



**UNIVERSITY  
OF TURKU**

This is a self-archived – parallel published version of an original article. This version may differ from the original in pagination and typographic details. When using please cite the original.

This is a pre-copied, author-produced version of an article accepted for publication in

**JOURNAL**                      The Quarterly Journal of Mathematics

following peer review. The version of record:

**CITATION**                      Kaisa Matomäki, Maksym Radziwiłł, Xuancheng Shao, Terence Tao, Joni Teräväinen, Singmaster's Conjecture In The Interior Of Pascal's Triangle, The Quarterly Journal of Mathematics, 2022;, haac006

is available online at:

**DOI**                                <https://doi.org/10.1093/qmath/haac006>

# SINGMASTER'S CONJECTURE IN THE INTERIOR OF PASCAL'S TRIANGLE

KAISA MATOMÄKI, MAKSYM RADZIWIŁŁ, XUANCHENG SHAO, TERENCE TAO,  
AND JONI TERÄVÄINEN

ABSTRACT. Singmaster's conjecture asserts that every natural number greater than one occurs at most a bounded number of times in Pascal's triangle; that is, for any natural number  $t \geq 2$ , the number of solutions to the equation  $\binom{n}{m} = t$  for natural numbers  $1 \leq m < n$  is bounded. In this paper we establish this result in the interior region  $\exp(\log^{2/3+\varepsilon} n) \leq m \leq n - \exp(\log^{2/3+\varepsilon} n)$  for any fixed  $\varepsilon > 0$ . Indeed, when  $t$  is sufficiently large depending on  $\varepsilon$ , we show that there are at most four solutions (or at most two in either half of Pascal's triangle) in this region. We also establish analogous results for the equation  $(n)_m = t$ , where  $(n)_m := n(n-1)\dots(n-m+1)$  denotes the falling factorial.

## 1. INTRODUCTION

In 1971, Singmaster [24] conjectured that any natural number greater than one only appeared in Pascal's triangle a bounded number of times. In asymptotic notation<sup>1</sup>, we can express this conjecture as

**Conjecture 1.1** (Singmaster's conjecture). For any natural number  $t \geq 2$ , the number of integer solutions  $1 \leq m < n$  to the factorial Diophantine equation

$$(1.1) \quad \binom{n}{m} = t$$

is  $O(1)$ .

Note that we can exclude the edges  $m = 0, m = n$  of Pascal's triangle from consideration since  $\binom{n}{m} = 1$  in these cases. Currently the largest known number of solutions to (1.1) for a given  $t$  is eight, arising from  $t = 3003$  and

$$(1.2) \quad (n, m) = (3003, 1), (78, 2), (15, 5), (14, 6), (14, 8), (15, 10), (78, 76), (3003, 3002).$$

For the purposes of attacking this conjecture, we may of course assume  $t$  to be larger than any given absolute constant, which we shall implicitly do in the sequel. In particular we can assume that the iterated logarithms

$$\log_2 t := \log \log t; \quad \log_3 t := \log \log \log t$$

are well-defined and positive.

In view of the symmetry

$$(1.3) \quad \binom{n}{m} = \binom{n}{n-m}$$

---

<sup>1</sup>Our conventions for asymptotic notation are set out in Section 1.5.

we may restrict attention to the left half

$$(1.4) \quad \{(m, n) \in \mathbb{N} \times \mathbb{N} : 1 \leq m \leq n/2\}$$

of Pascal's triangle. For solutions to (1.1) in this half (1.4) of the triangle, we have

$$t = \binom{n}{m} \geq \binom{2m}{m} \asymp 4^m / \sqrt{m}$$

by Stirling's approximation (2.4), and thus we have the upper bound

$$(1.5) \quad m \leq \frac{1}{\log 4} \log t + O(\log_2 t).$$

Since  $n \mapsto \binom{n}{m}$  is an increasing function of  $n$  for fixed  $m \geq 1$ ,  $n$  is uniquely determined by  $m$  and  $t$ . Thus by (1.5) we have at most  $O(\log t)$  solutions to the equation  $\binom{n}{m} = t$ , a fact already observed in the original paper [24] of Singmaster. This bound was improved to  $O(\log t / \log_2 t)$  by Abbott, Erdős, and Hansen [1], to  $O(\log t \log_3 t / \log_2^2 t)$  by Kane [16], and finally to  $O(\log t \log_3 t / \log_2^3 t)$  in a followup work of Kane [17]. This remains the best known unconditional bound for the total number of solutions, although it was observed in [1] that the improved bound  $O_\varepsilon(\log^{2/3+\varepsilon} t)$  was available for any  $\varepsilon > 0$  assuming the conjecture of Cramér [10].

From the elementary inequalities

$$\frac{(n-m)^m}{m!} < \binom{n}{m} \leq \frac{n^m}{m!}$$

and some rearranging we see that any solution to  $\binom{n}{m} = t$  obeys the bounds

$$(tm!)^{1/m} \leq n < (tm!)^{1/m} + m.$$

Applying Stirling's approximation (2.4) (and also  $n \geq m$ ) we can thus obtain the order of magnitude of  $n$  as a function of  $m$  and  $t$ :

$$(1.6) \quad n \asymp mt^{1/m}$$

or equivalently

$$(1.7) \quad \frac{n}{m} \asymp \exp\left(\frac{\log t}{m}\right).$$

In particular we see that  $n$  grows extremely rapidly when the ratio  $m/\log t$  becomes small. This makes the difficulty of the problem increase as  $m/\log t$  approaches zero, and indeed treating the case of small values of  $m/\log t$  is the main obstruction to making further progress on bounding the total number of solutions.

**Remark 1.2.** In the left half (1.4) of Pascal's triangle, a finer application of Stirling's approximation in [16, (3.1)] gave the more precise estimate

$$n = (tm!)^{1/m} + \frac{m-1}{2} + O(mt^{-1/m}).$$

We will not explicitly use this estimate here.

In this paper we study the opposite regime in which  $m/\log t$  is relatively large, or equivalently (by (1.7))  $n$  and  $m$  are somewhat comparable (in the doubly logarithmic sense  $\log_2 n \asymp \log_2 m$ ). More precisely, we have the following result:

**Theorem 1.3** (Singmaster’s conjecture in the interior of Pascal’s triangle). *Let  $0 < \varepsilon < 1$ , and assume that  $t$  is sufficiently large depending on  $\varepsilon$ . Then there are at most two solutions to (1.1) in the region  $\exp(\log^{2/3+\varepsilon} n) \leq m \leq n/2$ . By (1.3), we thus have at most four solutions to (1.1) in the region  $\exp(\log^{2/3+\varepsilon} n) \leq m \leq n - \exp(\log^{2/3+\varepsilon} n)$ . Furthermore, in the smaller region  $\exp(\log^{2/3+\varepsilon} n) \leq m \leq n / \exp(\log^{1-\varepsilon'} n)$  there is at most one solution, whenever  $0 < \varepsilon' < \frac{\varepsilon}{2/3+\varepsilon}$  and  $t$  is sufficiently large depending on both  $\varepsilon$  and  $\varepsilon'$ .*

**Remark 1.4.** The bound of two (or four) solutions is absolutely sharp, in view of the infinite family of solutions observed in [20], [25], [29] to the equation

$$\binom{n+1}{m+1} = \binom{n}{m+2}$$

given by

$$(1.8) \quad n = F_{2j+2}F_{2j+3} - 1; \quad m = F_{2j}F_{2j+3} - 1,$$

where  $F_j$  denotes the  $j^{\text{th}}$  Fibonacci number. See also [15] for further analysis of equations of this type. Besides this infinite family of collisions, and the “trivial” ones generated by (1.3),  $\binom{n}{0} = 1$ , and  $\binom{n}{m} = \binom{n}{1}$ , the only further known collisions between binomial coefficients arise from the identities  $\binom{n}{2} = \binom{n'}{m'}$  for

$$(n, n', m') = (16, 10, 3), (21, 2, 4), (56, 22, 3), (120, 36, 3), (153, 19, 5), (221, 17, 8)$$

as well as the example in (1.2). It was conjectured by de Weger [13] that these above examples generate all the non-trivial collisions  $\binom{n}{m} = \binom{n'}{m'} = t$ ; this would of course imply Singmaster’s conjecture. This conjecture has been verified for  $(m, m') = (2, 3)$  [3], for  $(m, m') = (2, 4)$  [23], [12], for  $(m, m') = (2, 5)$  [9], for  $(m, m') = (3, 4)$  [22], [13], and  $(m, m') = (2, 6), (2, 8), (3, 6), (4, 6), (4, 8)$  [28], and for  $n \leq 10^6$  or  $t \leq 10^{60}$  in [5].

**Remark 1.5.** In view of Theorem 1.3, we now see that to prove Conjecture 1.1, we may restrict attention without loss of generality to the region  $2 \leq m \leq \exp(\log^{2/3+\varepsilon} n)$  for any fixed  $\varepsilon > 0$ , or equivalently (by (1.7)) to  $2 \leq m \leq \frac{\log t}{\log_2^{3/2-\varepsilon} t}$  for any fixed  $\varepsilon > 0$ . (In particular, the infinite family (1.8) of collisions is of no further relevance to Singmaster’s conjecture, as it does not lie in this region.) It follows from the conjecture of de Weger [13] mentioned in Remark 1.4 that for  $t$  sufficiently large there is only at most one solution in this region, that is to say all but a finite number of binomial coefficients  $\binom{n}{m}$  for  $2 \leq m \leq \exp(\log^{2/3+\varepsilon} n)$  are distinct. In this direction, the number of solutions to the equation  $\binom{n}{m} = \binom{n'}{m'}$  for fixed  $2 \leq m < m'$  has been shown (via Siegel’s theorem on integral points) to be finite in [4] (see also the earlier result [18] treating the case  $(m, m') = (2, p)$  for an odd prime  $p$ ). This implies that there are no collisions in the regime  $2 \leq m \leq w(n)$  if  $w$  is a function of  $n$  that goes to infinity sufficiently slowly as  $n \rightarrow \infty$ . Unfortunately, due to the reliance on Siegel’s theorem, the function  $w$  given by these arguments is completely ineffective.

**Remark 1.6.** For some previous bounds of this type, in [1] it was shown that the number of solutions to (1.1) in the range  $n^{5/6} \leq m \leq n/2$  was  $O(\log^{3/4} t)$ , while the arguments in [16, §7], after some manipulation, show that the number of solutions to (1.1) in the range  $\exp(\log^{1/2+\varepsilon} n) \leq m \leq n^{5/6}$  is  $O_\varepsilon(\log t / \log_2^3 t)$ . However, to our knowledge Theorem 1.3 is the first result that establishes a bounded number of solutions to (1.1) even in the region  $\varepsilon n \leq m \leq n/2$ .

**Remark 1.7.** The implied quantitative bounds in the hypothesis “ $t$  is sufficiently large depending on  $\varepsilon$ ” are effective; however, we have made no attempt whatsoever to optimize them in this paper, and will likely be too large to be of use in numerical verification of Singmaster’s conjecture in their current form.

**Remark 1.8.** The analogue of Singmaster’s conjecture for multinomial coefficients  $\binom{n}{m_1, \dots, m_k}$  was studied very recently in [11]. It seems plausible that the methods in this paper may also be used to obtain some partial results on this multinomial formulation of the conjecture.

**1.1. An analog for falling factorials.** The methods used to handle the equation (1.1) can be modified to treat some other factorial Diophantine equations, and in particular the variant equation

$$(1.9) \quad (n)_m = t$$

for integers  $1 \leq m < n$  and  $t \geq 2$ , where  $(n)_m$  denotes the falling factorial

$$(n)_m := n(n-1)\dots(n-m+1) = m! \binom{n}{m}.$$

We exclude the cases  $m = 0, m = n$  since  $(n)_0 = 1$  and  $(n)_n = (n)_{n-1} = n!$ . In [1, Theorem 4] it was shown that for any  $t \geq 2$  the number of integer solutions  $(m, n)$  to (1.9) with  $1 \leq m \leq n-1$  is  $O(\sqrt{\log t})$ . We do not directly improve upon this bound here, but can obtain an analogue of Theorem 1.3:

**Theorem 1.9** (Falling factorial multiplicity in the interior). *Let  $0 < \varepsilon < 1$ , and assume that  $t$  is sufficiently large depending on  $\varepsilon$ . Then there are at most two integer solutions to (1.9) in the region  $\exp(\log^{2/3+\varepsilon} n) \leq m < n$ .*

We establish this result in Section 5. Note that the bound of two is best possible, as can be seen from the infinite family of solutions

$$(a^2 - a)_{a^2 - 2a} = (a^2 - a - 1)_{a^2 - 2a + 1}$$

for any integer  $a > 2$ , and more generally

$$((a)_b)_{(a)_b - a} = ((a)_b - 1)_{(a)_b - a + b - 1}$$

whenever  $2 \leq b < a$  are integers.

**1.2. Strategy of proof.** Theorem 1.3 is a consequence of two Propositions that we now describe. The proof of Theorem 1.9 will follow a similar pattern as described here and we refer the reader to Section 5 for details.

**Proposition 1.10** (Distance estimate). *Let  $\varepsilon > 0$ . Suppose we have two solutions  $(n, m), (n', m')$  to (1.1) in the left half (1.4) of Pascal’s triangle. Then one has*

$$m' - m \ll_{\varepsilon} \exp(\log^{2/3+\varepsilon}(n + n'))$$

for any  $\varepsilon > 0$ . Furthermore, if

$$m, m' \geq \exp(\log^{2/3+\varepsilon}(n + n'))$$

then we additionally have

$$n' - n \ll_{\varepsilon} \exp(\log^{2/3+\varepsilon}(n + n')).$$

Note how this proposition is consistent with the example in Remark 1.4. We shall discuss the proof of Proposition 1.10 in Section 1.3. For the application to Theorem 1.3, Proposition 1.10 localizes all solutions to (1.1) to a region of small diameter. To conclude Theorem 1.3, we can now proceed by adapting the Taylor expansion arguments of Kane [16], [17], in which one views  $n$  as an analytic function of  $m$  (keeping  $t$  fixed) and exploits the non-vanishing of certain derivatives of this function; see Section 2. This is what the proposition below accomplishes. In fact in our analysis only two derivatives of this function are needed (i.e., we only need to exploit the convexity properties of  $n$  as a function of  $m$ ).

