ON A NONINTEGRALITY CONJECTURE

FLORIAN LUCA AND CARL POMERANCE

Abstract. It is conjectured that the sum

$$S_r(n) = \sum_{k=1}^{n} \frac{k}{k+r}\binom{n}{k}$$

for positive integers r, n is never integral. This has been shown for $r \le 22$. In this note we study the problem in the "n aspect" showing that the set of n such that $S_r(n) \in \mathbb{Z}$ for some $r \ge 1$ has asymptotic density 0. Our principal tools are some deep results on the distribution of primes in short intervals.

1. Introduction

For positive integers r, n let

$$S_r(n) = \sum_{k=1}^{n} \frac{k}{k+r}\binom{n}{k}.$$

Motivated by some cases with small r, López-Aguayo [4] asked if $S_r(n)$ is ever an integer, showing for $r \in \{1,2,3,4\}$ that $S_r(n)$ is not integral for all n. In [5] it was conjectured that $S_r(n)$ is never integral, and they proved the conjecture for $r \le 6$. In [3] it was proved for $r \le 22$. Also in [3], using a deep theorem of Montgomery and Vaughan [6], it was shown for a fixed r that the set of n such that $S_r(n) \in \mathbb{Z}$ has upper density bounded by $O_k(1/r^k)$ for any $k \ge 1$. In fact, this density is 0, as we shall show. Actually we prove a stronger result. Let

$$\mathcal{S} := \{n : S_r(n) \in \mathbb{Z} \text{ for some } r \ge 1\}.$$

Theorem 1. *The set $\mathcal{S}$ has zero density as a subset of the integers. Further, every member of $\mathcal{S}$ is greater than 10^6.*

It follows from our argument that if we put $\mathcal{S}(x) = \mathcal{S} \cap [1, x]$ then $\#\mathcal{S}(x) = O_A(x/(\log x)^A)$ for every fixed A. In particular, taking $A = 2$, we see that the reciprocal sum of $\mathcal{S}$ is finite.

2. The proof

We let x be large and assume that $x/2 < n \le x$. Let

$$S(r, n) := \sum_{k=0}^{n} \frac{r}{k+r}\binom{n}{k},$$

Date: August 25, 2021.
2010 *Mathematics Subject Classification.* 11N37.
Key words and phrases. Primes in short intervals.

so that $S(r,n) + S_r(n) = \sum_{k=0}^{n} \binom{n}{k} = 2^n \in \mathbb{Z}$. We conclude that $S_r(n)$ is integral if and only if $S(r,n)$ is integral. It is shown in [5] that

$$S(r,n) = \sum_{j=1}^{r} (-1)^{r-j} r \binom{r-1}{j-1} \frac{2^{n+j}-1}{n+j}. \tag{1}$$

For an odd prime q we write $\ell_2(q)$ for the multiplicative order of 2 modulo q.

Lemma 1. *If there is a prime $p > n$ that divides some $k+r$ with $1 \le k \le n$, then $S_r(n)$ is not integral. Also, if there is a prime $q > r$ that divides some $n+j$ with $1 \le j \le r$ and $S_r(n)$ is integral, then $q\ell_2(q) \mid n+j$.*

Proof. For the first assertion, say $p \mid k_0 + r$, where $1 \le k_0 \le n$. Since $p > n$, we have that p does not divide any other $k+r$ for $1 \le k \le n$. So the term $(k_0/(k_0+r))\binom{n}{k_0}$ in the definition of $S_r(n)$, in reduced form, has a factor p in the denominator, and no other terms $(k/(k+r))\binom{n}{k}$ have this property. We deduce that $S_r(n)$ is nonintegral. For the second assertion, say $q = n + j_0$, $1 \le j_0 \le r$. Since $q > r$, it does not divide any other $n+j$, $1 \le j \le r$. For $S_r(n)$ to be integral (and so for $S(r,n)$ to be integral) it follows then from (1) that $q \mid 2^{n+j} - 1$, so $\ell_2(q) \mid n+j$. But also $\ell_2(q) \mid q-1$, so $\gcd(q, \ell_2(q)) = 1$. This implies that $q\ell_2(q) \mid n+j$, which proves our second assertion. □

We now show that $\mathcal{S}$ has asymptotic density 0. Suppose that $S_r(n)$ is integral and $x/2 < n \le x$. We distinguish various cases.

Case 1. $r \ge n$.

