

## ON POLYNOMIAL-FACTORIAL DIOPHANTINE EQUATIONS

DANIEL BEREND AND JØRGEN E. HARMSE

ABSTRACT. We study equations of the form  $P(x) = n!$  and show that for some classes of polynomials  $P$  the equation has only finitely many solutions. This is the case, say, if  $P$  is irreducible (of degree greater than 1) or has an irreducible factor of “relatively large” degree. This is also the case if the factorization of  $P$  contains some “large” power(s) of irreducible(s). For example, we can show that the equation  $x^r(x+1) = n!$  has only finitely many solutions for  $r \geq 4$ , but not that this is the case for  $1 \leq r \leq 3$  (although it undoubtedly should be). We also study the equation  $P(x) = H_n$ , where  $(H_n)$  is one of several other “highly divisible” sequences, proving again that for various classes of polynomials these equations have only finitely many solutions.

### 1. INTRODUCTION

The diophantine equation

$$(1.1) \quad x^2 - 1 = n!$$

seems to have been first posed by Brocard [Br1], [Br2], and it is believed that the only solutions are  $x = 5$ ,  $x = 11$  and  $x = 71$  (with  $n = 4$ ,  $n = 5$  and  $n = 7$ , respectively). In [EO], Erdős and Obláth considered, more generally, the diophantine equations  $x^k \pm y^k = n!$  and  $m! \pm n! = x^k$ . They were able to show that most of these have only finitely many solutions. In particular, the equation  $x^k - 1 = n!$  has no solutions for  $k > 1$ , except possibly for  $k = 2$  and  $k = 4$ . In the case  $k = 4$ , namely for the equation  $x^4 - 1 = n!$ , they showed that the number of solutions is finite, and Pollack and Shapiro [PolS] eventually proved that there are actually no solutions. However, the seemingly simplest case, of equation (1.1), remained open. Further evidence that (1.1) has only finitely many solutions was provided by Overholt [O], who showed that this would follow from the weak form of Szpiro’s conjecture (which is itself a special case of the well-known ABC conjecture). (See Dickson [Di, pp. 680-682] and Guy [Gu1, Sec. D25] for more details and [BernG] for computational results.) A related equation,

$$(1.2) \quad x(x+1) = n!,$$

was posed by Erdős at one of the Western Number Theory Conferences. He opined that it is no doubt true that (1.2) has only finitely many solutions, but that the problem is “quite hopeless in our lifetime or perhaps the lifetime of this miserable universe”. It was noted by Spiro, solving a simpler version of the problem, that

---

Received by the editors July 10, 2002 and, in revised form, July 9, 2004.

2000 *Mathematics Subject Classification*. Primary 11D99; Secondary 11B65.

The first author’s research was supported in part by the Israel Science Foundation (Grant #186/01).

©2005 American Mathematical Society  
Reverts to public domain 28 years from publication

there are infinitely many positive integers  $n$  for which the number  $x$  defined by (1.2) is not an integer (see [Gu2, Problems 301–305]). Erdős then asked [Gu3, Problem 302] to show that the set of integers  $n$  yielding an integer  $x$  in (1.2) is of 0 density, and this was accomplished in [BereO]. More generally, for any polynomial  $P$  of degree 2 or more with integer coefficients, the equation

$$(1.3) \quad P(x) = n!$$

has only a density 0 set of solutions  $n$ . Luca [Luca] has recently shown that the ABC conjecture implies that (1.3) has only finitely many solutions.

Of course, on probabilistic grounds (namely, the easy part of the Borel-Cantelli Lemma) one would actually expect (1.3) to have only finitely many solutions, but that does not follow from the machinery used in [BereO]. The starting point for this paper is a study of this equation with the intent of proving the stronger result, the finiteness of the number of solutions. (This objective is only partly accomplished.)

Our approach relies heavily on the fact that the numbers  $n!$  are “highly” divisible by many primes; moreover, any integer eventually divides all of them. It is thus natural to consider the equation resulting from (1.3) upon replacing  $n!$  by other sequences of such numbers. The equation (1.3) is replaced by the more general equation

$$(1.4) \quad P(x) = H_n,$$

where  $(H_n)$  is a highly divisible sequence. In addition to the choice

$$(1.5.a) \quad H_n = n!,$$

which gives (1.3), two other sequences immediately come to mind,

$$(1.5.b) \quad H_n = [1, 2, \dots, n],$$

(where  $[1, 2, \dots, n]$  is the least common multiple of all positive integers  $\leq n$ ), and

$$(1.5.c) \quad H_n = p_1 p_2 \dots p_n,$$

(where  $p_1 < p_2 < \dots$  is the sequence of all primes). With respect to the third sequence, it was noted by Lucas [Lucas2], [Lucas3, p. 351] that the equation  $x^k \pm y^k = p_1 p_2 \dots p_n$  has no solutions for  $k \geq 2$ ,  $n \geq 3$  (see also [Ba, pp. 44–46]). In a somewhat lighter vein, Nelson, Penney and Pomerance [NPP] considered the equation  $x(x+1) = p_1 p_2 \dots p_n$  and conjectured that it has only the solutions  $x = 1$ ,  $x = 2$ ,  $x = 5$ ,  $x = 14$  and  $x = 714$  (with  $n = 1$ ,  $n = 2$ ,  $n = 3$ ,  $n = 4$  and  $n = 7$ , respectively).

All three sequences (1.5.a)–(1.5.c) are *divisor sequences*, i.e., each term divides its successor. Another sequence, also consisting of highly divisible numbers, is that of the binomial coefficients  $\binom{2n}{n}$ , or, more generally, of multinomial coefficients

$$(1.5.d) \quad H_n = \binom{an}{n, n, \dots, n} = \frac{(an)!}{(n!)^a}$$

for a fixed integer  $a \geq 2$ . This is not a divisor sequence. While every fixed positive integer divides most elements of the sequence, every prime  $p > a$  is relatively prime to infinitely many of them.

Most of the paper is devoted to the case of polynomials of degree 2 or more. In Section 2, however, we study the case of linear polynomials, which turns out (for  $H_n$  as in (1.5.c) or (1.5.d)) to be non-trivial. In the other sections, all dealing only with  $\deg P \geq 2$ , we employ several techniques to deal with (1.4). In Section 3 we start

with the immediate observation that, for (1.4) to have infinitely many solutions,  $P(x)$  must be highly divisible for appropriate choices of  $x$ . This already proves the finiteness of the set of solutions for several classes of polynomials. In particular, this is the case for “most” polynomials (see Remarks 3.1 and 4.2 *infra*). In Section 4 we note that, in some cases, a (large) solution of (1.4) requires some factor of  $P$  to account for a smaller portion of the size of  $H_n$  than it should according to its degree, which gives an upper bound on the size of the solutions. Section 5 is devoted to polynomials with multiple factors. The results here depend on estimates regarding the number of primes in short intervals, and we indicate what we obtain with known results on this question and how improved estimates can strengthen our results. In Section 6 we combine the ideas of the two preceding sections, showing that in some cases we can take advantage of the appearance of some multiple irreducible divisors of  $P$ . Most of the proofs, especially the lengthier and more technical, are presented at the end of the paper; Section 7 deals mostly with auxiliary results, and in Section 8 we present all remaining proofs. Two appendices contain approximate values of some constants and a summary of notation.

We are grateful to J.-P. Allouche, I. Efrat, P. Erdős, H. Furstenberg, R. Heath-Brown, D. Hensley, A. Ivić, J. Lagarias and the referee of this paper for helpful discussions on this problem and important information on related questions, to L. Sapir for verifying by Maple the calculations in the proof of Proposition 7.4, and to Y. Caro for his many comments on an early draft of the paper.

## 2. LINEAR POLYNOMIALS

In this section we study (1.4) for the various sequences  $(H_n)$  in the case  $P$  is a linear polynomial,  $P = rX + s$  ( $r \neq 0$ ). Note that in the sequel we shall use  $X$  when we have a polynomial  $P(X)$  and  $x$  to denote an integer variable, supposed to satisfy an equation of the form  $P(x) = H_n$ .

Before stating our main results we recall a conjecture of Hardy and Littlewood [HaL].

**Prime  $k$ -Tuple Conjecture.** *Let  $a_1, a_2, \dots, a_k$  be integers which, for every prime  $p$ , do not represent all  $p$  congruence classes modulo  $p$ . Then there exist infinitely many positive integers  $n$  for which all the numbers  $n + a_1, n + a_2, \dots, n + a_k$  are prime.*

When we subsequently refer to this conjecture, we shall mean its assertion for all integers  $k$  simultaneously.

The *density* of a set  $A \subseteq \mathbf{N}$  is given by  $\lim_{N \rightarrow \infty} \frac{\#(A \cap [1, N])}{N}$  (where  $\#(B)$  denotes the cardinality of a finite set  $B$ ) if the limit exists.

For a prime  $p$  and a positive integer  $m$ , denote by  $\nu_p(m)$  the maximal  $k$  for which  $p^k | m$ . By  $\mathbf{P}$  we shall denote the set of all primes.

**Theorem 2.1.** *Let  $r \neq 0$  and  $s$  be integers. Then:*

- a) *The equation  $rx + s = n!$  has only finitely many solutions unless  $r | s$ , in which case every sufficiently large  $n$  yields a solution.*
- b) *The equation  $rx + s = [1, \dots, n]$  has only finitely many solutions unless  $r | s$ , in which case every sufficiently large  $n$  yields a solution.*
- c) *Consider the equation  $rx + s = p_1 \dots p_n$ . Let  $d = (r, s)$  and  $r = dr'$ ,  $s = ds'$ .*

Then:

1. The following conditions are necessary for the equation to have infinitely many solutions:

i. The number  $d$  is square-free.

ii. All prime divisors of  $r'$  are divisors of  $d$  as well.

2. If conditions 1.i-1.ii hold, then out of the  $\phi(r')$  possible choices for  $s'$  (modulo  $r'$ ), for at least  $\frac{1+\sqrt{4\phi(r')-3}}{2}$  the equation has infinitely many solutions. In particular, for  $r' = 1, 2, 3, 4, 6$ , conditions 1.i-1.ii are also sufficient for the equation to have infinitely many solutions (for any  $s'$ ).

3. Under the prime  $k$ -tuple conjecture, conditions 1.i-1.ii above are sufficient for the equation to have infinitely many solutions.

d) Consider the equation  $rx + s = \binom{an}{n, n, \dots, n}$ .

1. If  $r|s$ , then this equation has a density 1 set of solutions  $n$ . Otherwise, the set of solutions is of density 0.

2. Let  $F = \{p \in \mathbf{P} : \nu_p(r) > \nu_p(s)\}$ . If  $F$  contains two primes not exceeding  $a$ , then the equation has only finitely many solutions.

The non-trivial parts of the theorem are c) and d). While the main body of the proof will be presented only in Section 8, we state at this point several conjectures and results, relevant to both the proof and various refinements of the theorem. The following two conjectures are related to the prime  $k$ -tuple conjecture.

**Prime-Composite  $(k, l)$ -Tuple Conjecture.** Let  $a_1, a_2, \dots, a_k$  be as in the prime  $k$ -tuple conjecture and let  $b_1, b_2, \dots, b_l$  be any integers, such that  $a_i \neq b_j$  for every  $i$  and  $j$ . Then there exist infinitely many positive integers  $n$  for which all the numbers  $n + a_1, n + a_2, \dots, n + a_k$  are prime while all the numbers  $n + b_1, n + b_2, \dots, n + b_l$  are composite.

Again the conjecture is universal in the variables mentioned in its name.

**Conjecture 2.1.** Let  $M$  and  $e$  be any positive integers and let  $m_1, m_2, \dots, m_e$  be integers relatively prime to  $M$ . Then there exist infinitely many numbers  $n$  for which

$$(2.1) \quad p_{n+i} \equiv m_i \pmod{M}, \quad i = 1, 2, \dots, e.$$

The last conjecture is of course a multi-dimensional analogue of Dirichlet's Theorem on primes in arithmetic progressions. One may sharpen the conjecture, stating that the density of the set of those numbers  $n$  satisfying (2.1) is  $\frac{1}{\phi(M)^e}$ .

**Proposition 2.1.** The prime  $k$ -tuple conjecture implies the prime-composite  $(k, l)$ -tuple conjecture.

**Proposition 2.2.** The prime-composite  $(k, l)$ -tuple conjecture implies Conjecture 2.1.

Thus, taking into account Erdős's opinion [E2] regarding the validity of the prime  $k$ -tuple conjecture (and even of a strengthened version thereof; see also [BatH], for example) that "it is of course clear to every 'right thinking person' that this conjecture must be true", we may be on safe ground with the prime-composite  $(k, l)$ -tuple conjecture and Conjecture 2.1.

**Theorem 2.2.** Under Conjecture 2.1, conditions c)1.i.-c)1.ii. in Theorem 2.1 are sufficient for the equation  $rx + s = p_1 \dots p_n$  to have infinitely many solutions.

Now we discuss another conjecture, related to part d) of Theorem 2.1.

**Conjecture 2.2.** Consider the equation

$$(2.2) \quad rx + s = \binom{an}{n, n, \dots, n},$$

where  $r, s$  and  $a$  are non-zero integers,  $a \geq 2$ . Let  $F = \{p \in \mathbf{P} : \nu_p(r) > \nu_p(s)\}$ , as in Theorem 2.1d)2.

1. If

$$\sum_{p \in F} \left(1 - \log_p \left\lceil \frac{p}{a} \right\rceil\right) > 1,$$

then (2.2) has only finitely many solutions.

2. If

$$\sum_{p \in F} \left(1 - \log_p \left\lceil \frac{p}{a} \right\rceil\right) < 1,$$

then (2.2) has infinitely many solutions.

**Example 2.1.** In the case

$$(2.3) \quad \sum_{p \in F} \left(1 - \log_p \left\lceil \frac{p}{a} \right\rceil\right) = 1,$$

more information is needed to decide whether (2.2) has finitely many or infinitely many solutions. In fact, for both of the equations

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

and

$$4x + 2 = \binom{2n}{n},$$

one has equality in (2.3), yet the first is easily seen to have no solutions, while  $n = 2^m$  yields a solution of the second for every  $m$ .

We note that it is not clear whether there are less trivial instances where (2.3) holds, and we raise the following

**Question.** Can (2.3) hold with  $\#(F) \geq 2$ ?

A similar question may be asked with respect to the borderline case in Conjecture 2.3 *infra*. One is tempted to conjecture that the answer is negative, but it seems that current techniques in transcendental number theory are not strong enough to yield this result.

The value of a binomial coefficient  $\binom{n}{m}$  modulo a prime  $p$  is easily calculated in terms of the base  $p$  expansions of  $n$  and  $m$ . It was observed by Lucas [Lucas1] that, writing

$$n = \sum_{i=1}^k n_i p^i, \quad m = \sum_{i=1}^k m_i p^i \quad (0 \leq n_i, m_i < p),$$

we have  $p \mid \binom{n}{m}$  unless  $m_i \leq n_i$  for each  $i$ , in which case

$$(2.4) \quad \binom{n}{m} \equiv \prod_{i=1}^k \binom{n_i}{m_i} \pmod{p}.$$

Thus it is no surprise that Conjecture 2.2 is closely related to certain conjectures about the expansions of integers in various bases. As these conjectures are of independent interest, we formulate them here. (Heuristic arguments for their validity will be provided in Section 7.) For non-negative integers  $n, C$  and  $b \geq 2$ , denote by  $d_{b,C}(n)$  the number of distinct digits appearing in the base  $b$  expansion of  $n$  more than  $C$  times.

**Conjecture 2.3.** *Let  $b_1, b_2, \dots, b_g \geq 2$  be pairwise prime integers, and let  $C \geq 0$  be an arbitrary constant. Then:*

1.  $\sum_{j=1}^g \left(1 - \log_{b_j} d_{b_j, C}(n)\right) \leq 1$  for all sufficiently large integer  $n$ .
2. Suppose  $B_j \subseteq \{0, 1, \dots, b_j - 1\}$  for  $1 \leq j \leq g$ . If

$$(2.5) \quad \sum_{j=1}^g \left(1 - \log_{b_j} \#(B_j)\right) < 1,$$

*then there exist infinitely many positive integers  $n$  such that for each  $1 \leq j \leq g$  the base  $b_j$  expansion of  $n$  contains at most  $C$  occurrences of digits not belonging to  $B_j$  (and arbitrarily many occurrences of digits belonging to  $B_j$ ).*

Conjecture 2.3 is closely related to a conjecture of Furstenberg [FU, Conj. 2'].

*Remark 2.1.* It is conceivable that the first part of Conjecture 2.3 holds even if  $b_1, b_2, \dots, b_g$  are required to be only pairwise multiplicatively independent (i.e.,  $\log b_i / \log b_j$  is irrational for  $i \neq j$ ). However, in the second part one cannot relax the condition in this way, as the example  $b_1 = 4, b_2 = 6, B_1 = \{0, 2\}, B_2 = \{1, 3, 5\}$  demonstrates. It is possible to modify the conjecture so as to relate to arbitrary pairwise multiplicatively independent  $b_1, b_2, \dots, b_g$ . We shall refrain from doing that, as our interest here in the conjecture is restricted to the case where  $b_1, b_2, \dots, b_g$  are all prime.

