DIVISIBILITY OF BINOMIAL COEFFICIENTS BY POWERS OF PRIMES

LUKAS SPIEGELHOFER AND MICHAEL WALLNER

Abstract. For a prime p and nonnegative integers j and n let $\vartheta_p(j,n)$ be the number of entries in the n-th row of Pascal's triangle that are exactly divisible by p^j. Moreover, for a finite sequence $w = (w_{r-1} \cdots w_0) \neq (0,\ldots,0)$ in $\{0,\ldots,p-1\}$ we denote by $|n|_w$ the number of times that w appears as a factor (contiguous subsequence) of the base-p expansion $n = (n_{\nu-1} \cdots n_0)_p$ of n. It follows from the work of Barat and Grabner (*Digital functions and distribution of binomial coefficients*, J. London Math. Soc. (2) 64(3), 2001), that $\vartheta_p(j,n)/\vartheta_p(0,n)$ is given by a polynomial P_j in the variables X_w, where w are certain finite words in $\{0,\ldots,p-1\}$, and each variable X_w is set to $|n|_w$. This was later made explicit by Rowland (*The number of nonzero binomial coefficients modulo* p^α, J. Comb. Number Theory 3(1), 2011), independently from Barat and Grabner's work, and Rowland described and implemented an algorithm computing these polynomials P_j. In this paper, we express the coefficients of P_j using generating functions, and we prove that these generating functions can be determined explicitly by means of a recurrence relation. Moreover, we prove that P_j is uniquely determined, and we note that the proof of our main theorem also provides a new proof of its existence. Besides providing insight into the structure of the polynomials P_j, our results allow us to compute them in a very efficient way.

1. Introduction

The history of binomial coefficients in congruence classes modulo m begins not later than in the middle of the 19th century, when Kummer [28] stated his famous theorem on the highest prime power p^m dividing a binomial coefficient $\binom{n}{t}$: m is the number of *borrows* occurring in the subtraction $n-t$ in base p. In other words, this is the number of indices k such that $n \bmod p^k < t \bmod p^k$. Kummer's theorem was generalised to multinomial

2010 *Mathematics Subject Classification.* Primary 11B65; Secondary 11A63, 11B50, 05A16.

Key words and phrases. binomial coefficients modulo powers of primes.

The first author acknowledges support by the project MuDeRa (Multiplicativity, Determinism, and Randomness), which is a joint project between the ANR (Agence Nationale de la Recherche) and the FWF (Austrian Science Fund), and also by project F5502-N26 (FWF), which is a part of the Special Research Program "Quasi Monte Carlo Methods: Theory and Applications". The second is supported by Project SFB F50-03 (FWF), which is a part of the Special Research Program "Algorithmic and Enumerative Combinatorics".

and q-multinomial coefficients by Fray [18], and to generalised binomial coefficients by Knuth and Wilf [27].

A complete list of results related to Pascal's triangle modulo powers of primes would go beyond the scope of any research paper; we refer the reader to the surveys [21, 34] by Granville and Singmaster respectively for an overview of the topic. The question also attracts other areas of research: in [3, Section 14.6] and [1], connections with automatic sequences and combinatorics on words are highlighted. Moreover, the paper [4] considers the related question of counting coefficients equal to a given value of a polynomial over a finite field.

In this paper we restrict ourselves to questions concerning *exact divisibility* of binomial coefficients by powers of primes. This means that we are only concerned with the residue class p^j modulo p^{j+1}, in other words, we study the case $\nu_p\binom{n}{t} = j$, where $\nu_p(m)$ denotes the largest k such that $p^k \mid m$.