By Sylvester's theorem, one of the integers $k+r$ with $1 \le k \le n$ is divisible by a prime $p > n$. It follows from the first part of Lemma 1 that $S_r(n)$ is nonintegral.

Case 2. $n > r > (x/2)^{1/10}$.

By a result of Jia (see [2]) for every fixed $\varepsilon > 0$, the interval $[n+1, n+n^{1/20+\varepsilon}]$ contains a prime number p for almost all n, with the number of exceptional values of $n \le x$ being $\ll_{\epsilon,A} x/(\log x)^A$ for every fixed $A > 0$. If $r > (x/2)^{1/10} \ge (n/2)^{1/10}$, then $r > n^{1/11}$ holds for all $x > x_0$. If n is not exceptional in the sense of Jia's theorem, then the interval $[n+1, n+r]$ contains the interval $[n+1, n+n^{1/11}]$ and hence a prime $p > n$, so $S_r(n)$ cannot be an integer by the first part of Lemma 1. Hence, n must be exceptional in the sense of Jia's theorem and the set of such n has counting function $O_A(x/(\log x)^A)$ for any fixed $A > 0$.

Case 3. $y \le r \le (x/2)^{1/10}$, where $y := x^{1/\log\log x}$.

This is the most interesting part. We prove the following lemma.

Lemma 2. *There exists r_0 such that if $r > r_0$, then the interval $I = [r, r + r^{0.61}]$ contains 6 primes $p_1, \dots, p_6$ such that each $\ell_2(p_i) > r^{0.3}$ for $1 \le i \le 6$ and each $\gcd(p_i - 1, p_j - 1) < r^{0.001}$ for $1 \le i < j \le 6$.*

Proof. Let $\pi(I)$ be the number of primes in I. From Baker, Harman, and Pintz [1] we have for large r that

$$\pi(I) \gg r^{0.61}/\log r.$$

(Actually, this follows from earlier results, but [1] holds the record currently for primes in short intervals.) Let $\mathcal{Q}$ be the subset of primes $p \in I$ such that $\ell_2(p) \le r^{0.3}$. By a classical argument, $\#\mathcal{Q} \ll r^{0.6}/\log r$. Indeed,

$$r^{\#\mathcal{Q}} \le \prod_{p\in\mathcal{Q}} p \le \prod_{t\le r^{0.3}} (2^t-1) < 2^{\sum_{t\le r^{0.3}} t} < 2^{r^{0.6}},$$

from which we deduce the desired upper bound on $\#\mathcal{Q}$. Since

$$r^{0.6}/\log r = o(r^{0.61}/\log r) = o(\pi(I)), \qquad \text{as} \qquad r\to\infty,$$

we deduce that most primes p in I have $\ell_2(p) \ge r^{0.3}$. Let $\mathcal{P}$ denote this set of primes in I, so that $\#\mathcal{P} \gg r^{0.61}/\log r$. For any positive integer d the number of pairs of distinct primes p, q in $\mathcal{P}$ with $d \mid p-1$ and $d \mid q-1$ is $\ll r^{2\times 0.61}/d^2$ even ignoring the primality condition. Summing over $d \ge r^{0.001}$ we see that the number of pairs $p, q \in \mathcal{P}$ with $\gcd(p-1, q-1) \ge r^{0.001}$ is $\ll r^{2\times 0.61-0.001}$, so that most pairs of primes $p, q \in \mathcal{P}$ have $\gcd(p-1, q-1) < r^{0.001}$. In fact, the number of 6-tuples of primes $p_1, \dots, p_6 \in \mathcal{P}$ with some $\gcd(p_i-1, p_j-1) \ge r^{0.001}$ is $\ll r^{6\times 0.61-0.001}$, so we may deduce that most 6-tuples of primes in $\mathcal{P}$ satisfy the gcd condition of the lemma. Of course "6" may be replaced with any fixed positive integer, only affecting the choice of r_0. $\square$

Let $\{p_1, \dots, p_6\}$ be the 6 primes in I which exist for $x > x_0$ (such that $y > r_0$). Either there are 4 of these primes such that the interval $[n+1, n+r]$ contains a multiple of each, or there are 3 of these primes which do not have multiples in $[n+1, n+r]$. Take the case of 4 of the primes having a multiple in $[n+1, n+r]$ and without essential loss of generality, say they are p_1, p_2, p_3, p_4. They determine integers j_1, j_2, j_3, j_4 with $1 \le j_i \le r$ and $p_i \mid n + j_i$. Further, by the second part of Lemma 1 we have each $\ell_2(p_i) \mid n + j_i$. We conclude that n is in a residue class modulo

$$M := \operatorname{lcm}\{p_1, p_2, p_3, p_4, \ell_2(p_1), \ell_2(p_2), \ell_2(p_3), \ell_2(p_4)\}.$$

