$. It is easy to conclude that, there is a positive integer k such that $p = 6k - 1$ and $q = 6k + 1$. Hence it can be shown that

$$\left[\begin{array}{c} n \\ 2 \end{array}\right]_6 \equiv \sum_{\substack{0<i<q \\ i\equiv 4 \pmod 6}} \binom{n}{pi} + \sum_{\substack{0<j<p \\ j\equiv 2 \pmod 6}} \binom{n}{qj} \pmod n.$$

By Lemma 4.3, we know $p \mid \binom{n}{qj}$ for each j with $0 < j < p$. Moreover, for each i with $0 < i < q$ and $i \equiv 4 \pmod 6$, we have $4 \leq i \leq q - 3 = p - 1$, which implies $\binom{n}{pi} \equiv \binom{q}{i} = \binom{p+2}{i} \equiv \binom{1}{0}\binom{2}{i} = 0 \pmod p$ by Lucas' theorem. Thus

$$p \mid \left[\begin{array}{c} n \\ 2 \end{array}\right]_6.$$

By Lemma 4.3, we have $q \mid \binom{n}{pi}$ for each i with $0 < i < q$ similarly. By Lucas' theorem, we have $\binom{n}{qj} \equiv \binom{p}{j} \pmod{q}$ for each j with $0 < j < p$. Hence

$$\left[\begin{array}{c} n \\ 2 \end{array} \right]_6 \equiv \left[\begin{array}{c} p \\ 2 \end{array} \right]_6 \pmod{q}.$$

Notice that it can be concluded from [9] that

$$\left[\begin{array}{c} p \\ 2 \end{array} \right]_6 = \frac{1}{6}(1 + 2^p - 3^{\frac{p+1}{2}}), \quad \text{for an even } k,$$
$$\left[\begin{array}{c} p \\ 2 \end{array} \right]_6 = \frac{1}{6}(1 + 2^p + 3^{\frac{p+1}{2}}), \quad \text{for an odd } k.$$

By Euler's criterion and the Quadratic Reciprocity Law of the Legendre symbol, we know

$$3^{\frac{p+1}{2}} = 3^{\frac{q-1}{2}} \equiv \left(\frac{3}{q}\right) = (-1)^{(q-1)/2} \left(\frac{q}{3}\right) = (-1)^k \left(\frac{1}{3}\right) = (-1)^k \pmod{q}.$$

Thus, in both cases, we have

$$\left[\begin{array}{c} p \\ 2 \end{array} \right]_6 \equiv \frac{1}{6} \cdot 2^p \pmod{q},$$

which implies $q \nmid \left[\begin{array}{c} p \\ 2 \end{array} \right]_6$. Therefore, we have

$$q \nmid \left[\begin{array}{c} n \\ 2 \end{array} \right]_6.$$

Finally, we get

$$\gcd\left(n, \left[\begin{array}{c} n \\ 2 \end{array} \right]_6\right) = p.$$

□

It is interesting that we can improve the bound in Theorem 4.11 for some n's.

Theorem 4.12. *Let $n = pq$ be an RSA modulus. Suppose $q = p + 2$, i.e., p and q are twin primes. If $p > 5$ and $p \equiv \pm 1 \pmod{5}$, then $FAC(n, 1) \leq 5$.*

Proof. We first consider the case when $p \equiv 1 \pmod{5}$, $q \equiv 3 \pmod{5}$. It is easy to conclude that

$$\left[\begin{array}{c} n \\ 4 \end{array} \right]_5 \equiv \left[\begin{array}{c} q \\ 4 \end{array} \right]_5 \pmod{p},$$
$$\left[\begin{array}{c} n \\ 4 \end{array} \right]_5 \equiv \left[\begin{array}{c} p \\ 3 \end{array} \right]_5 \pmod{q}.$$

Define two sequences $\{u_i\}_{i\geq 0}, \{v_i\}_{i\geq 0}$ as follows:

$$u_0 = 0, u_1 = 1, u_{i+1} = u_i + u_{i-1} \text{ for } i \geq 1,$$
$$v_0 = 2, v_1 = 1, v_{i+1} = v_i + v_{i-1} \text{ for } i \geq 1.$$