**Example 2.2.** Let  $g = 2$ . If  $\log_{b_1} \#(B_1) + \log_{b_2} \#(B_2) > 1$ , then according to Conjecture 2.3.2 there should exist infinitely many positive integers  $n$  whose base  $b_1$  expansion consists only of digits in  $B_1$  and whose base  $b_2$  expansion consists only of digits in  $B_2$ . Erdős *et al.* [EGRS] proved that this is the case if  $B_1 = \{0, 1, \dots, c_1\}$  and  $B_2 = \{0, 1, \dots, c_2\}$ , where  $c_1$  and  $c_2$  satisfy the condition  $\frac{c_1}{b_1-1} + \frac{c_2}{b_2-1} \geq 1$ . For example, Conjecture 2.3.2 would imply the result if  $c_1 = \lfloor \sqrt{b_1} \rfloor$  and  $c_2 = \lfloor \sqrt{b_2} \rfloor$ , while [EGRS] would imply the same only for the much larger  $c_1 = \lfloor \frac{b_1-1}{2} \rfloor$ ,  $c_2 = \lfloor \frac{b_2-1}{2} \rfloor$ .

**Proposition 2.3.** *Conjecture 2.3.1 implies Conjecture 2.2.1.*

Let us state here a few more results related to Theorem 2.1d). It seems that the special case  $a = 2$  of our general sequence, namely the sequence  $\binom{2n}{n}$ , attracted the most attention, and we shall confine ourselves almost solely to this case. For a prime power  $r$  we have

**Proposition 2.4.**

1. For every odd prime  $p$  and positive integer  $l$ , the equation

$$p^l x + s = \binom{2n}{n}$$

*has infinitely many solutions for every  $s$ .*

2. For  $l \geq 2$  and  $0 \leq s \leq 2^l - 1$ , the equation

$$2^l x + s = \binom{2n}{n}$$

has infinitely many solutions for  $s = 0$  and for exactly one of the choices  $s = 2, 6, 10, \dots, 2^l - 2$ . The equation has no solutions for  $s$  odd (except  $n = 0$ ). One can effectively determine, given any  $s = 4, 8, 12, \dots, 2^l - 4$ , whether or not the equation has infinitely many solutions. In either case, one can effectively characterize the solutions.

**Example 2.3.** One might expect the first part of the proposition to remain valid if  $\binom{2n}{n}$  is replaced by  $\binom{an}{n, n, \dots, n}$  as long as  $p > a$ . However, one easily verifies (employing (2.4) and the fact that  $\binom{3n}{n, n, n} = \binom{3n}{n} \binom{2n}{n}$ ) that, say, the equations

$$5x + s = \binom{3n}{n, n, n}$$

and

$$7x + s = \binom{3n}{n, n, n}$$

have solutions only for  $s \equiv 0, 1 \pmod{5}$  and  $s \equiv 0, \pm 1 \pmod{7}$ , respectively. It is quite possible, though, that the result is true for all sufficiently large  $p$ .

The result of Erdős *et al.* [EGRS] mentioned in Example 2.2 served them to prove that the sequence  $\binom{2n}{n}$  assumes infinitely often values which are relatively prime to 15, or, more generally, relatively prime to  $pq$ , where  $p$  and  $q$  are distinct odd primes. In terms of the problem at hand, this yields

**Proposition 2.5** ([EGRS]). *Let  $p$  and  $q$  be distinct odd primes. Then the equation*

$$pqx + s = \binom{2n}{n}$$

*has infinitely many solutions for at least one  $s$  relatively prime to  $pq$ .*

Let us finally mention that Conjecture 2.2 seems more difficult when  $r$  is divisible by three or more primes. In fact, Graham ([Gu1, §B23], [Gr]) offers \$1000 for a solution of the following

**Question.** Is  $\left(\binom{2n}{n}, 105\right) = 1$  for infinitely many positive integers  $n$ ?

Note that, as

$$\log_3 2 + \log_5 3 + \log_7 4 \approx 0.631 + 0.683 + 0.712 = 2.026 > 2,$$

Conjecture 2.2 would imply that the answer to the last question is affirmative (and the same should be true if the number 105 is replaced by any product of three distinct odd primes).

### 3. INTRODUCTORY IDEAS – NO SOLUTIONS MODULO SOME INTEGER

One of the simplest ways of showing that a diophantine equation has no solutions is to find an integer so that the equation does not even have a solution modulo that integer. Our results in this section are based on this observation. These results will be superseded by those in the following sections, but they will serve us in presenting some of the ideas to be used later in a quantitative form.

Suppose the equation  $P(x) = H_n$  has infinitely many solutions. For the first two sequences, namely  $H_n = n!$  and  $H_n = [1, 2, \dots, n]$ , this implies that the congruence

$$(3.1) \quad P(x) \equiv 0 \pmod{m}$$

has a solution  $x$  for every positive integer  $m$ . An equivalent condition is that (3.1) has a solution for every prime power  $m = p^k$ , or, again equivalently, that  $P$  has a root in the ring  $\mathbf{Z}_p$  of  $p$ -adic integers for every prime  $p$ . For  $H_n = p_1 p_2 \dots p_n$ , the existence of infinitely many solutions implies that (3.1) is solvable for every square-free integer  $m$  (or, equivalently, every prime). We shall now present a few examples showing that these conditions are not that easy to fulfill.

**Example 3.1.** The congruence

$$(3.2) \quad x^2 - 2 \equiv 0 \pmod{m}$$

has no solution for  $m = 4$ , so that the equations  $x^2 - 2 = n!$  and  $x^2 - 2 = [1, 2, \dots, n]$  have only finitely many solutions. Also, (3.2) has no solution for  $m = 3$ , and therefore  $x^2 - 2 = p_1 p_2 \dots p_n$  has only finitely many solutions as well. One easily finds suitable moduli to show the same for the polynomials  $X^2 - 3$ ,  $X^2 - 5$  and  $X^2 - 6$  (but certainly not for  $X^2 - 1$  or  $X^2 - 4$ ). Note that the above does not apply to the equation  $x^2 - 2 = \binom{an}{n, n, \dots, n}$ . (The finiteness of the number of solutions of the last equation will follow from Theorem 4.1 *infra*.)

The example is but a special case of

**Theorem 3.1.** *If  $P \in \mathbf{Z}[X]$  is irreducible over  $\mathbf{Q}$  and  $\deg P \geq 2$ , then the equation  $P(x) = H_n$ , where  $(H_n)$  is any of the three sequences (1.5.a)–(1.5.c), has only finitely many solutions.*

The theorem follows straightforwardly from the fact that the conditions ensure that the congruence  $P(x) \equiv 0 \pmod{p}$  has no solutions modulo some prime (actually, infinitely many primes)  $p$  [J, pp. 138–139].

*Remark 3.1.* It will follow from Theorem 4.1 in the next section that the conclusion of Theorem 3.1 is true for the sequence (1.5.d) as well. Thus, our basic equation  $P(x) = H_n$  admits only finitely many solutions for “most” (see Remark 4.2 *infra*) polynomials of degree 2 or more for each of the four sequences considered in this paper.

The method of Theorem 3.1 also applies to many reducible polynomials. Given any polynomial  $P$  one may ask whether it has a root modulo every integer. If the answer is negative, then the equations  $P(x) = n!$  and  $P(x) = [1, 2, \dots, n]$  have only finitely many solutions. Similarly, if  $P$  fails to have a root modulo some prime, then the equation  $P(x) = p_1 p_2 \dots p_n$  has only finitely many solutions. In this connection we are naturally led to ask for an algorithm which, given a polynomial  $P \in \mathbf{Z}[X]$ , decides whether or not it has a root modulo every integer (or, alternatively, modulo every prime). The question is easy when the irreducible factors of  $P$  are all either linear or quadratic. More generally, let  $K$  be the splitting field of  $P$  over  $\mathbf{Q}$ . If the extension  $K/\mathbf{Q}$  is abelian, then the set of primes modulo which  $P$  has a root is given by certain congruence conditions (see, for example, [Wy] and the references there), and the question is easy. If the extension is non-abelian, the ensuing set of primes is in general quite complicated. An algorithm for testing whether a given polynomial has a root modulo every prime is given in [BereB]. (That this problem,

even for several polynomials in several variables, is decidable follows from [A]; see also [FrS].)

**Example 3.2.** Consider the congruence

$$(3.3) \quad (x^2 - r)(x^2 - s)(x^2 - rs) \equiv 0 \pmod{m},$$

where  $r$  and  $s$  are any integers. If  $m$  is prime, then the congruence must have a solution. In fact, this is clear if either  $r$  or  $s$  is a quadratic residue modulo  $m$ , while otherwise  $rs$  is a quadratic residue and again we are done. Thus the method of this section (as well as those of later sections) gives rise to

**Problem A.** Does the equation

$$(x^2 - 2)(x^2 - 7)(x^2 - 14) = p_1 p_2 \dots p_n$$

have infinitely many solutions?

**Example 3.3.** The equations

$$(x^2 - 2)(x^2 - 7)(x^2 - 14) = n!$$

and

$$(x^2 - 2)(x^2 - 7)(x^2 - 14) = [1, 2, \dots, n]$$

have only finitely many solutions. In fact, although the associated congruence has a solution for every prime, it fails to have solutions modulo some high powers, say modulo 8. (It has solutions modulo any prime power  $p^k$  for  $p \neq 2$ .)

However, some choices of  $r$  and  $s$  in (3.3) lead to polynomials with roots modulo every integer. This is the case, for example, with  $r = 13$ ,  $s = 17$  [BoS, p. 3].

**Problem B.** Do the equations

$$(x^2 - 13)(x^2 - 17)(x^2 - 221) = n!$$

and

$$(x^2 - 13)(x^2 - 17)(x^2 - 221) = [1, 2, \dots, n]$$

have infinitely many solutions?

#### 4. OVERLOADED FACTORS

We explain the basic idea of this section by means of

**Example 4.1.** Consider the equation

$$(4.1) \quad x(x^2 + 1) = n!.$$

(Note that the method employed in the preceding section obviously fails in this case.) For large  $n$ , a solution  $x$  should be of the order of magnitude  $\sqrt[3]{n!}$ , so that  $x^2 + 1$  is approximately  $\sqrt[3]{n!^2}$ . Now remember that the congruence

$$x^2 + 1 \equiv 0 \pmod{p},$$

where  $p$  is a prime, has a solution if and only if  $p \equiv 1 \pmod{4}$  (or  $p = 2$ ). Hence the factor  $x^2 + 1$  can account only for these primes. Since these primes are, roughly speaking, only half of all primes, we should obtain a contradiction for sufficiently large  $n$ , which hints that (4.1) has only finitely many solutions.

Erdős-Obláth's method for dealing with the equation  $x^p \pm y^p = n!$  [EO] is basically this idea, applied in cases where the set of prime divisors of the polynomial in question is defined in terms of certain congruences. As mentioned in the preceding

section, for general polynomials the ensuing sets of primes are not so easily describable. To deal with the general case, we first recall the notion(s) of density of a set of primes (see, for example, [Go] for more details). Take  $T \subseteq \mathbf{P}$ . The *Dirichlet density* (or *analytic density*) of  $T$  is defined by

$$D(T) = \lim_{s \rightarrow 1^+} \frac{\log \prod_{p \in T} \left(1 - \frac{1}{p^s}\right)^{-1}}{\log \zeta(s)} = \lim_{s \rightarrow 1^+} \frac{\sum_{p \in T} p^{-s}}{\log \frac{1}{s-1}},$$

provided that the limit exists. More natural (but perhaps less useful) is the notion of *natural density* (sometimes called *ordinary density* [La], *primitive density* [Pow] or *normal density* [R]), defined by

$$(4.2) \quad d(T) = \lim_{x \rightarrow \infty} \frac{\pi(x, T)}{\pi(x)},$$

where  $\pi(x)$  is the number of primes not exceeding  $x$  and  $\pi(x, T)$  is the number of those belonging to  $T$ . (Of course, in view of the Prime Number Theorem, one may replace the denominator on the right-hand side of (4.2) by  $\frac{x}{\log x}$ .) If  $d(T)$  exists, then so does  $D(T)$ , and the two densities coincide [Go, Th. 14-1-2]. (The converse is false; see, for example [Ser, p. 76].) In this paper we shall use only natural densities. The set  $T$  is *naturally regular* if  $d(T)$  exists. We note that some results which are often formulated for Dirichlet densities are in fact valid for natural densities as well (for example, the Chebotarev Density Theorem; see [N, Th. 7.11, 7.11\*]).

In the sequel we shall use the notions of the *upper natural density*  $d^*(T)$  and of the *lower natural density*  $d_*(T)$  of a set  $T$  of primes. These are defined as in (4.2), except that the limit is replaced by  $\limsup$  and by  $\liminf$ , respectively.

Our main result in this section is

**Theorem 4.1.** *Consider the equation*

$$P(x) = H_n,$$

where  $(H_n)$  is any of the four sequences (1.5.a)–(1.5.d). Let  $Q \in \mathbf{Z}[X]$  be any factor (irreducible or not) of  $P$ . Denote by  $S(Q) \subseteq \mathbf{P}$  the set of all primes  $p$  for which the congruence  $Q(x) \equiv 0 \pmod{p}$  has a solution. If  $d(S(Q)) < \frac{\deg Q}{\deg P}$ , then (1.4) has only finitely many solutions.

*Remark 4.1.* The set  $S(Q)$  is naturally regular [Ser, p. 76], so we do not have to introduce upper and lower natural densities in the formulation of the theorem.

Now return to our intuitive discussion in Example 4.1. The conclusion of the argument for the finiteness of the set of solutions depended on the implicit assumption that in the factorization of  $n!$ , for large  $n$ , the primes  $\equiv 1 \pmod{4}$  and the primes  $\equiv -1 \pmod{4}$  have roughly equal contributions. A generalization of this assertion is used in the proof of Theorem 4.1. For  $N \in \mathbf{N}$  and any set  $S$ , let

$$\psi_1(N, S) = \prod_{p \in S \cap \mathbf{P}} p^{\nu_p(N)}.$$

The main ingredient in the proof of Theorem 4.1 is given in

**Proposition 4.1.** *Let  $(H_n)$  be any of the four sequences (1.5.a)–(1.5.d). If  $T \subseteq \mathbf{P}$  is a naturally regular set of primes, then  $\psi_1(H_n, T) = H_n^{d(T)+o(1)}$ .*

For the first three of the sequences  $(H_n)$ , the following (even stronger) result is valid.

**Proposition 4.2.** *Let  $(H_n)$  be any of the three sequences (1.5.a)–(1.5.c). If  $T \subseteq \mathbf{P}$  is any set of primes, then  $H_n^{d^*(T)+o(1)} \leq \psi_1(H_n, T) \leq H_n^{d^*(T)+o(1)}$ .*

The corresponding result for the fourth sequence is not true. Indeed, let  $(n_k)_{k=1}^\infty$  be a rapidly increasing sequence, and let  $S = \bigcup_{k=1}^\infty (\mathbf{P} \cap (n_k, 2n_k])$ , so that  $d^*(S) = 1/2$ . According to the argument of Proposition 7.8 *infra* we have

$$\psi_1 \left( \binom{2n_k}{n_k}, S \right) = \binom{2n_k}{n_k}^{\frac{1}{2 \log 2} + o(1)},$$

whereas the upper bound predicted by Proposition 4.2 is the much smaller  $\binom{2n_k}{n_k}^{\frac{1}{2} + o(1)}$ .

The factor  $Q$  in Theorem 4.1 must sometimes be chosen reducible for the theorem to be applicable. We see this phenomenon in

**Example 4.2.** Consider the equation

$$x(x^2 + 1)(x^2 + 2) = H_n,$$

where  $(H_n)$  is any of the four sequences (1.5.a)–(1.5.d). For the set  $S(Q)$  not to be the whole of  $\mathbf{P}$ , the factor  $Q$  must not contain  $X$ . Taking  $Q(x) = x^2 + 1$  we get  $S(Q) = \{2\} \cup \{p \in \mathbf{P} : p \equiv 1 \pmod{4}\}$ , whereas  $Q(x) = x^2 + 2$  gives  $S(Q) = \{2\} \cup \{p \in \mathbf{P} : p \equiv 1, 3 \pmod{8}\}$ . Thus in each of these cases  $d(S(Q)) = \frac{1}{2}$ , while  $\frac{\deg Q}{\deg P} = \frac{2}{5}$ , and Theorem 4.1 cannot be put to use. Choosing, however,  $Q(x) = (x^2 + 1)(x^2 + 2)$ , we get  $S(Q) = \{2\} \cup \{p \in \mathbf{P} : p \equiv 1, 3, 5 \pmod{8}\}$ , so that  $d(S(Q)) = \frac{3}{4}$ , while  $\frac{\deg Q}{\deg P} = \frac{4}{5}$ , and by Theorem 4.1 the equation has only finitely many solutions.

**Example 4.3.** Consider the equation

$$(4.3) \quad P(x) = H_n,$$

where  $P$  is divisible by the  $r$ th cyclotomic polynomial  $\Phi_r$ . As is well known (see, for example, [Wa, p. 13]), the prime divisors of  $\Phi_r$  are (with finitely many exceptions) those primes which are 1 modulo  $r$ . Hence  $d(S(\Phi_r)) = \frac{1}{\phi(r)}$ , so that if  $\deg P < \phi(r)^2$ , then (4.3) has only finitely many solutions. In particular, if  $r < \phi(r)^2$ , which is the case unless  $r = 1, 2, 4$  or  $6$ , then the equation

$$(4.4) \quad x^r - 1 = H_n$$

has only finitely many solutions. For  $r = 6$  we arrive at the same conclusion by considering the divisor  $X^4 + X^2 + 1$  of  $X^6 - 1$  (the divisor  $\Phi_6 = X^2 - X + 1$  would not suffice), and noting that  $d(S(X^4 + X^2 + 1)) = \frac{1}{2}$ . The case  $r = 4$  (and certainly  $r = 2$ ) of (4.4) cannot be dealt with by means of Theorem 4.1, which explains why additional tools were required in [EO] and [PolS] to tackle it.