**Proposition 1.11** (Kane-type estimate). *Let  $\varepsilon > 0$ , and suppose  $t$  is sufficiently large depending on  $\varepsilon$ . Suppose that  $(n, m)$  is a solution to (1.1) in the left-half (1.4) of Pascal's triangle. Then there exists at most one other solution  $(n', m') \neq (n, m)$  to (1.1) with  $m' < m$ ,  $n' > n$  and*

$$|m - m'| + |n - n'| \ll \exp((\log_2 t)^{1-\varepsilon}).$$

With these two Propositions at hand it is easy to deduce Theorem 1.3.

*Deduction of Theorem 1.3.* Let  $\varepsilon > 0$ , let  $t$  be sufficiently large depending on  $\varepsilon$ , and let  $(n, m)$  be the solution to (1.1) in the region

$$(1.10) \quad \{(n, m) : \exp(\log^{2/3+\varepsilon} n) \leq m \leq n/2\}$$

with the maximal value of  $m$  (if there are no such solutions then of course Theorem 1.3 is trivial). For brevity we allow all implied constants in the following arguments to depend on  $\varepsilon$ . If  $(n', m')$  is any other solution in this region, then  $m' < m$  and  $n' > n$ . From (1.7) we have

$$m \gg \frac{\log t}{\log n} \geq \frac{\log t}{\log^{1/2/3+\varepsilon} m} \gg \frac{\log t}{\log_2^{1/2/3+\varepsilon} t}$$

thanks to (1.5). From further application of (1.7) we then have

$$n \ll \exp(O(\log_2^{1/2/3+\varepsilon} t)).$$

Similarly for  $n'$ . Applying Proposition 1.10 (with  $\varepsilon$  replaced by a sufficiently small quantity), we conclude that

$$(1.11) \quad m - m', n' - n \ll_{\varepsilon'} \exp(O(\log_2^{1-\varepsilon'} t))$$

whenever  $1 - \varepsilon' > \frac{2/3}{2/3+\varepsilon}$ , or equivalently  $\varepsilon' < \frac{\varepsilon}{2/3+\varepsilon}$ . The result now follows from Proposition 1.11.  $\square$

**Remark 1.12.** The above arguments showed that for  $t$  sufficiently large depending on  $\varepsilon$ , there were at most four solutions to (1.1) in the region  $\exp(\log^{2/3+\varepsilon} n) \leq m \leq n - \exp(\log^{2/3+\varepsilon} n)$ . A modification of the argument also shows that there cannot be exactly *three* such solutions. For if this were the case, we see from (1.3) that there must be a solution  $(n, m)$  with  $n = 2m$ , so that  $m \asymp \log t$  by Stirling's approximation. For all other solutions  $(n', m')$  to (1.1) we have  $n' \geq n + 1$ , hence

$$\binom{n}{n/2} = t = \binom{n'}{m'} \geq \binom{n+1}{m'}$$

and hence (by Stirling's approximation)

$$\binom{n+1}{m'} \leq \left(\frac{1}{2} + O\left(\frac{1}{n}\right)\right) \binom{n+1}{(n+1)/2}.$$

By Stirling's approximation (or the central limit theorem of de Moivre and Laplace) this forces  $|m' - \frac{n+1}{2}| \gg \sqrt{n}$ , thus  $|m' - m| \gg m^{1/2}$ . But this contradicts (1.11).

**1.3. Proof methods.** We now discuss the method of proof of Proposition 1.10, which is our main new contribution. In contrast to the ‘‘Archimedean’’ arguments of Kane (such as Proposition 1.11) that use real and complex analysis of the binomial coefficients  $\binom{n}{m}$ , the proof of Proposition 1.10 relies more on ‘‘non-Archimedean’’ arguments, based on evaluating the  $p$ -adic valuations  $v_p\left(\binom{n}{m}\right)$  for various primes  $p$ , defined as the number of times  $p$  divides  $\binom{n}{m}$ . From the classical Legendre formula

$$(1.12) \quad v_p(n!) = \sum_{j=1}^{\infty} \left\lfloor \frac{n}{p^j} \right\rfloor,$$

where  $\lfloor x \rfloor$  is the integer part of  $x$ , we obtain Kummer's theorem

$$(1.13) \quad \begin{aligned} v_p\left(\binom{n}{m}\right) &= \sum_{j=1}^{\infty} \left( \left\lfloor \frac{n}{p^j} \right\rfloor - \left\lfloor \frac{m}{p^j} \right\rfloor - \left\lfloor \frac{n-m}{p^j} \right\rfloor \right) \\ &= \sum_{j=1}^{\infty} \left( \left\{ \frac{m}{p^j} \right\} + \left\{ \frac{n-m}{p^j} \right\} - \left\{ \frac{n}{p^j} \right\} \right) \end{aligned}$$

where  $\{x\} := x - \lfloor x \rfloor$  denotes the fractional part of  $x$ . Note that the summands here vanish whenever  $p^j > n$ . From this identity we see that if  $(n, m), (n', m')$  are two solutions to (1.1) then we must have

$$(1.14) \quad \sum_{j=1}^{\infty} \left( \left\{ \frac{m}{p^j} \right\} + \left\{ \frac{n-m}{p^j} \right\} - \left\{ \frac{n}{p^j} \right\} \right) = \sum_{j=1}^{\infty} \left( \left\{ \frac{m'}{p^j} \right\} + \left\{ \frac{n'-m'}{p^j} \right\} - \left\{ \frac{n'}{p^j} \right\} \right)$$

for all primes  $p$ . Our strategy will be to apply this equation with  $p$  set equal to a *random* prime  $\mathbf{p}$  drawn uniformly amongst all primes in the interval  $[P, P + P \log^{-100} P]$  where the scale  $P$  is something like  $\exp(\log^{2/3+\varepsilon/2}(n+n'))$ , and inspect the distribution of the resulting random variables on the left and right-hand sides of (1.14) in order to obtain a contradiction when  $m, m'$  or  $n, n'$  are sufficiently well separated. In order to do this we need some information concerning the equidistribution of fractional parts such as  $\{\frac{n}{\mathbf{p}^j}\}$ . This will be provided by the following estimate, proven in Section 4. There and later the letter  $p$  always denotes a prime.

**Proposition 1.13** (Equidistribution estimate). *Let  $\varepsilon > 0$  and  $P \geq 2$  and let  $I$  be an interval contained in  $[P, 2P]$ . Let  $M, N$  be real numbers with  $M, N = O(\exp(\log^{3/2-\varepsilon} P))$ , and let  $j$  be a natural number.*

(i) *For all  $A > 0$ ,*

$$\sum_{p \in I} e\left(\frac{N}{p} + \frac{M}{p^j}\right) = \int_I e\left(\frac{N}{t} + \frac{M}{t^j}\right) \frac{dt}{\log t} + O_{\varepsilon, A}(P \log^{-A} P).$$

(ii) *Let  $W: \mathbb{R}^2 \rightarrow \mathbb{C}$  be a smooth  $\mathbb{Z}^2$ -periodic function. Then, for all  $A > 0$ ,*

$$\sum_{p \in I} W\left(\frac{N}{p}, \frac{M}{p^j}\right) = \int_I W\left(\frac{N}{t}, \frac{M}{t^j}\right) \frac{dt}{\log t} + O_{\varepsilon, A}(\|W\|_{C^3} P \log^{-A} P),$$

where

$$\|W\|_{C^3} := \sum_{j=0}^3 \sup_{x \in \mathbb{R}^2} |\nabla^j W(x)|.$$

One can generalize this proposition to control the joint equidistribution of any bounded number of expressions of the form  $\{\frac{n}{p^j}\}$ , but for our applications it will suffice to understand the equidistribution of pairs  $\{\frac{N}{p}\}, \{\frac{M}{p^j}\}$ .

When it comes to the proof of Proposition 1.13, the first step is to use Fourier expansion to reduce part (ii) of the proposition to part (i). For part (i), the case where  $\frac{|N|}{P} + \frac{|M|}{P^j}$  is small (say  $\leq \log^{O(A)} P$ ) is easily handled using the prime number theorem with classical error term. In the regime where  $\frac{|N|}{P} + \frac{|M|}{P^j}$  is large, we use Vaughan's identity to decompose the sum in (i) into type I and II sums, and assert that these exhibit cancellation; the type I and II bounds are given in (4.9) and (4.11).

Both type I and type II sums can be handled using Vinogradov's bound for sums of the form  $\sum_{n \in I} e(f(n))$  with  $f$  smooth, although we need to first cut from  $I$  small intervals around zeros of the first  $\log P$  derivatives of  $N/t + M/t^j$ . This way we obtain that the sum in (i) exhibits cancellation. It is here that the restriction  $N, M = O(\exp(\log^{3/2-\varepsilon} P))$  arises; even under the Riemann hypothesis we do not know how to relax this requirement<sup>2</sup>.

Once the equidistribution estimate, Proposition 1.13, is established, the analysis of the distribution of both sides of (1.14) is relatively straightforward, as long as the scale  $P$  is chosen so that the powers  $P^j$  do not lie close to various integer combinations of  $m, n, m', n'$ . However, there are some delicate cases when two of the numbers  $n, m, n-m, n', m', n'-m'$  are "commensurable" in the sense that one of them is close to a rational multiple of the other, where the rational multiplier has small height. Commensurable integers are also known to generate some exceptional examples of integer factorial ratios [6], [7], [27]. Fortunately, we can handle these cases in our context by an analysis of covariances between various fractional parts  $\{\frac{n_1}{\mathbf{p}}\}, \{\frac{n_2}{\mathbf{p}}\}$ , in particular taking advantage of the fact that these covariances are non-negative up to small errors, and small unless  $n_1, n_2$  are very highly commensurable.

**1.4. Acknowledgments.** KM was supported by Academy of Finland grant no. 285894. MR acknowledges the support of NSF grant DMS-1902063 and a Sloan Fellowship. XS was supported by NSF grant DMS-1802224. TT was supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1764034. JT was supported by a Titchmarsh Fellowship. We thank William Verrault for drawing attention to the recent preprint [11], Antoine Deleforge for pointing out a typo in the original version of this manuscript, and Jordan Ellenberg for noting a connection between Lemma 2.2 and the Bombieri–Pila determinant method.

---

<sup>2</sup>Using standard randomness heuristics one could tentatively conjecture that this restriction  $N, M = O(\exp(\log^{3/2-\varepsilon} P))$  could be relaxed to  $N, M = O(\exp(P^c))$  for some constant  $c > 0$ ; this would improve the range  $\exp(\log^{2/3+\varepsilon} n) \leq m \leq n/2$  in Theorem 1.3 to  $\log^C n \leq m \leq n/2$  for some constant  $C > 0$ . In the opposite direction, if we replaced Vinogradov's exponential sum estimates with the simpler estimates of van der Corput or Weyl, one could recover a weaker version of our main theorem in which the range of  $m$  is narrowed to  $n^\varepsilon \leq m \leq n - n^\varepsilon$ . We leave the details of these variants to the interested reader.

**1.5. Notation.** We use  $X \ll Y$ ,  $X = O(Y)$ , or  $Y \gg X$  to denote the estimate  $|X| \leq CY$  for some constant  $C$ . If we wish to permit this constant to depend on one or more parameters we shall indicate this by appropriate subscripts, thus for instance  $O_{\varepsilon,A}(Y)$  denotes a quantity bounded in magnitude by  $C_{\varepsilon,A}Y$  for some quantity  $C_{\varepsilon,A}$  depending only on  $\varepsilon, A$ . We write  $X \asymp Y$  for  $X \ll Y \ll X$ .

We use  $1_E$  to denote the indicator of an event  $E$ , thus  $1_E$  equals 1 when  $E$  is true and 0 otherwise.

We let  $e$  denote the standard real character  $e(x) := e^{2\pi ix}$ .

## 2. DERIVATIVE ESTIMATES

We generalize the binomial coefficient  $\binom{n}{m}$  to real  $0 \leq m \leq n$  by the formula

$$\binom{n}{m} := \frac{\Gamma(n+1)}{\Gamma(m+1)\Gamma(n-m+1)}$$

where

$$\Gamma(x) := \frac{e^{-\gamma x}}{x} \prod_{n=1}^{\infty} \left(1 + \frac{x}{n}\right)^{-1} e^{x/n}$$

is the Gamma function (with  $\gamma$  the Euler–Mascheroni constant). This is of course consistent with the usual definition of the binomial coefficient. Observe that the digamma function

$$\psi(x) := \frac{\Gamma'}{\Gamma}(x) = -\gamma + \sum_{n=0}^{\infty} \frac{1}{n+1} - \frac{1}{n+x}$$

is a smooth increasing concave function on  $(0, +\infty)$ , with

$$\psi'(x) = \sum_{n=0}^{\infty} \frac{1}{(n+x)^2}$$

positive and decreasing, and

$$\psi''(x) = -\sum_{n=0}^{\infty} \frac{2}{(n+x)^3}$$

negative. For future reference we also observe the standard asymptotics

$$(2.1) \quad \psi(x) = \log x + O\left(\frac{1}{x}\right)$$

$$(2.2) \quad \psi'(x) = \frac{1}{x} + O\left(\frac{1}{x^2}\right)$$

$$(2.3) \quad \psi''(x) = -\frac{1}{x^2} + O\left(\frac{1}{x^3}\right)$$

and the Stirling approximation

$$(2.4) \quad \log \Gamma(x) = x \log x - x - \frac{1}{2} \log x + \log \sqrt{2\pi} + O\left(\frac{1}{x}\right)$$

for any  $x \geq 1$ ; see e.g., [2, §6.1, 6.3, 6.4]. One could also extend these functions meromorphically to the entire complex plane, but we will not need to do so here.

