The exponents in the prime decomposition of factorials

Jan-Christoph Schlage-Puchta

Abstract. Let $\nu_p(n)$ be the exponent of p in the prime decomposition of n. We show that for different primes p, q satisfying some mild constraints the integers $\nu_p(n!)$ and $\nu_q(n!)$ cannot both be of a rather special form.

Mathematics Subject Classification (2010). Primary 11B65; Secondary 11A63.

Keywords. Factorial, prime decomposition, Diophantine Approximation.

1. Introduction and results

For an integer n and a prime number p let $\nu_p(n)$ be the exponent of p in the prime decomposition of n, i.e. $\nu_p(n)$ is the largest k with $p^k|n$. The distribution of the sequences $\nu_p(n!)$ has received some attention, a large part of which has been stimulated by the question whether for each finite set of primes π there exist infinitely many n such that $\nu_p(n!)$ is even for all $p \in \pi$, which was posed by Erdős and Graham[3, p. 77]. This question was answered in the affirmative by Berend[2]. Shevelev[5] showed that for all primes $p < q$ such that $p = 2$ or $\frac{q-1}{p-1}$ is not a power of 2 there are only finitely many n such that $\nu_p(n!)$ and $\nu_q(n!)$ are simultaneously powers of 2. In this note we generalize this statement. We prove the following.

Theorem 1.1. *Let p, q be distinct primes, d, α, β be integers with $d \geq 3$. Suppose that $\frac{(q-1)\alpha}{(p-1)\beta}$ is not a perfect d-th power. Then the system of equations $\nu_p(n!) = \alpha x^d$, $\nu_q(n!) = \beta y^d$ has only finitely many solutions.*

Note that this statement is in fact a generalization of Shevelev's result. To see this assume that $\frac{q-1}{p-1}$ is not a power of 2. Then there is some d, such that $\frac{(p-1)\alpha}{(q-1)\beta}$ is not a d-th power whenever α, β are powers of 2. We then apply Theorem 1.1 to all pairs $(\alpha, \beta) = (2^a, 2^b)$, where $1 \leq a, b \leq d$. Since every power of 2 can be written as $2^a x^d$, where $1 \leq a \leq d$, we deduce Shevelev's result.

This result is not effective, since it is based on the Thue-Siegel-Roth-theorem. Using Baker's method, we can prove the following.

Theorem 1.2. *Let p, q be distinct primes, π_1, π_2 be finite sets of primes. Write $\frac{p-1}{q-1} = \frac{a}{b}$ with a, b coprime. Assume that one of the following holds true.*

1. *a has a prime divisor which is not in π_2, or b has a prime divisor which is not in π_1;*
2. *$\pi_1 \cap \pi_2 \subseteq \{p\}$, or $\pi_1 \cap \pi_2 \subseteq \{q\}$.*

Then there is an effectively computable integer n_0, such that all integers n for which all prime divisors of $\nu_p(n!)$ are in π_1, and all prime divisors of $\nu_q(n!)$ are in π_2, satisfy $n < n_0$.

The constant n_0 is too large to check all possible n, however, if $\pi_1 \cup \pi_2$ consist of only two primes p_1, p_2, then we can discard all n which are not close to a power of p_1 and p_2 immediately, and search the remaining space using continued fractions. As an example, we prove the following.

Proposition 1.3. *Let n be an integer, such that $\nu_2(n!)$ is a power of 2, and $\nu_3(n!)$ contains only the prime divisors 2 and 3. Then $n \in \{1, 2, 3, 6, 7, 10, 11, 18, 19\}$.*

The conditions of Theorem 1.2 are probably not optimal. In fact, if neither of the two conditions of Theorem 1.2 is satisfied, we still obtain $s_p(n) = s_q(n)$ (see Lemma 2.6 below), and solutions of this equation are probably quite rare. Together with the fact that n must be close to an integer, which has prime factors in a fixed finite set, we are led to believe that Theorem 1.2 holds true under much weaker conditions, it may well be possible that (1) and (2) can simply be deleted from the theorem.