Now we give some examples involving non-abelian extensions.

**Example 4.4.** Consider the equation

$$(4.5) \quad x(x^3 - 2) = H_n.$$

We have  $d(S(X^3 - 2)) = \frac{2}{3}$  (see, for example, [CaF, p. 354]), so that Theorem 4.1, applied with  $Q(x) = x^3 - 2$ , shows that (4.5) has only finitely many solutions for

each of the four sequences (1.5.a)–(1.5.d). We note in passing that in this case the set of primes dividing the polynomial in question can be “explicitly” characterized; in fact, it was shown by Gauss that

$$S(X^3 - 2) = \{p \in \mathbf{P} : p \equiv -1 \pmod{3}\} \cup \{p \in \mathbf{P} : p \equiv 1 \pmod{3}, p = a^2 + 27b^2\}.$$

**Example 4.5.** Consider the equation

$$(4.6) \quad R(x)(x^5 - x - 1) = H_n.$$

In view of [Wy], the Galois group of the splitting field of  $X^5 - X - 1$  over  $\mathbf{Q}$  is the symmetric group on 5 symbols  $S_5$  and  $d(S(X^5 - X - 1)) = \frac{19}{30}$ . It follows easily from Theorem 4.1 that, if  $R$  is linear or quadratic, then (4.6) has only finitely many solutions. For cubic  $R$  we have

$$d(S(X^5 - X - 1)) = \frac{19}{30} > \frac{5}{8} = \frac{\deg(X^5 - X - 1)}{\deg(R \cdot (X^5 - X - 1))},$$

so that Theorem 4.1 fails (by a small margin).

**Problem C.** Do the equations

$$x(x+1)(x+2)(x^5 - x - 1) = H_n,$$

with  $H_n$  as in (1.5.a)–(1.5.d), have infinitely many solutions?

**Example 4.6.** More generally than (4.6), consider the equation

$$(4.7) \quad P(x) = Q(x)R(x) = H_n,$$

where  $Q$  is an irreducible polynomial such that the Galois group of its splitting field over  $\mathbf{Q}$  is the symmetric group on  $q = \deg Q$  symbols. As in [Wy], one can verify that  $S(Q)$  corresponds to the set of those permutations having a 1-cycle. Hence

$$d(S(Q)) = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + \frac{(-1)^{q-1}}{q!}.$$

For large  $q$  we have  $d(S(Q)) \sim 1 - \frac{1}{e}$ . Hence if  $\deg Q > (1 - \frac{1}{e} + \varepsilon) \deg P$ , then (4.7) can have only finitely many solutions.

*Remark 4.2.* The last example can be rephrased to assert that, for “most” polynomials having a “large” irreducible factor, our basic equation has only finitely many solutions. (See [vW], where it is proved that most polynomials of degree  $d$  over  $\mathbf{Q}$  have the symmetric group on  $d$  symbols as their Galois group.) This strengthens the observation of Remark 3.1.

## 5. MULTIPLE FACTORS

It is easy to see that each of the four sequences  $(H_n)$  in (1.5.a)–(1.5.d) has the property that all its terms from some place on satisfy  $p|H_n, p^2 \nmid H_n$  for some prime  $p$  depending on  $n$ . Moreover, if  $F$  is any finite subset of  $\mathbf{P}$ , then for sufficiently large  $n$  we may choose  $p \in \mathbf{P} \setminus F$ . This simple observation already yields

**Theorem 5.1.** *Let  $P(x) = P_0 P_1(x)^{e_1} P_2(x)^{e_2} \dots P_u(x)^{e_u}$  be the decomposition of  $P$  as a product of an integer  $P_0$  and primitive irreducible polynomials. If  $e_i \geq 2$  for every  $1 \leq i \leq k$ , then the equation  $P(x) = H_n$  has only finitely many solutions for each of the four sequences (1.5.a)–(1.5.d).*

In this section we shall see that one can often prove that our equation has only finitely many solutions by the existence of just one multiple factor of  $P$ . To present the main result, put

$$\lambda_r = \begin{cases} 0.8431, & r = 2, \\ 0.7907, & r = 3, \\ 0.7642, & r = 4, \\ 0.7405, & r = 5, \\ 0.5 + \frac{0.5099}{r} + \frac{\log(r-1)}{2r}, & r \geq 6, \end{cases}$$

$$\theta_{a,k} = \sum_{j=0}^{\infty} \left( \frac{1}{ja+k} - \frac{1}{ja+k+1} \right), \quad a \geq 2, 1 \leq k.$$

**Theorem 5.2.** Let  $P(x) = P_0P_1(x)^{e_1}P_2(x)^{e_2} \dots P_u(x)^{e_u}$ , as in Theorem 5.1.

a) If

$$\frac{1}{\deg P} \sum_{i:r|e_i} e_i \deg P_i > \lambda_r$$

for some  $r \geq 2$ , then the equation  $P(x) = n!$  has only finitely many solutions.

b) If  $e_i \geq 2$  for some  $1 \leq i \leq u$ , then the equation  $P(x) = [1, 2, \dots, n]$  has only finitely many solutions.

c) If  $e_i \geq 2$  for some  $1 \leq i \leq u$ , then the equation  $P(x) = p_1p_2 \dots p_n$  has only finitely many solutions.

d) If

$$\frac{1}{\deg P} \sum_{i:r|e_i} e_i \deg P_i > \frac{r}{\log a} \sum_{k=1}^{\lfloor \frac{a-1}{r} \rfloor} k\theta_{a,kr}$$

for some  $r \geq 2$ , then the equation  $P(x) = \binom{an}{n, n, \dots, n}$  has only finitely many solutions. In particular, this is the case if  $e_i \geq a$  for some  $1 \leq i \leq u$ .

Parts b) and c) of the theorem are very simple. We shall now shed some light on the other two parts. The idea of part a) is seen in

**Example 5.1.** Consider the equation

$$(5.1) \quad x^2(x+1) = n!.$$

One would expect the part of  $n!$  coming from primes  $p$  with  $\nu_p(n!)$  even to account (for large  $n$ ) for about  $\sqrt{n!}$ . Since the factor  $x^2$  of  $x^2(x+1)$  can contribute only to these primes, however, they should account for at least  $\sqrt[3]{n!^2}$ . This contradiction should prove that (5.1) has only finitely many solutions. Unfortunately, our results regarding the portion of  $n!$  under investigation are most probably not the best possible, and the finiteness of the number of solutions of (5.1) does not follow from Theorem 5.2. However, Theorem 5.2a) yields the finiteness for the equation

$$x^r(x+1) = n!$$

if  $r \geq 4$ . In fact,

$$\frac{1}{\deg P} \sum_{i:r|e_i} e_i \deg P_i = \frac{r}{r+1}.$$

For  $r = 4$  the last expression assumes the value  $4/5$ , which exceeds  $\lambda_4 = 0.7648$ . As  $r/(r + 1)$  increases with  $r$ , whereas  $\lambda_r$  decreases with  $r$ , we may conclude our assertion for  $r \geq 4$ . Since for  $r = 3$  we have

$$\frac{1}{\deg P} \sum_{i:r|e_i} e_i \deg P_i = \frac{3}{4} < 0.7907 = \lambda_3,$$

Theorem 5.2 is not strong enough to conclude the finiteness for  $r = 2, 3$ .

For  $N \in \mathbf{N}$  and  $A \subseteq \mathbf{N}$ , let

$$\psi_2(N, A) = \prod_{\substack{p \in \mathbf{P} \\ \nu_p(N) \in A}} p^{\nu_p(N)}.$$

Then the portion of  $n!$  due to primes appearing with exponents divisible by a certain integer  $r$  is  $\psi_2(n!, r\mathbf{N})$ . The problem with making the idea of Example 5.1 precise is that we do not know how to show that  $\psi_2(n!, r\mathbf{N})$  is about  $\sqrt[r]{n!}$ . The difficulty lies especially with small primes. For example, Erdős and Graham [EG, p. 77] ask whether for every  $k$  there exists some  $n$  with all the exponents  $\nu_2(n!), \nu_3(n!), \nu_5(n!), \dots, \nu_{p_k}(n!)$  even. (This question was settled affirmatively in [Be].) The contribution of the big primes (say,  $p > \frac{n}{c}$  for some constant  $c$ ) satisfying this condition is easy to estimate, but they account in any case only for a negligible part of  $n!$ , namely  $n!^{o(1)}$ . The hope for using the idea is based on primes  $p > \sqrt{n}$ , and it is useful to divide even these primes according to their size.

For  $N \in \mathbf{N}$ ,  $A \subseteq \mathbf{N}$  and any set  $S$ , let

$$\psi(N, A, S) = \prod_{\substack{p \in S \cap \mathbf{P} \\ \nu_p(N) \in A}} p^{\nu_p(N)}.$$

(Then  $\psi_2(N, A)$  defined above is  $\psi(N, A, \mathbf{P})$ , and  $\psi_1(N, S)$  defined in Section 4 is  $\psi(N, \mathbf{N}, S)$ .) The *upper density*  $\delta^*(A)$  of a set  $A \subseteq \mathbf{N}$  is given by

$$\delta^*(A) = \limsup_{N \rightarrow \infty} \frac{\#(A \cap [1, N])}{N}.$$

The *lower density*  $\delta_*(A)$  is similarly defined, with  $\limsup$  replaced by  $\liminf$ , and the *density* of  $A$  is

$$\delta(A) = \delta^*(A) = \delta_*(A)$$

if  $\delta^*(A)$  and  $\delta_*(A)$  coincide. For  $0 < \alpha \leq 1$  set

$$\gamma_*(\alpha) = \liminf_{x \rightarrow \infty} \frac{\#(\mathbf{P} \cap (x, x + x^\alpha])}{\frac{x^\alpha}{\log x}}, \quad \gamma^*(\alpha) = \limsup_{x \rightarrow \infty} \frac{\#(\mathbf{P} \cap (x, x + x^\alpha])}{\frac{x^\alpha}{\log x}}.$$

**Proposition 5.1.** *Let  $A \subseteq \mathbf{N}$  and  $0 < \alpha < 1$ . Set*

$$\beta = \beta(\alpha) = \frac{1 - \alpha}{2 - \alpha} = 1 - \frac{1}{2 - \alpha} < \frac{1}{2}.$$

*Then for  $0 \leq \beta_1 < \beta$ ,*

$$(n!)^{(\beta - \beta_1)\gamma^*(\alpha)\delta_*(A) + o(1)} \leq \psi(n!, A, (n^{1-\beta}, n^{1-\beta_1}]) \leq (n!)^{(\beta - \beta_1)\gamma^*(\alpha)\delta^*(A) + o(1)}.$$

This is a special case of Proposition 6.1. In Section 7 we present a self-refinement of the proposition.

We now present the results known regarding  $\gamma_*(\alpha)$  and  $\gamma^*(\alpha)$ . By the Prime Number Theorem we have  $\gamma_*(1) = \gamma^*(1) = 1$ . One expects that actually  $\gamma_*(\alpha) = \gamma^*(\alpha) = 1$  for every  $0 < \alpha \leq 1$ . However, even the inequality

$$(5.2) \quad \gamma^*(\alpha) \leq \frac{2}{\alpha}, \quad 0 < \alpha \leq 1$$

(which follows, for example, from [Mon, (4.11)]), is non-trivial. We also have

**Lemma 5.1.**  $\gamma^*(\alpha)$  is non-increasing with  $\alpha$  and  $\gamma_*(\alpha)$  is non-decreasing with  $\alpha$ .

A problem closely related to that of calculating  $\gamma_*(\alpha)$  and  $\gamma^*(\alpha)$  is that of the existence of primes in short intervals. Trying to prove this kind of result by means of the Prime Number Theorem,

$$\pi(x) = \text{li}x + O(xe^{-c(\log x)^{3/5}}(\log \log x)^{-1/5})$$

([Mi], [So1], [So2]) is of little use, as the best currently known estimate for the error term is too large. (Note that, in view of the lower bound for the error term [I, p. 307]

$$\pi(x) - \text{li}x = \Omega_{\pm} \left( \frac{x^{\frac{1}{2}} \log \log \log x}{\log x} \right),$$

even the sharpest form of the Prime Number Theorem which may be valid would require intervals of size  $x^{\frac{1}{2}+o(1)}$  to ensure the existence of a prime, but it is generally believed that much smaller intervals suffice.) Following Hoheisel [Ho], who proved that, for  $x > x_0(\varepsilon)$ , the interval  $(x, x + x^{1-\frac{1}{33000}+\varepsilon}]$  must contain a prime, there is a long chain of improvements. The best unconditional result known to date is due to Baker, Harman and Pintz [BakHP], who showed the existence of a prime in the interval  $(x, x + x^{0.525})$ . We mention that the best known result assuming the Riemann hypothesis is

$$p_{n+1} - p_n = O(\sqrt{p_n} \log p_n)$$

(cf. [I, p. 299]; for slight improvements, under additional hypothesis, see [Mu], [He1], [HeaG]), and both are very far from Cramér's 1937 conjecture [Cr], according to which (it follows in particular that) the interval  $(x, x + (1 + \varepsilon) \log^2 x]$  must contain a prime for large  $x$  (see also [Sh]).

However, all these results are not quite what we need, as they do not prove that the part of  $n!$  consisting of primes with exponents divisible by a fixed integer is non-negligible. For this we need to know that the intervals in question contain not one but "many" (i.e., approximately what we expect, or at least a fixed fraction of that) primes. Davenport [Dave, p. 174] mentions, with respect to all results known at the time on the existence of primes in short intervals, that they are easily modified to show that the number of primes in those intervals is asymptotically what one would expect. The best result of this sort is due to Heath-Brown [He3]

$$(5.3) \quad \pi(x + x^{\frac{7}{12}}) - \pi(x) \sim \frac{x^{\frac{7}{12}}}{\log x}.$$

When reducing the exponent from  $\frac{7}{12}$  the situation regarding the number of primes in the interval becomes worse. In fact, the results of Lou and Yao [LoY] give only

$$(5.4) \quad 0.99 \frac{x^{\frac{11}{20}+\varepsilon}}{\log x} < \pi(x + x^{\frac{11}{20}+\varepsilon}) - \pi(x) < 1.01 \frac{x^{\frac{11}{20}+\varepsilon}}{\log x}$$

and

$$(5.5) \quad 0.969 \frac{x^{\frac{6}{11} + \varepsilon}}{\log x} < \pi(x + x^{\frac{6}{11} + \varepsilon}) - \pi(x) < 1.031 \frac{x^{\frac{6}{11} + \varepsilon}}{\log x}$$

for  $\varepsilon > 0$  and sufficiently large  $x$ , those of Baker and Harman imply

$$(5.6) \quad (0.4 - \varepsilon) \frac{x^{0.54}}{\log x} < \pi(x + x^{0.54}) - \pi(x),$$

and those of Baker, Harman and Pintz [BakHP] give

$$(5.7) \quad 0.09 \frac{x^{0.525}}{\log x} < \pi(x + x^{0.525}) - \pi(x).$$

With our notations, (5.3)–(5.7) amount to

$$(5.8) \quad \gamma_* \left( \frac{7}{12} \right) = \gamma^* \left( \frac{7}{12} \right) = 1,$$

$$(5.9) \quad \gamma_* \left( \frac{11}{20} + \varepsilon \right) \geq 0.99, \quad \gamma^* \left( \frac{11}{20} + \varepsilon \right) \leq 1.01,$$

$$(5.10) \quad \gamma_* \left( \frac{6}{11} + \varepsilon \right) \geq 0.969, \quad \gamma^* \left( \frac{6}{11} + \varepsilon \right) \leq 1.031,$$

$$(5.11) \quad \gamma_*(0.54) \geq 0.4,$$

and

$$(5.12) \quad \gamma_*(0.525) \geq 0.09,$$

respectively. Maier (see [E3]) showed that a quantitative version of Cramér’s conjecture mentioned above (namely, that the interval  $(x, x + \log^2 x]$  contains approximately  $\log x$  primes) is certainly not true. However, Erdős [E3] opines that perhaps

$$\frac{\pi(x + y) - \pi(x)}{\frac{y}{\log x}} \rightarrow 1$$

as  $x \rightarrow \infty$  if  $y \rightarrow \infty$  faster than any power of  $\log x$  (see also [E1]). This result would mean in particular that  $\gamma_*(\alpha) = \gamma^*(\alpha) = 1$  for every  $\alpha > 0$ .

Now we discuss some aspects related to part d) of Theorem 5.2. First we provide some more information on the constants  $\theta_{a,k}$ , and in particular a closed form formula for them. Approximate values of  $\theta_{a,k}$  for  $1 \leq k < a \leq 9$  are provided in Appendix A.

**Lemma 5.2.** *For any  $a \geq 2$ :*

1.  $\theta_{a,k} = \frac{1}{a} \sum_{\xi \in R_a} (\xi^{k+1} - \xi^k) \log(1 - \xi^{-1})$  for  $1 \leq k < a$ , where  $R_a$  is the set of all  $a$ th roots of unity. (On the right-hand side we take  $0 \log 0 = 0$ .)
2.  $k\theta_{a,k} > (k + 1)\theta_{a,k+1}$ ,  $1 \leq k < a$ .
3.  $\sum_{k=1}^{a-1} k\theta_{a,k} = \log a$ .

