Binomial collisions and near collisions

Aart Blokhuis, Andries Brouwer, Benne de Weger
Eindhoven University of Technology
a.blokhuis@tue.nl, aeb@cwi.nl, b.m.m.d.weger@tue.nl

July 21, 2017

Abstract

We describe efficient algorithms to search for cases in which binomial coefficients are equal or almost equal, give a conjecturally complete list of all cases where two binomial coefficients differ by 1, and give some identities for binomial coefficients that seem to be new.

1 Introduction

Let us call a quadruple (n, k, m, l) with $2 \le k \le \frac{1}{2}n$ and $2 \le l \le \frac{1}{2}m$ a (binomial) *collision* when $k < l$ and $\binom{n}{k} = \binom{m}{l}$, and a *near collision* when $\binom{m}{l} - \binom{n}{k} = d > 0$ with $\binom{m}{l} \ge d^3$. The exponent 3 is somewhat arbitrary. Maybe 5 is a more natural exponent, see the end of this paper.

Collisions have been studied by many authors. Some references will be given below. In this note we report on computer searches for collisions and near collisions, and give seven infinite families of near collisions.

2 Collisions

We list the known collisions. There are the double collision

$$\binom{78}{2} = \binom{15}{5} = \binom{14}{6} = 3003,$$

six further sporadic examples given in the table below:

n	k	m	l	$\binom{m}{l} = \binom{n}{k}$
16	2	10	3	120
21	2	10	4	210
56	2	22	3	1540
120	2	36	3	7140
153	2	19	5	11628
221	2	17	8	24310

and a miraculous infinite family given by

$$\binom{F_{2i+2}F_{2i+3}}{F_{2i}F_{2i+3}} = \binom{F_{2i+2}F_{2i+3} - 1}{F_{2i}F_{2i+3} + 1} \quad \text{for} \quad i = 1, 2, \ldots,$$

where F_i is the ith Fibonacci number (defined by $F_0 = 0$, $F_1 = 1$, $F_{i+1} = F_i + F_{i-1}$ for $i \geq 1$). The infinite family is due to Lind [9], and was rediscovered by several others such as Singmaster [15] and Tovey [18]. Examples are

$$\binom{15}{5} = \binom{14}{6}, \quad \binom{104}{39} = \binom{103}{40}, \quad \binom{714}{272} = \binom{713}{273}, \quad \binom{4895}{1869} = \binom{4894}{1870}.$$

Twenty years ago one of us conjectured

Conjecture 2.1 ([20]) *There are no other collisions than those given above.*

The current status is as follows.

Theorem 2.2 *There are no unknown collisions in the following cases:*

- $(k, l) = (2, 3), (2, 4), (2, 5), (2, 6), (2, 8), (3, 4), (3, 6), (4, 6), (4, 8)$,
- $(m, l) = (n - 1, k + 1), (n - 1, k + 2), (n - 2, k + 1)$,
- $n \leq 10^6$,
- $\binom{n}{k} \leq 10^{60}$.

Proof. The first two parts can be found in the literature.

The case $(k, l) = (2, 3)$ was settled in [2]. The case $(k, l) = (2, 4)$ was settled in [12], and also in [19]. The case $(k, l) = (2, 5)$ was settled in [5]. The cases $(k, l) = (2, 6), (2, 8), (3, 6), (4, 6), (4, 8)$ were settled in [16]. The case $(k, l) = (3, 4)$ was settled in Mordell [11] (actually, he solved an equivalent equation and seems not to have noted the relation to binomial coefficients).

The case $(m, l) = (n - 1, k + 1)$ was settled in [18] (and yields the infinite family). The cases $(m, l) = (n - 1, k + 2), (n - 2, k + 1)$ were settled in [17].

The last two parts are the results of computer searches we report on in this paper. Some details are given below. $\square$

Earlier computer searches handled $n \leq 10^3$ and $\binom{n}{k} \leq 10^{30}$ ([20]). In the literature one also finds finiteness results ([7], [4]), and results on the number of times an integer may occur as binomial coefficient ([14], [1], [8]).

2.1 Settling $n \leq 10^6$

In order to find all collisions with $n \leq N$ for some fixed N, generate a list of all values $\binom{n}{k}$ with $2 \leq k \leq n/2$ and $n \leq N$. Sort it, and compare successive elements to find duplicates.

Now the list has length about $\frac{1}{4}N^2$, and probably does not fit into memory. One approach is to split the list into parts, e.g. into the sublists consisting of all binomial coefficients between 10^{e-1} and 10^e, for all relevant e. We tried this in Mathematica and did $N = 34000$ in 23h30m on a 2.6 GHz Intel i7.

