INTEGERS: ELECTRONIC JOURNAL OF COMBINATORIAL NUMBER THEORY 7 (2007), #A53

IMPROVED BOUNDS ON THE NUMBER OF WAYS OF EXPRESSING t AS A BINOMIAL COEFFICIENT

Daniel M. Kane
Department of Mathematics, Harvard University, Cambridge, MA 02139, USA
dankane@math.harvard.edu

Received: 7/5/07, Revised: 9/5/07, Accepted: 10/30/07, Published: 12/3/07

Abstract

Let $N(t)$ denote the number of ways of writing t as a binomial coefficient. We show that $N(t) = O\left(\frac{(\log t)(\log\log\log t)}{(\log\log t)^3}\right)$.

1. Introduction

As in [2], we define $N(t) = \left|\left\{(n,m) \in \mathbb{Z}^2 : \binom{n}{m} = t\right\}\right|$ to be the number of ways of writing an integer $t > 1$ as a binomial coefficient. $N(3003) = 8$, and $N(t) \geq 6$ for infinitely many t, but essentially no other lower bounds on $N(t)$ are known. Singmaster conjectured in [2] that $N(t) = O(1)$. Although no one has yet managed to achieve this bound (or even gotten particularly close), there has been some work on bounding the size of $N(t)$ (see [1, 2, 3]). The record was that $N(t) = O\left(\frac{(\log t)(\log\log\log t)}{(\log\log t)^2}\right)$ proved by the author in [1]. Using a refinement of this argument we improve this bound by a factor of $\log\log t$.

2. Overview of Our Technique

We recall the basics of the argument from [1]. First we note that it suffices to consider only solutions of the form $t = \binom{n}{m}$ where $n > 2m$, since for any other solution (n,m) with $n < 2m$, we have the solution $(n, n-m)$ with $n > 2m$ (there is at most one solution with $n = 2m$).

Next we consider the implicitly defined function $f(x)$ given by $\binom{f(x)}{x} = t$. By interpolating the binomial coefficient using the Γ-function, we make $f(x)$ smooth. We now are trying to bound that number of solutions to $f(m) = n$, or in other words the number of lattice points on the graph of the smooth function f.

We will use the fact that f has derivatives (of appropriate order) that are small but non-zero to bound the number of integral points on its graph.

3. Review of Previous Results

With $f(x)$ defined as above, we have from [1] that f can be extended to a complex analytic function, so that

$$f(z) = \exp\left(\frac{\log t + \Gamma(z+1)}{z}\right) + \frac{z-1}{2} + O\left(\frac{z^2}{f(z)}\right) \tag{1}$$

uniformly where $|f(z)| > |2z|$, which holds as long as

$$\left|\exp\left(\frac{\log t + \Gamma(z+1)}{z}\right)\right| > |6z|.$$

We define the function $\alpha(x) = \frac{\log f(x)}{\log x}$, so $f(x) = x^{\alpha(x)}$. Using Equation 1 and Sterling's formula, we obtain that as long as $\alpha > 1 + \epsilon$ (for some constant $\epsilon > 0$) that

$$\alpha(x) \sim \frac{\log t}{x \log x} + 1 \tag{2}$$

uniformly as $t \to \infty$.

Also in [1] it is shown that for t sufficiently large, and k and integer more than 1, if

$$x^{7\alpha/4-2} > 3^{k+1}k! \tag{3}$$

then

$$0 < \left|\frac{1}{k!}\frac{\partial^k}{\partial x^k} f(x)\right| < 2x^{\alpha-k}e^{2\alpha}(\log x)^k. \tag{4}$$

Note that for k large, this will imply that the k^{th} derivative of f is small but non-zero.

In order to relate derivatives of f to integer points on its graph, we use the following lemma from [1]:

Lemma 1. *If $F(x) : \mathbb{R} \to \mathbb{R}$ is an infinitely differentiable function and if $F(x) = 0$ for $x = x_1, x_2, ..., x_{n+1}$ (where $x_1 < x_2 < ... < x_{n+1}$), then $F^{(n)}(y) = 0$ for some $y \in (x_1, x_{n+1})$.*

Proof. We proceed by induction on n. The case of $n = 1$ is Rolle's Theorem. Given the statement of Lemma 2.1 for $n-1$, if there exists such an F with $n+1$ zeroes, $x_1 < x_2 < ... < x_{n+1}$, then by Rolle's theorem, there exist points $y_i \in (x_i, x_{i+1})$ $(1 \le i \le n)$ so that $F'(y_i) = 0$. Then since F' has at least n roots, by the induction hypothesis there exists a y with $x_1 < y_1 < y < y_n < x_{n+1}$, and $F^{(n)}(y) = (F')^{(n-1)}(y) = 0$. $\square$