Like the proof of part a) of Theorem 5.2, the proof of part d) depends mainly on knowing the part of  $\binom{an}{n, n, \dots, n}$  consisting of primes appearing with various exponents in its factorization. Fortunately, here the situation is much simpler. (When a set argument to  $\psi$  or  $\psi_2$  is a singleton, we may replace it with its single element.)

**Proposition 5.2.** *For any  $a \geq 2$ :*

1.  $\psi_2 \left( \binom{an}{n, n, \dots, n}, k \right) = \binom{an}{n, n, \dots, n}^{\frac{k\theta_{a,k}}{\log a} + o(1)}, \quad k = 1, 2, \dots, a - 1,$
2.  $\psi_2 \left( \binom{an}{n, n, \dots, n}, \{a, a + 1, \dots\} \right) = \binom{an}{n, n, \dots, n}^{o(1)}.$

*Remark 5.1.* For  $a = 2$ , Sárközy [Sár] proved (with a totally different purpose from ours) a more refined version of part 2.

Now, while Theorem 5.2 looks stronger than Theorem 5.1, this is not always the case for  $H_n = n!$  and  $H_n = \binom{an}{n, n, \dots, n}$ . For example, by Theorem 5.1 the equation

$$x^2(x + 1)^3 = n!$$

has only finitely many solutions, but Theorem 5.2 does not yield the same result. The following theorem will include both Theorem 5.1 and Theorem 5.2a), 5.2d), and is the best one can obtain employing the method of this section. As in other cases, it is sometimes best to focus on “general” divisors of  $P(x)$ . For  $r_1, r_2, \dots, r_s \in \mathbf{N}$  let  $h(r_1, r_2, \dots, r_s) = \delta(r_1\mathbf{N} \cup r_2\mathbf{N} \cup \dots \cup r_s\mathbf{N})$  (the density of a union of arithmetic progressions). It is readily verified by inclusion-exclusion that

$$\begin{aligned} h(r_1, r_2, \dots, r_s) &= \sum_{1 \leq i \leq s} \frac{1}{r_i} - \sum_{1 \leq i_1 < i_2 \leq s} \frac{1}{[r_{i_1}, r_{i_2}]} \\ &\quad + \sum_{1 \leq i_1 < i_2 < i_3 \leq s} \frac{1}{[r_{i_1}, r_{i_2}, r_{i_3}]} - \dots + \frac{(-1)^{s-1}}{[r_1, r_2, \dots, r_s]}. \end{aligned}$$

For  $0 \leq h \leq 1$  put

$$\eta_h = \frac{1}{2} + \int_0^{1/2} \min \left\{ h\gamma^* \left( \frac{1 - 2\beta}{1 - \beta} \right), 1 - (1 - h)\gamma_* \left( \frac{1 - 2\beta}{1 - \beta} \right) \right\} d\beta.$$

Clearly, for  $r_1, r_2, \dots, r_s \geq 2$  we have  $h(r_1, r_2, \dots, r_s) < 1$ . Also,  $\eta_h < 1$  for  $h < 1$ , and therefore  $\eta_{h(r_1, r_2, \dots, r_s)} < 1$  for any  $r_1, r_2, \dots, r_s \geq 2$ .

**Theorem 5.3.** *Let  $P(x) = P_0P_1(x)^{e_1}P_2(x)^{e_2} \dots P_u(x)^{e_u}$ , as in Theorem 5.1.*

a) *If for some  $r_1, r_2, \dots, r_s \in \mathbf{N}$*

$$\frac{1}{\deg P} \sum_{i: r_1|e_i \vee \dots \vee r_s|e_i} e_i \deg P_i > \eta_{h(r_1, r_2, \dots, r_s)},$$

*then the equation  $P(x) = n!$  has only finitely many solutions.*

d) *If for some  $r_1, r_2, \dots, r_s \in \mathbf{N}$*

$$\frac{1}{\deg P} \sum_{i: r_1|e_i \vee \dots \vee r_s|e_i} e_i \deg P_i > \frac{1}{\log a} \sum_{l < a: r_1|l \vee \dots \vee r_s|l} l\theta_{a,l},$$

*then the equation  $P(x) = \binom{an}{n, n, \dots, n}$  has only finitely many solutions.*

*Remark 5.2.* In view of Theorem 5.2, there is no need for parts b) and c).

To be able to apply Theorem 5.3a), we give an upper bound for  $\eta_h$ .

**Proposition 5.3.**

$$\eta_h \leq \begin{cases} 0.5 + 0.5099h + \frac{h}{2} \log \frac{1-h}{h}, & 0 \leq h \leq \frac{81}{446}, \\ 0.3891 + 0.3678h + h \log \frac{1-h}{h}, & \frac{81}{446} \leq h \leq \frac{9}{49}, \\ 0.5016 + 0.5012h + \frac{h}{2} \log \frac{1-h}{h}, & \frac{9}{49} \leq h \leq \frac{1911}{7811}, \\ 0.3396 + 0.5997h + h \log \frac{1-h}{h}, & \frac{1911}{7811} \leq h \leq \frac{2457}{9757}, \\ 0.5079 + 0.4758h + \frac{h}{2} \log \frac{1-h}{h}, & \frac{2457}{9757} \leq h \leq \frac{21}{80}, \\ 0.6859 + 0.3143h, & \frac{21}{80} \leq h \leq \frac{1}{2}, \\ 0.6861 + 0.3139h, & \frac{1}{2} \leq h \leq 1. \end{cases}$$

In particular,  $\eta_{1/r} \leq \lambda_r$ .

We conclude this section with conditional results, indicating how the result of Theorem 5.2a) would be strengthened under various conjectures. Specifically, we consider the conjectures

$$(5.13) \quad \gamma_*(\alpha) = \gamma^*(\alpha) = 1, \quad \frac{1}{2} < \alpha \leq 1,$$

which seems to be the best possible result by means of current techniques (and would follow from the Riemann hypothesis),

$$(5.14) \quad \gamma_*(\alpha) = \gamma^*(\alpha) = 1, \quad 0 < \alpha \leq 1,$$

and

**Conjecture 5.1.** *If  $A \subseteq \mathbf{N}$  is a union of finitely many arithmetic progressions and  $0 \leq \beta_1 < \beta \leq 1$ , then*

$$\psi(n!, A, (n^{1-\beta}, n^{1-\beta_1}]) = (n!)^{(\beta-\beta_1)\delta(A)+o(1)}.$$

*Remark 5.3.* It should be possible to replace the union of arithmetic progressions by any “reasonable” set. However, one can construct a 0-density set  $A$  with

$$\psi(n!, A, (1, \sqrt{n}]) = \Omega((n!)^{1/2+o(1)}).$$

**Theorem 5.4.** *Consider the equation  $P(x) = n!$ , where*

$$P(x) = P_0 P_1(x)^{e_1} P_2(x)^{e_2} \dots P_u(x)^{e_u},$$

*as in Theorem 5.1.*

i) *Under (5.13), if*

$$\frac{1}{\deg P} \sum_{i:r|e_i} e_i \deg P_i > \begin{cases} \frac{2}{3} + \frac{1}{3r}, & r = 2, 3, \\ \frac{1}{2} + \frac{2-\log 3}{2r} + \frac{\log(r-1)}{2r}, & r \geq 4, \end{cases}$$

*for some  $r \geq 2$ , then the equation has only finitely many solutions.*

ii) Under (5.14), if

$$\frac{1}{\deg P} \sum_{i:r|e_i} e_i \deg P_i > \frac{1}{2} + \frac{1}{2r}$$

for some  $r \geq 2$ , then the equation has only finitely many solutions.

iii) Under Conjecture 5.1, if

$$\frac{1}{\deg P} \sum_{i:r|e_i} e_i \deg P_i > \frac{1}{r}$$

for some  $r \geq 2$ , then the equation has only finitely many solutions.

### 6. MULTIPLE OVERLOADED FACTORS

In this section we combine the techniques of Sections 4 and 5. The idea is to take advantage both of the fact that the polynomial  $P$  has a multiple factor and of the fact that this factor may be irreducible of degree  $\geq 2$  and therefore cannot account for the part of the factorization of  $H_n$  coming from many primes. Here, in view of Theorem 5.2, we have to deal only with  $H_n$  as in (1.5.a) and (1.5.d). We illustrate this by means of

**Example 6.1.** Consider the equation

$$(6.1) \quad Q(x)(x^2 + 1)^2 = n!,$$

where  $Q$  is any polynomial (over  $\mathbf{Z}$ ) relatively prime to  $X^2 + 1$  over  $\mathbf{Q}$ . The factor  $(x^2 + 1)^2$  can account only for primes  $p \equiv 1 \pmod{4}$ . However, we expect that for large  $n$  only about half of the contribution of these primes towards the factorization of  $n!$  is with even powers. Since only finitely many primes can divide both  $x^2 + 1$  and  $Q(x)$ , we may expect that the  $(x^2 + 1)^2$  factor will be able to account only for about  $\sqrt[4]{n!}$ . Hence we would expect that if  $\deg Q < 12$ , then (6.1) has only finitely many solutions.

For  $S \subseteq \mathbf{P}$  and  $0 < \alpha \leq 1$  set

$$\gamma_*(S, \alpha) = \liminf_{x \rightarrow \infty} \frac{\#(S \cap (x, x + x^\alpha])}{\frac{x^\alpha}{\log x}}, \quad \gamma^*(S, \alpha) = \limsup_{x \rightarrow \infty} \frac{\#(S \cap (x, x + x^\alpha])}{\frac{x^\alpha}{\log x}}.$$

Thus,  $\gamma_*(\alpha)$  and  $\gamma^*(\alpha)$ , defined in Section 5, are  $\gamma_*(\mathbf{P}, \alpha)$  and  $\gamma^*(\mathbf{P}, \alpha)$ . Unfortunately, there currently seem to be no results regarding these numbers. One might expect that  $\gamma_*(S, \alpha) = d(S)\gamma_*(\alpha)$  and  $\gamma^*(S, \alpha) = d(S)\gamma^*(\alpha)$  for “reasonable” sets  $S \subseteq \mathbf{P}$ . This would make Theorem 6.1a) below apply to cases for which Theorem 5.3a) cannot.

The following two results generalize Lemma 5.1 and Proposition 5.1, respectively. It can also be shown that  $\gamma_*(S, 1) \leq d_*(S) \leq d^*(S) \leq \gamma^*(S, 1)$ .

**Lemma 6.1.**  $\gamma^*(S, \alpha)$  is non-increasing with  $\alpha$  and  $\gamma_*(S, \alpha)$  is non-decreasing with  $\alpha$ .

**Proposition 6.1.** In the setup of Proposition 5.1, for  $S \subseteq \mathbf{P}$ ,

$$\begin{aligned} (n!)^{(\beta - \beta_1)\gamma_*(S, \alpha)\delta_*(A) + o(1)} &\leq \psi(n!, A, S \cap (n^{1-\beta}, n^{1-\beta_1}]) \\ &\leq (n!)^{(\beta - \beta_1)\gamma^*(S, \alpha)\delta^*(A) + o(1)}. \end{aligned}$$

**Proposition 6.2.** *Let  $S$  be a naturally regular set of primes,  $a \geq 2$  and  $k \in \{1, 2, \dots, a - 1\}$ . Then*

$$\psi \left( \binom{an}{n, n, \dots, n}, k, S \right) = \binom{an}{n, n, \dots, n}^{\frac{kd(S)\theta_{a,k}}{\log a} + o(1)}.$$

For  $S \subseteq \mathbf{P}$  and  $0 \leq h \leq 1$  put  $\alpha(\beta) = \frac{1-2\beta}{1-\beta}$  and

$$\eta_{S,h} = \frac{d^*(S)}{2} + \int_0^{1/2} \min\{h\gamma^*(S, \alpha(\beta)), d^*(S) - (1-h)\gamma_*(S, \alpha(\beta)), 1 - (1-h)\gamma_*(\alpha(\beta)) - h\gamma_*(\mathbf{P} \setminus S, \alpha(\beta))\} d\beta.$$

**Theorem 6.1.** *Let  $P(x) = P_0P_1(x)^{e_1}P_2(x)^{e_2} \dots P_u(x)^{e_u}$  be the decomposition of  $P$  into a product of an integer and primitive irreducible polynomials.*

a) *Let  $A = r_1\mathbf{N} \cup r_2\mathbf{N} \cup \dots \cup r_s\mathbf{N}$  for some  $r_1, r_2, \dots, r_s \in \mathbf{N}$ . Let  $B \subseteq \{i : 1 \leq i \leq u, e_i \in A\}$  and put  $S = \bigcup_{i \in B} S(P_i)$ . If*

$$\frac{1}{\deg P} \sum_{i \in B} e_i \deg P_i > \eta_{S, \delta(A)},$$

*then the equation  $P(x) = n!$  has only finitely many solutions.*

d) *Choose  $B \subseteq \{1, 2, \dots, u\}$ . If*

$$\frac{1}{\deg P} \sum_{i \in B} e_i \deg P_i > \frac{1}{\log(a)} \sum_{k=1}^{a-1} d \left( \bigcup_{\substack{i \in B \\ e_i | k}} S(P_i) \right) k\theta_{a,k},$$

*then the equation  $P(x) = \binom{an}{n, n, \dots, n}$  has only finitely many solutions.*

*Remark 6.1.* Again, as in Theorem 5.3, there is no need for parts b) and c).

Theorem 6.1d) is the best possible with our techniques, but it is not clear how best to formulate a). Apparently we could do better by breaking  $A$  into residue classes modulo  $[r_1, r_2, \dots, r_s]$  and letting  $S$  vary as in d). However, this would produce a more complicated expression, and (since hardly anything is known about  $\gamma_*(S, \alpha)$  and  $\gamma^*(S, \alpha)$ ) the more complicated expression might in principle be worse than that given above.

**Example 6.2.** Consider the equation

$$(6.2) \quad Q(x)x^6(x^2 + 1)^4 = \binom{7n}{n, n, \dots, n},$$

where  $Q$  is any polynomial relatively prime to both the polynomials  $X$  and  $X^2 + 1$  over  $\mathbf{Q}$ . Employing Theorem 5.2 for  $r = 2, r = 4$  and  $r = 6$ , we can infer that (6.2) has only finitely many solutions if

$$\frac{14}{\deg(Q) + 14} > \frac{2\theta_{7,2} + 4\theta_{7,4} + 6\theta_{7,6}}{\log 7} \approx 0.4473,$$

if

$$\frac{8}{\deg(Q) + 14} > \frac{4\theta_{7,4}}{\log 7} \approx 0.1376,$$

and if

$$\frac{6}{\deg(Q) + 14} > \frac{6\theta_{7,6}}{\log 7} \approx 0.1159,$$

respectively. Thus we would arrive at the desired conclusion so long as  $\deg(Q) \leq 17$ ,  $\deg(Q) \leq 44$  or  $\deg(Q) \leq 37$ , respectively. Choosing  $B$  to correspond to the factor  $(X^2 + 1)^4$  in Theorem 6.1 we obtain the same under the requirement

$$\frac{8}{\deg(Q) + 14} > \frac{2\theta_{7,4}}{\log 7} \approx 0.0688.$$

Thus we can conclude that (6.2) has only finitely many solutions if  $\deg(Q) \leq 102$ .

### 7. AUXILIARY RESULTS

In this section we prove results about  $\psi(H_n, A, S)$ , including some propositions stated in previous sections. We also deal with some related matters, like the constants  $\theta_{a,k}$ . In Section 8 we use these results to prove the assertions in Sections 4, 5 and 6.

For later reference we note at this point the factorizations of the general terms of the sequences (1.5.a)–(1.5.d). Recall that  $\nu_p(m)$  denotes the maximal  $k$  for which  $p^k | m$ . For  $x$  real, let  $\{x\}$  denote the fractional part of  $x$ . The various parts of the following lemma are all either trivial or well known.

**Lemma 7.1.** *For every  $n \in \mathbf{N}$  and  $p \in \mathbf{P}$ :*

- a)  $\nu_p(n!) = \sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor \left( = \sum_{k=1}^{\lfloor \log_p n \rfloor} \left\lfloor \frac{n}{p^k} \right\rfloor \right)$ .
- b)  $\nu_p([1, \dots, n]) = \lfloor \log_p n \rfloor$ .
- c)  $\nu_p(p_1 \dots p_n) = \begin{cases} 1, & p \leq p_n, \\ 0, & p > p_n. \end{cases}$
- d)  $\nu_p\left(\binom{an}{n, \dots, n}\right) = \sum_{k=1}^{\infty} \left( \left\lfloor \frac{an}{p^k} \right\rfloor - a \left\lfloor \frac{n}{p^k} \right\rfloor \right) = \sum_{k=1}^{\infty} \left( a \left\{ \frac{n}{p^k} \right\} - \left\{ \frac{an}{p^k} \right\} \right)$ .

Proposition 4.2 will follow from a few lemmas. The inequality for  $p_1 p_2 \dots p_n$  is easy, and Lemma 7.4 below shows that  $[1, 2, \dots, n]$  is essentially the same. For  $n!$ , Lemmas 7.2 and 7.3 allow us to replace  $\nu_p(n!)$  with a simpler expression and manage any ‘irregularity’ of  $S$ .