We therefore introduce the following notion, which is central in our paper. Let j and n be nonnegative integers and p a prime number, and define

$$\vartheta_p(j,n) = \left|\left\{t \in \{0,\ldots,n\} : \nu_p\binom{n}{t} = j\right\}\right|.$$

Put into words, $\vartheta_p(j,n)$ is the number of entries in the n-th row of Pascal's triangle that are exactly divisible by p^j. The case $j = 0$ can be reduced to properties of the base-p expansion of the row number n by appealing to Lucas' congruence [29]. This well-known congruence asserts that for $t \le n$ having the (not necessarily proper) base-p representations $n = (n_{\nu-1}\cdots n_0)_p$ and $t = (t_{\nu-1}\cdots t_0)_p$, we have

$$\binom{n}{t} \equiv \binom{n_{\nu-1}}{t_{\nu-1}}\cdots\binom{n_0}{t_0} \bmod p.$$

Since p is a prime number, we have $p \nmid \binom{n}{t}$ if and only if none of the factors is divisible by p, which in turn is equivalent to $t_i \le n_i$ for all $i < \nu$. We obtain, denoting by $|n|_a$ the number of times the digit $a \ne 0$ occurs in the base-p expansion of n,

$$\vartheta_2(0,n) = 2^{|n|_1}$$

for the case $p = 2$ (Glaisher [19]) and more generally (Fine [15])

$$\vartheta_p(0,n) = \prod_{0\le i<\nu}(n_i+1) = 2^{|n|_1}3^{|n|_2}4^{|n|_3}\cdots p^{|n|_{p-1}}. \tag{1.1}$$

Lucas' congruence has been generalised and extended in different directions, see for example [18], [26] (reproved in [32]), [9, 20, 21]; moreover [10] for an account of less recent results. In order to be able to formulate our results concerning general $j \geq 0$, we need some notation.

Notation. The letter p always denotes a prime number; we use typewriter font to indicate digits in the base-p expansion, except for variables representing digits. For the $(p-1)$-st digit we write $\mathtt{q}$, a letter supposed to be a mnemonic relating to $\mathtt{9}$ in the decimal expansion. If v is an infinite word over the alphabet $\{\mathtt{0}, \ldots, \mathtt{q}\}$ such that $v_i \neq \mathtt{0}$ for only finitely many $i \geq 0$, let $(v)_p = \sum_{i \geq 0} v_i p^i$ be the integer represented by v in base p. Moreover, if $w = (w_{\nu-1} \cdots w_0) \in \{\mathtt{0}, \ldots, \mathtt{q}\}^\nu$ contains at least one nonzero digit and v is as above, let $|v|_w$ be the number of times that w occurs as a factor of v. More precisely,

$$|v|_w = |\{i \geq 0 : (v_{i+\nu-1}, \ldots, v_i) = (w_{\nu-1}, \ldots, w_0)\}|.$$

For finite words v we extend the above notions by padding with zeros. Moreover, if n is a nonnegative integer and $n = (v)_p$, we set $|n|_w \coloneqq |v|_w$. Occurrences of factors may overlap: for example, for $p = 2$ we have $|42|_{\mathtt{1010}} = |\mathtt{101010}|_{\mathtt{1010}} = 2$. Moreover, as a consequence of the padding with zeros we have $|\mathtt{1}|_{\mathtt{1}} = |\mathtt{1}|_{\mathtt{01}} = |\mathtt{1}|_{\mathtt{001}} = \cdots = 1$, while $|\mathtt{1}|_{\mathtt{10}} = 0$.

The following statement is an easy reformulation of [31, Theorem 2]. The method used for proving this theorem is very similar to the method used in the older paper [5, Theorem 5], which proves a less detailed form of the result, but can be adapted to yield the full statement. See also Remark 1.

Theorem 0 (Rowland [31]–Barat–Grabner [5]). *Let p be a prime and $j \geq 0$. Then $\vartheta_p(j,n)/\vartheta_p(0,n)$ is given by a polynomial P_j of degree j in the variables X_w, where w ranges over the set*

$$W_j = \left\{w \in \{\mathtt{0}, \ldots, \mathtt{q}\}^\nu : 2 \leq \nu \leq j+1, w_{\nu-1} \neq \mathtt{0}, w_0 \neq \mathtt{q}\right\}, \tag{1.2}$$

and X_w is set to $|n|_w$.

Note that $W_0 = \emptyset$ and $P_0(x) = 1$. Determining $\vartheta_p(j,n)/\vartheta_p(0,n)$ by means of this theorem is a two-step procedure:

$$n \mapsto \left(|n|_w\right)_{w \in W_j} \mapsto P_j\Big(\left(|n|_w\right)_{w \in W_j}\Big) = \frac{\vartheta_p(j,n)}{\vartheta_p(0,n)}. \tag{1.3}$$

Barat and Grabner [5, Theorem 5] used a representation of $\vartheta_p(j,n)/\vartheta_p(0,n)$ of this kind in order to establish an asymptotic formula for the partial sums $\sum_{0 \leq n < N} \vartheta_p(j,n)$. Their Theorem 5 generalises the case $j = 0$ [16] (see also [6, 36]), and yields a quantitative version of the statement "any integer divides almost all binomial coefficients" [33].