From the increasing nature of  $\psi$  we see that  $n \mapsto \binom{n}{m}$  is strictly increasing on  $[m, +\infty)$  for fixed real  $m > 0$ , and from Stirling's approximation (2.4) we see that it goes to infinity

as  $n \rightarrow \infty$ . Thus for given  $t > 1$ , we see from the inverse function theorem that there exists a unique smooth function  $f_t: [0, +\infty) \rightarrow [0, +\infty)$  with  $f_t(m) > m$  for all  $m$ , such that

$$(2.5) \quad \binom{f_t(m)}{m} = t.$$

In particular, the equation (1.1) holds for given integers  $1 \leq m \leq n$  and  $t \geq 2$  if and only if  $n = f_t(m)$ . This function  $f_t$  was analyzed by Kane [16], who among other things was able to extend  $f_t$  holomorphically to a certain sector, which then allowed him to estimate high derivatives of this function. However, for our analysis we will only need to control the first few derivatives of  $f_t$ , which can be estimated by hand:

**Proposition 2.1** (Estimates on the first few derivatives). *Let  $t, m$  be sufficiently large with  $m \leq f_t(m)/2$ . Then*

$$(2.6) \quad f_t(m) \asymp mt^{1/m}$$

and

$$(2.7) \quad -f_t'(m) \asymp (f_t(m) - 2m) \frac{\log t}{m^2}$$

and

$$(2.8) \quad f_t''(m) \asymp f_t(m) \left( \frac{\log t}{m^2} \right)^2.$$

In particular,  $f_t$  is convex and decreasing in this regime.

The bound (2.6) can be viewed as a generalization of (1.6) to non-integer values of  $n, m, t$ .

*Proof.* Taking logarithms in (2.5) we have

$$(2.9) \quad \log \Gamma(f_t(m) + 1) - \log \Gamma(f_t(m) - m + 1) - \log \Gamma(m + 1) = \log t.$$

Writing  $n = f_t(m) \geq 2m$ , we thus see from the mean value theorem that

$$m\psi(n - \theta m + 1) - \log \Gamma(m + 1) = \log t$$

for some  $0 \leq \theta \leq 1$  depending on  $t, m$ . Applying (2.1), we conclude that

$$\log(n - \theta m) = \frac{1}{m}(\log t + \log \Gamma(m + 1)) + O\left(\frac{1}{n}\right)$$

which implies that

$$n \asymp n - \theta m \asymp \exp\left(\frac{1}{m}(\log t + \log \Gamma(m + 1))\right)$$

and the claim (2.6) then follows from Stirling's approximation (2.4).

If we differentiate (2.9) we obtain

$$(2.10) \quad f_t'(m)\psi(f_t(m) + 1) - (f_t'(m) - 1)\psi(f_t(m) - m + 1) - \psi(m + 1) = 0.$$

In particular we obtain the first derivative formula

$$(2.11) \quad f_t'(m) = \frac{\psi(m + 1) - \psi(n - m + 1)}{\psi(n + 1) - \psi(n - m + 1)}.$$

From (2.2) and the mean value theorem we have

$$(2.12) \quad \psi(n + 1) - \psi(n - m + 1) \asymp \frac{m}{n}$$

while from either the mean-value theorem and (2.2) (if  $m \asymp n$ ) or from (2.1) (if say  $m \leq n/4$ ) we see that

$$\psi(n-m+1) - \psi(m+1) \asymp \frac{n-2m}{n} \log \frac{n}{m}.$$

We conclude that

$$-f'_t(m) \asymp \frac{n-2m}{m} \log \frac{n}{m}$$

and the claim (2.7) follows from (2.6).

Differentiating (2.10) again, we conclude

$$f''_t(m)\psi(n+1) + (f'_t(m))^2\psi'(n+1) - f''_t(m)\psi(n-m+1) - (f'_t(m)-1)^2\psi'(n-m+1) - \psi'(m+1) = 0.$$

which we can rearrange using (2.11) as

$$\begin{aligned} f''_t(m)(\psi(n+1) - \psi(n-m+1))^3 &= (\psi(n+1) - \psi(n-m+1))^2\psi'(m+1) \\ &\quad + (\psi(n+1) - \psi(m+1))^2\psi'(n-m+1) \\ &\quad - (\psi(n-m+1) - \psi(m+1))^2\psi'(n+1). \end{aligned}$$

From (2.12), (2.6) it thus suffices to show that

$$\begin{aligned} &(\psi(n+1) - \psi(n-m+1))^2\psi'(m+1) \\ &+ (\psi(n+1) - \psi(m+1))^2\psi'(n-m+1) \\ &- (\psi(n-m+1) - \psi(m+1))^2\psi'(n+1) \asymp \frac{m}{n^2} \log^2(n/m). \end{aligned}$$

The quantity  $(\psi(n+1) - \psi(n-m+1))^2\psi'(m+1)$  is non-negative and is of size  $O(m/n^2)$  by (2.12), (2.2). Thus it will suffice to show that

$$(\psi(n+1) - \psi(m+1))^2\psi'(n-m+1) - (\psi(n-m+1) - \psi(m+1))^2\psi'(n+1) \asymp \frac{m}{n^2} \log^2(n/m).$$

We split the left-hand side as the sum of

$$(\psi(m+1) - \psi(n+1))^2(\psi'(n-m+1) - \psi'(n+1))$$

and

$$\begin{aligned} &\psi'(n+1)[(\psi(n+1) - \psi(m+1))^2 - (\psi(n-m+1) - \psi(m+1))^2] \\ &= (\psi(n+1) - \psi(m+1) + \psi(n-m+1) - \psi(m+1))(\psi(n+1) - \psi(n-m+1))\psi'(n+1). \end{aligned}$$

From (2.1), (2.3), and the mean value theorem the first term is positive and comparable to  $\frac{m}{n^2} \log^2 \frac{n}{m}$ ; similarly, from (2.1), (2.2), and (2.12) the second term is positive and bounded above by  $O(\frac{m}{n^2} \log \frac{n}{m})$ . The claim follows.  $\square$

To apply these derivative bounds, we use the following lemma that implicitly appears in [16], [17] (compare also with the determinant method of Bombieri and Pila [8]):

**Lemma 2.2** (Small non-zero derivative implies few integer values). *Let  $k \geq 1$  be a natural number, and suppose that  $f : I \rightarrow \mathbb{R}$  is a smooth function on an interval  $I$  of some length  $|I|$  such that one has the derivative bound*

$$(2.13) \quad 0 < \left| \frac{1}{k!} f^{(k)}(x) \right| < |I|^{-k(k+1)/2}$$

for all  $x \in I$ . Then there are at most  $k$  integers  $m \in I$  for which  $f(m)$  is also an integer.

*Proof.* Suppose for contradiction that there are  $k + 1$  distinct integers  $m_1, \dots, m_{k+1} \in I$  with  $f(m_1), \dots, f(m_{k+1})$  an integer. By Lagrange interpolation, the function

$$(2.14) \quad P(x) := \sum_{i=1}^{k+1} \prod_{1 \leq j \leq k+1: j \neq i} \frac{x - m_j}{m_i - m_j} f(m_i)$$

is a polynomial of degree at most  $k$  such that  $f(x) - P(x)$  vanishes at  $m_1, \dots, m_{k+1}$ . By many applications of Rolle's theorem (see [16, Corollary 2.1]), there must then exist  $x_* \in I$  such that  $f^{(k)}(x_*) - P^{(k)}(x_*)$  vanishes. From (2.14),  $\frac{1}{k!}P^{(k)}(x)$  (which is the degree  $k$  coefficient of  $P(x)$ ) is an integer multiple of  $\frac{1}{\prod_{1 \leq i < j \leq k+1} |m_i - m_j|} \geq |I|^{-k(k+1)/2}$ , and thus either vanishes or has magnitude at least  $|I|^{-k(k+1)/2}$ . But this contradicts (2.13).  $\square$

As an application of these bounds, we can locally control the number of solutions (1.1) in the region  $n^{1/2+\varepsilon} \leq m \leq n/2$ , thus giving a version of Theorem 1.3 in a small interval:

**Corollary 2.3.** *Let  $0 < \varepsilon < 1$ , let  $t$  be sufficiently large depending on  $\varepsilon$ , and suppose that  $(n, m)$  is a solution to (1.1) in the left half (1.4) of Pascal's triangle with  $m \geq n^{1/2+\varepsilon}$ . Then there is at most one other solution  $(n', m')$  to (1.1) in the interval  $m' \in [m - m^{\varepsilon/10}, m]$ .*

*Proof.* From (1.7) and the hypothesis  $n^{1/2+\varepsilon} \leq m \leq n/2$  we have

$$(2.15) \quad \frac{\log t}{\log_2 t} \ll m \ll \log t.$$

For  $x$  in the interval  $I := [m - m^{\varepsilon/10}, m]$ , we then have  $\frac{\log t}{x} = \frac{\log t}{m} + O(m^{-2+\varepsilon/10} \log t) = \frac{\log t}{m} + O(1)$ , and so we see from Proposition 2.1 and (2.15) that  $f_t(x) \asymp n$  and

$$0 < |f_t''(x)| \ll n \left( \frac{\log t}{m^2} \right)^2 \ll \frac{n}{m^2} \log_2^2 t \ll \frac{n}{m^2} \log^2 m$$

for all  $x \in I$ . Since  $m \geq n^{1/2+\varepsilon}$  and  $t$  is sufficiently large depending on  $\varepsilon$ ,  $m$  is also sufficiently large depending on  $\varepsilon$ , and we have

$$0 < |f_t''(x)| < |I|^{-3}$$

for all  $x \in I$ . Applying Lemma 2.2, there are at most two integers  $m' \in I$  with  $f_t(m')$  an integer. Since  $m$  is already one of these integers, the claim follows.  $\square$

The same method, using higher derivative estimates on  $f_t$ , also gives similar results (with weaker bounds on the number of solutions) for  $m < n^{1/2+\varepsilon}$ ; see [16], [17]. However, we will only need to apply this method in the  $m \geq n^{1/2+\varepsilon}$  regime here.

We are now ready to prove Proposition 1.11.

*Proof of Proposition 1.11.* Let  $\varepsilon > 0$ , let  $t$  be sufficiently large depending on  $\varepsilon$ , and let  $(n, m)$  be a solution to (1.1) in the region

$$(2.16) \quad \{(n, m) : \exp(\log^{2/3+\varepsilon} n) \leq m \leq n/2\}$$

For brevity we allow all implied constants in the following arguments to depend on  $\varepsilon$ . Suppose  $(n', m')$  is another solution in this region with  $m' < m$ ,  $n' > n$  and

$$m - m', n' - n \ll_{\varepsilon'} \exp(O(\log_2^{1-\varepsilon'} t)).$$

From (2.7) and convexity (and the bounds  $m \ll \log t$  and  $m - m' \geq 1$ ) we have

$$\begin{aligned} n' - n &= f_t(m') - f_t(m) \\ &\geq f'_t(m)(m' - m) \\ &\gg (n - 2m) \frac{\log t}{m^2} (m - m') \\ &\gg \frac{n - 2m}{m} \\ &= \frac{n}{m} - 2 \end{aligned}$$

and thus

$$n/m \ll_{\varepsilon'} \exp(O(\log_2^{1-\varepsilon'} t))$$

From (1.7) we have  $n \gg \log t$ , hence  $\log_2^{1-\varepsilon'} t \ll \log^{1-\varepsilon'} n$ , and so for some constant  $C > 0$ ,  $m \geq n / \exp(C \log^{1-\varepsilon'} n) \geq n^{9/10}$  (shrinking  $\varepsilon'$  slightly if necessary) if  $t$  is sufficiently large depending on  $\varepsilon'$ . The result now follows from Corollary 2.3.  $\square$

It remains to establish Proposition 1.10. This will be the objective of the next two sections of the paper.

### 3. THE DISTANCE BOUND

In this section we assume Proposition 1.13 and use it to establish Proposition 1.10.

Throughout this section  $0 < \varepsilon < 1$  will be fixed; we can assume it to be small. We may assume that  $t$  is sufficiently large depending on  $\varepsilon$ , as the claim is trivial otherwise. We may assume that  $m' < m$ , hence also  $n' > n$ . We assume for sake of contradiction that at least one of the claims

$$(3.1) \quad m - m' \geq \exp(\log^{2/3+\varepsilon} n')$$

and

$$(3.2) \quad m, m', n' - n \geq \exp(\log^{2/3+\varepsilon} n')$$

is true, as the claim is trivial otherwise. This allows us to select a “good” scale:

**Lemma 3.1** (Selection of scale). *With the above assumptions, there exists  $P > 1$  obeying the following axioms:*

- (i) ( *$m, m', n, n'$  not too large*) We have  $m, m', n, n' \leq \exp(\log^{\frac{3}{2}-\frac{\varepsilon}{10}} P)$ . (In particular,  $P$  will be sufficiently large depending on  $\varepsilon$ , since otherwise  $t = O_\varepsilon(1)$ .)
- (ii) (*Dichotomy*) If  $a, a', b, b'$  are integers with  $|a|, |a'|, |b|, |b'| \leq \log^{1/100} P$ , and  $j$  is a natural number, then either

$$(3.3) \quad |am + a'm' + bn + b'n'| \leq P^j / \log^{1000} P$$

or

$$|am + a'm' + bn + b'n'| \geq P^j \log^{1000} P.$$

- (iii) (*Separation*) At least one of the statements

$$m - m' \geq P \log^{100} P$$

and

$$m, m', n' - n \geq P \log^{100} P$$

is true.

*Proof.* We restrict  $P$  to be a power of two in the range

$$\exp(\log^{2/3+\varepsilon/2} n') \leq P \leq \exp(2 \log^{2/3+\varepsilon/2} n');$$

such a choice will automatically obey (i) since  $n' > n > m > m'$  and (iii) since we assumed that either (3.1) or (3.2) holds. There are  $\gg \log^{2/3+\varepsilon/2} n'$  choices for  $P$ . Some of these will not obey (ii), but we can control the number of exceptions as follows. Firstly, observe that the conclusion (3.3) will hold unless  $j = O(\log^{1/3} n')$ , so we may restrict attention to this range of  $j$ . The number of possible tuples  $(a, a', b, b', j)$  for any such  $P$  is then  $O(\log^{4/100} n' \log^{1/3} n')$ . For each such tuple, we see from the restriction on  $P$  that the number of  $P$  with

$$P^j / \log^{1000} P < |am + a'm' + bn + b'n'| < P^j \log^{1000} P$$

is at most  $O(\log_2 n')$  (since  $am + a'm' + bn + b'n'$  is of size  $O((n')^2)$ , say). Thus we see that the total number of  $P$  which fail to obey (ii) is at most

$$O(\log^{4/100} n' \log^{1/3} n' \log_2 n')$$

which is negligible compared to the total number of choices, which is  $\gg \log^{2/3+\varepsilon/2} n'$ . Thus we can find a choice of  $P$  which obeys all of (i), (ii), and (iii), giving the claim.  $\square$

Henceforth we fix a scale  $P$  obeying the properties in Lemma 3.1. We now introduce a relation  $\approx$  on the reals by declaring  $x \approx y$  if  $|x - y| \leq P / \log^{1000} P$ . Thus, by Lemma 3.1(ii), if  $am + a'm' + bn + b'n' \not\approx 0$  for  $a, a', b, b'$  as in Lemma 3.1(ii) then  $|am + a'm' + bn + b'n'| \geq P \log^{1000} P$ . Also, from Lemma 3.1(iii), at least one of the statements

$$m \not\approx m'$$

and

$$m, m', n' - n \not\approx 0$$

is true.