2. Proof of Theorem 1.1 and 1.2

We begin by collecting some known results. For a prime p denote by $s_p(n)$ the sum of digits of n written to base p. Our first two results are well known and proved by simple counting arguments.

Lemma 2.1. *We have $\nu_p(n!) = \frac{n - s_p(n)}{p-1}$.*

Lemma 2.2. *We have $s_p(n) \leq (p-1) + \frac{(p-1)\log n}{\log p}$.*

The following is the Thue-Siegel-Roth-theorem.

Theorem 2.3. *Let α be an irrational algebraic number. Then for any $\epsilon > 0$ there are only finitely many rational numbers $\frac{p}{q}$ satisfying $|\alpha - \frac{p}{q}| < \frac{1}{q^{2+\epsilon}}$.*

We use this theorem to prove Theorem 1.1.

Proof of Theorem 1.1. Suppose that $\nu_p(n!) = \alpha x^d$, $\nu_q(n!) = \beta y^d$. Then from Lemma 2.1 we obtain

$$\frac{(q-1)\alpha x^d}{(p-1)\beta y^d} = \frac{n - s_p(n)}{n - s_q(n)} = 1 + \mathcal{O}(\frac{\log n}{n}),$$

and therefore

$$\left|\frac{x}{y} - \sqrt[d]{\frac{(q-1)\alpha}{(p-1)\beta}}\right| \ll \frac{\log n}{n} \ll \frac{\log y}{y^d} \ll \frac{1}{y^{2.5}}.$$

Now Theorem 2.3 implies that either there are only finitely many choices for y and therefore for n, or $\sqrt[d]{\frac{(q-1)\alpha}{(p-1)\beta}}$ is rational. However, the latter condition is excluded by the assumptions of the theorem, and our claim follows. $\square$

The following is a version of Baker's estimate for linear forms in logarithms, confer[1] for the size of the constants.

Theorem 2.4. *Let $\alpha_1, \dots, \alpha_k$ be algebraic numbers generating a number field of degree $\leq d$, such that α_i is of height A_i, $n_1, \dots, n_k$ be integers in $[-B, B]$, and assume that*

$$\Lambda = n_1 \log \alpha_1 + \cdots + n_k \log \alpha_k \neq 0.$$

Then

$$|\Lambda| \geq \exp\left(-(16nd^2)^{2(n+2)} \log A_1 \log A_2 \dots \log A_k \log B\right)$$

One consequence of Baker's theorem is the following.

Lemma 2.5. *Let p, q be distinct primes, m an integer. Then there exists a constant $c = c(p, q)$ such that there are at most finitely many n such that $s_p(n)$ and $s_q(n)$ are both $\leq \frac{c \log\log n}{\log\log\log n}$.*

Proof. Suppose that $s_p(n), s_q(n) \leq m$. Write $n = \sum p^{e_i} = \sum q^{f_i}$ with $e_1 \geq e_2 \geq \dots e_k$, $f_1 \geq f_2 \geq \dots f_\ell$. Suppose there exists an integer x, such that $x > n^{2/3}$ and $[x^{1-\delta}, x]$ contains none of the powers p^{e_i}, q^{f_i}, where $\delta > 0$.

Let i_0 be the largest index with $p^{e_{i_0}} > x$, j_0 the largest index with $q^{f_{j_0}} > x$. Put $a = p^{e_1 - e_{i_0}} + \dots + p^{e_{i_0 - 1} - e_{i_0}} + 1$, $b = q^{f_1 - f_{i_0}} + \dots + q^{f_{i_0 - 1} - f_{i_0}} + 1$. Then we have

$$|ap^{e_{i_0}} - bq^{f_{j_0}}| \leq m(p^{e_{i_0+1}} + q^{f_{j_0+1}}) \leq 2mx^{1-\delta}. \tag{2.1}$$