**Lemma 7.2.** *For  $n \in \mathbf{N}$  and  $p \in \mathbf{P}$  we have  $\nu_p(n!) \leq \frac{n}{p-1}$ . On the other hand, for  $n \in \mathbf{N}$*

$$\prod_{\substack{p \in \mathbf{P} \\ p \leq n}} p^{\frac{n}{p-1} - \nu_p(n!)} = n!^{o(1)}.$$

*Proof.* Certainly,

$$\nu_p(n!) = \sum_{j=1}^{\infty} \left\lfloor \frac{n}{p^j} \right\rfloor \leq \sum_{j=1}^{\infty} \frac{n}{p^j} = \frac{n}{p-1}.$$

On the other hand,

$$\begin{aligned} \frac{n}{p-1} - \nu_p(n!) &= \sum_{j=1}^{\infty} \left( \frac{n}{p^j} - \left\lfloor \frac{n}{p^j} \right\rfloor \right) \\ &= \sum_{j \leq \log_p n} \left( \frac{n}{p^j} - \left\lfloor \frac{n}{p^j} \right\rfloor \right) + \sum_{j > \log_p n} \left( \frac{n}{p^j} - \left\lfloor \frac{n}{p^j} \right\rfloor \right) \\ &\leq \sum_{j \leq \log_p n} 1 + \sum_{j > \log_p n} \frac{n}{p^j} \\ &\leq \log_p n + 2. \end{aligned}$$

It follows that  $p^{\frac{n}{p-1}-\nu_p(n!)} \leq np^2 \leq n^3$  for  $p \leq n$ . Thus

$$\prod_{\substack{p \in \mathbf{P} \\ p \leq n}} p^{\frac{n}{p-1}-\nu_p(n!)} \leq n^{3\pi(n)}.$$

Since  $\pi(n) = o(n)$  by the Prime Number Theorem and  $n^n = (n!)^{1+o(1)}$  by Stirling’s Formula, the lemma follows.

Proposition 7.1 below depends upon more than a count of primes in  $S$  less than  $n$  because different primes contribute different amounts to the factorization of  $n!$ . The natural approach is to divide  $[2, n]$  into subintervals and count primes in each subinterval. Even in the case that  $S$  is naturally regular, this involves unnecessary complications. (The complications are unnecessary in this context but must be faced in other cases.) It is much easier to employ the following simple observation.

**Lemma 7.3.** *Let  $\mu_1, \mu_2$  be non-negative Borel measures on  $(1, \infty)$  such that  $\mu_1((1, t]) \leq \mu_2((1, t])$  for every  $t$ . Then for every non-negative decreasing function  $f$  on  $(1, \infty)$ ,*

$$\int_{(1, \infty)} f \, d\mu_1 \leq \int_{(1, \infty)} f \, d\mu_2 .$$

*Proof.* This follows immediately once the integrals are written in terms of distribution functions. Indeed

$$\int_{(1, \infty)} f \, d\mu_1 = \int_0^\infty \mu_1(f^{-1}(\tau, \infty)) \, d\tau \leq \int_0^\infty \mu_2(f^{-1}(\tau, \infty)) \, d\tau = \int_{(1, \infty)} f \, d\mu_2 .$$

**Proposition 7.1.** *Let  $S \subseteq \mathbf{P}$ . Then for  $0 \leq \beta_1 < \beta \leq 1$ ,*

$$(n!)^{d_*(S)(\beta-\beta_1)+o(1)} \leq \psi_1(n!, (n^{1-\beta}, n^{1-\beta_1}] \cap S) \leq (n!)^{d_*(S)(\beta-\beta_1)+o(1)} .$$

*In particular,  $(n!)^{d_*(S)+o(1)} \leq \psi_1(n!, S) \leq (n!)^{d_*(S)+o(1)}$ .*

*Proof.* It suffices to prove the first inequality, because the second will then follow from the lower bounds on  $\psi_1(n!, (n^0, n^{1-\beta}])$ ,  $\psi_1(n!, (n^{1-\beta}, n^{1-\beta_1}] \cap (\mathbf{P} \setminus S))$  and  $\psi_1(n!, (n^{1-\beta_1}, n^1])$ .

We may assume  $d_*(S)$  to be positive. Choose  $\varepsilon$  strictly between 0 and  $\min\{d_*(S), \beta-\beta_1\}$ . By the Prime Number Theorem, there is  $N_0$  such that the number of primes not exceeding  $x$  and belonging to  $S$  satisfies  $\pi(x, S) \geq (d_*(S) - \varepsilon) \int_2^x \frac{dt}{\log t}$  for  $n \geq N_0$  and  $x \geq n^{1-\beta+\varepsilon}$ . There is also  $N_1$  such that for  $n \geq N_1$ ,

$$\pi(n^{1-\beta}, S) \leq n^{1-\beta} \leq \frac{(n^{1-\beta+\varepsilon} - 2)(d_*(S) - \varepsilon)}{\log(n^{1-\beta+\varepsilon})} \leq (d_*(S) - \varepsilon) \int_2^{n^{1-\beta+\varepsilon}} \frac{dt}{\log t} .$$

Now take  $n > \max\{N_0, N_1, 2\}$ , and define Borel measures  $\mu_j$  by

$$\mu_1(A) = (d_*(S) - \varepsilon) \int_{A \cap (n^{1-\beta+\varepsilon}, n^{1-\beta_1}]} \frac{dt}{\log t}$$

and

$$\mu_2(A) = \#(A \cap S \cap (n^{1-\beta}, n^{1-\beta_1}]) .$$

Of course  $\mu_1((1, x]) = 0 \leq \mu_2((1, x])$  for  $x \leq n^{1-\beta+\varepsilon}$ , and  $\mu_1((1, x]) \leq \mu_2((1, x])$  for  $x \geq n^{1-\beta+\varepsilon}$  by choice of  $n$ . Thus by Lemma 7.3,

$$\begin{aligned} \log \left( \prod_{p \in (n^{1-\beta}, n^{1-\beta_1}] \cap S} p^{\frac{n}{p-1}} \right) &= \int_{t \in (1, \infty)} \log(t^{\frac{n}{t-1}}) d\mu_2 \\ &\geq \int_{t \in (1, \infty)} \log(t^{\frac{n}{t-1}}) d\mu_1 \\ &= (d_*(S) - \varepsilon) \int_{n^{1-\beta+\varepsilon}}^{n^{1-\beta_1}} \log(t^{\frac{n}{t-1}}) \frac{dt}{\log t} \\ &= (d_*(S) - \varepsilon)n \int_{n^{1-\beta+\varepsilon}}^{n^{1-\beta_1}} \frac{dt}{t-1} \\ &= (d_*(S) - \varepsilon)n \log \left( \frac{n^{1-\beta_1} - 1}{n^{1-\beta+\varepsilon} - 1} \right) \\ &\geq (d_*(S) - \varepsilon)n \log \left( \frac{n^{1-\beta_1}}{n^{1-\beta+\varepsilon}} \right) \\ &= (d_*(S) - \varepsilon)n(\beta - \beta_1 - \varepsilon) \log n \\ &\geq ((d_*(S) - \varepsilon)(\beta - \beta_1 - \varepsilon)) \log n!. \end{aligned}$$

The result follows by Lemma 7.2.

**Proposition 7.2.** *Let  $S \subseteq \mathbf{P}$  with  $d^*(S) = d$ . Then*

$$\psi_1(p_1 p_2 \dots p_n, S) \leq (p_1 p_2 \dots p_n)^{d+o(1)}.$$

*Proof.* Clearly

$$\prod_{p \in S} p^{\nu_p(p_1 p_2 \dots p_n)} = \prod_{\substack{p \in S \\ p \leq p_n}} p \leq \prod_{\substack{p \in S \\ p \leq p_n}} p_n \leq p_n^{(d+o(1))n}.$$

On the other hand, for every  $\varepsilon > 0$

$$p_1 \dots p_n \geq (p_n^{1-\varepsilon})^{n-p_n^{1-\varepsilon}} = p_n^{(1-\varepsilon)(n-o(n))}.$$

This completes the proof.

**Lemma 7.4.**

1.  $\nu_p([1, 2, \dots, n]) \geq \nu_p(p_1 p_2 \dots p_{\pi(n)})$  for  $n \in \mathbf{N}$  and  $p \in \mathbf{P}$ .
2.  $[1, 2, \dots, n] = (p_1 p_2 \dots p_{\pi(n)})^{1+o(1)}$ .

*Proof.* It is obvious that  $\nu_p([1, 2, \dots, n]) \geq \nu_p(p_1 p_2 \dots p_{\pi(n)})$ , and it follows that  $[1, 2, \dots, n] \geq (p_1 p_2 \dots p_{\pi(n)})$ . To obtain the upper bound for  $[1, 2, \dots, n]$ , observe that

$$\begin{aligned} \frac{[1, 2, \dots, n]}{p_1 p_2 \dots p_{\pi(n)}} &= \prod_{p \in \mathbf{P} \cap [1, \sqrt{n}]} p^{\lfloor \log_p n \rfloor - 1} \leq \prod_{p \in \mathbf{P} \cap [1, \sqrt{n}]} n \\ &= n^{\pi(\sqrt{n})} \leq n^{\sqrt{n}} = e^{o(n)} = (p_1 \dots p_{\pi(n)})^{o(1)}. \end{aligned}$$

(Using the Prime Number Theorem we could of course replace  $n^{\sqrt{n}}$  by  $e^{O(\sqrt{n})} = O(1)^{\sqrt{n}}$ .)

**Proposition 7.3.** *Let  $S \subseteq \mathbf{P}$  with  $d^*(S) = d$ . Then*

$$\psi_1([1, 2, \dots, n], S) \leq [1, 2, \dots, n]^{d+o(1)}.$$

*Proof.* By Proposition 7.2 and Lemma 7.4,

$$\begin{aligned} \psi_1([1, 2, \dots, n], S) &= \frac{[1, 2, \dots, n]}{\psi_1([1, 2, \dots, n], \mathbf{P} \setminus S)} \\ &\leq \frac{(p_1 p_2 \dots p_{\pi(n)})^{1+o(1)}}{\psi_1(p_1 p_2 \dots p_{\pi(n)}, \mathbf{P} \setminus S)} \\ &= (p_1 p_2 \dots p_{\pi(n)})^{o(1)} \psi_1(p_1 p_2 \dots p_{\pi(n)}, S) \\ &\leq (p_1 p_2 \dots p_{\pi(n)})^{d+o(1)} \\ &\leq [1, 2, \dots, n]^{d+o(1)}. \end{aligned}$$

We now prove results related to  $\gamma_*(S, \alpha)$  and  $\gamma^*(S, \alpha)$  needed for Sections 5 and 6.

**Lemma 7.5.** *Suppose  $S \subseteq \mathbf{P}$  and  $0 < \alpha < 1$ . Then*

$$\begin{aligned} \gamma_*(S, \alpha) &\leq \liminf_{\substack{(x, x^{-\alpha}t) \rightarrow (\infty, \infty) \\ t \leq x}} \frac{\#(S \cap (x, x+t])}{\frac{t}{\log x}} \\ &\leq \limsup_{\substack{(x, x^{-\alpha}t) \rightarrow (\infty, \infty) \\ t \leq x}} \frac{\#(S \cap (x, x+t])}{\frac{t}{\log x}} \leq \gamma^*(S, \alpha). \end{aligned}$$

*In particular,  $\gamma^*(S, \alpha)$  is non-increasing with  $\alpha$  and  $\gamma_*(S, \alpha)$  is non-decreasing with  $\alpha$ .*

*Proof.* Let  $\varepsilon > 0$  be given. Choose  $M > 0$  so that for  $x \geq M$ ,

$$\gamma_*(S, \alpha) - \varepsilon \leq \frac{\#(S \cap (x, x+x^\alpha])}{\frac{x^\alpha}{\log x}} \leq \gamma^*(S, \alpha) + \varepsilon.$$

Take  $x, t$  with  $x > \max\{M, 2^{1/\varepsilon}\}$ ,  $x^{-\alpha}t > 1/\varepsilon$ , and  $t \leq x$ . Define a sequence  $(x_k)$  as follows:  $x_0 = x$  and  $x_{k+1} = x_k + x_k^\alpha$ . Choose  $n$  so that  $x_n \leq x+t < x_{n+1}$ . Then

$$\begin{aligned} \#(S \cap (x, x+t]) &\leq \#(S \cap (x_0, x_{n+1}]) = \sum_{k=0}^n \#(S \cap (x_k, x_{k+1}]) \\ &\leq (\gamma^*(S, \alpha) + \varepsilon) \sum_{k=0}^n \frac{x_{k+1} - x_k}{\log x_k} \\ &\leq (\gamma^*(S, \alpha) + \varepsilon) \sum_{k=0}^n \frac{x_{k+1} - x_k}{\log x} \\ &= (\gamma^*(S, \alpha) + \varepsilon) \frac{x_{n+1} - x_0}{\log x} \\ &\leq (\gamma^*(S, \alpha) + \varepsilon) \frac{(x+t) + (x+t)^\alpha - x}{\log x} \\ &\leq (\gamma^*(S, \alpha) + \varepsilon) \frac{(1+2\varepsilon)t}{\log x}. \end{aligned}$$

Similarly (provided  $\varepsilon < \min\{\gamma_*(S, \alpha), \frac{1}{2}\}$ ),

$$\begin{aligned} \#(S \cap (x, x+t]) &\geq \#(S \cap (x_0, x_n]) = \sum_{k=0}^{n-1} \#(S \cap (x_k, x_{k+1}]) \\ &\geq (\gamma_*(S, \alpha) - \varepsilon) \sum_{k=0}^{n-1} \frac{x_{k+1} - x_k}{\log x_k} \\ &\geq (\gamma_*(S, \alpha) - \varepsilon) \sum_{k=0}^{n-1} \frac{x_{k+1} - x_k}{\log 2x} \\ &= (\gamma_*(S, \alpha) - \varepsilon) \frac{x_n - x_0}{\log x + \log 2} \\ &\geq (\gamma_*(S, \alpha) - \varepsilon) \frac{(x+t) - (x+t)^\alpha - x}{(1+\varepsilon)\log x} \\ &\geq (\gamma_*(S, \alpha) - \varepsilon) \frac{(1-2\varepsilon)t}{(1+\varepsilon)\log x}. \end{aligned}$$

The monotonicity follows from the first part: for  $0 < \alpha < \beta \leq 1$  take  $t = x^\beta$  to prove that  $\gamma_*(S, \alpha) \leq \gamma_*(S, \beta) \leq \gamma^*(S, \beta) \leq \gamma^*(S, \alpha)$ .

**Proposition 7.4.**

$$\eta_h \leq \begin{cases} 0.5 + 0.5099h + \frac{h}{2} \log \frac{1-h}{h}, & 0 \leq h \leq \frac{81}{446}, \\ 0.3891 + 0.3678h + h \log \frac{1-h}{h}, & \frac{81}{446} \leq h \leq \frac{9}{49}, \\ 0.5016 + 0.5012h + \frac{h}{2} \log \frac{1-h}{h}, & \frac{9}{49} \leq h \leq \frac{1911}{7811}, \\ 0.3396 + 0.5997h + h \log \frac{1-h}{h}, & \frac{1911}{7811} \leq h \leq \frac{2457}{9757}, \\ 0.5079 + 0.4758h + \frac{h}{2} \log \frac{1-h}{h}, & \frac{2457}{9757} \leq h \leq \frac{21}{80}, \\ 0.6859 + 0.3143h, & \frac{21}{80} \leq h \leq \frac{1}{2}, \\ 0.6861 + 0.3139h, & \frac{1}{2} \leq h \leq 1. \end{cases}$$

*Proof.* Denote

$$f(\beta) = \min \left\{ h\gamma^* \left( \frac{1-2\beta}{1-\beta} \right), 1 - (1-h)\gamma_* \left( \frac{1-2\beta}{1-\beta} \right) \right\}.$$

By Lemma 7.5,  $f$  is increasing, and hence Riemann integrable. By (5.8)

$$(7.1) \quad f(\beta) = h, \quad 0 \leq \beta \leq \frac{5}{17}.$$

For  $\frac{5}{17} \leq \beta < \frac{9}{29}$ , (5.9) yields

$$(7.2) \quad f(\beta) \leq \begin{cases} 1.01h, & 0 \leq h \leq 1/2, \\ 0.01 + 0.99h, & 1/2 \leq h \leq 1. \end{cases}$$

For  $\frac{9}{29} \leq \beta < \frac{5}{16}$ , we have by (5.10)

$$(7.3) \quad f(\beta) \leq \begin{cases} 1.031h, & 0 \leq h \leq 1/2, \\ 0.031 + 0.969h, & 1/2 \leq h \leq 1. \end{cases}$$

By (5.11), in the region  $\frac{5}{16} \leq \beta < \frac{23}{73}$ ,

$$(7.4) \quad f(\beta) \leq \min \left\{ \frac{2h(1-\beta)}{1-2\beta}, 1 - (1-h) \cdot 0.4 \right\} \\ = \begin{cases} \frac{2h(1-\beta)}{1-2\beta}, & 0 \leq h \leq \frac{81}{446}, \\ \frac{2h(1-\beta)}{1-2\beta}, & \frac{81}{446} < h \leq \frac{9}{49}, \frac{5}{16} \leq \beta < \frac{3-8h}{6-6h}, \\ 0.6 + 0.4h, & \frac{81}{446} < h \leq \frac{9}{49}, \frac{3-8h}{6-6h} \leq \beta < \frac{23}{73}, \\ 0.6 + 0.4h, & \frac{9}{49} < h. \end{cases}$$