We introduce a random variable  $\mathbf{p}$ , which is drawn uniformly from the primes in the interval  $I := [P, P + P \log^{-100} P]$  (note that there is at least one such prime thanks to the prime number theorem). From (1.14) we surely have

$$\sum_{j=1}^{\infty} \left( \left\{ \frac{m}{\mathbf{p}^j} \right\} + \left\{ \frac{n-m}{\mathbf{p}^j} \right\} - \left\{ \frac{n}{\mathbf{p}^j} \right\} \right) = \sum_{j=1}^{\infty} \left( \left\{ \frac{m'}{\mathbf{p}^j} \right\} + \left\{ \frac{n'-m'}{\mathbf{p}^j} \right\} - \left\{ \frac{n'}{\mathbf{p}^j} \right\} \right).$$

We can restrict attention to those  $j$  with  $j \leq \log^{1/2} P$ , since the summands vanish otherwise. For any real number  $N$ , we may take covariances of both sides of this identity with the random variable  $\left\{ \frac{N}{\mathbf{p}} \right\}$  to conclude that

$$(3.4) \quad \sum_{j \leq \log^{1/2} P} (c_j(N, m) + c_j(N, n - m) - c_j(N, n)) = \sum_{j \leq \log^{1/2} P} (c_j(N, m') + c_j(N, n' - m') - c_j(N, n'))$$

for any real number  $N$ , where the covariances  $c_j(N, M)$  are defined as

$$\begin{aligned} c_j(N, M) &:= \mathbf{E} \left\{ \frac{N}{\mathbf{p}} \right\} \left\{ \frac{M}{\mathbf{p}^j} \right\} - \mathbf{E} \left\{ \frac{N}{\mathbf{p}} \right\} \mathbf{E} \left\{ \frac{M}{\mathbf{p}^j} \right\} \\ &:= \mathbf{E} \left( \frac{1}{2} - \left\{ \frac{N}{\mathbf{p}} \right\} \right) \left( \frac{1}{2} - \left\{ \frac{M}{\mathbf{p}^j} \right\} \right) - \mathbf{E} \left( \frac{1}{2} - \left\{ \frac{N}{\mathbf{p}} \right\} \right) \mathbf{E} \left( \frac{1}{2} - \left\{ \frac{M}{\mathbf{p}^j} \right\} \right). \end{aligned}$$

We now compute these covariances:

**Proposition 3.2** (Covariance estimates). *Let  $N, M \in \{m, n, m - n, m', n', n' - m'\}$ , and  $j$  be a natural number with  $1 \leq j \leq \log^{1/2} P$ .*

- (i) *If  $j \geq 2$ , then  $c_j(N, M) \ll \log^{-10} P$ .*
- (ii) *If  $j = 1$  and  $N \approx 0$  or  $M \approx 0$ , then  $c_j(N, M) \ll \log^{-1000} P$ .*
- (iii) *If  $j = 1$ ,  $N, M \not\approx 0$  and there exist coprime natural numbers  $1 \leq a, b \leq \log^{1/100} P$  such that  $aN \approx bM$ , then  $c_j(N, M) = \frac{1}{12ab} + O(\log^{-1/1000} P)$ .*
- (iv) *If  $j = 1$  and  $N, M$  are not of the form in (ii) or (iii), then  $c_j(N, M) \ll \log^{-1/1000} P$ .*

**Remark 3.3.** The term  $\frac{1}{12ab}$  appearing in Proposition 3.2(iii) is also the covariance between  $\{n\mathbf{x}\}$  and  $\{m\mathbf{x}\}$  for  $\mathbf{x}$  drawn randomly from the unit interval whenever  $n, m$  are natural numbers with  $an = bm$  for some coprime  $a, b$ ; see [26, Section 2]. Indeed, both assertions are proven by the same Fourier-analytic argument, and Proposition 3.2 endows the linear span of the six functions  $\{\frac{N}{\mathbf{p}}\}$  for  $N \in \{m, n, m - n, m', n', n' - m'\}$  with an inner product closely related to the norm  $N(\cdot)$  studied in [26], the structure of which is the key to obtaining a contradiction from our separation hypotheses on  $n - n', m - m'$ .

*Proof of Proposition 3.2 assuming Proposition 1.13.* We first dispose of the easy case (ii). If  $N \approx 0$ , then  $\{\frac{N}{\mathbf{p}}\} \leq \log^{-1000} P$ , and the claim follows from the triangle inequality; similarly if  $M \approx 0$  or actually if  $M \leq P^j / \log^{1000} P$ . Hence by Lemma 3.1(ii), we may from now on assume that

$$N \geq P \log^{1000} P \quad \text{and} \quad M \geq P^j \log^{1000} P.$$

To handle the remaining cases we use the truncated Fourier expansion

$$\begin{aligned} \frac{1}{2} - \{x\} &= \sum_{0 < |n| \leq N_0} \frac{e(nx)}{2\pi in} + O\left(\frac{1}{1 + N_0 \text{dist}(x, \mathbb{Z})}\right) \\ (3.5) \quad &= \sum_{0 < |n| \leq N_0} \frac{e(nx)}{2\pi in} + O\left(1_{\text{dist}(x, \mathbb{Z}) \leq N_0^{-1/2}} + \frac{1}{N_0^{1/2}}\right) \end{aligned}$$

that holds for any  $N_0 \geq 1$  (see e.g. [14, Formula (4.18)]).

Our primary tool is Proposition 1.13. Note that, for  $t \in I$ ,  $\log t = \log P + O(\log^{-99} P)$ , so that together with the prime number theorem Proposition 1.13 implies that

$$(3.6) \quad \mathbf{E}W\left(\frac{N}{\mathbf{p}}, \frac{M}{\mathbf{p}^j}\right) = \frac{1}{|I|} \int_I W\left(\frac{N}{t}, \frac{M}{t^j}\right) dt + O_\varepsilon(\|W\|_{C^3} \log^{-99} P)$$

for any smooth  $\mathbb{Z}^2$ -periodic  $W: \mathbb{R}^2 \rightarrow \mathbb{C}$  and that, for any  $M', N' = O(\exp(\log^{3/2 - \varepsilon/2} P))$ ,

$$(3.7) \quad \mathbf{E}e\left(\frac{N'}{\mathbf{p}} + \frac{M'}{\mathbf{p}^j}\right) = \frac{1}{|I|} \int_I e\left(\frac{N'}{t} + \frac{M'}{t^j}\right) dt + O_\varepsilon(\log^{-99} P)$$

Applying (3.6) with  $W$  a suitable cutoff localized to the region  $\{(x, y) : \text{dist}(x, \mathbb{Z}) \leq 2N_0^{-1/2}\}$  that equals one on  $\{(x, y) : \text{dist}(x, \mathbb{Z}) \leq N_0^{-1/2}\}$  chosen so that  $\|W\|_{C^3} \ll N_0^{3/2}$ , we see that, for any  $N_0 \in [1, \log^{20} P]$  we have

$$\mathbf{P}\left(\text{dist}\left(\frac{N}{\mathbf{p}}, \mathbb{Z}\right) \leq N_0^{-1/2}\right) \ll \frac{1}{|I|} \int_I 1_{\text{dist}(\frac{N}{t}, \mathbb{Z}) \leq 2N_0^{-1/2}} dt + N_0^{-1/2}.$$

Since  $N \geq P \log^{1000} P$ , the first term on the right-hand side can be computed to be  $O(N_0^{-1/2})$ . Thus

$$(3.8) \quad \mathbf{P} \left( \text{dist} \left( \frac{N}{\mathbf{p}}, \mathbb{Z} \right) \leq N_0^{-1/2} \right) \ll N_0^{-1/2}$$

and a similar argument gives

$$(3.9) \quad \mathbf{P} \left( \text{dist} \left( \frac{M}{\mathbf{p}^j}, \mathbb{Z} \right) \leq N_0^{-1/2} \right) \ll N_0^{-1/2}.$$

To prepare for the proofs of parts (i), (iii) and (iv), let us first show that, for  $1 \leq j \leq \log^{1/2} P$ , we have

$$(3.10) \quad \mathbf{E} \left( \frac{1}{2} - \left\{ \frac{M}{\mathbf{p}^j} \right\} \right) \ll \log^{-10} P.$$

We use the Fourier expansion (3.5) with  $N_0 = \log^{20} P$ . Averaging over  $p \in I$  and applying (3.9) to handle the first error term, we see that

$$\mathbf{E} \left( \frac{1}{2} - \left\{ \frac{M}{\mathbf{p}^j} \right\} \right) = \sum_{0 < |m| \leq \log^{20} P} \frac{1}{2\pi i m} \mathbf{E} e \left( m \frac{M}{\mathbf{p}^j} \right) + O(\log^{-10} P).$$

By the triangle inequality and (3.7), it suffices to show that, for every non-zero integer  $m = O(\log^{20} P)$ ,

$$\frac{1}{|I|} \int_I e \left( m \frac{M}{t^j} \right) dt \ll \log^{-11} P.$$

Recalling that  $M \geq P^j \log^{1000} P$ , this estimate follows from a standard integration by parts (see e.g. [14, Lemma 8.9]). Similarly

$$(3.11) \quad \mathbf{E} \left( \frac{1}{2} - \left\{ \frac{N}{\mathbf{p}} \right\} \right) \ll \log^{-10} P.$$

Furthermore, using similarly (3.5), (3.8), (3.9) and (3.7), we see that, whenever  $1 \leq N_0 \leq \log^{20} P$ ,

$$(3.12) \quad \mathbf{E} \left( \frac{1}{2} - \left\{ \frac{N}{\mathbf{p}} \right\} \right) \left( \frac{1}{2} - \left\{ \frac{M}{\mathbf{p}^j} \right\} \right) = - \sum_{0 < |m|, |n| < N_0} \frac{1}{4\pi^2 m n} \frac{1}{|I|} \int_I e \left( n \frac{N}{t} + m \frac{M}{t^j} \right) dt + O \left( \frac{1}{N_0^{1/2}} \right).$$

Now we are ready to prove (i), (iii), and (iv). Let us start with (i). In light of (3.10), (3.11) and (3.12) with  $N_0 = \log^{20} P$ , it suffices to show that

$$\frac{1}{|I|} \int_I e \left( n \frac{N}{t} + m \frac{M}{t^j} \right) dt \ll \log^{-11} P.$$

whenever  $n, m = O(\log^{20} P)$  are non-zero integers. Applying a change of variables  $t = P/s$ , we reduce to showing that

$$(3.13) \quad \int_{1/(1+\log^{-100} P)}^1 e(as + bs^j) ds \ll \log^{-200} P$$

(say), where  $a := nN/P$  and  $b := mM/P^j$ . By hypothesis, we have  $|a|, |b| \geq \log^{1000} P$ . Since  $2 \leq j \leq \log^{1/2} P$ , the derivative  $a + jbs^{j-1}$  of the phase  $as + bs^j$  is at least  $\log^{200} P$

outside of an interval of length at most  $O(\log^{-200} P)$ , and (3.13) now follows from a standard integration by parts (see e.g. [14, Lemma 8.9]). This concludes the proof of (i).

Let us now turn to (iv). In light of (3.10), (3.11) and (3.12) with  $N_0 = \log^{1/500} P$ , it suffices to show that

$$\frac{1}{|I|} \int_I e\left(\frac{nN + mM}{t}\right) dt \ll \log^{-1/500} P$$

whenever  $n, m = O(\log^{1/500} P)$  are non-zero integers. From the hypothesis (iv) and Lemma 3.1(ii) (after factoring out any common multiple of  $n$  and  $m$ ), we have  $|nN + mM| \geq P \log^{1000} P$ . The claim (iv) now follows from integration by parts.

Finally we show (iii). In light of (3.10), (3.11) and (3.12) with  $N_0 = \log^{1/500} P$ , it suffices to show that

$$-\sum_{0 < |n|, |m| \leq \log^{1/500} P} \frac{1}{4\pi^2 mn} \frac{1}{|I|} \int_I e\left(\frac{nN + mM}{t}\right) dt = \frac{1}{12ab} + O(\log^{-1/1000} P).$$

Let us first consider those  $n, m = O(\log^{1/500} P)$  for which  $nN + mM \not\approx 0$ . By Lemma 3.1(ii)  $|nN + mM| \geq P \log^{1000} P$  and similarly to case (iv), the contribution of such pairs  $(n, m)$  is acceptable.

Consider now the case  $nN \approx -mM$  for some non-zero integers  $n, m = O(\log^{1/500} P)$ . By assumption also  $aN \approx bM$  for some co-prime positive integers  $a, b \leq \log^{1/100} P$ . and hence by Lemma 3.1(ii)  $-amM \approx bnM$  which contradicts the assumption  $M \not\approx 0$  unless  $(n, m)$  is a multiple of  $(a, -b)$ . On the other hand if  $(n, m)$  is a multiple of  $(a, -b)$ , then  $nN \approx -mM$  by Lemma 3.1(ii).