Suppose now that f has $k+1$ integer points on its graph, $f(m_i) = n_i$, for $1 \le i \le k+1$. We let

$$g(x) = \sum_{i=1}^{k+1} n_i \prod_{j \neq i} \frac{(x - m_j)}{(m_i - m_j)}$$

be the polynomial of degree k that interpolates f at these points. By letting $F(x) = f(x) - g(x)$ and applying Lemma 1 we get that for some y between the largest and smallest of the m_i, that

$$\frac{1}{k!}\frac{\partial^k}{\partial x^k} f(y) = \frac{1}{k!}\frac{\partial^k}{\partial x^k} g(y) = \sum_{i=1}^{k+1} \frac{n_i}{\prod_{j \neq i}(m_i - m_j)} = \frac{A}{B(m_1, \ldots, m_{k+1})} \tag{5}$$

for some integer A and $B(m_1, \ldots, m_{k+1}) = \mathrm{LCM}_i \left(\prod_{j \neq i}(m_i - m_j)\right)$. Our strategy will be to show that B is small and thus that the k^{th} derivative of f is either 0 or a multiple of B (which is large), leading to a contradiction.

4. The New Bound

Here we prove the new result that will give us the improvement over [1].

Proposition 2. *If m_i are integers where the largest and smallest differ by at most S,*

$$\log(B(m_1, \ldots, m_k)) = O\left(S \max(1, \log\left(\frac{k^2 \log S}{S}\right))\right).$$

Proof. We first show that $\log(B(m_1, \ldots, m_k)) = O(S \log(k))$, thus proving our bound for $k > S^{2/3}$. We note that B is at most

$$\mathrm{LCM} \prod_{i=1}^{k-1} r_i$$

where the LCM is over all sequences of $k-1$ distinct non-zero numbers of absolute value at most S. We compute this by counting the number of multiples of each prime p. Each power of a prime, p^n, can divide at most $\max(k-1, 2\left\lfloor \frac{S}{p^n} \right\rfloor)$ of the r_i ($k-1$ being the number of r_i and $2\left\lfloor \frac{S}{p^n} \right\rfloor$ the number of non-zero terms of absolute value at most S divisible by p^n). Therefore we have that

$$\log(B) \le \sum_{p^n} \max(k-1, 2\left\lfloor \frac{S}{p^n} \right\rfloor) \log p \le \sum_{p^n < S/k} (k-1)\log p + 2 \sum_{S \ge p^n \ge S/k} S \frac{\log p}{p^n}.$$

Using integration by parts we find that this is at most

$$(k-1)\psi\left(\frac{S}{k}\right) + 2S\left(\int_{S/k}^{S} \frac{\psi(x)dx}{x^2} + \frac{\psi(S)}{S} - \frac{\psi(S/k)}{S/k}\right)$$

where $\psi(x)$ is Chebyshev's function $\sum_{p^n<x} \log p$, the sum being over powers of primes, p^n that are less than x. Using the prime number theorem, this is at most

$$O\left((k-1)\frac{S}{k} + 2S\left(\int_{S/k}^{S} \frac{dx}{x} + \frac{S}{S}\right)\right) = O(S + 1 + S\log k) = O(S\log k).$$

We now assume that $k < S^{2/3}$. We note that since B does not decrease when we add more m_i's, that it suffices to show that

$$\log(B(m_1, \ldots, m_k)) = O(S(1 + \log(k^2 \log S/S)))$$

when $k > 2\sqrt{\frac{S}{\log S}}$.

Consider first the contribution to $\log(B)$ from powers of primes less than $\frac{S}{k}$. There are $\pi\left(\frac{S}{k}\right)$ such primes. The power of such a prime dividing any $(m_i - m_j)$ is at most S. Therefore, the power of such a prime dividing B is at most S^k. Hence the contribution to $\log(B)$ from such primes is at most

$$\pi\left(\frac{S}{k}\right) k\log(S) = O\left(S\frac{\log S}{\log\left(\frac{S}{k}\right)}\right)$$

by the prime number theorem. This in turn is $O(S)$ if $k < S^{2/3}$ and hence is $O(S(1 + \log(k^2 \log S/S)))$.