Next, for  $\frac{23}{73} \leq \beta < \frac{19}{59}$ ,

$$(7.5) \quad f(\beta) \leq \min \left\{ \frac{2h(1-\beta)}{1-2\beta}, 1 - (1-h) \cdot 0.09 \right\} \\ = \begin{cases} \frac{2h(1-\beta)}{1-2\beta}, & 0 \leq h \leq \frac{1911}{7811}, \\ \frac{2h(1-\beta)}{1-2\beta}, & \frac{1911}{7811} < h \leq \frac{2457}{9757}, \frac{23}{73} \leq \beta \leq \frac{91-191h}{182-182h}, \\ 0.91 + 0.09h, & \frac{1911}{7811} < h \leq \frac{2457}{9757}, \frac{91-191h}{182-182h} < \beta \leq \frac{19}{59}, \\ 0.91 + 0.09h, & \frac{2457}{9757} < h. \end{cases}$$

Finally, for  $\frac{19}{59} \leq \beta \leq \frac{1}{2}$ ,

$$(7.6) \quad f(\beta) \leq \min \left\{ \frac{2h(1-\beta)}{(1-2\beta)}, 1 \right\} \\ = \begin{cases} \frac{2h(1-\beta)}{(1-2\beta)}, & 0 \leq h \leq \frac{21}{80}, \frac{19}{59} \leq \beta \leq \frac{1-2h}{2-2h}, \\ 1, & 0 \leq h \leq \frac{21}{80}, \frac{1-2h}{2-2h} \leq \beta \leq \frac{1}{2}, \\ 1, & \frac{21}{80} \leq h \leq 1. \end{cases}$$

From (7.1) – (7.6) we infer

$$\eta_h \leq \begin{cases} \frac{1}{2} + \frac{7889807h}{7888000} - \frac{h}{2} \log \frac{8}{3} + \frac{h}{2} \log \frac{1-h}{h}, & 0 \leq h \leq \frac{81}{446}, \\ \frac{142}{365} + \frac{927760711h}{575824000} - \frac{h}{2} \log \frac{2920}{243} + h \log \frac{1-h}{h}, & \frac{81}{446} \leq h \leq \frac{9}{49}, \\ \frac{2929}{5840} + \frac{575068511h}{575824000} - \frac{h}{2} \log \frac{73}{27} + \frac{h}{2} \log \frac{1-h}{h}, & \frac{9}{49} \leq h \leq \frac{1911}{7811}, \\ \frac{585049}{1722800} + \frac{56417848469h}{33973616000} - \frac{h}{2} \log \frac{430700}{51597} + h \log \frac{1-h}{h}, & \frac{1911}{7811} \leq h \leq \frac{2457}{9757}, \\ \frac{34999}{68912} + \frac{33713699749h}{33973616000} - \frac{h}{2} \log \frac{59}{21} + \frac{h}{2} \log \frac{1-h}{h}, & \frac{2457}{9757} \leq h \leq \frac{21}{80}, \\ \frac{47263}{68912} + \frac{10680739749h}{33973616000}, & \frac{21}{80} \leq h \leq \frac{1}{2}, \\ \frac{23308441749}{33973616000} + \frac{10665174251h}{33973616000}, & \frac{1}{2} \leq h \leq 1, \end{cases}$$

and the proposition follows.

Lemma 7.6 connects a sum which will arise in the proof of Proposition 6.1 to  $\delta(A)$ .

**Lemma 7.6.** *Let  $(a_N)_{N=1}^\infty, (b_N)_{N=1}^\infty$  be sequences in  $\mathbf{N}$  with  $a_N = o(b_N)$ . Define  $F$  on  $\mathcal{P}(\mathbf{N}) \times \mathbf{N}$  by*

$$F(A, N) = \sum_{\substack{a_N < j \leq b_N \\ j \in A}} \frac{1}{j}, \quad A \subseteq \mathbf{N}, N \in \mathbf{N}.$$

Then

$$\delta_*(A) \leq \liminf_{N \rightarrow \infty} \frac{F(A, N)}{F(\mathbf{N}, N)} \leq \limsup_{N \rightarrow \infty} \frac{F(A, N)}{F(\mathbf{N}, N)} \leq \delta^*(A), \quad A \subseteq \mathbf{N}.$$

*Proof.* The last inequality is of course the result of applying the first to  $\mathbf{N} \setminus A$ , so it suffices to prove the first.

Take any  $\varepsilon$  strictly between 0 and  $\delta_*(A)$ , and take an integer  $M > 1/\varepsilon$  such that  $\#(A \cap [1, x]) \geq \varepsilon \#(\mathbf{N} \cap [1, x])$  for  $x \geq M$ . Suppose  $N \in \mathbf{N}$  is so large that  $Ma_N < b_N$ . Then for  $x \geq Ma_N$ ,  $\#(A \cap (a_N, x]) = \#(A \cap [1, x]) - \#(A \cap [1, a_N]) \geq \varepsilon \cdot \#(\mathbf{N} \cap [1, x]) - a_N \geq (\varepsilon - \frac{1}{M}) \cdot \#(\mathbf{N} \cap [1, x]) > (\varepsilon - \frac{1}{M}) \cdot \#(\mathbf{N} \cap (a_N, x])$ . We

now apply Lemma 7.3 with  $\mu_1(X) = (\varepsilon - \frac{1}{M}) \cdot \#(X \cap \mathbf{N} \cap (a_N, b_N])$ ,  $\mu_2(X) = \#(X \cap A \cap (a_N, b_N]) + \#(X \cap \mathbf{N} \cap (a_N, Ma_N])$ , and  $f(t) = \frac{1}{t}$ . This shows that

$$F(A, N) + \ln M \geq \left(\varepsilon - \frac{1}{M}\right) F(\mathbf{N}, N).$$

Since  $F(\mathbf{N}, N) \rightarrow \infty$  as  $N \rightarrow \infty$ , it follows that

$$\liminf_{N \rightarrow \infty} \frac{F(A, N)}{F(\mathbf{N}, N)} \geq \left(\varepsilon - \frac{1}{M}\right).$$

$M$  may be taken arbitrarily large, and  $\varepsilon$  can be arbitrarily close to  $\delta_*(A)$  (unless  $\delta_*(A) = 0$ ), so this completes the proof.

We now prove Proposition 6.1. To make it clear in the sequel which results have been proved, we restate the proposition.

**Proposition 7.5.** *In the setup of Proposition 5.1, for  $S \subseteq \mathbf{P}$ ,*

$$\begin{aligned} (n!)^{(\beta-\beta_1)\gamma_*(S,\alpha)\delta_*(A)+o(1)} &\leq \psi(n!, A, S \cap (n^{1-\beta}, n^{1-\beta_1}]) \\ &\leq (n!)^{(\beta-\beta_1)\gamma^*(S,\alpha)\delta^*(A)+o(1)}. \end{aligned}$$

*Proof.* Choose  $\beta_2$  strictly between  $\beta_1$  and  $\beta$ . Since, by Proposition 7.1,

$$\psi_1(n!, (n^{1-\beta}, n^{1-\beta_2}]) = (n!)^{\beta-\beta_2+o(1)},$$

it suffices to prove the inequalities with  $\beta_2$  in place of  $\beta$ .

For  $p \in \mathbf{P}$ , if  $n^{1-\beta_2} < p \leq n$ , then  $\nu_p(n!) = \lfloor \frac{n}{p} \rfloor < n^{\beta_2}$ . Thus  $\nu_p(n!)$  equals a specific positive integer  $c (< n^{\beta_2}) \iff p \in (\frac{n}{c+1}, \frac{n}{c}]$ . Now

$$\begin{aligned} \frac{n}{c(c+1)} &= \frac{c}{c+1} \frac{n^{1-\alpha}}{c^{2-\alpha}} \left(\frac{n}{c}\right)^\alpha \geq \left(\frac{c}{c+1}\right)^{1-\alpha} n^{1-\alpha-(2-\alpha)\beta_2} \left(\frac{n}{c+1}\right)^\alpha \\ &= \left(\frac{c}{c+1}\right)^{1-\alpha} n^{(2-\alpha)(\beta-\beta_2)} \left(\frac{n}{c+1}\right)^\alpha. \end{aligned}$$

Thus for such  $c$  by Lemma 7.5,

$$\gamma_*(S, \alpha) + o(1) \leq \frac{\#(S \cap (\frac{n}{c+1}, \frac{n}{c}])}{\frac{\frac{n}{c(c+1)}}{\log \frac{n}{c+1}}} \leq \gamma^*(S, \alpha) + o(1),$$

as  $n \rightarrow \infty$ . (For fixed  $n$ , let  $E_n$  be the largest error term in the inequalities as  $c$  ranges over the set of admissible values. Then  $E_n$  tends to zero as  $n$  tends to infinity.)

Of course, for  $p \in (\frac{n}{c+1}, \frac{n}{c}]$ ,

$$\log \frac{n}{c+1} < \log p \leq \log \frac{n}{c}.$$

By these results and Lemma 7.6,

$$\begin{aligned}
 \log (\psi(n!, A, S \cap (n^{1-\beta_2}, n^{1-\beta_1}))) &\geq \log \prod_{\substack{n^{\beta_1} < c < \lfloor n^{\beta_2} \rfloor - 1 \\ c \in A}} \prod_{\substack{\frac{n}{c+1} < p \leq \frac{n}{c} \\ p \in S}} p^c \\
 &\geq \sum_{\substack{n^{\beta_1} < c < \lfloor n^{\beta_2} \rfloor - 1 \\ c \in A}} (\gamma_*(S, \alpha) + o(1)) \\
 &\quad \cdot \frac{\frac{n}{c(c+1)}}{\log \frac{n}{c+1}} \log \left( \frac{n}{c+1} \right) c \\
 &= (\gamma_*(S, \alpha) + o(1)) \sum_{\substack{n^{\beta_1} < c < \lfloor n^{\beta_2} \rfloor - 1 \\ c \in A}} \frac{n}{c+1} \\
 &\geq (\gamma_*(S, \alpha) + o(1)) (\delta_*(A) + o(1)) \\
 &\quad \cdot \sum_{c=\lfloor n^{\beta_1} \rfloor + 1}^{\lfloor n^{\beta_2} \rfloor - 2} \frac{n}{c+1} \\
 &= (\gamma_*(S, \alpha) \delta_*(A) + o(1)) n (\log n^{\beta_2} - \log n^{\beta_1}) \\
 &= ((\beta_2 - \beta_1) \gamma_*(S, \alpha) \delta_*(A) + o(1)) \log(n!).
 \end{aligned}$$

The upper bound is similar.

Applying Proposition 7.5 with  $S = \mathbf{P}$  to both the set  $A$  and its complement  $\mathbf{N} \setminus A$  and using Proposition 7.1, we obtain the following self-refinement.

**Proposition 7.6.** *In the setup of Proposition 5.1,*

$$\psi(n!, A, (n^{1-\beta}, n^{1-\beta_1})) \leq (n!)^{(\beta-\beta_1) \min\{\gamma^*(\alpha)\delta^*(A), 1-\gamma_*(\alpha)(1-\delta^*(A))\}} + o(1).$$

Of course, in the same way, one establishes an analogous lower bound for the product in question. In this formulation, however, the upper bound (applied to the complement of  $A$ ) implies the lower bound.

The last proposition gives rise to a further self-refinement, obtained by using it on small subintervals of  $[\beta_1, \beta]$  and applying the definition of the Riemann integral (and Lemma 7.5). This gives the best refinement (short of a different approach) for any possible behaviour of  $\gamma_*$  and  $\gamma^*$ . Note that the inverse of the function defining  $\beta$  in terms of  $\alpha$  in Proposition 5.1 is given by

$$\alpha = \alpha(\beta) = \frac{1 - 2\beta}{1 - \beta}, \quad 0 < \beta < \frac{1}{2}.$$

A limiting process and an application of Proposition 7.1 give

**Proposition 7.7.** *For  $A \subseteq \mathbf{N}$  and  $0 \leq \beta_1 < \beta \leq \frac{1}{2}$ ,*

$$\psi(n!, A, (n^{1-\beta}, n^{1-\beta_1})) \leq (n!)^{\int_{\beta_1}^{\beta} \min\{\gamma^*\left(\frac{1-2\beta_2}{1-\beta_2}\right)\delta^*(A), 1-\gamma_*\left(\frac{1-2\beta_2}{1-\beta_2}\right)(1-\delta_*(A))\} d\beta_2 + o(1)}.$$

We now consider  $\binom{an}{n, n, \dots, n}$ . As before, we approximate  $\nu_p$  with something simpler.

**Lemma 7.7.** *Suppose  $a, n \in \mathbf{N}$  and  $p \in \mathbf{P}$ . Then:*

1. *If  $p > \sqrt{an}$ , then  $\nu_p \left( \binom{an}{n, n, \dots, n} \right) = \lfloor an/p \rfloor - a \lfloor n/p \rfloor \in \{0, 1, 2, \dots, a - 1\}$ .*
2. *In any case,  $\lfloor an/p \rfloor - a \lfloor n/p \rfloor \leq \nu_p \left( \binom{an}{n, n, \dots, n} \right) \leq (a - 1) \log_p an$ .*

*Proof.* Observe that for  $x \geq 0$ ,  $\lfloor ax \rfloor - a\lfloor x \rfloor \in \{0, 1, 2, \dots, a - 1\}$ . Both parts of the lemma now follow from the further observation that for  $k > \log_p an$  we have  $\lfloor an/p^k \rfloor - a\lfloor n/p^k \rfloor = 0 - 0 = 0$ .

**Lemma 7.8.**

$$\prod_{\substack{p \in \mathbf{P} \\ \nu_p\left(\binom{an}{n, n, \dots, n}\right) \neq \lfloor an/p \rfloor - a\lfloor n/p \rfloor}} p^{\nu_p\left(\binom{an}{n, n, \dots, n}\right)} = e^{o(n)}.$$

*Proof.* Now

$$\begin{aligned} \prod_{p \in \mathbf{P} \cap [1, \sqrt{an}]} p^{a \log_p(an)} &= \prod_{p \in \mathbf{P} \cap [1, \sqrt{an}]} (an)^a = (an)^{a\pi(\sqrt{an})} \\ &\leq (an)^{a\sqrt{an}} = \exp((\ln a + \ln n)a^{3/2}\sqrt{n}) = e^{o(n)}. \end{aligned}$$

In view of Lemma 7.7, this completes the proof. (We could obtain a better estimate by using the Prime Number Theorem to approximate  $\pi(\sqrt{an})$ .)

It is now convenient to restate (and prove) Proposition 6.2.

**Proposition 7.8.** *Let  $S$  be a naturally regular set of primes, let  $a \geq 2$  and let  $k \in \{1, 2, \dots, a - 1\}$ . Then*

$$\psi\left(\left(\binom{an}{n, n, \dots, n}\right), k, S\right) = \left(\binom{an}{n, n, \dots, n}\right)^{\frac{kd(S)\theta_{a,k}}{\log a} + o(1)}.$$

*Proof.* By Lemma 7.8, we may replace  $\nu_p\left(\binom{an}{n, n, \dots, n}\right)$  by  $\lfloor an/p \rfloor - a\lfloor n/p \rfloor$ . This quantity equals  $k$  if, and only if, there is  $j \in \mathbf{N} \cup \{0\}$  such that  $j + \frac{k}{a} \leq \frac{n}{p} < j + \frac{k+1}{a}$ , i.e.,  $p/n \in I_j = \left(\frac{a}{ja+k+1}, \frac{a}{ja+k}\right]$ . Clearly,

$$\log\left(\prod_{\substack{p \in S \\ \frac{n}{p} \in I_j}} p\right) = n\left(\frac{a}{ja+k} - \frac{a}{ja+k+1}\right)d(S) + o(n).$$

It remains to check that the limiting processes (limit in  $n$  and infinite summation in  $j$ ) commute. This is clear since for  $J \in \mathbf{N}$ ,

$$\begin{aligned} \limsup_{n \in \mathbf{N}} \frac{1}{n} \sum_{j=J+1}^{\infty} \log\left(\prod_{\substack{p \in S \\ \frac{n}{p} \in I_j}} p\right) &\leq \limsup_{n \in \mathbf{N}} \frac{1}{n} \sum_{\substack{p \in P \\ p \leq \frac{n}{J}}} \log p \\ &= \limsup_{n \in \mathbf{N}} \frac{1}{n} \left(\frac{n}{J} + o(n)\right) = \frac{1}{J}. \end{aligned}$$

**Lemma 7.9.** *For any  $a \geq 2$ :*

1.  $\theta_{a,k} = \frac{1}{a} \sum_{\xi \in R_a} (\xi^{k+1} - \xi^k) \log(1 - \xi^{-1})$  for  $1 \leq k < a$ , where  $R_a$  is the set of all  $a$ th roots of unity. (On the right-hand side we take  $0 \log 0 = 0$ .)
2.  $k\theta_{a,k} > (k+1)\theta_{a,k+1}$ ,  $1 \leq k < a$ .

*Proof.* 1. This part is certainly well known, and we shall only outline the proof. Put

$$S_{a,k}(x) = \sum_{j=0}^{\infty} \left( \frac{x^{ja+k}}{ja+k} - \frac{x^{ja+k+1}}{ja+k+1} \right).$$

Then  $\theta_{a,k} = S_{a,k}(1)$ . Now  $S'_{a,k}(x)$  is readily seen to be a difference of two geometric series, and thus is easily computed:

$$S'_{a,k}(x) = \frac{x^k - x^{k-1}}{x^a - 1}.$$

The denominator is a product of distinct linear factors (over  $p_1 p_2 \dots p_n$ ), and the representation of the function as a sum of partial fractions is therefore routine. (We use the fact that  $\xi \lim_{x \rightarrow \xi} \frac{x^a - 1}{x - \xi} = a$  for  $\xi \in R_a$ .) Integration yields the formula required.