Thus it remains to show that

$$\sum_{0 < |k| \leq \frac{\log^{1/500} P}{\max\{a, b\}}} \frac{1}{4\pi^2 k^2 ab} \frac{1}{|I|} \int_I e\left(\frac{kaN - kbM}{t}\right) dt = \frac{1}{12ab} + O(\log^{-1/1000} P).$$

Since  $aN \approx bM$  we have, for every  $k \leq \log^{1/500} P$ ,

$$\frac{1}{|I|} \int_I e\left(\frac{kaN - kbM}{t}\right) dt = 1 - O(\log^{-100} P)$$

and so it suffices to show that

$$\sum_{0 < |k| \leq \frac{\log^{1/500} P}{\max\{a, b\}}} \frac{1}{4\pi^2 k^2 ab} = \frac{1}{12ab} + O(\log^{-1/1000} P)$$

This is trivial for  $ab \geq \log^{1/1000} P$ . For  $ab \leq \log^{1/1000} P$  the claim follows from the Basel identity

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

and the tail bound

$$\sum_{k \geq \log^{1/1000} P} \frac{1}{k^2} \ll \log^{-1/1000} P.$$

□

Now we can get back to proving Proposition 1.10 assuming Proposition 1.13. From Proposition 3.2(i) and (3.4) we see that

$$(3.14) \quad c_1(N, m) + c_1(N, n - m) - c_1(N, n) = c_1(N, m') + c_1(N, n' - m') - c_1(N, n') + O(\delta)$$

for  $N \in \{m, n, n - m, m', n', m' - n'\}$ , where for brevity we introduce the error tolerance

$$\delta := \log^{-1/1000} P.$$

We can now arrive at the desired contradiction by some case analysis (reminiscent of that in [26, 27]) using the remaining portions of Proposition 3.2, as follows.

**Case  $m' \approx 0$ .** Applying (3.14) with  $N = m$ , we conclude from Proposition 3.2(ii) that

$$(3.15) \quad c_1(m, m) + c_1(m, n - m) - c_1(m, n) = c_1(m, n' - m') - c_1(m, n') + O(\delta).$$

From Lemma 3.1(iii) we have  $m \not\approx 0$  (and hence also  $n - m, n' - m', n' \not\approx 0$ , since these quantities are greater than or equal to  $m$ ), hence by Proposition 3.2(iii) we have  $c_1(m, m) = \frac{1}{12} + O(\delta)$ . Furthermore, since  $m' \approx 0$ , we see from Lemma 3.1(ii) that, for  $1 \leq a, b \leq \log^{1/100} P$ ,  $am \approx b(n' - m')$  if and only if  $am \approx bn'$ . Hence Proposition 3.2(iii) (iv) implies that

$$c_1(m, n' - m') = c_1(m, n') + O(\delta).$$

Plugging these facts into (3.15) and rearranging, we obtain

$$\frac{1}{12} + c_1(m, n - m) = c_1(m, n) + O(\delta).$$

But by Proposition 3.2(iii), (iv) we know that  $c_1(m, n - m) \geq -O(\delta)$ , so that

$$c_1(m, n) \geq \frac{1}{12} + O(\delta)$$

But since  $m \not\approx n$  (because  $m \leq n/2$  and  $m \not\approx 0$ ), another application of Proposition 3.2(iii), (iv) gives

$$c_1(m, n) \leq \frac{1}{2} \frac{1}{12} + O(\delta),$$

which is a contradiction.

Since  $m'$  was the smallest element of  $\{m, n, n - m, m', n', m' - n'\}$ , we now thus have  $N \not\approx 0$  for all  $N \in \{m, n, n - m, m', n', m' - n'\}$ , and case (ii) of Proposition 3.2 no longer applies.

**Case  $m \not\approx m'$  and  $m' \not\approx 0$ .** We apply (3.14) with  $N = m'$  to conclude that

$$(3.16) \quad c_1(m', m) + c_1(m', n - m) - c_1(m', n) = c_1(m', m') + c_1(m', n' - m') - c_1(m', n') + O(\delta).$$

Now if there are no co-prime positive integers  $a, b \leq \log^{1/100} P$  such that  $am' \approx bn'$  or  $am' \approx b(n' - m')$ , then by Proposition 3.2(iv) we have

$$c_1(m', n' - m') - c_1(m', n') = O(\delta)$$

On the other hand, if such co-prime integers exist, then  $am' \approx bn'$  if and only if  $(a - b)m' \approx b(n' - m')$  and necessarily  $a > b$ , so that by Proposition 3.2(iii) we have in this case

$$(3.17) \quad c_1(m', n' - m') - c_1(m', n') = \frac{1}{12(a - b)b} - \frac{1}{12ab} - O(\delta) \geq -O(\delta).$$

Since Proposition 3.2(iii) also gives  $c_1(m', m') \geq 1/12 + O(\delta)$ , combining with (3.16) we obtain that

$$(3.18) \quad c_1(m', m) + c_1(m', n - m) - c_1(m', n) \geq \frac{1}{12} - O(\delta).$$

On the other hand, since  $m' \not\approx m$ , we also have  $m' \not\approx n - m$  since  $n - m \geq m > m'$ . By Proposition 3.2(iii), (iv), we have

$$c_1(m', m) + c_1(m', n - m) \leq \frac{1}{12} \cdot \frac{1}{2} + \frac{1}{12} \cdot \frac{1}{2} + O(\delta),$$

which can be improved to

$$(3.19) \quad c_1(m', m) + c_1(m', n - m) \leq \frac{1}{12} \cdot \frac{1}{3} + \frac{1}{12} \cdot \frac{1}{2} + O(\delta),$$

unless both  $m \approx 2m'$  and  $n - m \approx 2m'$ . Since by Proposition 3.2(iii), (iv) we have  $c_1(m', n) \geq -O(\delta)$ , the estimate (3.19) contradicts (3.18).

Hence we must have both  $m \approx 2m'$  and  $n - m \approx 2m'$ . But then Lemma 3.1(ii) forces  $n \approx 4m'$ , hence by Proposition 3.2(iii)

$$c_1(m', m) + c_1(m', n - m) - c_1(m', n) = \frac{1}{12} \cdot \frac{1}{2} + \frac{1}{12} \cdot \frac{1}{2} - \frac{1}{12} \cdot \frac{1}{4} + O(\delta),$$

and we again contradict (3.18).

**Case  $m \approx m'$  and  $m' \not\approx 0$ .** By Lemma 3.1(iii), we must have  $n \not\approx n'$ . We apply (3.14) for  $N = n$  to obtain

$$(3.20) \quad c_1(n, m) + c_1(n, n - m) - c_1(n, n) = c_1(n, m') + c_1(n, n' - m') - c_1(n, n') + O(\delta).$$

Since  $m \approx m'$ , we have by Proposition 3.2(iii), (iv) (using also Lemma 3.1(ii)) that  $c_1(n, m) = c_1(n, m') + O(\delta)$ . Proposition 3.2(iii) also gives  $c_1(n, n) = 1/12 + O(\delta)$ . Plugging these into (3.20) and rearranging, we obtain

$$(3.21) \quad c_1(n, n - m) + c_1(n, n') = \frac{1}{12} + c_1(n, n' - m') + O(\delta).$$

Since  $n \not\approx n'$  and  $m \not\approx 0$ , we see from Proposition 3.2(iii), (iv) that

$$(3.22) \quad c_1(n, n - m) + c_1(n, n') \leq \frac{1}{12} \cdot \frac{1}{2} + \frac{1}{12} \cdot \frac{1}{2} + O(\delta)$$

which can be improved to

$$(3.23) \quad c_1(n, n - m) + c_1(n, n') \leq \frac{1}{12} \cdot \frac{1}{3} + \frac{1}{12} \cdot \frac{1}{2} + O(\delta)$$

unless  $2(n - m) \approx n$  and  $n' \approx 2n$ . Now (3.23) contradicts (3.21) since by Proposition 3.2(iii), (iv)  $c_1(n, n' - m') \geq -O(\delta)$ .

Hence we can assume that  $2(n - m) \approx n$  and  $n' \approx 2n$ . But using  $m \approx m'$  and Lemma 3.1(ii) this implies that  $2(n' - m') \approx 3n$ , so that by (3.21) and Proposition 3.2(iii) we obtain

$$c_1(n, n - m) + c_1(n, n') = \frac{1}{12} + c_1(n, n' - m') + O(\delta) = \frac{1}{12} + \frac{1}{12} \cdot \frac{1}{2 \cdot 3} + O(\delta).$$

contradicting (3.22).

**Remark 3.4.** Morally speaking, the ability to obtain a contradiction here reflects the fact that one cannot have an identity of the form

$$(3.24) \quad \{mx\} + \{(n-m)x\} - \{nx\} = \{m'x\} + \{(n'-m')x\} - \{n'x\}$$

for almost all real numbers  $x$  and some integers  $1 \leq m \leq n/2$ ,  $1 \leq m' \leq n'/2$  unless one has both  $m = m'$  and  $n = n'$  (this type of connection goes back to Landau [19, p. 116]). This latter fact is easily established by inspecting the jump discontinuities of both sides of (3.24), but it is also possible to establish it by computing the covariances of both sides of (3.24) with  $\{Nx\}$  for various choices of  $N$ , and the arguments above can be viewed as an adaptation of this latter method.

It remains to establish Proposition 1.13. This will be established in the next section.

#### 4. EQUIDISTRIBUTION

In this section we prove Proposition 1.13. Fix  $\varepsilon, A$ . We may assume that  $P$  is sufficiently large depending on  $\varepsilon, A$ , as the claim is trivial otherwise. If we have  $P^j \geq M \log^A P$  then we can replace in both parts of the proposition  $\frac{M}{P^j}$  by 0 with negligible error, so we may assume that either  $M = 0$  or  $P^j < M \log^A P$ . In either event we may thus assume that  $j \leq \log^{1/2} P$ . Next, by partitioning  $I$  into at most  $\log^{100} P$  intervals of length at most  $P \log^{-100} P$  and using the triangle inequality, it suffices (after suitable adjustment of  $P, A$ ) to assume that  $I \subset [P, P + P \log^{-100} P]$ . In particular we have

$$(4.1) \quad P^{j-1} \leq t^{j-1} \leq 2P^{j-1}$$

for all  $t \in I$ .

Let us first reduce Proposition 1.13(ii) to Proposition 1.13(i). We perform a Fourier expansion

$$W(x, y) = \sum_{n, m \in \mathbb{Z}} c_{n, m} e(nx + my)$$

where by integration by parts the Fourier coefficients

$$c_{n, m} = \int_{\mathbb{R}^2/\mathbb{Z}^2} W(x, y) e(-nx - my) dx dy$$

obey the bounds

$$|c_{n, m}| \ll \|W\|_{C^3} (1 + |n| + |m|)^{-3}.$$

By the triangle inequality, the contributions of those frequencies  $n, m$  with  $|n| + |m| \geq \log^{2A} P$  is then acceptable. By a further application of the triangle inequality, Proposition 1.13(ii) follows from showing that

$$\sum_{p \in I} e\left(n \frac{N}{p} + m \frac{M}{p^j}\right) = \int_I e\left(n \frac{N}{t} + m \frac{M}{t^j}\right) \frac{dt}{\log t} + O_{\varepsilon, A}(P \log^{-10A} P)$$

whenever  $n, m$  are integers with  $|n| + |m| \leq \log^{2A} P$ . But this follows from Proposition 1.13(i) by adjusting the values of  $\varepsilon, A, M, N$  suitably.

The proof of part (i) will use the standard tools of Vaughan's identity and Vinogradov's exponential sum estimates. We state a suitable form of the latter tool here:

**Lemma 4.1** (Vinogradov's exponential sum estimate). *Let  $X \geq 2$ ,  $F \geq X^4$ , and  $\alpha \geq 1$ . Let  $I \subset [X, 2X]$  be an interval. Let  $f(x)$  be a smooth function on  $I$  satisfying for all  $t \in I$*

$$(4.2) \quad \alpha^{-r^3} F \leq \frac{t^r}{r!} |f^{(r)}(t)| \leq \alpha^{r^3} F$$

for all integers  $1 \leq r \leq 10 \lceil \log F / (\log X) \rceil + 1$ . Assume further that

$$(4.3) \quad (\log \alpha) \frac{(\log F)^2}{(\log X)^3} < 10^{-3}.$$

Then we have

$$(4.4) \quad \sum_{n \in I} e(f(n)) \ll \alpha X \exp(-2^{-18} (\log X)^3 / (\log F)^2),$$

where the implied constant is absolute.

*Proof.* This is essentially [14, Theorem 8.25] with minor modifications (the modification needed is that we only assume (4.2) for  $r$  in a certain range, not all integers  $r \geq 1$ ).

Let  $R := 10 \lceil \log F / (\log X) \rceil$ , and as in [14, p. 217], let

$$F_n(q) := \sum_{0 \leq r \leq R} \alpha_r(n) q^r, \quad \alpha_r(n) := \frac{f^{(r)}(n)}{r!}.$$

Let  $S_f(I)$  denote the sum in (4.4). By Taylor's formula, for any  $q \geq 1$  we have

$$S_f(I) = \sum_{n \in I} e(F_n(q)) + O\left(q + X q^{R+1} \frac{\max_{t \in I} |f^{(R+1)}(t)|}{(R+1)!}\right).$$

Let  $\mathcal{Q} := \{xy : 1 \leq x \leq V, 1 \leq y \leq V\} \cap \mathbb{N}$ , where  $\mathcal{Q}$  is interpreted as a multiset. Also let  $Q = |\mathcal{Q}| = V^2$ . Then

$$S_f(I) = \sum_{n \in I} |\mathcal{Q}|^{-1} \sum_{q \in \mathcal{Q}} e(F_n(q)) + O\left(Q + X Q^{R+1} \frac{\max_{t \in I} |f^{(R+1)}(t)|}{(R+1)!}\right).$$

We take  $V = X^{1/4}$  in which case by (4.2) the error term is

$$(4.5) \quad \begin{aligned} &\ll X^{1/2} + X \cdot X^{(R+1)/2} \alpha^{(R+1)^3} F / X^{R+1} \\ &\ll X^{1/2} + X^{-(R+1)/4} \alpha^{(R+1)^3} \cdot (F X^{1-(R+1)/4}). \end{aligned}$$

The term in the parenthesis is  $\leq F X^{3/4} F^{-10/4} \leq 1$ . Using also (4.3) we see that (4.5) is  $\ll X^{1/2}$  which is in particular smaller than the right-hand side of (4.4). The sum  $\sum_{q \in \mathcal{Q}} e(F_n(q))$  is precisely the one estimated in [14, pp. 217–225]. The only assumption needed of  $f$  in that argument is (4.2), and the only restriction on  $F$  and  $X$  there is  $F \geq X^4$ . Hence, we conclude that the lemma holds by following the analysis there verbatim.  $\square$

We now apply this estimate to obtain an estimate for an exponential sum over integers.