If $ap^{e_{i_0}} - bq^{f_{j_0}} = 0$, then a is divisible by $q^{f_{j_0}}$, since p and q are distinct. But $a \leq p^{e_1}/x \leq x^{1/2}$, and $q^{f_{j_0}} \geq x$, and we obtain a contradiction. Hence, $\Lambda = e_{i_0} \log p - f_{j_0} \log q + \log a/b$ does not vanish. From Baker's theorem we obtain

$$\begin{aligned} |\Lambda| &> \exp\left(-(48)^{10} \log p \log q \log \frac{\log n}{\log 2} \max(\log a, \log b)\right) \\ &> \exp\left(-C(p,q) \log n/x \log\log n\right). \end{aligned}$$

Without loss we may assume that $ap^{e_{i_0}} - bq^{f_{j_0}}$ is positive. Using (2.1) we obtain

$$\Lambda \leq \frac{ap^{e_{i_0}} - bq^{f_{j_0}}}{bq^{f_{j_0}}} \leq \frac{2mx^{1-\delta}}{n - 2mx^{1-\delta}} \leq 2m \exp\left(-\log \frac{n}{x} - \delta \log x\right),$$

and we obtain a contradiction provided that $\delta \log x > C \log \frac{n}{x} \log\log n$.

If we define the sequence x_i by

$$x_1 = n/p, \qquad \log \frac{x_i}{x_{i+1}} = C \log \frac{n}{x_i} \log\log n$$

we conclude that as long as $x_i > n^{2/3}$ we have that the interval $[x_{i-1}, x_i]$ contains at least one of the powers p^{e_i}, q^{f_j}. Hence, after at most $2m$ steps this sequence drops below $n^{2/3}$. If we put $y_i = n/x_i$, the recursion becomes $y_{i+1} = y_i^{1+C\log\log n}$, and we obtain $y_i = p^{(1+C\log logn)^i}$. Hence $y_i < n^{1/3}$ for $i < \frac{c\log\log n}{\log\log\log n}$, and our claim follows. □

For the proof of Theorem 1.2 we use Baker's theorem also to prove the following.

Lemma 2.6. *Let p, q be distinct primes, π_1, π_2 be finite sets of primes. Then there exists an effectively computable n_0, such that for $n > n_0$ we have that if all prime divisors of $\nu_p(n!)$ are in π_1, and all prime divisors of $\nu_q(n!)$ are in π_2, then*

$$\frac{p-1}{q-1} \cdot \frac{\nu_p(n!)}{\nu_q(n!)} = 1 \qquad \textit{and} \qquad s_p(n) = s_q(n). \tag{2.2}$$

Proof. Suppose that $\nu_p(n!)$ and $\nu_q(n!)$ have prime divisors in π_1 and π_2, respectively. Then we obtain for certain non-negative integers e_i, f_i

$$\frac{p-1}{q-1} \prod_{p_i\in\pi_1} p_i^{e_i} \prod_{p_i\in\pi_2} p_i^{-f_i} = \frac{n - s_p(n)}{n - s_q(n)} = 1 + \mathcal{O}(\frac{\log n}{n}). \tag{2.3}$$

Assume first that the left hand side is not equal to 1. Define Λ to be the logarithm of the left hand side. Then Λ is a non-vanishing linear combination of logarithms of algebraic numbers, hence we can apply Theorem 2.4. Note that $p_i^{e_i}$ and $p_i^{f_i}$ are bounded above by n, hence all coefficients are $\leq \frac{\log n}{\log 2}$. Putting $\pi = \pi_1 \cup \pi_2$ and assuming $p < q$ we now obtain

$$\begin{aligned} |\Lambda| &\geq \exp\left(-(16|\pi|+16)^{2|\pi|+4} \prod_{p_i\in\pi} \log p_i \log \frac{\log n}{\log 2} \log(q-1)\right) \\ &\geq (\log n)^{-C(p,q,\pi)} \end{aligned}$$

On the other hand we have $\Lambda \ll \frac{\log n}{n}$. Comparing these estimates we obtain an effective upper bound for n.