2. The proof of the preceding part gives

$$\begin{aligned} k\theta_{a,k} - (k+1)\theta_{a,k+1} &= \int_0^1 k \frac{x^k - x^{k-1}}{x^a - 1} dx - (k+1) \int_0^1 \frac{x^{k+1} - x^k}{x^a - 1} dx \\ &= \int_0^1 \frac{kx^{k-1} - (k+1)x^k}{1+x+\dots+x^{a-1}} dx \\ &= \left[ \frac{x^k - x^{k+1}}{1+x+\dots+x^{a-1}} \right]_0^1 \\ &\quad + \int_0^1 \frac{(x^k - x^{k+1})(1+2x+\dots+(a-1)x^{a-2})}{(1+x+\dots+x^{a-1})^2} dx \\ &> 0. \end{aligned}$$

**Lemma 7.10.**  $\sum_{k=1}^{a-1} k\theta_{a,k} = \log a, a \geq 2$ .

*Proof.* This follows from part 1 of the preceding lemma when we change the order of summation and observe that  $\sum_{\xi \in R_a \setminus \{1\}} \log(1 - \xi^{-1}) = \log a$ .

**Proposition 7.9.** *Let  $S$  be a naturally regular set of primes. Then*

$$\prod_{p \in S} p^{\nu_p((n, n, \dots, n))} = \binom{an}{n, n, \dots, n}^{d(S)+o(1)}.$$

The proposition follows from Proposition 7.8 and Lemma 7.10.

The reader can no doubt guess the inequality required for Theorem 6.1.

**Proposition 7.10.** *Take  $A \subseteq \mathbf{N}$  with  $\delta^*(A) = \delta_*(A)$  and  $S \subseteq \mathbf{P}$ . Then*

$$\psi(n!, A, S) \leq (n!)^{\eta_{S, \delta(A)} + o(1)}.$$

*Proof.* By Proposition 7.1,

$$\psi(n!, A, S \cap [2, \sqrt{n}]) \leq (n!)^{d^*(S)/2 + o(1)}.$$

For  $0 \leq \beta_1 \leq \beta < 1/2$ ,

$$\psi(n!, A, S \cap (n^{1-\beta}, n^{1-\beta_1}]) \leq (n!)^{(\beta-\beta_1)\gamma^*(S, \alpha(\beta))\delta(A) + o(1)}$$

by Proposition 7.5. Using Proposition 7.1 and the lower bound in Proposition 7.5, we see that

$$\psi(n!, A, S \cap (n^{1-\beta}, n^{1-\beta_1}]) \leq (n!)^{(\beta-\beta_1)(d^*(S) - (1-\delta(A))\gamma_*(S, \alpha(\beta))) + o(1)}$$

and

$$\psi(n!, A, S \cap (n^{1-\beta}, n^{1-\beta_1}]) \leq (n!)^{(\beta-\beta_1)(1-(1-\delta(A))\gamma_*(\alpha(\beta))-\delta(A)\gamma_*(\mathbf{P} \setminus S, \alpha(\beta))) + o(1)}.$$

The bound for  $\psi(n!, A, S \cap (\sqrt{n}, n])$  follows as in the proof of Proposition 7.7.

## 8. PROOFS

We now prove the outstanding assertions from Sections 2, 4, 5 and 6 in approximately the order stated. Except for results from Section 2, this is a matter of citing results proved in Section 7 and using ideas already discussed.

*Proof of Proposition 2.1.* The proof is carried out by induction on  $l$ . The case  $l = 0$  reduces to the prime  $k$ -tuple conjecture. Assuming our claim to be true with  $l - 1$  instead of  $l$ , we shall prove it for  $l$ . Given  $a_1, \dots, a_k, b_1, \dots, b_l$ , select a prime  $p$  not dividing any of the numbers  $a_i - b_l$ ,  $1 \leq i \leq k$ . Choose numbers  $a_{k+1}, \dots, a_K$  so that the following requirements are satisfied:

1. The numbers  $a_1, \dots, a_K$  represent all congruence classes modulo  $p$ , except for the class of  $b_l$ .
2. For every  $1 \leq i \leq K$ ,  $1 \leq j \leq l$  we have  $a_i \not\equiv b_j$ .
3. The numbers  $a_1, \dots, a_K$  do not form a complete system of residues modulo any prime.

Such a choice is indeed possible since  $K - k$  can be taken as  $p - 1 - k'$ , where  $k'$  is the number of distinct congruence classes modulo  $p$  determined by  $a_1, \dots, a_k$ . Once  $K$  is chosen there are only finitely many primes which must be considered in requirement 3. Thus  $a_{k+1}, \dots, a_K$  may be chosen congruent to  $a_1$  modulo each of these primes satisfying requirements 1 and 2.

By the induction hypothesis there are infinitely many  $n$  such that  $n + a_1, n + a_2, \dots, n + a_K$  are all prime and  $n + b_1, n + b_2, \dots, n + b_{l-1}$  are all composite. Choose  $n$  satisfying these conditions so large that  $n + a_1, n + a_2, \dots, n + a_K, n + b_l > p$ . Then the first condition implies  $n + a_i \not\equiv 0 \pmod{p}$ . By requirement 1 this implies that  $n + b_l \equiv 0 \pmod{p}$ , so  $n + b_l$  is composite. The proposition follows.

*Proof of Proposition 2.2.* Let  $M, e, m_1, \dots, m_e$  satisfy the hypotheses of Conjecture 2.1. Adding  $m_i$ 's to the list, we may assume that the  $m_i$ 's yield (perhaps with repetitions) all residue classes modulo  $M$  which are relatively prime to  $M$ . Let  $d_0, d_1, \dots, d_{s-1}$  be all distinct numbers modulo  $M$  such that  $m_i + d_j$  is relatively prime to  $M$  for every  $1 \leq i \leq e$ ,  $0 \leq j \leq s - 1$ . (Thus, if  $M = q_1^{g_1} q_2^{g_2} \dots q_r^{g_r}$  is the prime power factorization of  $M$ , then  $d_j = jq_1 q_2 \dots q_r$  for  $0 \leq j \leq s - 1$ , where  $s = q_1^{g_1-1} q_2^{g_2-1} \dots q_r^{g_r-1}$ .) Using the Chinese remainder theorem we can find positive integers  $a_{ij}$  with  $a_{ij} \equiv m_i + d_j \pmod{M}$ ,  $1 \leq i \leq e$ ,  $0 \leq j \leq s - 1$ , satisfying the assumptions of the prime  $k$ -tuple conjecture, and such that

$$a_{10} < a_{20} < \dots < a_{e0} < a_{11} < a_{21} < \dots < a_{e1} < \dots < a_{e,s-1}.$$

Under the prime-composite  $(k, l)$ -tuple conjecture, there exist infinitely many positive integers  $n$  for which all numbers  $n + a_{ij}$  are prime, while all other integers between  $n + a_{10}$  and  $n + a_{e,s-1}$  are composite. For each such  $n$ , the numbers  $n + a_{1j}, n + a_{2j}, \dots, n + a_{ej}$  satisfy the conclusion of Conjecture 2.1 for some  $0 \leq j \leq s - 1$ .

*Proof of Theorem 2.2.* Set  $M = r'$  and choose  $m_1, m_2, \dots, m_e$  so that the products  $m_1, m_1m_2, m_1m_2m_3, \dots, m_1m_2 \dots m_e$  represent all the residues modulo  $M$  of numbers relatively prime to  $M$ . Choose  $n$  so that  $d \mid p_1 \dots p_n$  and  $p_{n+j} \equiv m_j \pmod{r'}$  for  $j \leq e$ . (Conjecture 2.1 and the assumptions imply that there are infinitely many such  $n$ .) Set  $t = p_1 \dots p_n/d$ , and observe that  $s' \equiv tp_{n+1} \dots p_{n+k} \pmod{r'}$  for some  $k, 1 \leq k \leq e$ . It follows that  $rx + s = p_1 \dots p_{n+k}$  for some integer  $x$ .

*Proof of Theorem 2.1.* Parts a), b) and c)1 are immediate. Part c)3 follows from Propositions 2.1 and 2.2 and Theorem 2.2.

c)2. Rewrite our equation in the form

$$r'x + s' = tp_{j+1} \dots p_n$$

for sufficiently large  $n$  (that is, such that no prime divisors of  $d$  exceed  $p_n$ ). Note that if  $\phi(r') = 1$ , then every large value of  $n$  yields a solution since  $s'$  and the right-hand side are both relatively prime to  $r'$ , and that the inequality is therefore correct in this case. We therefore assume that  $\phi(r') > 1$ . For large  $n$ , denote by  $s'_n$  the least non-negative residue of  $tp_{j+1} \dots p_n$  modulo  $r'$ . Let  $l$  be the number of values  $v$  such that  $s'_n = v$  for infinitely many  $n$ . We also consider the number  $L$  of pairs  $(v, w)$  with  $v \neq w$  such that  $s'_n = v$  and  $s'_{n+1} = w$  for infinitely many  $n$ . Clearly  $l(l-1) \geq L$ . On the other hand, each congruence class modulo  $r'$ , relatively prime to  $r'$ , contains infinitely many primes, so for  $2 \leq u < r', (u, r') = 1$ , there exist  $1 \leq v, w \leq r'$  with  $w \equiv uv \pmod{r'}$  such that  $s'_n = v$  and  $s'_{n+1} = w$  for infinitely many integers  $n$ . It follows that  $L \geq \phi(r') - 1$ , so  $l(l-1) \geq \phi(r') - 1$ . Solving the inequality we see that either  $l \geq \frac{1 + \sqrt{4\phi(r') - 3}}{2}$  or  $l \leq \frac{1 - \sqrt{4\phi(r') - 3}}{2}$ . Since  $\phi(r') > 1$ , the latter inequality implies that  $l < 0$  and may be excluded.

In the cases  $r' = 1, 2, 3, 4, 6$ , we have  $\phi(r') = 1$  or  $2$ , so the lower bound exceeds  $\phi(r') - 1$ , and hence for all  $s'$  there exist infinitely many solutions.

d)1. We may assume  $0 \leq s \leq r - 1$ . Suppose first  $s > 0$ . Take a prime  $p$  for which  $\nu_p(s) < \nu_p(r)$ . Put  $l = \nu_p(s) + 1$ . For any solution of our equation we have  $p^l \nmid \binom{an}{n, n, \dots, n}$ . Since  $\binom{2n}{n} \mid \binom{an}{n, n, \dots, n}$ , this means that  $p^l \nmid \binom{2n}{n}$ . By [San],

$$\# \left( \left\{ 1 \leq n \leq N : p^l \nmid \binom{2n}{n} \right\} \right) = O \left( N^{1 - \frac{\log 2}{\log p} + \frac{1}{p \log p}} \cdot (\log N)^{l-1} \right).$$

This clearly shows that the set of solutions  $n$  is of density 0. Since every  $n$  is a solution of exactly one of the equations obtained as  $s$  varies from 0 to  $r - 1$ , this shows that for  $s = 0$  the set of solutions is of density 1.

d)2. Select two primes  $p, q \in F$  not exceeding  $a$ . Set  $u = \nu_p(s), v = \nu_q(s)$ . From Lemma 7.1 we infer that, given any solution  $n$  of our equation, the base  $p$  expansion of  $n$  contains at most  $u$  non-zero digits and the base  $q$  expansion of  $n$  contains at most  $v$  non-zero digits. In view of [SenS], this implies that there exist only finitely many solutions  $n$ .

*Heuristic justification of Conjecture 2.3.* It will be more convenient to start with the second part. Since the  $b_j$ 's are relatively prime, one would expect the digits of a "randomly chosen" integer to be (statistically) independent. Now if a number  $n$  is to be chosen with up to  $l$  digits in base  $b$ , then the probability that all digits will belong to a certain set  $B \subseteq \{0, 1, \dots, b - 1\}$  is  $(\#(B)/b)^l$ , which for most numbers

$n$  in the range is approximately  $(\#(B)/b)^{\log_b n}$ . Thus the probability that for each  $j$  its base  $b_j$  expansion will consist of digits belonging to  $B_j$  only is about

$$\prod_{j=1}^g \left( \frac{\#(B_j)}{b_j} \right)^{\log_{b_j} n} = n^{\sum_{j=1}^g \log_{b_j} \#(B_j) - g} .$$

Summing these probabilities over all  $n$ , the series diverges if (2.5) is satisfied. Hence the Borel-Cantelli Lemma suggests that if (2.5) holds, then there exist infinitely many integers  $n$  with the property given.

The first part of the conjecture is similarly justified.

*Proof of Proposition 2.3.* Assume the condition in item 1 of Conjecture 2.2. Let  $n$  be a solution of (2.2). Then for every  $p \in F$  we have  $p^{\nu_p(s)+1} \nmid \binom{an}{n, n, \dots, n}$ . By Lemma 7.1d), this implies that the base  $p$  expansion of  $n$  contains at most  $\nu_p(s)$  occurrences of digits greater than or equal to  $p/a$ . Letting  $C = \max\{\nu_p(s) : p \in F\}$ , this yields

$$d_{p,C}(n) \leq \left\lceil \frac{p}{a} \right\rceil, \quad p \in F .$$

It follows that

$$\sum_{p \in F} \log_p d_{p,C}(n) \leq \sum_{p \in F} \log_p \left\lceil \frac{p}{a} \right\rceil < \#(F) - 1 .$$

Assuming item 1 of Conjecture 2.3, the set of numbers  $n$  satisfying this inequality is necessarily finite, and therefore (2.2) has only finitely many solutions.

*Proof of Proposition 2.4.* The first part is the contents of Theorem 1.1 of [BereH]. For the second part, we first recall that, by an old result of Kummer [K],  $\nu_2 \left( \binom{2n}{n} \right)$  is exactly the number of ones in the binary expansion of  $n$ . Hence:

- i) All numbers  $n$  with at least  $l$  ones in their binary expansion solve the equation with  $s = 0$ .
- ii)  $\binom{2n}{n}$  is even for every  $n > 0$ , and therefore there are no solutions with odd values of  $s$ .
- iii) Only numbers  $n$  which are powers of 2 solve the equation for values of  $s$  which are divisible by 2 but not by 4. Since by [BereH, Prop. 2.1] the sequence  $\left( \binom{2^{k+1}}{2^k} \right)_{k=1}^\infty$  converges in the ring of 2-adic integers, for each  $l$  only one of the values  $s = 2, 6, 10, \dots, 2^l - 2$  is obtained infinitely often modulo  $2^l$ .

It remains to deal with the case  $l > \nu_2(s) \geq 2$ . Put  $t = \nu_2(s)$ . According to the above, the only numbers  $n$  which may solve our equation are of the form  $n = 2^{e_1} + 2^{e_2} + \dots + 2^{e_t}$ , where  $e_1 > e_2 > \dots > e_t \geq 0$ . Denote  $D(n) = (e_1 - e_2, e_2 - e_3, \dots, e_t - e_{t+1})$ , where we put  $e_{t+1} = 0$ . Call numbers  $n$  and  $n'$  with  $D(n) = (d_1, d_2, \dots, d_t)$  and  $D(n') = (d'_1, d'_2, \dots, d'_t)$  equivalent if for each  $1 \leq i \leq t$  we have either  $d_i = d'_i$  or  $d_i, d'_i \geq l - 3$ . According to [DaviW, Th. 2], if  $n$  and  $n'$  are equivalent, then  $\binom{2n}{n} \equiv \binom{2n'}{n'} \pmod{2^l}$ . (In fact, the equivalence means that one can get from  $n$  to  $n'$  by successive changes, in each of which we insert a zero to a block of zeros of length at least  $l - 3$  or delete a zero from a block of zeros of length at least  $l - 2$ . The formula in the mentioned theorem ensures that each such change does not change the value of the binomial coefficient in question modulo  $2^l$ .) If the equation  $2^l x + s = \binom{2n}{n}$  has infinitely many solutions, then it has a solution  $n$ , with at least one of the components of  $D(n)$  being  $l - 3$  or more. Hence it suffices to go over all numbers  $n$  with exactly  $t$  ones in their binary expansion, and with the

maximal component of  $D(n)$  being exactly  $l - 3$ . If one of them solves the equation, then there exist infinitely many solutions, while otherwise there are at most finitely many solutions. This proves the proposition.

*Proof of Proposition 4.2.* The second inequality follows from Propositions 7.1, 7.2 and 7.3. The first is proved by passing to complements.

*Proof of Proposition 4.1.* For the three sequences  $(H_n)$  given in (1.5.a)–(1.5.c), this follows from Proposition 4.2. For  $H_n = \binom{an}{n, n, \dots, n}$ , this is the content of Proposition 7.9.

*Proof of Theorem 4.1.* Clearly,  $|Q(x)| \sim |P(x)|^{\deg(Q)/\deg(P)}$  as  $|x| \rightarrow \infty$ . If  $P(x) = H_n$ , then  $Q(x) | \prod_{p \in S(Q)} p^{\nu_p(H_n)} = \psi_1(H_n, S(Q))$ . The result follows from Proposition 4.1.

*Proof of Lemma 5.1.* This is contained in Lemma 7.5.

*Proof of Lemma 5.2.* Follows from Lemma 7.9 and Lemma 7.10.

*Proof of Proposition 5.1.* This is a special case of Proposition 7.5.

*Proof of Proposition 5.2.* Part 1 is a special case of Proposition 7.8. Part 2 follows from Part 1 as well as Part 3 of Lemma 5.2.