Theorem 0 implies, as noted by Rowland, that $n \mapsto \vartheta_p(j,n)/\vartheta_p(0,n)$ is a *p-regular sequence* in the sense of Allouche and Shallit [2, 3]. We will however not follow this line of research in this paper.

In Proposition 2.1 we will prove that a polynomial P_j as in Theorem 0 is uniquely determined, so that we may talk about the coefficients of P_j without ambiguity. These polynomials are the main object of study in this paper, and want to obtain a better understanding of its coefficients. Our main theorem (restated in Section 2) concerns the behaviour of the coefficients of a single monomial in the sequence $(P_j)_{j\geq 0}$ of polynomials.

Theorem. *Let W be the set of all words $(w_{\nu-1},\ldots,w_0) \in \{\mathtt{0},\ldots,\mathtt{q}\}^\nu$ such that $\nu \geq 2$, $w_{\nu-1} \neq \mathtt{0}$ and $w_0 \neq \mathtt{q}$. Assume that $w^{(1)},\ldots,w^{(\ell)} \in W$, and $k_1,\ldots,k_\ell$ are positive integers. Let c_j be the coefficient of the monomial*

$$X_{w^{(1)}}^{k_1} \cdots X_{w^{(\ell)}}^{k_\ell}$$

in the polynomial P_j. Then

$$\sum_{j\geq 0} c_j x^j = \frac{1}{k_1!}\bigl(\log r_{w^{(1)}}(x)\bigr)^{k_1} \cdots \frac{1}{k_\ell!}\bigl(\log r_{w^{(\ell)}}(x)\bigr)^{k_\ell},$$

where r_w is a rational function defined at 0 such that $r_w(0) = 1$.

The rational function r_w can be determined explicitly by means of a recurrence, see Section 2. The easiest nontrivial example is $r_{\mathtt{10}}(x) = 1 + x/2$ ($p = 2$). Note that the coefficients c_j always belong to a fixed monomial $X_{w^{(1)}}^{k_1} \cdots X_{w^{(\ell)}}^{k_\ell}$. However, in order to increase readability we will not emphasize this relationship by additional sub- or superscripts. It will always be clear from the context which monomial is referred to.

As a direct consequence of our results we will obtain the following corollary.

Corollary. *Let $p = 2$. The coefficient c_j of the monomial $X_{\mathtt{10}}$ in P_j equals $[x^j]\log(1 + x/2)$. In particular,*

$$\sum_{j\geq 0} c_j = \log(3/2).$$

This special case confirms an observation by Rowland [31], who noted that a plot of the first few partial sums $c'_j = c_0 + \cdots + c_{j-1}$ "suggests that the limit of this sequence exists". He computed the first seven polynomials

$$P'_j = P_0 + \cdots + P_{j-1}$$

with the help of his Mathematica package BinomialCoefficients, which is based on his paper [31] and available from his website, and determined the coefficients c'_j that way. By the above corollary the limit does exist indeed, and its value is $\log(3/2)$. It is however not true for each monomial M that the sequence of coefficients of M in P'_j converges as $j \to \infty$, nor is it the case that all coefficients of P'_j are nonnegative. A simultaneous counterexample for both questions is given by X_{1010} (see the examples after Corollary 2.10). The sequence of coefficients of this monomial has the generating function

$$\log\Bigl(1 + \tfrac{1}{2}x^3/\bigl(1 + x/2\bigr)^2\Bigr),$$

which has a unique dominant singularity $x_0 \sim -0.86408$. Therefore negative signs occur infinitely often and the sequence of coefficients diverges to ∞ in absolute value (this is true for the coefficients in P_j as well as in P'_j).

While the above results concern the behaviour of a single monomial in different polynomials P_j, we will also prove an "orthogonal" result, namely an asymptotic estimate of the number of nonzero coefficients in P_j and P'_j (Corollary 2.7).