Now assume that the left hand side equals 1. Then both fractions in (2.3) equal 1, and we see that the conditions (2.2) hold. □

We now deduce Theorem 1.2. In each case it suffices to consider the case that (2.2) holds. Obviously

$$\frac{a}{b} \cdot \frac{\nu_p(n!)}{\nu_q(n!)} = \frac{p-1}{q-1} \cdot \frac{\nu_p(n!)}{\nu_q(n!)} = 1$$

implies that every prime divisor of a also divides $\nu_q(n!)$, and therefore all prime divisors of a are in π_2. Similarly we find that all prime divisors of b are in π_1. Hence the first condition of Theorem 1.2 suffices to imply our claim.

In the second case we may assume that $\pi_1 \cap \pi_2 = \{p\}$, since we can add p to both π_1 and π_2, if necessary. Write $\nu_p(n!) = xp^u$, $\nu_q(n!) = yp^v$, where $p \nmid xy$. Then we have $axp^u = byp^v$. Since x and y are coprime, and not divisible by p, we deduce $x|b$, $y|a$. Hence, there are only finitely many choices for x and y, and it suffices that for fixed values x_0, y_0 there are only finitely many n.

We estimate $s_p(n)$. We have $\nu_p(n!) = x_0p^u$, thus $n = (p-1)x_0p^u + s_p(n)$. Since the sum of digits is subadditive, we conclude $s_p(n) \leq s_p((p-1)x_0) + s_p(s_p(n))$. Lemma 2.2 now implies that $s_p(n)$ is bounded by some constant. On the other hand from (2.2) we deduce that $s_p(n) = s_q(n)$, thus $s_q(n)$ is bounded as well. But then Lemma 2.5 implies that n is bounded, and the second case of Theorem 1.2 is proven as well.

3. Explicit computations

We now prove the Proposition. By direct inspection we check that our claim holds true for $n \leq 1000$.

We first have to make the Landau symbol in (2.3) explicit. We have

$$\frac{n - s_2(n)}{n - s_3(n)} \geq \frac{n - \frac{\log n}{\log 2} - 1}{n} = 1 - \frac{\log 2n}{n \log 2},$$

and

$$\frac{n - s_2(n)}{n - s_3(n)} \leq \frac{n}{n - 2\frac{\log n}{\log 3} - 2} \leq 1 + 1.9\frac{\log n}{n},$$

provided that $n > 1000$. Looking at the proof of Lemma 2.6 we here have that all prime factors of $p - 1$ and $q - 1$ are already contained in $\pi_1 \cup \pi_2$, that is, the linear form Λ actually has the form $\Lambda = a \log 2 + b \log 3$. For linear forms in two logarithm we have far better bounds then for the general case. We use the following, which is a special case of a result due to Laurent, Mignotte and Nesterenko[4].

Theorem 3.1. *Let α_1, α_2 be multiplicatively independent positive integers, $b_1, b_2 \in \mathbb{Z}$ with $b_1, b_2 \neq 0$, and put $\Lambda = b_1 \log \alpha_1 + b_2 \log \alpha_2$. Then*

$$\log |\Lambda| \geq -24.34 \max(\log b' + 0.14, 21)^2 \log \alpha_1 \log \alpha_2,$$

where $b' = \frac{b_1}{\log \alpha_2} + \frac{b_2}{\log \alpha_1}$.

In the notation of the proof of Lemma 2.6 we have $\Lambda = (e_1 - f_1 - 1) \log 2 + f_2 \log 3$, and $2^{e_1} \leq n$, $2^{f_1+1}3^{f_2} \leq n$, thus $b' \leq \frac{2 \log n}{\log 2 \log 3} \leq 2.63 \log n$.