*Proof of Proposition 5.3.* Follows from Proposition 7.4, above.

*Proof of Theorem 5.2.* a) By Proposition 7.1 and Proposition 7.7 (with  $A = r\mathbf{N}$ ),

$$\psi_2(n!, r\mathbf{N}) \leq (n!)^{\eta_{1/r} + o(1)}.$$

From Proposition 5.3 it follows that

$$(8.1) \quad \psi_2(n!, r\mathbf{N}) \leq (n!)^{\lambda_r + o(1)}.$$

On the other hand, since the polynomials  $\prod_{i:r|e_i} P_i^{e_i}$  and  $\prod_{i:r \nmid e_i} P_i^{e_i}$  are relatively prime, there exist only finitely many primes that can divide the values of both of these polynomials at any point  $x$ . Hence there exists a finite set  $F \subset \mathbf{P}$  such that the equality  $P(x) = n!$  implies

$$\prod_{i:r|e_i} P_i(x)^{e_i} \mid \psi_2(n!, r\mathbf{N}) \psi_1(n!, F).$$

Now  $\left| \prod_{i:r|e_i} P_i(x)^{e_i} \right| \sim |P(x)|^{\frac{1}{\deg P} \sum_{i:r|e_i} e_i \deg P_i}$  as  $|x| \rightarrow \infty$ , so that using Proposition 7.1 we have

$$(8.2) \quad \psi_2(n!, r\mathbf{N}) \geq (n!)^{\frac{1}{\deg P} \sum_{i:r|e_i} e_i \deg P_i + o(1)}.$$

The required result follows from (8.1) and (8.2).

b) Follows immediately from Lemma 7.4.

c) Trivial.

d) By Proposition 7.8 we may neglect a finite set of primes. Our claim then follows from Proposition 5.2 (by estimating the contribution from primes which appear to a power which is a multiple of  $r$ ).

The theorem follows.

*Proof of Theorem 5.3.* The result follows as in the proof of Theorem 5.2. Part a) is simpler because we do not need Proposition 5.3.

*Proof of Theorem 5.4.* Under (5.13) or (5.14) we obtain an improved version of Proposition 5.3. The first two parts follow from Theorem 5.3 (in the special case  $s = 1$ ). The third part follows from the ideas used to prove Theorem 5.3.

*Proof of Lemma 6.1.* This follows from Lemma 7.5.

*Proof of Proposition 6.1.* This is Proposition 7.5, above.

*Proof of Proposition 6.2.* This is Proposition 7.8, above.

*Proof of Theorem 6.1.* This is similar to the proof of Theorem 5.3. The upper bounds on parts of  $H_n$  follow from Proposition 7.10 and Proposition 6.2.

#### APPENDIX A. SOME PARAMETERS

TABLE 1. Approximate Values of  $\theta_{a,k}$

	1	2	3	4	5	6	7	8
2	.6931							
3	.6046	.2470						
4	.5660	.2194	.1272					
5	.5455	.2042	.1151	.0776				
6	.5333	.1948	.1075	.0713	.0523			
7	.5255	.1885	.1024	.0669	.0485	.0376		
8	.5201	.1842	.0988	.0639	.0459	.0352	.0284	
9	.5163	.1810	.0961	.0615	.0438	.0334	.0267	.0221

#### APPENDIX B. NOTATION

In this appendix,  $x$  is a real number,  $S$  a set of primes,  $T$  any set,  $h, \alpha \in [0, 1]$ ,  $A$  a set of positive integers,  $p$  a prime, and  $a, k$  and  $N$  positive integers.

$\{x\}$  is the fractional part of  $x$ .

$\#(T)$  is the cardinality of a finite set  $T$ .

$\mathbf{P}$  is the set of all primes.  $\pi(x, S) = \#(S \cap [1, x])$  and  $\pi(x) = \pi(x, \mathbf{P})$ .

$$d_*(S) = \liminf_{x \rightarrow \infty} \frac{\pi(x, S)}{\pi(x)},$$

$$\gamma_*(S, \alpha) = \liminf_{x \rightarrow \infty} \frac{\#(S \cap (x, x + x^\alpha])}{\frac{x^\alpha}{\log x}},$$

$$\gamma_*(\alpha) = \gamma_*(\mathbf{P}, \alpha),$$

$$\delta_*(A) = \liminf_{N \rightarrow \infty} \frac{\#(A \cap [1, N])}{N}.$$

The functions  $d^*$ ,  $\gamma^*$  and  $\delta^*$  are defined similarly, with lim sup in place of lim inf. Where the limit exists, the ornament may be omitted.

$$\theta_{a,k} = \sum_{j=0}^{\infty} \left( \frac{1}{ja+k} - \frac{1}{ja+k+1} \right).$$

$\nu_p(N) = k - 1$  if  $p^{k-1} \mid N$  and  $p^k \nmid N$ .

$$\eta_{S,h} = \frac{d^*(S)}{2} + \int_0^{1/2} \min\{h\gamma^*(S, \alpha(\beta)), d^*(S) - (1-h)\gamma_*(S, \alpha(\beta)), 1 - (1-h)\gamma_*(\alpha(\beta)) - h\gamma_*(\mathbf{P} \setminus S, \alpha(\beta))\} d\beta,$$

where  $\alpha(\beta) = \frac{1-2\beta}{1-\beta}$ .

$$\eta_h = \eta_{\mathbf{P},h} = \frac{1}{2} + \int_0^{1/2} \min\left\{h\gamma^*\left(\frac{1-2\beta}{1-\beta}\right), 1 - (1-h)\gamma_*\left(\frac{1-2\beta}{1-\beta}\right)\right\} d\beta,$$

$$\psi(N, A, T) = \prod_{\substack{p \in T \cap \mathbf{P} \\ \nu_p(N) \in A}} p^{\nu_p(N)},$$

$$\psi_1(N, T) = \psi(N, \mathbf{N}, T),$$

$$\psi_2(N, A) = \psi(N, A, \mathbf{P}).$$

If  $k$  is the only element of  $A$ , we write  $\psi(N, k, T)$  or  $\psi_2(N, k)$ .

## REFERENCES

- [A] J. Ax, Solving diophantine problems modulo every prime, *Ann. of Math.* **85** (1967), 161–183. MR0209224 (35:126)
- [Ba] P. Bachmann, *Niedere Zahlentheorie*, Vol. I, B. G. Teubner, Leipzig – Berlin, 1921. MR0238661 (39:25)
- [BakH] R. C. Baker and G. Harman, The difference between consecutive primes, *Proc. London Math. Soc.* (3) **72** (1996), 261–280. MR1367079 (96k:11111)
- [BakHP] R. C. Baker, G. Harman and J. Pintz, The difference between consecutive primes, II, *Proc. London Math. Soc.* (3) **83** (2001), 532–562. MR1851081 (2002f:11125)
- [BatH] P. T. Bateman and R. A. Horn, A heuristic asymptotic formula concerning the distribution of prime numbers, *Math. Comp.* **16** (1962), 363–367. MR0148632 (26:6139)
- [Be] D. Berend, On the parity of exponents in the factorization of  $n!$ , *J. Number Theory* **64** (1997), 13–19. MR1450483 (98g:11019)
- [BereB] D. Berend and Y. Bilu, Polynomials with roots modulo every integer, *Proc. Amer. Math. Soc.* **124** (1996), 1663–1671. MR1307495 (96h:11107)
- [BereH] D. Berend and J. E. Harmse, On some arithmetical properties of middle binomial coefficients, *Acta Arithmetica* **84** (1998), 31–41. MR1613294 (99a:11018)
- [BereO] D. Berend and C. F. Osgood, On the equation  $P(x) = n!$  and a question of Erdős, *J. Number Theory* **42** (1992), 189–193. MR1183375 (93e:11016)
- [BernG] B. Berndt and W. F. Galway, On the Brocard-Ramanujan Diophantine equation  $n! + 1 = m^2$ , *Ramanujan J.* **4** (2000), 41–42. MR1754629 (2001a:11044)
- [BoS] Z. I. Borevich and I. R. Shafarevich, *Number Theory*, Academic Press, New York, 1966. MR0195803 (33:4001)
- [Br1] H. Brocard, Question 166, *Nouv. Corresp. Math.* **2** (1876), 287.
- [Br2] H. Brocard, Question 1532, *Nouv. Ann. Math.* (3) **4** (1885), 391.
- [CaF] J. W. S. Cassels and A. Fröhlich, *Algebraic Number Theory (Proceedings of an Instructional Conference, University of Sussex, 1965)*, Academic Press and St. Edmundsbury Press, Suffolk, 1990. MR0911121 (88h:11073)
- [Cr] H. Cramér, On the order of magnitude of the difference between consecutive prime numbers, *Acta Arithmetica* **2** (1937), 23–46.
- [Dave] H. Davenport, *Multiplicative Number Theory*, 2nd ed., Springer-Verlag, New York, 1980. MR0606931 (82m:10001)
- [DaviW] K. Davis and W. Webb, A binomial coefficient congruence modulo prime powers, *J. Number Theory* **43** (1993), 20–23. MR1200804 (93m:11016)
- [Di] L. E. Dickson, *History of the Theory of Numbers*, Vol. II, Chelsea Pub. Co., New York, 1966. MR0245500 (39:6807b)

- [E1] P. Erdős, Some new problems and results in number theory, *Number Theory, Proceedings, Mysore 1981*, K. Alladi – ed., Springer-Verlag Lecture Notes #938, Berlin, 1981, 50–74. MR0665438 (84g:10002)
- [E2] P. Erdős, On some of my problems in number theory I would most like to see solved, *Number Theory, Proceedings, Ootacamund, India 1984*, Springer-Verlag Lecture Notes #1122, Berlin, 1985, 74–84. MR0797781
- [E3] P. Erdős, Some problems and results in number theory, *Number Theory and Combinatorics, Japan 1984*, World Scientific, Singapore, 1985, 65–87. MR0827779 (87g:11003)
- [EG] P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, L'Enseignement Mathématique, Imprimerie Kundig, Geneva, 1980. MR0592420 (82j:10001)
- [EGRS] P. Erdős, R. L. Graham, I. Z. Ruzsa and E. G. Straus, On the prime factors of  $\binom{2n}{n}$ , *Math. Comp.* **29** (1975), 83–92. MR0369288 (51:5523)
- [EO] P. Erdős and R. Obláth, Über diophantische Gleichungen der Form  $n! = x^p \pm y^p$  und  $n! \pm m! = x^p$ , *Acta Szeged* **8** (1937), 241–255.
- [FrS] M. Fried and G. Sacerdote, Solving diophantine problems over all residue class fields of an algebraic number field and all finite fields, *Ann. Math.* **104** (1976), 203–233. MR0491477 (58:10722)
- [FU] H. Furstenberg, Intersections of Cantor sets and transversality of semigroups, *Problems in Analysis* (R. C. Gunning, general ed.), Princeton University Press, Princeton, NJ, 1970, pp. 41–59. MR0354562 (50:7040)
- [Go] L. J. Goldstein, *Analytic Number Theory*, Prentice-Hall, Englewood Cliffs, 1971. MR0498335 (58:16471)
- [Gr] R. Graham, Personal communication.
- [Gu1] R. K. Guy, *Unsolved Problems in Number Theory*, Springer-Verlag, New York, 1981. MR0656313 (83k:10002)
- [Gu2] R. K. Guy, Problems from Western Number Theory Conferences, 1981.
- [Gu3] R. K. Guy, Problems from Western Number Theory Conferences, 1982.
- [HaL] G. H. Hardy and J. E. Littlewood, Some problems of ‘partitio numerorum’, III. On the expression of a number as a sum of primes, *Acta Math.* **44** (1923), 1–70.
- [He1] D. R. Heath-Brown, Gaps between primes, and the pair correlation of zeros of the zeta function, *Acta Arith.* **41** (1982), 85–99. MR0667711 (83m:10078)
- [He2] D. R. Heath-Brown, Sieve identities and gaps between primes, *Astérisque* **94** (1982), 61–65.
- [He3] D. R. Heath-Brown, The number of primes in a short interval, *J. Reine Angew. Math.* **389** (1988), 22–63. MR0953665 (89i:11099)
- [HeaG] D. R. Heath-Brown and D. A. Goldston, A note on the difference between consecutive primes, *Math. Ann.* **266** (1984), 317–320. MR0730173 (85e:11064)
- [HenR] D. Hensley and I. Richards, Primes in intervals, *Acta Arithmetica* **25** (1974), 375–391. MR0396440 (53:305)
- [Ho] G. Hoheisel, Primzahlprobleme in der Analysis, *Sitz. Preuss. Akad. Wiss.* **33** (1930), 3–11.
- [I] A. Ivić, *The Riemann Zeta-Function*, John Wiley & Sons, New York, 1985. MR0792089 (87d:11062)
- [J] G. J. Janusz, *Algebraic Number Fields*, Academic Press, New York and London, 1973. MR0366864 (51:3110)
- [K] E. E. Kummer, Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen, *J. Reine Angew. Math.* **44** (1852), 93–146.
- [La] S. Lang, *Algebraic Number Theory*, Addison-Wesley, Reading, MA, 1970. MR0282947 (44:181)
- [LoY] S. Lou and Q. Yao, A Chebychev’s type of prime number theorem in a short interval, II, *Hardy-Ramanujan J.* **15** (1992), 1–33. MR1215589 (95d:11115)
- [Luca] F. Luca, The Diophantine equation  $P(x) = n!$  and a result of M. Overholt, *Glas. Mat. Ser. III* **37(57)** (2002), 269–273. MR1951531 (2003i:11045)
- [Lucas1] E. Lucas, Théorie des fonctions numériques simplement périodiques, *Amer. J. Math.* **1**(1878), 184–240, 289–321.
- [Lucas2] E. Lucas, Question 301, *Nouv. Corresp. Math.* **4** (1878), 123.
- [Lucas3] E. Lucas, *Théorie des Nombres*, Gauthier-Villars, Paris, 1891.

- [Mi] T. Mitsui, On the prime ideal theorem, *J. Math. Soc. Japan* **20** (1968), 233–247. MR0223314 (36:6362)
- [Mon] H. L. Montgomery, *Topics in Multiplicative Number Theory*, Springer-Verlag Lecture Notes #227, 1971. MR0337847 (49:2616)
- [Mu] J. H. Mueller, On the difference between consecutive primes, *Recent Progress in Analytic Number Theory*, Vol. 1, H. Halberstam and C. Hooley – eds., Academic Press, London, 1981, pp. 269–273. MR0637352 (83c:10062)
- [N] W. Narkiewicz, *Elementary and Analytic Theory of Numbers*, 2nd edition, Springer-Verlag, and PWN-Polish Scientific Publishers, Warsaw, 1990. MR1055830 (91h:11107)
- [NPP] C. Nelson, D. E. Penney and C. Pomerance, 714 and 715, *J. Recreational Math.* **7** (1974), 87–89.
- [O] M. Overholt, The Diophantine equation  $n!+1 = m^2$ . *Bull. London Math. Soc.* **25** (1993), 104. MR1204060 (93m:11026)
- [PolS] R. M. Pollack and H. N. Shapiro, The next to last case of a factorial diophantine equation, *Comm. Pure Appl. Math.* **26** (1973), 313–325. MR0360465 (50:12915)
- [Pow] B. J. Powell, Primitive densities of certain sets of primes, *J. Number Theory* **12** (1980), 210–217. MR0578814 (81k:10093)
- [R] I. Richards, On the normal density of primes in short intervals, *J. Number Theory* **12** (1980), 378–384. MR0586467 (82c:10049)
- [San] J. W. Sander, Prime power divisors of  $\binom{2n}{n}$ , *J. Number Theory* **39** (1991), 65–74. MR1123169 (92i:11097)
- [Sár] A. Sárközy, On divisors of binomial coefficients, I, *J. Number Theory* **20** (1985), 70–80. MR0777971 (86c:11002)
- [SenS] H. G. Senge and E. G. Strauss, PV-numbers and sets of multiplicity, *Period. Math. Hungar.* **3** (1973), 93–100. MR0340185 (49:4941)
- [Ser] J. P. Serre, *A Course in Arithmetic*, Springer-Verlag, New York, 1973. MR0344216 (49:8956)
- [Sh] D. Shanks, On the maximal gaps between successive primes, *Math. Comp.* **18** (1964), 646–651. MR0167472 (29:4745)
- [So1] A. V. Sokolovskii, The distance between “neighbouring” prime ideals, *Doklady Akademii Nauk SSSR* **172** (1967), 1273–1275 (Russian). MR0205947 (34:5772)
- [So2] A. V. Sokolovskii, Theorems on zeros of Dedekind’s zeta-function and the distance between “neighbouring” prime ideals, *Acta Arith.* **13** (1968), 321–334 (Russian). MR0223332 (36:6380)
- [vW] B. L. van der Waerden, Die Seltenheit der Gleichungen mit Affekt, *Math. Ann.* **109** (1933), 13–16.
- [Wa] L. C. Washington, *Introduction to Cyclotomic Fields*, Springer-Verlag, New York, 1982. MR0718674 (85g:11001)
- [Wy] B. F. Wyman, What is a reciprocity law?, *Amer. Math. Monthly* **79** (1972), 571–586. MR0308084 (46:7199)

DEPARTMENTS OF MATHEMATICS AND OF COMPUTER SCIENCE, BEN-GURION UNIVERSITY, BEER-SHEVA 84105, ISRAEL

ANALYSIS AND APPLIED RESEARCH DIVISION, BAE SYSTEMS, BUILDING 27-16, 6500 TRACOR LANE, AUSTIN, TEXAS 78725