Since the Legendre symbol $\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right) = \left(\frac{1}{5}\right) = 1$, we have $v_{p-1} \equiv 2 \pmod{p}, u_{p-1} \equiv 0 \pmod{p}$ and $u_p \equiv 1 \pmod{p}$ by a well-known result of the Fibonacci sequence (see [4]). It follows that $v_q = v_{p+2} \equiv 4 \pmod{p}$. Moreover, by a result of Sun [9], we have

$$-2v_q = 5 \left[\begin{array}{c} q \\ 4 \end{array} \right]_5 - 2^q.$$

Hence we have $p \mid \left[\begin{array}{c} q \\ 4 \end{array} \right]_5$ since $2^q \equiv 8 \pmod{p}$, which implies

$$p \mid \left[\begin{array}{c} n \\ 4 \end{array} \right]_5.$$

On the other hand, since the Legendre symbol $\left(\frac{5}{q}\right) = -1$, similarly we have $v_{q+1} \equiv -2 \pmod{q}, u_{q+1} \equiv 0 \pmod{q}$ and $u_q \equiv -1 \pmod{q}$, and it follows that

$$v_p = v_{q-2} \equiv 4 \pmod{q}.$$

By a result of Sun [9], we have

$$-2v_p = 5 \left[\begin{array}{c} p \\ 3 \end{array} \right]_5 - 2^p.$$

Hence $q \nmid \left[\begin{array}{c} p \\ 3 \end{array} \right]_5$, which implies

$$q \nmid \left[\begin{array}{c} n \\ 4 \end{array} \right]_5.$$

Therefore,

$$\gcd\left(n, \left[\begin{array}{c} n \\ 4 \end{array} \right]_5\right) = p.$$

For the case $p \equiv -1 (\text{mod } 5)$, the proof is similar. □

Notice that for the case $p \equiv 2 (\text{mod } 5)$, the experiments show that the bound 6 can not be replaced by 5.

5. Experimental Results

We have done numerous experiments using NTL library [8]. These experiments show the remarkable fact that FAC(n, a)'s, even FAC$(n, 1)$'s, are usually much smaller than n, and they grow very slowly as n increases. In Table 1 we list some values of FAC$(n, 1)$'s for $n = pq$ where p and q have three digits.

$n = pq$	$F(1)$	$n = pq$	$F(1)$	$n = pq$	$F(1)$
10403=101*103	5	10807=101*107	8	11009=101*109	13
11413=101*113	17	12827=101*127	21	13231=101*131	8
13837=101*137	22	14039=101*139	12	15049=101*149	21
15251=101*151	17	15857=101*157	18	16463=101*163	9
16867=101*167	15	17473=101*173	21	18079=101*179	12
18281=101*181	22	19291=101*191	20	19493=101*193	9
19897=101*197	13	20099=101*199	12	21311=101*211	9
251659=359*701	38	254531=359*709	17	258121=359*719	5
235247=367*641	20	235981=367*643	9	237449=367*647	12
255067=379*673	33	256583=379*677	51	258857=379*683	36
409763=593*691	25	415693=593*701	7	420437=593*709	15
563903=607*929	24	571187=607*941	43	586969=607*967	60
621787=701*887	52	536713=709*757	59	750187=757*991	72
812909=853*953	17	756731=857*883	33	782549=859*911	61
921551=953*967	17	936799=953*983	34	988027=991*997	9

Table 1

For a fixed n, different a's will generally lead to distinct FAC(n, a)'s. Usually, there exists some a such that the corresponding FAC(n, a) is remarkably less than FAC$(n, 1)$, which indicates that we can obtain that combinatorial sum much more quickly when choosing such a. We also list some experimental results in Table 2.

$n = pq$	FAC$(n, 1)$	a	FAC(n, a)
323910211=16453*19687	266	224606094	25
401112223=16487*24329	266	254658360	62
481118119=18371*26189	260	447652040	16
556453211=20333*27367	39	501040105	23
580839353=20201*28753	209	494398594	17
712415273=25237*28229	113	395527894	47
89441974637=276839*323083	712	49599857930	68
91457375567=300721*304127	584	41380446395	123
154709636971=332933*464687	1161	16603703892	162
408187969489=531911*767399	1025	142966224429	67
702358343579=733003*958193	2467	413854661934	245
1039342803007=1012751*1026257	1771	176505030244	198

Table 2

6. Conclusion and Open Problems

It is well-known that integer factorization is a very important computational problem. However, there has been no substantial progress on solving this problem since the invention of the general number field sieve method in 1993 [6, 5, 3]. We propose a new method to factor integers based on combinatorial sums of binomial coefficients in this paper. As we know, it is the first time to connect the combinatorial sum with integer factorization. We believe that our method yields new and important idea, which makes it worthwhile to study further.