Next consider the contribution from primes larger than $\frac{S^2}{k^2 \log S}$. For each such prime, p, we note that in any term, $\prod_{j\neq i}(m_i - m_j)$, since the $(m_i - m_j)$ are distinct, non-zero integers of absolute value at most S, p divides at most $O(S/p)$ of them. Furthermore, since $k < S^{2/3}$, none are divisible by p^3. Therefore, B is divisible by $O(S/p)$ powers of p. Hence the contribution to $\log(B)$ of these primes is (using integration by parts)

$$\begin{aligned} O\left(\sum_p \frac{S}{p}\log p\right) =& O\left(\sum_{S^2/(k^2\log S)<p^n<S} \frac{S}{p^n}\log p\right) \\ =& O\left(S\left(\int_{S^2/(k^2\log S)}^{S} \frac{\psi(x)}{x^2}dx + \frac{\psi(S)}{S}\right)\right), \end{aligned}$$

Using the prime number theorem, this is

$$O\left(S\left(\int_{S^2/(k^2\log S)}^{S} \frac{dx}{x} + \frac{S}{S}\right)\right) = O\left(S(1 + \log(k^2 \log S/S))\right).$$

Lastly, we consider the contribution to B from primes between S/k and $\frac{S^2}{k^2 \log S}$. The contribution to $\log B$ from each such prime, p is at most $\log S$ times the maximum (over i) of the number of terms $m_i - m_j$ divisible by p. Note that since each such p is bigger than $S^{1/3}$, no $m_i - m_j$ is divisible by more than 2 of them. Let l be the number of such primes.

Let $d_1, d_2, \ldots, d_l$ be defined by letting d_a be the maximum number of $m_i - m_j$ (for some i fixed) divisible by the a^{th} of these primes. Therefore the contribution to $\log B$ by these primes is $O(\log S \sum d_i)$. Next note that there are $d_a + 1$ $m's$ congruent modulo the a^{th} of these primes. Hence $d_a(d_a+1)/2 > d_a^2/2$ of the $m_i - m_j$ are divisible by this prime. Hence since there are at most $k^2/2$ pairs, each divisible by at most two primes, $\sum d_a^2 < 2k^2$. Hence by Cauchy-Schwartz,

$$\sum_a d_a \leq \sqrt{\left(\sum_a d_a^2\right)\left(\sum_a 1\right)} \leq \sqrt{2k^2 l} = O(k\sqrt{l})$$

Now since l is clearly at most $\pi\left(\frac{S^2}{k^2 \log S}\right)$ and since $\frac{S^2}{k^2 \log S} > S^{1/3}$, the prime number theorem implies that $l = O\left(\frac{S^2}{k^2 \log^2 S}\right)$. Therefore, the contribution to $\log B$ from these primes is

$$O(\log S k\sqrt{l}) = O(S).$$

This completes the proof. □

5. Cases

Let

$$D(t) = \left|\left\{(n,m) \in \mathbb{Z}^2 : \binom{n}{m} = t, n > 2m, n < m^{\frac{\log\log t}{24 \log\log\log t}}\right\}\right|,$$

$$E(t) = \left|\left\{(n,m) \in \mathbb{Z}^2 : \binom{n}{m} = t, n > m^{\frac{\log\log t}{24 \log\log\log t}}, n < m^{(\log\log t)^3}\right\}\right|,$$

$$F(t) = \left|\left\{(n,m) \in \mathbb{Z}^2 : \binom{n}{m} = t, n > m^{(\log\log t)^3}\right\}\right|.$$

Recalling that we can restrict our attention to solutions where $n > 2m$, we find that

$$N(t) = O(D(t) + E(t) + F(t)). \tag{6}$$

6. The Easy Cases

From [1] we know that

$$D(t) = O\left(\frac{\log t}{(\log\log t)^3}\right). \tag{7}$$

Furthermore, if $\alpha > (\log\log t)^3$, then by Equation 2, we have that $m = O\left(\frac{\log t}{\alpha}\right) = O\left(\frac{\log t}{(\log\log t)^3}\right)$. Since each solution has a distinct value of m, this implies that

$$F(t) = O\left(\frac{\log t}{(\log\log t)^3}\right). \tag{8}$$

7. The Bound on $E(t)$

Let $\alpha_0 = \frac{\log\log t}{24\log\log\log t}$. Let $E_i(t)$ be the number of solutions with $2^i\alpha_0 \leq \alpha \leq 2^{i+1}\alpha_0$. Let $k_i = 2^{i+2}\alpha_0$. Suppose that we have $k_i + 1$ integer points on the graph of f, in the range where $2^i\alpha_0 \leq \alpha \leq 2^{i+1}\alpha_0$ ($\alpha < (\log\log t)^3$). Suppose that these points are separated by a total distance of S. Notice that by Equation 2 that in this range, $\log x = \Theta(\log\log t)$. In this range, Equation 3 holds since

$$\log\left(x^{7/4\alpha-2}\right) = \log x((7/4)\alpha - 2) \gg k_i(\log\log t) > k_i \log k_i \gg \log(3^{k_i+1}k_i!).$$

Therefore, Equation 4 holds and

$$0 < \left|\frac{1}{k_i!}\frac{\partial^{k_i}}{\partial x^{k_i}}f(x)\right| < 2e^{2\alpha}x^{\alpha-k_i}(\log x)^{k_i} = \exp\left(-\Omega\left(k_i(\log\log t)\right)\right).$$