The results that we have outlined above provide answers to questions posed by Rowland [31] at the end of his paper. For more details, we refer to Section 2. Finally, we want to note that our main theorem together with the recurrence for r_w enables us to compute the polynomials P_j very efficiently (see Remark 5).

We will also use the following notations in this article. The integer $s_2(n) := |n|_1$ is the *sum of digits* of n in base 2, more generally $s_p(n) := |n|_1 + 2\,|n|_2 + \cdots + (p-1)\,|n|_q$ is the sum of digits of n in base p. For a finite word w we denote by $|w|$ the length of w. Finally, $\mathbb{N}$ denotes the set of nonnegative integers.

Plan of the paper. In Section 1.1 we will meet the fundamental recurrence relation for the values $\vartheta_p(j,n)$, found by Carlitz [7], while in Section 1.2 we list some of the polynomials P_j for the case $p = 2$. In Sections 2.1 and 2.2, we will state in detail the results we announced above, and study the rational functions r_w more carefully. Section 2.3 gives an alternative form of the fundamental recurrence relation for $\vartheta_p(j,n)$, which can be written as an elegant but enigmatic infinite product. This also yields a new proof of Carlitz' recurrence relation. Finally, we note in Section 2.4 that we can reuse the polynomials P_j for the columns in Pascal's triangle. Proofs not given in the main section are stated in Section 3.

1.1. **A recurrence for the values $\vartheta_p(j,n)$, and the case $j = 1$.** Carlitz [7] gave a recurrence relation for the values $\vartheta_p(j,n)$, which also involves another family ψ_p defined by[1]

$$\psi_p(j,n) = \left|\left\{t \in \{0,\ldots,n\} : \nu_p\binom{n}{t} = j - \nu_p(n+1)\right\}\right|.$$

He then obtains [7, Equations (1.7)–(1.9)] for $n \geq 0$ and $j \geq 1$, using the convention $\psi_p(j,-1) = 0$,

$$\begin{aligned} \vartheta_p(j,pn+a) &= (a+1)\vartheta_p(j,n) \\ &\quad + (p-a-1)\psi_p(j-1,n-1), \qquad 0 \leq a < p; \\ \psi_p(j,pn+a) &= (a+1)\vartheta_p(j,n) \\ &\quad + (p-a-1)\psi_p(j-1,n-1), \quad 0 \leq a < p-1; \\ \psi_p(j,pn+p-1) &= p\psi_p(j-1,n). \end{aligned} \tag{1.4}$$

Rewriting the recurrence (1.4) using the obvious identity

$$\psi_p(j,n) = \begin{cases} \vartheta_p(j-\nu_p(n+1),n), & j \geq \nu_p(n+1); \\ 0, & j < \nu_p(n+1), \end{cases}$$

we obtain for $0 \leq a < p$

$$\begin{aligned} \vartheta_p(j,pn+a) &= (a+1)\vartheta_p(j,n) \\ &\quad + \begin{cases} (p-a-1)\vartheta_p(j-1-\nu_p(n),n-1), & j > \nu_p(n); \\ 0, & j \leq \nu_p(n). \end{cases} \end{aligned} \tag{1.5}$$

Among other things, Carlitz evaluates $\vartheta_p(j,n)$ for special values of n, using associated generating functions. Moreover, he proves the explicit formula [7, Equation (2.5)], saying that for the base-p expansion $n = \sum_{i=0}^{\nu-1} n_i p^i$ we have

$$\vartheta_p(1,n) = \sum_{0 \leq i < \nu-1} (n_{\nu-1}+1)\cdots(n_{i+2}+1)n_{i+1}(p-n_i-1)(n_{i-1}+1)\cdots(n_0+1).$$

By (1.1) this implies that

$$\frac{\vartheta_p(1,n)}{\vartheta_p(0,n)} = \sum_{0 \leq i < \nu-1} \frac{n_{i+1}}{n_{i+1}+1} \cdot \frac{p-n_i-1}{n_i+1}.$$