**Proposition 4.2** (Exponential sums over integers). *Let  $\varepsilon > 0$ ,  $A \geq 1$ ,  $X \geq 2$ ,  $2 \leq j \ll \log^{1/2} X$ , and let  $N, M$  be real numbers with  $N, M \ll \exp(O(\log^{3/2-\varepsilon} X))$ . Let  $I$  be an interval in  $[X, X + X \log^{-100} X]$ . Then*

$$(4.6) \quad \sum_{n \in I} e\left(\frac{N}{n} + \frac{M}{n^j}\right) \ll_{\varepsilon, A} X(1+F)^{-c} \log^{O(A)} X + X \log^{-A} X$$

for some absolute constant  $c > 0$ , where

$$F := \frac{|N|}{X} + \frac{|M|}{X^j}.$$

*Proof.* We may assume without loss of generality that  $A$  is sufficiently large, and  $X$  is sufficiently large depending on  $\varepsilon, A$ . By hypothesis we have  $F \ll \exp(O(\log^{3/2-\varepsilon} X))$ . We may assume that  $F \geq \log^{CA} X$  for a large absolute constant  $C$ , since the claim is trivial otherwise.

Let  $f: I \rightarrow \mathbb{R}$  denote the phase function

$$f(t) := \frac{N}{t} + \frac{M}{t^j}.$$

Then for any  $r \geq 1$  and  $t \in I$  we have

$$(4.7) \quad \frac{t^r}{r!} |f^{(r)}(t)| = \left| \frac{N}{t} + \frac{M_r}{t^j} \right| \asymp X^{-1} |N + M_r/t^{j-1}|$$

where

$$M_r := \binom{r+j-1}{j-1} M.$$

Since

$$1 \leq \binom{r+j-1}{j-1} \leq (r+j)^r = \exp(r \log(r+j)) = \exp(O(r^2 \log_2 X))$$

we conclude that

$$M_r = \exp(O(r^2 \log_2 X)) M$$

and

$$\frac{|N|}{X} + \frac{|M_r|}{X^j} = \exp(O(r^2 \log_2 X)) F.$$

If  $|M_r| \leq |N|X^{j-1}/4$  then from the triangle inequality and (4.1) we have

$$X^{-1} |N + M_r/t^{j-1}| \asymp X^{-1} |N| = \exp(O(r^2 \log_2 X)) F.$$

Consider then the case  $|M_r| > |N|X^{j-1}/4$ . We have the upper bound

$$X^{-1} |N + M_r/t^{j-1}| \ll \frac{|M_r|}{X^j} \ll \exp(O(r^2 \log_2 X)) F$$

for all  $t \in I$  from the triangle inequality. Furthermore, since the function  $t \mapsto -1/t^{j-1}$  has derivative  $\asymp j/X^j$  on  $I$ , we also have, for all  $t$  outside of an interval of length  $O(X \log^{-2A} X)$ , the lower bound

$$X^{-1} |N + M_r/t^{j-1}| \gg \frac{|M_r|}{X^j} \log^{-3A} X \gg \exp(O(r^2 \log_2 X)) F \log^{-3A} X.$$

If we set  $\alpha := \log^{4A} X$  and  $A$  is sufficiently large, then we conclude from (4.7) and the bounds above that the estimate (4.2) holds for all  $1 \leq r \leq \log X$  and all  $t \in I$  outside the union of  $O(\log X)$  intervals of length  $O(X \log^{-2A} X)$ . The contribution of these exceptional intervals to (4.6) is negligible, and removing them splits  $I$  up into at most  $O(\log X)$  subintervals, so by the triangle inequality it suffices to show that

$$\sum_{n \in I'} e \left( \frac{N}{n} + \frac{M}{n^j} \right) \ll_{\varepsilon, A} X \log^{-2A} X$$

for any subinterval  $I'$  with the property that (4.2) holds for all  $t \in I'$  and  $1 \leq r \leq \log X$ . If  $F \geq X^4$ , we may apply Lemma 4.1 to conclude that

$$\sum_{n \in I'} e\left(\frac{N}{n} + \frac{M}{n^j}\right) \ll X \log^{4A} X \exp(-c \log^{2\varepsilon} X)$$

for some absolute constant  $c > 0$ , and the claim follows. If instead  $F < X^4$ , we can apply the Weyl inequality [14, Theorem 8.4] with  $k = 5$  to conclude that

$$\sum_{n \in I'} e\left(\frac{N}{n} + \frac{M}{n^j}\right) \ll \alpha^{O(1)}(F/X^5 + 1/F)^c X \log X$$

for some absolute constant  $c > 0$ ; since  $F \geq \log^{CA} X$ , we obtain the claim by taking  $C$  large enough.  $\square$

Now we prove Proposition 1.13(i). We may assume without loss of generality that  $j \geq 2$ , since for  $j = 1$  we can absorb the  $M$  terms into the  $N$  term (and add a dummy term with  $M = 0$  and  $j = 2$ , say). By summation by parts (see e.g. [21, Lemma 2.2]), and adjusting  $A$  as necessary, it suffices to show that

$$\sum_{p \in I} e\left(\frac{N}{p} + \frac{M}{p^j}\right) \log p = \int_I e\left(\frac{N}{t} + \frac{M}{t^j}\right) dt + O_{\varepsilon, A}(P \log^{-10A} P)$$

for all intervals  $I \subset [P, P + P \log^{-100} P]$ . This is equivalent to

$$\sum_{n \in I} e\left(\frac{N}{n} + \frac{M}{n^j}\right) \Lambda(n) = \int_I e\left(\frac{N}{t} + \frac{M}{t^j}\right) dt + O_{\varepsilon, A}(P \log^{-10A} P),$$

where  $\Lambda$  is the von Mangoldt function, since the contribution of the prime powers is negligible. We introduce the quantity

$$F := \frac{|N|}{P} + \frac{|M|}{P^j}.$$

If  $F \leq \log^{CA} P$  for some large absolute constant  $C > 0$ , then the total variation of the phase  $t \mapsto \frac{N}{t} + \frac{M}{t^j}$  is  $O(\log^{CA} P)$ , and the claim readily follows from a further summation by parts (see e.g. [21, Lemma 2.2]) and the prime number theorem (with classical error term). Thus we may assume that

$$(4.8) \quad F > \log^{CA} P.$$

In this case, a change of variables  $t = P/s$  gives

$$\int_I e\left(\frac{N}{t} + \frac{M}{t^j}\right) dt = -P \int_{P/I} e\left(\frac{N}{P}s + \frac{M}{P^j}s^j\right) \frac{ds}{s^2}.$$

The derivative of the phase here is  $N/P + js^{j-1}M/P^j$  which, once  $C$  is large enough, is  $\geq \log^{10A} P$  for all  $s \in P/I$  apart from an interval of length at most  $O(\log^{-10A} P)$ . Hence by partial integration we get that

$$\int_I e\left(\frac{N}{t} + \frac{M}{t^j}\right) dt \ll P \log^{-10A} P$$

if  $C$  is large enough, so it remains to establish the bound

$$\sum_{n \in I} e\left(\frac{N}{n} + \frac{M}{n^j}\right) \Lambda(n) \ll P \log^{-10A} P$$

under the hypothesis (4.8).

By Vaughan's identity in the form of [14, Proposition 13.4] (with  $y = z = P^{1/3}$ ), followed by a shorter-than-dyadic decomposition, we can write

$$\Lambda(n) = \sum_{r \leq R} (\alpha_r * 1(n) + \alpha'_r * \log(n) + \beta_r * \gamma_r(n))$$

for  $n \in [P, 2P]$ , where  $*$  denotes Dirichlet convolution, and

$$\begin{aligned} R &\ll \log^{O(1)} P, \\ |\alpha_r(n)|, |\alpha'_r(n)|, |\beta_r(n)|, |\gamma_r(n)| &\ll \log P, \\ \text{supp}(\alpha_r), \text{supp}(\alpha'_r) &\subset [M_r, (1 + \log^{-100} P)M_r], \\ \text{supp}(\beta_r) &\subset [K_r, (1 + \log^{-100} P)K_r], \\ \text{supp}(\gamma_r) &\subset [N_r, (1 + \log^{-100} P)N_r], \\ 1 &\leq M_r \ll P^{2/3}, \\ P^{1/3} &\ll K_r, N_r \ll P^{2/3} \end{aligned}$$

(the bound for the coefficients arising from Vaughan's identity is  $\ll \log P$  since  $1 * \Lambda = \log$ ). By the triangle inequality, it thus suffices to establish the Type I estimates

$$(4.9) \quad \sum_{n \in I} e\left(\frac{N}{n} + \frac{M}{n^j}\right) (\alpha_r * 1)(n) \ll_{\varepsilon, A} P \log^{-11A} P$$

and

$$(4.10) \quad \sum_{n \in I} e\left(\frac{N}{n} + \frac{M}{n^j}\right) (\alpha'_r * \log)(n) \ll_{\varepsilon, A} P \log^{-11A+1} P$$

as well as the Type II estimates

$$(4.11) \quad \sum_{n \in I} e\left(\frac{N}{n} + \frac{M}{n^j}\right) (\beta_r * \gamma_r)(n) \ll_{\varepsilon, A} P \log^{-11A} P$$

for all  $1 \leq r \leq R$  and  $I \subset [P, P + P \log^{-100} P]$ . The second Type I estimate (4.10) follows from the first Type I estimate (4.9) (replacing  $\alpha_r$  with  $\alpha'_r$ ) and a summation by parts (see e.g. [21, Lemma 2.2]), so it suffices to establish (4.9) and (4.11).

We begin with (4.9). By the triangle inequality, the left-hand side is bounded by

$$\ll \log P \sum_{m \in [M_r, (1 + \log^{-100} P)M_r]} \left| \sum_{n \in \frac{1}{m} \cdot I} e\left(\frac{N}{mn} + \frac{M}{m^j n^j}\right) \right|.$$

Applying Proposition 4.2 with  $X = P/m$  and  $N/m$  and  $M/m^j$  in place of  $N$  and  $M$ , we can bound this by

$$\ll_{\varepsilon, A} P \log P \left( (1 + F)^c \log^{O(A)} P + \log^{-20A} P \right)$$

for some constant  $c > 0$ , and the claim now follows from (4.8).

Now we establish (4.11). We can assume that  $K_r N_r \asymp P$ , as the sum vanishes otherwise. By the triangle inequality, the left-hand side is bounded by

$$\ll \log P \sum_{m \in [K_r, (1+\log^{-100} P)K_r]} \left| \sum_{n \in \frac{1}{m} \cdot I} \gamma_r(n) e \left( \frac{N}{mn} + \frac{M}{m^j n^j} \right) \right|.$$

By Cauchy–Schwarz it suffices to show that

$$\sum_{m \in [K_r, (1+\log^{-100} P)K_r]} \left| \sum_{n \in \frac{1}{m} \cdot I} \gamma_r(n) e \left( \frac{N}{mn} + \frac{M}{m^j n^j} \right) \right|^2 \ll K_r N_r^2 \log^{-30A} P$$

(say). Rearranging, it suffices to show that

$$(4.12) \quad \sum_{n, n' \in [N_r, (1+\log^{-100} P)N_r]} \gamma_r(n) \overline{\gamma_r(n')} X_{n, n'} \ll K_r N_r^2 \log^{-30A} P$$

where

$$X_{n, n'} := \sum_{m \in [K_r, (1+\log^{-100} P)K_r] \cap \frac{1}{n} \cdot I \cap \frac{1}{n'} \cdot I} e \left( \frac{N(n' - n)}{nn'm} + \frac{M((n')^j - n^j)}{n^j (n')^j m^j} \right).$$

By Proposition 4.2, we have

$$X_{n, n'} \ll_{\varepsilon, A} K_r \left( \left( 1 + \frac{|n' - n|}{N_r} F \right)^{-c} \log^{O(A)} P + \log^{-40A} P \right)$$

for some absolute constant  $0 < c < 1$ . Bounding  $\gamma_r(n) \overline{\gamma_r(n')} \ll \log^2 P$  and noting that

$$\sum_{n \in [N_r, (1+\log^{-100} P)N_r]} \left( 1 + \frac{|n' - n|}{N_r} F \right)^{-c} \ll N_r F^{-c}$$

for all  $n' \in [N_r, (1+\log^{-100} P)N_r]$ , we obtain the claim (4.12) from (4.8). This completes the proof of Proposition 1.13.

## 5. MULTIPLICITY OF THE FALLING FACTORIAL

In this section we establish Theorem 1.9. We first observe that if  $1 \leq m \leq n$  solves (1.9) for some sufficiently large  $t$ , then

$$t = (n)_m \geq (m)_m = m! \gg (m/e)^m$$

by Stirling's formula. Hence we have an analogue of (1.5):

$$(5.1) \quad m \ll \frac{\log t}{\log_2 t}$$

Next, since

$$(n - m)^m < (n)_m \leq n^m$$

we have

$$(5.2) \quad t^{1/m} \leq n < t^{1/m} + m$$

and we obtain an analogue

$$(5.3) \quad n \asymp t^{1/m} = \exp\left(\frac{\log t}{m}\right)$$

of (1.6), (1.7).

Next, we obtain the following analogue of Proposition 1.10.

**Proposition 5.1** (Distance estimate). *Suppose we have two solutions  $(n, m), (n', m')$  to (1.9) in region  $\{(m, n) \in \mathbb{N}^2 : 1 \leq m \leq n\}$ . Then one has*

$$(5.4) \quad m' - m \ll \log(n + n').$$

Furthermore, if

$$(5.5) \quad \exp(\log^{2/3+\varepsilon}(n + n')) \leq m, m' \leq (n + n')^{2/3}$$

for some  $\varepsilon > 0$ , then we additionally have

$$(5.6) \quad n' - n \ll_{A,\varepsilon} \frac{m + m'}{\log^A(m + m')}$$

for any  $A > 0$ .