On the other hand, if we have solutions with integer points (n_i, m_i) for $1 \leq i \leq k_i + 1$ in this range, where the m_i have maximum separation S, then this derivative is at least

$$\frac{1}{B(m_1, \ldots, m_{k_i+1})} = \exp(-O(S\max(1, \log(k_i^2(\log S)/S))))$$

by Proposition 2. Let $D = \frac{S}{k_i}$. Comparing these two bounds on the size of the k^{th} derivative of f, we have that

$$D\max\left(1, \log\left(\frac{k_i\log k_i}{D}\right)\right) > C\log\log t$$

where C is some positive constant. So either $D > C\log\log t$, or (substituting the value of k_i),

$$D\log\left(\left(\frac{\log\log t}{D}\right)2^{i+2}\right) > C\log\log t.$$

The latter implies that $D = \Omega((\log\log t)/(i+1))$. Hence

$$D = \Omega\left(\frac{\log\log t}{(i+1)}\right).$$

Note that by Equation 2 that for $2^i\alpha_0 \leq \alpha$ (assuming that $i = O(\log\log\log t)$) that

$$x = O\left(\frac{\log t}{2^i\alpha_0\log\log t}\right) = O\left(\frac{(\log t)(\log\log\log t)}{2^i(\log\log t)^2}\right).$$

By the above, any $k_i + 1$ solutions must be separated by a total distance of at least Dk_i. Therefore, since the total range of all solutions is $O\left(\frac{(\log t)(\log\log\log t)}{2^i(\log\log t)^2}\right)$, we have that

$$Dk_i\left\lfloor\frac{E_i(t)}{k_i}\right\rfloor = O\left(\frac{(\log t)(\log\log\log t)}{2^i(\log\log t)^2}\right).$$

Therefore,

$$E_i(t) = O\left(\frac{1}{D}\frac{(\log t)(\log\log\log t)}{2^i(\log\log t)^2} + k_i\right) = O\left(\left(\frac{i+1}{2^i}\right)\left(\frac{(\log t)(\log\log\log t)}{(\log\log t)^3}\right)\right).$$

Summing the above over all i from 0 to $\log\log\log t$ yields that

$$E(t) = O\left(\frac{(\log t)(\log\log\log t)}{(\log\log t)^3}\right). \tag{9}$$

Now by combining Equations 6,7,8,9 we get our result that

$$N(t) = O\left(\frac{(\log t)(\log\log\log t)}{(\log\log t)^3}\right).$$

8. Further Work

It should be noted that the bound we obtained can not be improved by much more using this technique. This is because if we have $k^2 = \Omega(S)$, then B can be as large as $\exp(\Omega(S))$. This comes from the fact that if we pick k elements of $\{1, 2, \ldots, S\}$ randomly and independently, there is a constant probability that any prime $p < S/2$ will divide a difference of some two elements. Since the product of these primes is $\exp(\Omega(S))$ by the prime number theorem, the expected size of $\log B$ is $\Omega(S)$.

Consider the region where $\alpha > \log\log t$. The k^{th} derivative of f over $k!$ has log of size about $(\log x)(\alpha - k)$. Therefore, to get any useful information we need to set $k > \alpha$. We then obtain a bound looking something like $\log(B) > k(\log\log t)$. By the above, this can be satisfied with S as small as $O(k(\log\log t))$. Therefore, we can only prove that the inverse density of solutions is $O(\log\log t)$ (but no better). Therefore, since there are $\Theta\left(\frac{\log t}{(\log\log t)^2}\right)$ values of m in this range, we cannot by this technique alone exclude the possibility of as many as $O\left(\frac{(\log t)}{(\log\log t)^3}\right)$ solutions.

It would be interesting to improve this gap some. This leads to the problem of finding the correct bounds on $\log(B)$ for given values of k and S. The known upper bounds are $O\left(S\max(1, \log\left(\frac{k^2\log S}{S}\right))\right)$ (Prop 2) and $O(k^2\log(S))$ (by $B < \prod_{i,j,i\neq j}(m_i - m_j)$). The randomized construction gives the lower bound of $\Omega(k^2(1+\log(S/k^2)))$ if $S > k^2$ and $\Omega(S)$ otherwise. It should be noted that the upper and lower bounds agree if $k < S^{1/2-\epsilon}$.

References

[1] D. Kane, *On the Number of Representations of t as a Binomial Coefficient*, Integers: Electronic Journal of Combinatorial Number Theory, **4**, (2004), #A07, pp. 1-10.

[2] D. Singmaster, *How often Does an Integer Occur as a Binomial Coefficient?*, American Mathematical Monthly, **78**, (1971) 385-386

[3] H. L. Abbott, P. Erdös and D. Hanson, *On the number of times an integer occurs as a binomial coefficient*, American Mathematical Monthly, **81**, (1974) 256-261.