A different approach is to have a table and a priority queue, both of size N. Both contain the same elements. Initially both contain the numbers $\binom{n+2}{2}$ for $n < N$. The priority queue is kept sorted. At each step the top two elements are compared for equality. Afterwards the top element is discarded. When $\binom{n+k}{k}$ is discarded, the new value $\binom{n+k+1}{k+1}$ is added, unless $k \geq n$. The new value needed is computed from the old one via $\binom{n+k+1}{k+1} = \binom{n+k}{k} + \binom{n+k}{k+1}$. Note that the value $\binom{n+k}{k+1}$ is present in the table at index $n-1$ at the moment it is needed.

Computation time for the algorithms as described is cubic in N if the precise value of the binomial coefficients is computed, since not only the length of the list grows, but also the size of the numbers. Bounded precision suffices to ensure that (almost) collisions are unlikely, and reduces the time needed to $O(N^2 \log N)$. Almost collisions still occur (for example, $\binom{102091}{12877} = 1.256839391954534 \cdot 10^{16800}$, $\binom{200954}{9642} = 1.256839391954529 \cdot 10^{16800}$). We used interval arithmetic to distinguish almost equal numbers, and full exact multiple length arithmetic in the few cases where the interval arithmetic did not suffice. We tried this in C, with a custom data type (since the usual data types do not handle large exponents, or are too slow), and did $N = 10^6$ in 56h14m on an old 2 GHz PC.

2.2 Settling $\binom{n}{k} \le 10^{60}$

In order to find all collisions with $\binom{n}{k} \le M$ we handle each relevant pair (k, l) separately. Let $l_{\max}$ be the largest l such that $\binom{2l}{l} \le M$. As we saw, the pairs (k, l) with $k < l \le 4$ have been done already, so it suffices to handle $5 \le l \le l_{\max}$, and for each l the values of k with $2 \le k \le l-1$, with $k \ge 3$ if $l = 5$.

Given a pair (k, l), let $m_{\max}$ be the largest m with $\binom{m}{l} \le M$. Make a list of all m with $2l \le m \le m_{\max}$, and discard the m for which $\binom{m}{l}$ cannot be of the form $\binom{n}{k}$. What is left are possible collisions, and in practice only actual collisions are left.

The discarding is done via a sieving process. The function $f(n) = \binom{n}{k}$ is a polynomial of degree k in the variable n, with rational coefficients. The denominators of the coefficients have only prime factors $\le k$. For any prime $p > k$ the function f induces a polynomial map from $\mathbb{F}_p$ to itself. Let $A(k, p)$ be the size of the image. Experience shows that $A(k, p) \approx (1-e^{-1})p$ when k is odd, and $A(k, p) \approx (1 - e^{-1/2})p$ when k is even. See below for more remarks on this function $A(k, p)$. Since $1 - e^{-1} = 0.63...$ and $1 - e^{-1/2} = 0.39...$, a significant fraction of all residues mod p cannot be of the form $\binom{n}{k}$. Now pick $p > l > k$, and look at $\binom{a}{l}$ for $0 \le a \le p-1$. Whenever $\binom{a}{l}$ is not in the mod p image of f, discard all $\binom{m}{l}$ with $m \equiv a \pmod{p}$ from the list.

Repeating this sieve action for all primes less than 500 (stopping earlier when the list has become empty) we found all collisions up to $M = 10^{60}$. The largest prime needed was $p = 401$. This took about 375 CPU hours total on a few old 2 GHz machines. For large l the upper bound $m_{\max}$ is small, and sieving is very quick. (In fact for $l \ge 10$ we sieved up to 10^{100}.) The main part of the work are the pairs $(k, l) = (3, 5), (4, 5)$, where the list has length roughly $M^{1/5}$.