We conclude that either $\Lambda = 0$, that is, (2.2) holds true, or

$$\exp\left(-18.54 \max(\log \log n + 1.11, 21)^2\right) \leq 1.9\frac{\log n}{n},$$

which implies $n < e^{8187}$, that is, $f_2 \leq 7452$. We could cover this range by an exhaustive search, however, in the case that $|\pi_1 \cup \pi_2| = 2$ it is better to use continued fractions. If Λ is very small, then $\frac{e_1 - f_1 - 1}{f_2}$ is a very good approximation to $\frac{\log 3}{\log 2}$. Since the continued fraction algorithm yields the best approximations we easily check that $|\frac{\log 3}{\log 2} - \frac{p}{q}| > \frac{0.04}{q^2}$ holds true for all $q \leq 20000$. More precisely we have that $\Lambda \leq 1.9\frac{\log n}{n}$ implies

$$\frac{1}{25f_2^2} < \left|\frac{e_1 - f_1 - 1}{f_2} - \frac{\log 3}{\log 2}\right| \leq 1.9\frac{\log n}{nf_2 \log 2} \leq \frac{3.02}{n} \leq 1.51 \cdot 2^{-f_1}3^{-f_2}.$$

We first neglect the factor 2^{-f_1} on the right and find that this inequality implies $f_2 \leq 7$. In this range we can replace 0.04 by 0.32 and obtain $n \leq 471$. We conclude that if n is an integer satisfying the assumptions of the Proposition, then either n is contained in the list given in that proposition, or n satisfies (2.2). If $\nu_2(n!)$ is a power of 2, and $n > 3$, then n is of the form $2^k + 2, 2^k + 3$, thus $s_2(n) = 2, 3$.

Hence we have to consider the solutions of the equations $2^k + 2 = 3^x + 3^y$ and $2^k + 3 = 3^x + 3^y + 3^z$. In the second case the left hand side is not divisible by 3, hence $z = 0$, and we are led to case 1. Suppose that $x \geq y$. Then the right hand side is divisible by 3^y, which implies that $3^{y-1}|k - 1$. Hence for a solution we have $\left|\frac{k}{x} - \frac{\log 3}{\log 2}\right| \leq \frac{k-2}{x \cdot 2^k}$, and we have seen before that this inequality has no solutions with $x > 5$. We conclude that (2.2) does not lead to further solutions ≤ 1000, and our proof is complete.

If one would try to obtain similar results for larger sets π_1, π_2 one would have to use numerically weaker bounds for linear forms in more than two logarithms. This would greatly increase the initial range for n. For example, if we would add the prime number 5 to the set π_1 in the example, the upper bound for n would increase to $e^{6.54 \cdot 10^{19}}$, thus $e_1 \leq 9.44 \cdot 10^{19}$, $e_2 \leq 4.07 \cdot 10^{19}$. This range could still efficiently be searched using continued fractions, however, since we are now looking for linear combination of more than 2 real numbers, we would have to use algorithms based on the LLL-algorithm, which are much more complicated.

References

[1] A. Baker, G. Wüstholz, Logarithmic forms and group varieties, *J. Reine Angew. Math.* **442** (1993), 19–62.

[2] D. Berend, On the parity of exponents in the factorization of $n!$, *J. Number Theory* **64** (1997), 13–19.

[3] P. Erdős, R, Graham, *Old and new problems and Results in Combinatorial Number Theory*, L'Enseignement Mathématique, Université de Genève, 1980.

[4] M. Laurent, M. Mignotte, Y. Nesterenko, Formes linéaires en deux logarithmes et déterminants d'interpolation, *J. Number Theory* **55** (1995), 285–321.

[5] V. Shevelev, Compact integers and factorials, *Acta Arith.* **126** (2007), 195–236.

Jan-Christoph Schlage-Puchta
Mathematisches Institut
Universität Rostock
Ulmenstrasse 69
Haus 3
18057 Rostock
Germany
e-mail: jan-christoph.schlage-puchta@uni-rostock.de