*Proof.* We begin with (5.4). We follow the arguments from [1, Proof of Theorem 4]. Taking 2-valuations  $v_2$  of both sides of (1.9) and using (1.12) we have

$$\sum_{j=1}^{\infty} \left( \left\lfloor \frac{n}{2^j} \right\rfloor - \left\lfloor \frac{n-m}{2^j} \right\rfloor \right) = \sum_{j=1}^{\infty} \left( \left\lfloor \frac{n'}{2^j} \right\rfloor - \left\lfloor \frac{n'-m'}{2^j} \right\rfloor \right).$$

The summands here vanish unless  $j \leq \log(n + n')$ . Writing  $\lfloor x \rfloor = x + O(1)$ , we conclude that

$$\sum_{1 \leq j \leq \log(n+n')} \frac{m}{2^j} + O(\log(n + n')) = \sum_{1 \leq j \leq \log(n+n')} \frac{m'}{2^j} + O(\log(n + n'))$$

and (5.4) follows.

Now we prove (5.6). Fix  $A, \varepsilon > 0$ . We may assume without loss of generality that  $m' < m$ , so that  $n' > n$  by (1.9). We may also assume  $t$  is sufficiently large depending on  $A, \varepsilon$ , as the claim is trivial otherwise; from (5.5) this also implies that  $m, m', n, n'$  are sufficiently large depending on  $A, \varepsilon$ . Henceforth all implied constants are permitted to depend on  $A, \varepsilon$ . By (5.5) we have

$$\log^{2/3+\varepsilon} n \leq \log m$$

while from (5.3) we have  $\log n \asymp \frac{\log t}{m}$ . From this and (5.1) we have

$$(5.7) \quad m \asymp \frac{\log t}{\log_2 t}$$

and then

$$\log n \ll \log_2^{\frac{1}{2/3+\varepsilon}} t.$$

Similarly for  $m', n'$ . From (5.4) we conclude that

$$(5.8) \quad m - m' \ll \log_2^{\frac{1}{2/3+\varepsilon}} t \ll \log^{\frac{1}{2/3+\varepsilon}} m.$$

In particular  $m \asymp m'$  and, combining (5.3) with (5.8) and (5.7), also  $n \asymp n't^{1/m-1/m'} \asymp n'$ . Hence from (5.5) we see that

$$(5.9) \quad n, n' \gg m^{3/2}.$$

Also we have

$$\frac{\log t}{m'} = \frac{\log t}{m} + O\left(\frac{\log t \log^{\frac{1}{2/3+\varepsilon}} m}{m^2}\right) = \frac{\log t}{m} + O(m^{-1/2})$$

(say), hence on exponentiating and using (5.2), (5.9)

$$(5.10) \quad n' = \exp\left(\frac{\log t}{m'}\right) + O(m) = n + O(m^{-1/2}n).$$

Suppose that we could find a prime  $p > m$  obeying the inequalities

$$(5.11) \quad \max\left(1 - \left\{\frac{n' - n}{p}\right\}, 1 - \frac{m}{p}\right) < \left\{\frac{n - m}{p}\right\} < 1; \quad \left\{\frac{n' - n}{p}\right\} < 1 - \frac{m}{p}.$$

These inequalities imply in particular that

$$\left\{\frac{n - m}{p}\right\} - 1 + \frac{m}{p} \in [0, 1) \text{ and } \left\{\frac{n - m}{p}\right\} + \left\{\frac{n' - n}{p}\right\} - 1 + \frac{m}{p} \in [0, 1),$$

so that these quantities respectively equal  $\{\frac{n}{p}\}$  and  $\{\frac{n'}{p}\}$ . Consequently, if (5.11) hold, then we would have

$$(5.12) \quad \left\{\frac{n}{p}\right\} = \left\{\frac{n - m}{p}\right\} - 1 + \frac{m}{p} < \frac{m}{p}$$

and (since  $m' < m$ )

$$(5.13) \quad \left\{\frac{n'}{p}\right\} = \left\{\frac{n - m}{p}\right\} + \left\{\frac{n' - n}{p}\right\} - 1 + \frac{m}{p} \geq \frac{m}{p} \geq \frac{m'}{p}.$$

Now (5.12) implies that  $p$  divides  $(n)_m$ , while (5.13) implies that  $p$  does not divide  $(n')_{m'}$ . This contradicts the assumption  $(n)_m = t = (n')_{m'}$ . Thus there cannot be any prime  $p \geq 2m$  obeying (5.11).

Let  $w_1: \mathbb{R} \rightarrow [0, 1]$  be a suitable smooth  $\mathbb{Z}$ -periodic function supported on the region

$$\{x \in \mathbb{R} : \{x\} \in (1 - \log^{-2A} m, 1)\}$$

chosen so that  $\int_0^1 w_1 \gg \log^{-2A} m$  and  $\|w_1\|_{C^3} \ll \log^{6A} m$ , and let  $w_2: \mathbb{R} \rightarrow [0, 1]$  similarly be a smooth  $\mathbb{Z}$ -periodic function supported on the region

$$\{y \in \mathbb{R} : \{y\} \in (\log^{-2A} m, 1/2)\}$$

chosen so that  $w_2(y) = 1$  when  $\{y\} \in [2\log^{-2A} m, 1/4]$  and  $\|w_2\|_{C^3} \ll \log^{6A} m$ . Let  $\mathbf{p}$  be a prime drawn uniformly from all the primes in  $[2m, 100m]$ . As  $\mathbf{p}$  does not obey (5.11), we have

$$\mathbf{E} w_1\left(\frac{n - m}{\mathbf{p}}\right) w_2\left(\frac{n' - n}{\mathbf{p}}\right) = 0$$

and hence by Proposition 1.13 (and dyadic decomposition)

$$\int_2^{100} w_1\left(\frac{n - m}{tm}\right) w_2\left(\frac{n' - n}{tm}\right) dt \ll \log^{-100A} m,$$

or on changing variables  $t = 1/s$

$$(5.14) \quad \int_{1/100}^{1/2} w_1 \left( \frac{n-m}{m} s \right) w_2 \left( \frac{n'-n}{m} s \right) ds \ll \log^{-100A} m.$$

On the other hand, by (5.10), (5.9) we have

$$(5.15) \quad \frac{n-m}{m} \gg \frac{n}{m} \gg m^{1/2} \frac{n'-n}{m} + m^{1/2}.$$

We perform a Fourier expansion

$$w_1(x) = \sum_{\ell \in \mathbb{Z}} c_\ell e(\ell x),$$

where by integration by parts the Fourier coefficients obey the bounds

$$|c_\ell| \ll (1 + |\ell|)^{-3} \log^{6A} m.$$

Thus (5.14) can then be rewritten as

$$(5.16) \quad \sum_{\ell \in \mathbb{Z}} c_\ell \int_{1/100}^{1/2} w_2 \left( \frac{n'-n}{m} s \right) e \left( \frac{n-m}{m} \ell s \right) ds \ll \log^{-100A} m.$$

By (5.15) and integration by parts, one readily establishes the bound

$$\int_{1/100}^{1/2} w_2 \left( \frac{n'-n}{m} s \right) e \left( \frac{n-m}{m} \ell s \right) ds \ll \frac{\log^{6A} m}{|\ell| m^{1/2}}$$

for  $\ell \neq 0$ . Thus the total contribution to the left-hand side of (5.16) from the terms with  $\ell \neq 0$  is negligible, and hence

$$c_0 \int_{1/100}^{1/2} w_2 \left( \frac{n'-n}{m} s \right) ds \ll \log^{-100A} m.$$

Since  $c_0 = \int_0^1 w_1 \gg \log^{-2A} m$  and  $w_2$  equals 1 on  $[2 \log^{-2A} m, 1/4]$ , we have

$$(5.17) \quad f \left( \frac{n'-n}{m} \right) \ll \log^{-98A} m$$

where

$$f(\theta) := \int_{1/100}^{1/2} 1_{2 \log^{-2A} m \leq \{\theta s\} \leq 1/4} ds.$$

However, direct calculation shows that when  $\theta \geq 3$ , we have

$$f(\theta) \geq \sum_{\frac{\theta}{16} \leq n \leq \frac{\theta}{2} - \frac{1}{4}} \int_{\mathbb{R}} 1_{n+1/100 \leq \theta s \leq n+1/4} ds \gg \theta \cdot \theta^{-1} = 1,$$

when  $1/2 < \theta < 3$ , we have

$$f(\theta) \geq \int_{\frac{1}{30\theta}}^{\frac{1}{20\theta}} ds \asymp 1,$$

and, when  $8 \log^{-A} m \leq \theta \leq 1/2$ , we have

$$f(\theta) \geq \int_{1/4}^{1/2} ds \asymp 1.$$

Hence (5.17) can only hold if

$$\frac{n' - n}{m} \ll \log^{-A} m,$$

giving the claim (5.6).  $\square$

Now we adapt the analysis from Section 2. We extend the falling factorial  $(n)_m$  to real  $n \geq m \geq 0$  by the formula

$$(n)_m := \frac{\Gamma(n+1)}{\Gamma(n-m+1)}.$$

From the increasing nature of the digamma function  $\psi$  we see that for fixed  $m$ ,  $(n)_m$  increases from  $\Gamma(m+2)$  when  $n$  goes from  $m+1$  to infinity. Applying the inverse function theorem, we conclude that for any sufficiently large  $t$  there is a unique smooth function  $g_t: \{m > 0 : \Gamma(m+2) \leq t\} \rightarrow \mathbb{R}$  such that for any  $m > 0$  with  $\Gamma(m+2) \leq t$ , one has  $g_t(m) \geq m$  and

$$(5.18) \quad (g_t(m))_m = t.$$

Indeed, one could simply set  $g_t(m) := f_{t/\Gamma(m+1)}(m)$ , where  $f_t$  is the function studied in Section 2.

We have an analogue of Proposition 2.1:

**Proposition 5.2** (Estimates on the first few derivatives). *Let  $C > 1$ , and let  $t, m$  be sufficiently large depending on  $C$  with  $\Gamma(m+2) \leq t$ . Then*

$$(5.19) \quad g_t(m) \asymp t^{1/m}.$$

In the range  $m \leq g_t(m)/2$ , we have

$$(5.20) \quad -g'_t(m) \asymp g_t(m) \frac{\log t}{m^2}$$

and in the range  $m \leq g_t(m) - C \log^2 g_t(m)$ , one has

$$(5.21) \quad 0 < g''_t(m) \ll g_t(m) \left( \frac{\log t}{m^2} \right)^2 + C^{-1} \log^{-3} m.$$

*Proof.* Write  $n = g_t(m) \geq m$ . First note that (5.19) is simply (5.3). Taking logarithms in (5.18) we have

$$(5.22) \quad \log \Gamma(g_t(m) + 1) - \log \Gamma(g_t(m) - m + 1) = \log t.$$

If we differentiate (5.22) we obtain

$$(5.23) \quad g'_t(m) \psi(g_t(m) + 1) - (g'_t(m) - 1) \psi(g_t(m) - m + 1) = 0.$$

In particular we obtain the first derivative formula

$$(5.24) \quad g'_t(m) = \frac{-\psi(n - m + 1)}{\psi(n + 1) - \psi(n - m + 1)}.$$

In the regime  $m \leq n/2$  we can then obtain (5.20) from (2.12), (2.1), (5.19).

Differentiating (5.23) again, we conclude

$$g''_t(m) \psi(n + 1) + (g'_t(m))^2 \psi'(n + 1) - g''_t(m) \psi(n - m + 1) - (g'_t(m) - 1)^2 \psi'(n - m + 1) = 0$$

which we can rearrange using (5.24) as

$$(5.25) \quad g''_t(m) (\psi(n + 1) - \psi(n - m + 1))^3 = \psi(n + 1)^2 \psi'(n - m + 1) - \psi(n - m + 1)^2 \psi'(n + 1).$$

Suppose first that  $m \leq n/2$ . Then (2.12) applies, and it suffices to show that

$$\psi(n+1)^2 \psi'(n-m+1) - \psi(n-m+1)^2 \psi'(n+1) \asymp \left(\frac{m}{n}\right)^3 n \left(\frac{\log t}{m^2}\right)^2.$$

By (5.19) the right-hand side is  $\asymp \frac{m \log^2 n}{n^2}$ . On the other hand, from the mean value theorem and (2.1), (2.2), (2.3) we have

$$0 < \psi(n+1)^2 (\psi'(n-m+1) - \psi'(n+1)) \asymp \frac{m \log^2 n}{n^2}$$

and

$$0 < (\psi(n+1)^2 - \psi(n-m+1)^2) \psi'(n+1) \ll \frac{m \log n}{n^2}$$

giving the claim.

Now suppose that  $n/2 \leq m \leq n - C \log^2 n$ . From (2.1), (2.2) we have

$$\begin{aligned} \psi(n+1) - \psi(n-m+1) &\asymp \log \frac{n}{n-m} \\ \psi(n+1)^2 (\psi'(n-m+1) - \psi'(n+1)) &\asymp \frac{\log^2 n}{n-m} \\ 0 < (\psi(n+1)^2 - \psi(n-m+1)^2) \psi'(n+1) &\ll \frac{\log^2 n}{n} \end{aligned}$$

and hence by (5.25)

$$g_t''(m) \asymp \frac{\log^2 n}{(n-m) \log^3 \frac{n}{n-m}}.$$

Since  $\frac{n}{2} \geq n-m \geq C \log^2 n$ , we have

$$(n-m) \log^3 \frac{n}{n-m} \gg C \log^5 n$$

(as can be seen by checking the cases  $n-m \leq \sqrt{n}$  and  $n-m > \sqrt{n}$  separately), and the claim follows.  $\square$

Now we can establish Theorem 1.9. Let  $C > 0$  be a large absolute constant, let  $\varepsilon > 0$ , and suppose that  $t$  is sufficiently large depending on  $\varepsilon, C$ . Let  $(n, m)$  be the integer solution to (1.9) in the region  $\exp(\log^{2/3+\varepsilon} n) \leq m \leq n-1$  with a maximal value of  $m$ ; we may assume that such a solution exists, since we are done otherwise. If  $(n', m')$  is any other solution in this region, then  $m' < m$  and  $n < n'$ . Note that  $n, n', m, m'$  are sufficiently large depending on  $\varepsilon, C$ . From Proposition 5.1 and (5.3) we have

$$m - m' \ll \log n' \ll \frac{\log t}{m'}$$

and from (5.3) and (5.1)

$$m' \asymp \frac{\log t}{\log n'} \geq \frac{\log t}{\log^{2/3+\varepsilon} \frac{1}{m'}} \gg \frac{\log t}{\log_2^{2/3+\varepsilon} t}$$

and thus  $m' \asymp m$  and  $\frac{\log t}{m} = \frac{\log t}{m'} + O(1)$ . Hence  $n \asymp n'$  and

$$(5.26) \quad m - m' \ll \log n.$$

First suppose that  $m \leq n^{1/2} \log^{10} n$ . Here we will exploit the fact that  $n$  grows rapidly as  $m$  decreases. From Proposition 5.1 we have

$$n' - n \ll_{\varepsilon} \frac{m}{\log^{200} m} \ll \frac{m}{\log^{100} n}.$$

On the other hand, from (5.20) and the mean value theorem we have

$$n' - n = g_t(m') - g_t(m) \gg \frac{n \log t}{m^2} (m - m') \geq \frac{n}{m}$$

thanks to (5.1) and the trivial bound  $m - m' \geq 1$ . Thus we have

$$\frac{n}{m} \ll \frac{m}{\log^{100} n}$$

but this contradicts the hypothesis  $m \leq n^{1/2} \log^{10} n$ .