On $A(k, p)$

There is a lot of literature on the size of the image of a polynomial on $\mathbb{F}_p$. For $k = 3$ and $k = 4$ the value of $A(k, p)$ was found by Daublebsky v. Sterneck [6]. One has $A(3, p) = (2p \pm 1)/3$ when $p \equiv \pm 1 \pmod 6$, and $A(4, p) = (3p + 4 + \chi(-1) + 2\chi(5) - 2\chi(10))/8$ for $p > 5$, where χ is the quadratic character. Birch & Swinnerton-Dyer [3] showed that 'general' polynomials of degree k on $\mathbb{F}_p$ have an image of size $a_k p + O(\sqrt{p})$ where $a_k = \sum_{i=1}^{k} (-1)^{i-1} \frac{1}{i!}$. We conjecture in our situation that the value $A_k = \lim_{p \to \infty} \frac{A(k,p)}{p}$ exists, and equals $A_k = a_k$ for odd

k, and $A_k = \sum_{i=1}^{k/2}(-1)^{i-1}\frac{1}{2^i i!}$ for even k. This is true for $k \leq 5$. Note that for even k there is the symmetry $f(x) = f(k+1-x)$ explaining the smaller image size.

3 Near collisions

3.1 Difference 1

We know about the following examples with $d = 1$:

n	k	m	l	$\binom{m}{l} = \binom{n}{k} + 1$
6	3	7	2	21
7	3	9	2	36
11	2	8	3	56
10	5	23	2	253
12	4	32	2	496
16	3	34	2	561
60	2	23	3	1771
27	3	77	2	2926
29	3	86	2	3655
34	3	21	4	5985
22	5	230	2	26335
260	3	2407	2	2895621
93	4	2417	2	2919736
62	5	3598	2	6471003
28	11	6554	2	21474181
665	3	9879	2	48792381
135	5	26333	2	346700278
139	5	28358	2	402073903
19630	3	1587767	2	1260501229261
160403633	2	425779	3	12864662659597529

The above table is complete for the cases (k, l), $(l.k) = (2, 3)$, $(2, 4)$, $(2, 6)$, $(3, 4)$, $(4, 6)$ (as one sees by finding all integral points on the corresponding elliptic curves), and for $\binom{n}{k} \leq 10^{30}$. We conjecture the following

Conjecture 3.1 *There are no other near collisions with difference* 1 *than those given above.*

and, more generally,

Conjecture 3.2 *Given a fixed difference d, the number of near collisions with difference d is finite.*

The latter conjecture can be backed by standard heuristic arguments. The infinite family of collisions seems like a miracle.

The cases mentioned above correspond to (double covers of) elliptic curves in cubic Weierstrass form or in quartic form, and can be solved completely using e.g. Sage [13] or Magma [10]. See [16], Table 1, for the transformations from the binomial equations to the elliptic equations.

3.2 Infinite families

When d is not fixed, there are a few infinite families.

$$\binom{12x^2 - 12x + 3}{3} + \binom{x}{2} = \binom{24x^3 - 36x^2 + 15x - 1}{2} \tag{1}$$

$$\binom{12x^2 - 12x + 5}{3} + \binom{x}{2} = \binom{24x^3 - 36x^2 + 21x - 4}{2} \tag{2}$$

$$\binom{60x^2 - 60x + 15}{5} + \binom{x}{2} = \binom{a}{2} \tag{3}$$

where $a = 3600x^5 - 9000x^4 + 8700x^3 - 4050x^2 + 905x - 77$,

$$\binom{60x^2 - 60x + 19}{5} + \binom{x}{2} = \binom{a}{2} \tag{4}$$

where $a = 3600x^5 - 9000x^4 + 9300x^3 - 4950x^2 + 1355x - 152$,

$$\binom{240x^2 - 240x + 62}{5} + \binom{3x - 1}{2} = \binom{a}{2} \tag{5}$$

where $a = 115200x^5 - 288000x^4 + 288000x^3 - 144000x^2 + 35995x - 3597$,

$$\binom{11340x^2 + 11340x + 2835}{9} + \binom{y}{2} = \binom{a}{2} \tag{6}$$

where $y = 22680x^3 + 34020x^2 + 17001x + 2831$ and $a = 41342070848400000x^9 + 186039318817800000x^8 + 372013015300920000x^7 + 433862065736820000x^6 + 325224326359359000x^5 + 162497395464547500x^4 + 54118008336955500x^3 + 11584437364095750x^2 + 1446265881317760x + 80234671844510$,

$$\binom{11340x^2 + 11340x + 2843}{9} + \binom{y}{2} = \binom{a}{2} \tag{7}$$

where $y = 22680x^3 + 34020x^2 + 17019x + 2840$ and $a = 41342070848400000x^9 + 186039318817800000x^8 + 372144259970280000x^7 + 434321422079580000x^6 + 325913360873499000x^5 + 163071590892997500x^4 + 54405106066489500x^3 + 11670566701326750x^2 + 1460620778510760x + 81260022737510$.