Of course, there are still some open problems left. One is to obtain a tighter upper bound of $\mathrm{FAC}(n,a)$ for some fixed a since the experiments show that $\mathrm{FAC}(n,a)$ is usually much smaller than our bounds in this paper. The other is to give a better theoretic estimate for $\min_a \mathrm{FAC}(n,a)$ when a runs over some specific set.

Acknowledgments We thank the anonymous referees for their many valuable suggestions on how to improve the presentation of this paper. The work of this paper was supported by the NNSF of China (Grants Nos. 11471314, 61572490), and the National Center for Mathematics and Interdisciplinary Sciences, CAS.

References

[1] M. Agrawal, N. Kayal and N. Saxena, *Primes is in P*, Ann. of Math. (2) **160** (2004), 781–793.

[2] P. J. Cameron, *Combinatorics: Topics, Techniques, Algorithms*, Cambridge University Press, Cambridge, 1994.

[3] H. Cohen, *A Course in Computational Algebraic Number Theory*, Graduate Texts in Mathematics, vol. **138**, Springer, Berlin, 1993.

[4] R. Crandall and C. Pomerance, *Prime Numbers, A Computational Perspective*, Second edition, Springer, New York, 2005.

[5] A. K. Lenstra, *Integer Factoring, Towards a Quarter-century of Public Key Cryptography*, Des. Codes Cryptogr. **19** (2000), 101–128.

[6] A. K. Lenstra and H. W. Lenstra, Jr.(Eds.), *The Development of the Number Field Sieve*, Lecture Notes in Mathematics, vol. **1554**, Springer, Berlin, 1993.

[7] R. L. Rivest, A. Shamir and L. Adleman, *A Method for Obtaining Digital Signatures and Public-key Cryptosystems*, Comm. ACM **21**(1978), 120–126.

[8] V. Shoup, *NTL: A Library for Doing Number Theory*, Available at http://www.shoup.net/ntl/

[9] Z.-H. Sun, *Combinatorial Sum* $\sum_{k=0,\ k\equiv r\pmod m}^{n}\binom{n}{k}$ *and its Applications in Number Theory (I)*, Nanjing Daxue Xuebao Shuxue Bannian Kan **9** (1992), 227–240.

[10] Z.-H. Sun, *Combinatorial Sum* $\sum_{k=0,k\equiv r\pmod m}^{n}\binom{n}{k}$ *and its Applications in Number Theory (II)*, Nanjing Daxue Xuebao Shuxue Bannian Kan **10**(1993), 105–118.

[11] Z.-H. Sun, *Combinatorial Sum* $\sum_{k=0,k\equiv r \pmod m}^{n} \binom{n}{k}$ *and its Applications in Number Theory (III)*, Nanjing Daxue Xuebao Shuxue Bannian Kan **12**(1995), 90–102.

[12] Z.-W. Sun, *On the Sum* $\sum_{k\equiv r \pmod m} \binom{n}{k}$ *and Related Congruences*, Israel J. Math. **128**(2002), 135–156.

[13] Z.-W. Sun, *Polynomial Extension of Fleck's Congruence*, Acta Arith. **122**(2006), 91–100.

[14] Z.-W. Sun, *On Sums of Binomial Coefficients and their Applications*, Discrete Math. **308**(2008), 4231–4245.

[15] Z.-W. Sun and D. Davis, *Combinatorial Congruences Modulo Prime Powers*, Trans. Amer. Math. Soc. **359**(2007), 5525–5553.

[16] Z.-W. Sun and R. Tauraso, *Congruences for Sums of Binomial Coefficients*, J. Number Theory **126**(2007), 287–296.

[17] J. von zur Gathen and J. Gerhard, *Modern Computer Algebra*, Cambridge University Press, Cambridge, 1999.

[18] C. S. Weisman, *On p-adic Differentiability*, J. Number Theory **9**(1977), 79–86.