Now p_1, p_2, p_3, p_4 are distinct primes in $[r+1, r+r^{0.61}]$, and each $\ell_2(p_i)$, since it divides $p_i - 1$, has all prime factors $\le r$, so is coprime to the other p_j's. Moreover, each $\ell_2(p_i) > r^{0.3}$ and being a divisor of $p_i - 1$, each $\gcd(\ell_2(p_i), \ell_2(p_j)) \le r^{0.001}$. Thus,

$$M > r^4 r^{1.2} r^{-0.006} = r^{5.194}.$$

Further, $M \ll r^8 < x$. Thus, the number of n in this residue class is $\ll x/M < x/r^{5.194}$. Summing over the different possibilities for j_1, j_2, j_3, j_4, our count is $\ll x/r^{1.194}$. Now summing over $r > y$, we have that the number of n in this case is $\ll x/y^{0.194}$.

We also must consider the possibility that 3 of our 6 primes do not divide any $n + j$ with $1 \le j \le r$. Again without essential loss of generality, assume they are p_1, p_2, p_3. Since each is in $[r+1, r+r^{0.61}]$, it follows that each p_i corners n in a set of $O(r^{0.61})$ residue classes mod p_i. With the Chinese Remainder Theorem, such n's are in a set of $O(r^{1.83})$ residues classes modulo $p_1p_2p_3$. Note that the modulus is small, at most $O(r^3) = o(x)$. Thus, the number of such n is at most

$$O\left(\frac{r^{1.83}x}{p_1p_2p_3}\right) = O\left(\frac{x}{r^{1.17}}\right).$$

Varying the 3 primes in $\binom{6}{3} = 20$ ways multiplies the above count by a constant factor. Summing on $r > y$ we deduce that the number of n in $(x/2, x]$ is $\ll x/y^{0.17}$.

With our above estimate, this puts the count in Case 3 at $O(x/y^{0.17}) = o(x)$ as $x \to \infty$.

Case 4. *We assume that* $r \in (22, y]$.

Here, we do the "regular" thing, where we distinguish between smooth numbers and numbers with a large prime factor. Let $P(m)$ denote the largest prime factor of m. If $P(n+1) \le y$, this puts n in a set of size $x/(\log x)^{(1+o(1))\log\log\log x}$ as $x \to \infty$, by standard estimates for smooth numbers. So, assume that $p = P(n+1) > y$. Since $r \le y$, it follows that p does not divide any other $n+j$ with $j \le r$, so that (1) and $S(r,n)$ integral imply that $\ell_2(p) \mid n+1$.

The number of primes $2 < q \le t$ with $\ell_2(q) \le q^{0.3}$ is by the argument in the previous case at most $t^{0.6}$. By a partial summation argument, the number of $n \in (x/2, x]$ with $n+1$ divisible by such a prime $q > y$ is $O(x/y^{0.4})$. So, assume that $\ell_2(p) > p^{0.3}$. The number of integers $n \in (x/2, x]$ with $n+1$ divisible by $p\ell_2(p)$ is $\le (x+1)/(p\ell_2(p)) \ll x/p^{1.3}$. Summing on $p > y$ our count is $\ll x/y^{0.3}$.

Putting together everything, we get that $\#\mathcal{S}(x)$ is $O_A(x/(\log x)^A)$ for every fixed $A > 0$. This completes the proof of the first part of the theorem.

We next show that all members of $\mathcal{S}$ are large. By the first part of Lemma 1, if one of $n+1, \dots, n+r$ is a prime, then $S_r(n)$ is nonintegral. From [3] we know that if $S_r(n)$ is integral, then $r \ge 23$. We conclude that if one of $n+1, \dots, n+23$ is a prime, then $S_r(n)$ is nonintegral. One can check that there is a prime in $[n+1, n+23]$ for every $n \le 1326$, so the least member of $\mathcal{S}$ (if there are any members at all) is at least 1327.