How does one find such identities? In order to get $\binom{b}{3} + \binom{x}{2} = \binom{a}{2}$, where x is small, one needs $\frac{1}{3}b(b-1)(b-2) = (a-x)(a+x-1)$, a product of two nearly equal numbers. If $b = 3e^2$, then $\frac{1}{3}b(b-1)(b-2) = (e(b-2))(e(b-1))$ and we can take $a - x = e(b-2)$, $a + x - 1 = e(b-1)$ and find $e = 2x - 1$, $b = 3(2x-1)^2$, the first family. The other families arise in a similar way.

Are there many such identities? Let us say that the quality of an identity $\binom{n(x)}{k} + d(x) = \binom{m(x)}{l}$ is the degree of x in $\binom{n(x)}{k}$ and $\binom{m(x)}{l}$ divided by that in $d(x)$. Then our identities (1)-(7) have qualities 3, 3, 5, 5, 5, 3, 3. These are the only identities of quality at least 3 that we know of. Maybe there are no others. Maybe there is a number α, supposedly ≤ 3, such that there are only finitely many identities of quality at least α.

It follows from the existence of the identities (1)-(7) that there are infinitely many near collisions, even with $\binom{m}{l} \geq d^5$. Maybe there is a number β, certainly $\beta > 5$, such that there are only finitely many near collisions with $\binom{m}{l} \geq d^\beta$.

References

[1] H.L. Abbott, P. Erdős & D. Hanson, *On the number of times an integer occurs as a binomial coefficient*, Am. Math. Monthly **81** (1974) 256–261.

[2] È.T. Avanesov, *Solution of a problem on figurate numbers* (Russian), Acta Arith. **12** (1966) 409–420.

[3] B.J. Birch & H.P.F. Swinnerton-Dyer, *Note on a problem of Chowla*, Acta Arith. **5** (1959) 417–423.

[4] B. Brindza, *On a special superelliptic equation*, Publ. Math. Debrecen **39** (1991) 159–162.

[5] Y. Bugeaud, M. Mignotte, S. Siksek, M. Stoll & S. Tengely, *Integral points on hyperelliptic curves*, Algebra and Number Theory **2** (2008) 859–885.

[6] R. Daublebsky v. Sterneck, *Über die Anzahl inkongruenter Werte, die eine ganze Funktion dritten Grades annimmt*, Sitzungsberichte der kaiserlichen Akademie der Wissenschaften, Wien, Bd. CXVI, Heft V, Abteilung IIa (1907) 895–904.

[7] G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983) 349–366.

[8] D. Kane, *New bounds on the number of representations of T as a binomial coefficient*, Integers **4** (2004) #A07.

[9] D.A. Lind, *The quadratic field $\mathbb{Q}\left(\sqrt{5}\right)$ and a certain diophantine equation*, Fibonacci Quart. **6** (1968) 86–93.

[10] Wieb Bosma, John Cannon & Catherine Playoust, The Magma algebra system. I. The user language, J. Symbolic Comput. **24** (1997) 235–265.

[11] L.J. Mordell, *On the integer solutions of $y(y+1) = x(x+1)(x+2)$*, Pacific J. Math. **13** (1963) 1347–1351.

[12] Á. Pintér, *A note on the diophantine equation $\binom{x}{4} = \binom{y}{2}$*, Publ. Math. Debrecen **47** (1995) 411–415.

[13] SageMath, the Sage Mathematics Software System (Version 7.5.1), The Sage Developers, 2017, `http://www.sagemath.org`.

[14] D. Singmaster, *How often does an integer occur as a binomial coefficient*, Am. Math. Monthly **78** (1971) 385–386.

[15] D. Singmaster, *Repeated binomial coefficients and Fibonacci numbers*, Fibonacci Quart. **13** (1975) 295–298.

[16] R.J. Stroeker & B.M.M. de Weger, *Elliptic binomial Diophantine equations*, Math. Comp. **68** No. 227 (1999) 1257–1281.

[17] R.J. Stroeker & B.M.M. de Weger, *Solving elliptic Diophantine equations: the general cubic case*, Acta Arith. **87** (1999) 339–365.

[18] C.A. Tovey, *Multiple occurrences of binomial coefficients*, Fibonacci Quart. **23** (1985) 356–358.

[19] B.M.M. de Weger, *A binomial diophantine equation*, Quart. J. Math. Oxford Ser. 2 **47** (1996]), 221–231.

[20] B.M.M. de Weger, *Equal Binomial Coefficients: Some Elementary Considerations*, J. Number Th. **63** (1997) 373–386.