Now suppose we are in the regime

$$n^{1/2} \log^{10} n < m \leq n - C \log^2 n.$$

Here we will take advantage of the convexity properties of  $g_t$ . From (5.26),  $m'$  lies in the interval  $[m - O(\log n), m]$ . By (5.19), for all  $x$  in this interval, we have

$$g_t(x) \asymp t^{1/x} \asymp t^{1/m} \asymp n$$

and by (5.21), we have

$$\begin{aligned} 0 < g_t''(x) &\ll g_t(x) \left( \frac{\log t}{x^2} \right)^2 + C^{-1} \log^{-3} x \\ &\ll n \left( \frac{\log t}{m^2} \right)^2 + C^{-1} \log^{-3} m \\ &\ll n \left( \frac{\log n}{m} \right)^2 + C^{-1} \log^{-3} n \\ &\ll C^{-1} \log^{-3} n \end{aligned}$$

since  $m > n^{1/2} \log^{10} n$ . Applying Lemma 2.2 with  $k = 2$ , we see (for  $C$  large enough) that there are at most two integers  $m'$  in this interval with  $g_t(m')$  an integer, giving Theorem 1.9 follows in this case.

It remains to handle the case

$$(5.27) \quad n - C \log^2 n < m \leq n - 1.$$

Recall from (5.26) that  $m'$  lies in the interval  $[m - O(\log n), m]$ . From (5.3), (5.27) we have

$$m \asymp n \asymp \frac{\log t}{\log_2 t}$$

so  $m' = m - O(\log_2 t)$ . From (5.3) again we thus also have

$$m' \asymp n' \asymp \frac{\log t}{\log_2 t}.$$

From (1.9) we have

$$\frac{n'}{n' - m'} \frac{n' - 1}{n' - 1 - m'} \cdots \frac{n + 1}{n + 1 - m'} = (n - m') \cdots (n - m + 1).$$

The right-hand side is at most  $\exp(O(\log_2 t \log_3 t))$ . This implies that  $n' - n \ll \log_3 t$ , since otherwise the left hand side would be, for any  $C \geq 1$ ,

$$\gg \left( \frac{n}{n - m' + 1 + C \log_3 t} \right)^{C \log_3 t} \gg \exp \left( \frac{C}{2} \log_3 t \log_2 t \right)$$

which contradicts the bound for the right hand side when  $C$  is sufficiently large.

In particular we have from the triangle inequality that

$$n - m, n' - m' \ll C \log_2^2 t.$$

Making the change of variables  $\ell := n - m$ , it now suffices to show that there are at most two integer solutions to the equation

$$(5.28) \quad (n)_{n-\ell} = t$$

in the regime  $1 \leq \ell \ll C \log_2^2 t$ . We write this equation (5.28) as

$$n! = t\ell!$$

or equivalently

$$n = h_t(\ell)$$

where  $h_t(x) := \Gamma^{-1}(t\Gamma(x+1)) - 1$ , and  $\Gamma^{-1}: [1, +\infty) \rightarrow [2, +\infty)$  is the inverse of the gamma function. Here we will exploit the very slowly varying nature of  $h_t$ . From Stirling's formula we have

$$h_t(x) \asymp \frac{\log t}{\log_2 t}$$

whenever  $1 \leq x \ll C \log_2^2 t$ . Taking the logarithmic derivative of the equation

$$\Gamma(h_t(x) + 1) = t\Gamma(x + 1)$$

we have

$$h'_t(x)\psi(h_t(x) + 1) = \psi(x + 1).$$

Hence by (2.1)

$$h'_t(x) \asymp \frac{\log x}{\log h_t(x)} \ll \frac{\log_3 t}{\log_2 t}$$

in the regime  $1 \leq x \ll C \log_2^2 t$ . In particular, for two solutions  $(n, \ell), (n', \ell')$  to (5.28) in this regime we have

$$(5.29) \quad n - n' \ll \frac{\log_3 t}{\log_2 t} |\ell - \ell'|.$$

For fixed  $n$  there is at most one  $\ell \geq 1$  solving (5.28). We conclude that for two distinct solutions  $(n, \ell), (n', \ell')$  to (5.28) in this regime, we have  $|n - n'| \geq 1$ , and hence the separation

$$|\ell - \ell'| \gg \frac{\log_2 t}{\log_3 t}.$$

Now suppose we have three solutions  $(n_1, \ell_1), (n_2, \ell_2), (n_3, \ell_3)$  to (5.28) in this regime. We can order  $\ell_1 < \ell_2 < \ell_3$ , so that  $n_1 < n_2 < n_3$ . From the preceding discussion we have

$$\frac{\log_2 t}{\log_3 t} \ll \ell_2 - \ell_1, \ell_3 - \ell_2 \ll C \log_2^2 t$$

and

$$1 \leq n_2 - n_1, n_3 - n_2 \ll C \log_2 t \log_3 t.$$

If  $2^j$  is a power of 2 that divides an integer in  $(n_1, n_2]$  as well as an integer in  $(n_2, n_3]$ , then we must therefore have  $2^j \ll C \log_2 t \log_3 t$ , so that  $j \ll \log_3 t$ . Thus, there must exist  $i = 1, 2$  such that the interval  $(n_i, n_{i+1}]$  only contains multiples of  $2^j$  when  $j \ll \log_3 t$ . Fix this  $i$ . Taking 2-adic valuations of (5.28) using (1.12) we have

$$\sum_{j=1}^{\infty} \left\lfloor \frac{n_i}{2^j} \right\rfloor = v_2(t) + \sum_{j=1}^{\infty} \left\lfloor \frac{\ell_i}{2^j} \right\rfloor$$

and

$$\sum_{j=1}^{\infty} \left\lfloor \frac{n_{i+1}}{2^j} \right\rfloor = v_2(t) + \sum_{j=1}^{\infty} \left\lfloor \frac{\ell_{i+1}}{2^j} \right\rfloor$$

and thus

$$(5.30) \quad \sum_{j=1}^{\infty} \left( \left\lfloor \frac{n_{i+1}}{2^j} \right\rfloor - \left\lfloor \frac{n_i}{2^j} \right\rfloor \right) = \sum_{j=1}^{\infty} \left( \left\lfloor \frac{\ell_{i+1}}{2^j} \right\rfloor - \left\lfloor \frac{\ell_i}{2^j} \right\rfloor \right).$$

Since

$$(5.31) \quad \ell_{i+1} - \ell_i \gg \frac{\log_2 t}{\log_3 t},$$

we certainly have  $\ell_{i+1} - \ell_i \geq 2$ , and the right-hand side of (5.30) is at least

$$\left\lfloor \frac{\ell_{i+1}}{2} \right\rfloor - \left\lfloor \frac{\ell_i}{2} \right\rfloor \gg \ell_{i+1} - \ell_i.$$

By construction, the terms on the left-hand side of (5.30) vanish unless  $j \ll \log_3 t$ , in which case they are equal to  $\frac{n_{i+1} - n_i}{2^j} + O(1)$ . Thus the left-hand side of (5.30) is at most  $O(n_{i+1} - n_i + \log_3 t)$ . Thus

$$\ell_{i+1} - \ell_i \ll n_{i+1} - n_i + \log_3 t.$$

But from (5.29) one has  $n_{i+1} - n_i \ll \frac{\log_3 t}{\log_2 t} (\ell_{i+1} - \ell_i)$ . Hence  $\ell_{i+1} - \ell_i \ll \log_3 t$ . But this contradicts (5.31). This concludes the proof of Theorem 1.9.

## REFERENCES

- [1] H. L. Abbott, P. Erdős, and D. Hanson. On the number of times an integer occurs as a binomial coefficient. *Amer. Math. Monthly*, 81:256–261, 1974.
- [2] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York City, ninth Dover printing, tenth GPO printing edition, 1964.
- [3] È. T. Avanesov. Solution of a problem on figurate numbers. *Acta Arith.*, 12:409–420, 1966/67.
- [4] F. Beukers, T. N. Shorey, and R. Tijdeman. Irreducibility of polynomials and arithmetic progressions with equal products of terms. In *Number theory in progress, Vol. 1 (Zakopane-Kościelisko, 1997)*, pages 11–26. de Gruyter, Berlin, 1999.
- [5] Aart Blokhuis, Andries Brouwer, and Benne de Weger. Binomial collisions and near collisions. *Integers*, 17:Paper No. A64, 8, 2017.
- [6] J. W. Bober. Factorial ratios, hypergeometric series, and a family of step functions. *J. Lond. Math. Soc. (2)*, 79(2):422–444, 2009.
- [7] J. William Bober. *Integer ratios of factorials, hypergeometric functions, and related step functions*. ProQuest LLC, Ann Arbor, MI, 2009. Thesis (Ph.D.)—University of Michigan.
- [8] E. Bombieri and J. Pila. The number of integral points on arcs and ovals. *Duke Math. J.*, 59(2):337–357, 1989.
- [9] Yann Bugeaud, Maurice Mignotte, Samir Siksek, Michael Stoll, and Szabolcs Tengely. Integral points on hyperelliptic curves. *Algebra Number Theory*, 2(8):859–885, 2008.

- [10] Harald Cramér. On the order of magnitude of the difference between consecutive prime numbers. *Acta Arith.*, 2:23–46, 1936.
- [11] Jean-Marie de Koninck, Nicolas Doyon, and William Verreault. Repetitions of multinomial coefficients and a generalization of Singmaster’s conjecture. *Integers*, 21:Paper No. A34, 11, 2021.
- [12] B. M. M. de Weger. A binomial Diophantine equation. *Quart. J. Math. Oxford Ser. (2)*, 47(186):221–231, 1996.
- [13] Benjamin M. M. de Weger. Equal binomial coefficients: some elementary considerations. *J. Number Theory*, 63(2):373–386, 1997.
- [14] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.
- [15] H. Jenkins. Repeated binomial coefficients and high-degree curves. *Integers*, 16:Paper No. A69, 14, 2016.
- [16] Daniel Kane. New bounds on the number of representations of  $T$  as a binomial coefficient. *Integers*, 4:A7, 10, 2004.
- [17] Daniel M. Kane. Improved bounds on the number of ways of expressing  $t$  as a binomial coefficient. *Integers*, 7:A53, 7, 2007.
- [18] P. Kiss. On the number of solutions of the Diophantine equation  $\binom{x}{p} = \binom{y}{2}$ . *Fibonacci Quart.*, 26(2):127–130, 1988.
- [19] Edmund Landau. *Collected works. Vol. 1*. Thales-Verlag, Essen, 1985. With a contribution in English by G. H. Hardy and H. Heilbronn, Edited and with a preface in English by L. Mirsky, I. J. Schoenberg, W. Schwarz and H. Wefelscheid.
- [20] D. A. Lind. The quadratic field  $Q(\sqrt{5})$  and a certain Diophantine equation. *Fibonacci Quart.*, 6(3):86–93, 1968.
- [21] Kaisa Matomäki, Maksym Radziwiłł, and Terence Tao. Correlations of the von Mangoldt and higher divisor functions I. Long shift ranges. *Proc. Lond. Math. Soc. (3)*, 118(2):284–350, 2019.
- [22] L. J. Mordell. On the integer solutions of  $y(y + 1) = x(x + 1)(x + 2)$ . *Pacific J. Math.*, 13:1347–1351, 1963.
- [23] Ákos Pintér. A note on the Diophantine equation  $\binom{x}{4} = \binom{y}{2}$ . *Publ. Math. Debrecen*, 47(3-4):411–415, 1995.
- [24] David Singmaster. Research Problems: How Often Does an Integer Occur as a Binomial Coefficient? *Amer. Math. Monthly*, 78(4):385–386, 1971.
- [25] David Singmaster. Repeated binomial coefficients and Fibonacci numbers. *Fibonacci Quart.*, 13(4):295–298, 1975.
- [26] K. Soundararajan. Integral Factorial Ratios. *arXiv e-prints*, page arXiv:1901.05133, January 2019.
- [27] K. Soundararajan. Integral factorial ratios: irreducible examples with height larger than 1. *Philos. Trans. Roy. Soc. A*, 378(2163):20180444, 13, 2020.
- [28] Roelof J. Stroeker and Benjamin M. M. de Weger. Elliptic binomial Diophantine equations. *Math. Comp.*, 68(227):1257–1281, 1999.
- [29] Craig A. Tovey. Multiple occurrences of binomial coefficients. *Fibonacci Quart.*, 23(4):356–358, 1985.

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF TURKU, 20014 TURKU, FINLAND  
*Email address:* ksmato@utu.fi

DEPARTMENT OF MATHEMATICS, CALTECH, 1200 E CALIFORNIA BLVD, PASADENA, CA, 91125, USA  
*Email address:* maksym.radziwill@gmail.com

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF KENTUCKY, 715 PATTERSON OFFICE TOWER, LEXINGTON, KY 40506, USA  
*Email address:* xuancheng.shao@uky.edu

DEPARTMENT OF MATHEMATICS, UCLA, 405 HILGARD AVE, LOS ANGELES CA 90095, USA  
*Email address:* tao@math.ucla.edu

MATHEMATICAL INSTITUTE, UNIVERSITY OF OXFORD, WOODSTOCK ROAD, OXFORD OX2 6GG, UNITED KINGDOM  
*Email address:* joni.teravainen@maths.ox.ac.uk