Let $n_0 = 1{,}349{,}533$. Below n_0 the largest gap between consecutive primes has length 114, see https://oeis.org/A002386. So, if $r > 114$ and $S_n(r)$ is integral, then $n \ge n_0$. Assume then that $23 \le r \le 114$ and that $1327 \le n < n_0$. If some $n+j$, $1 \le j \le 23$, is divisible by a prime $q \ge (n+j)/4$, then $n+j = aq$ for some $a \le 4$. Since $(n+j)/4 > 114$ and $q > 2^4$ (so that $\ell_2(q) > 4$), the second part of Lemma 1 shows that $S_r(n)$ is nonintegral. However, every $n < n_1 := 17{,}258$ has some $n+j$, $1 \le j \le 23$, divisible by a prime $q \ge (n+j)/4$. We deduce that each member of $\mathcal{S}$ is at least n_1.

Continuing, assume $n \ge n_1$. If some $n+j$, $1 \le j \le 23$, is of the form aq with q prime and $a \le 12$, then $n \notin \mathcal{S}$. Indeed, the greatest prime factor of any $2^i - 1$ for $i \le 12$ is 127 and $q > n_1/12 > 127$. We find that every $n < n_2 := 178{,}701$ has this property, so every member of $\mathcal{S}$ is at least n_2.

Now assume that $n_2 \le n \le 10^6$. If some $n+j$, $1 \le j \le 23$, is of the form aq with q prime and $a \le 41$, then $q > n_2/41 > 4300$. Further, the only values of $q > 4300$ with $\ell_2(q) \le 41$ are $q = 2^p - 1$ for $p = 13$ and some values of $q > 130{,}000$ with $\ell_2(q) \ge 17$, so that in these latter cases, $q\ell_2(q) > 2 \times 10^6$. So except possibly for those n with $13(2^{13} - 1) = 106{,}483 \mid n+j$ for some $1 \le j \le 23$, if n is such that some $n+j = aq$ with q prime and $a \le 41$, then $n \notin \mathcal{S}$. We have checked those n with $106{,}483 \mid n+j$ for some $1 \le j \le 23$ and $n_2 \le n \le 10^6$, and each has some $n+j'$ of the form aq with q prime, $a \le 30$, and $\ell_2(q) \nmid a$. It follows that these values of n are not in $\mathcal{S}$. This concludes our proof that $\mathcal{S}$ contains no numbers $\le 10^6$.

Remarks. Note that assuming Cramér's conjecture that for some constant c and for large x there is a prime in $[x, x+c(\log x)^2]$, the estimate in Case 2 is eliminated. By then optimizing the choice of y, our final count for $\mathcal{S}(x)$ would be of the shape

$O(x/\exp(c\sqrt{\log x \log\log x}))$ for some $c > 0$. The hardest cases to try and do better seem to be $r = O(1)$.

Let $s_r(m)$ be the largest r-smooth divisor of m and let $M_r(n) = \min\{s_r(n+j) : 1 \le j \le r\}$. It follows from [3, Proposition 3.1] that if $M_r(n) \le \log_2 r$, then $S_r(n)$ is nonintegral. Unfortunately, as discussed in [3, Remark 2], it is not always the case that $M_r(n) \le \log_2 r$. Nevertheless, it seems interesting to get estimates for $M(r) := \max\{M_r(n) : n > 0\}$.

References

[1] R. C. Baker, G. Harman, and J. Pintz, "The difference between consecutive primes. II", *Proc. London Math. Soc.* (3) **83** (2001), 532–562.

[2] C. Jia, "Almost all short intervals containing prime numbers", *Acta Arith.* **76** (1996), 21–84.

[3] S. Laishram, D. López-Aguayo, C. Pomerance and T. Thongjunthug, "Progress towards a non-integrality conjecture", *Eur. J. Math.* **6** (2020), 1496–1504.

[4] D. López-Aguayo, "Non-integrality of binomial sums and Fermat's Little Theorem", *Math. Mag.* **88** (2015), 231–234.

[5] F. López Aguayo and F. Luca, "Sylvester's theorem and the non-integrality of a certain binomial sum", *Fibonacci Quart.* **54** (2016), 44–48.

[6] H. L. Montgomery and R. C. Vaughan, "On the distribution of reduced residues", *Ann. of Math.* **123** (1986), 311–333.

School of Maths, Wits University, South Africa, Max Planck Institute for Mathematics, Bonn, Germany, Research Group in Algebraic Structures and Applications, King Abdulaziz University, Jeddah, Saudi Arabia

E-mail address: `florian.luca@wits.ac.za`

Mathematics Department, Dartmouth College, 27 N. Main Street, 6188 Kemeny Hall, Hanover, Hanover NH 03755-3551, USA

E-mail address: `carlp@math.dartmouth.edu`