In particular, counting identical summands, we obtain

$$\frac{\vartheta_p(1,n)}{\vartheta_p(0,n)} = \sum_{\substack{0 \leq c,a < p \\ c \neq 0, a \neq p-1}} \frac{c}{c+1} \cdot \frac{p-a-1}{a+1} \left|n\right|_{ca}. \tag{1.6}$$

Note that we defined the quantity $\left|n\right|_{ca}$ as the number of occurrences of $(c,a) = (n_{i+1}, n_i)$ in the base-p expansion $n = \sum_{i=0}^{\infty} n_i p^i$. Since c is nonzero,

[1] Our notation differs slightly from Carlitz' who wrote $\theta_j(n)$ instead of $\vartheta_p(j,n)$ and $\psi_j(n)$ instead of $\psi_p(j,n)$, omitting p altogether.

this is equal to the number of occurrences of this pattern for $0 \le i < \nu - 1$. For the prime $p = 2$ only one summand remains, yielding the formula

$$\frac{\vartheta_2(1,n)}{\vartheta_2(0,n)} = \frac{1}{2} |n|_{\mathtt{10}} .$$

This formula was observed by Howard [23, Equation (2.4)], see also [22, Theorem 2.2]. (The latter is however not correct if n is a power of 2.)

1.2. **The polynomials P_j for $j > 1$.** In 1971, Howard [23] also found formulas for $\vartheta_2(2,n)$, $\vartheta_2(3,n)$, and $\vartheta_2(4,n)$ in terms of factor counting functions $|n|_w$. In different notation, he obtained the formulas

$$\frac{\vartheta_2(2,n)}{\vartheta_2(0,n)} = -\frac{1}{8} |n|_{\mathtt{10}} + \frac{1}{8} |n|_{\mathtt{10}}^2 + |n|_{\mathtt{100}} + \frac{1}{4} |n|_{\mathtt{110}} ,$$

$$\begin{aligned}\frac{\vartheta_2(3,n)}{\vartheta_2(0,n)} &= \frac{1}{24} |n|_{\mathtt{10}} - \frac{1}{16} |n|_{\mathtt{10}}^2 - \frac{1}{2} |n|_{\mathtt{100}} - \frac{1}{8} |n|_{\mathtt{110}} + \frac{1}{48} |n|_{\mathtt{10}}^3 + \frac{1}{2} |n|_{\mathtt{10}} |n|_{\mathtt{100}} \\ &\quad + \frac{1}{8} |n|_{\mathtt{10}} |n|_{\mathtt{110}} + 2 |n|_{\mathtt{1000}} + \frac{1}{2} |n|_{\mathtt{1010}} + \frac{1}{2} |n|_{\mathtt{1100}} + \frac{1}{8} |n|_{\mathtt{1110}} ,\end{aligned}$$

$$\begin{aligned}\frac{\vartheta_2(4,n)}{\vartheta_2(0,n)} &= -\frac{1}{64} |n|_{\mathtt{10}} + \frac{11}{384} |n|_{\mathtt{10}}^2 - \frac{1}{4} |n|_{\mathtt{100}} + \frac{1}{32} |n|_{\mathtt{110}} - \frac{1}{64} |n|_{\mathtt{10}}^3 \\ &\quad - \frac{3}{8} |n|_{\mathtt{10}} |n|_{\mathtt{100}} - \frac{3}{32} |n|_{\mathtt{10}} |n|_{\mathtt{110}} - |n|_{\mathtt{1000}} - \frac{1}{2} |n|_{\mathtt{1010}} - \frac{1}{2} |n|_{\mathtt{1100}} \\ &\quad - \frac{1}{16} |n|_{\mathtt{1110}} + \frac{1}{384} |n|_{\mathtt{10}}^4 + \frac{1}{8} |n|_{\mathtt{10}}^2 |n|_{\mathtt{100}} + \frac{1}{32} |n|_{\mathtt{10}}^2 |n|_{\mathtt{110}} + \frac{1}{2} |n|_{\mathtt{100}}^2 \\ &\quad + \frac{1}{4} |n|_{\mathtt{100}} |n|_{\mathtt{110}} + \frac{1}{32} |n|_{\mathtt{110}}^2 + |n|_{\mathtt{10}} |n|_{\mathtt{1000}} + \frac{1}{4} |n|_{\mathtt{10}} |n|_{\mathtt{1010}} \\ &\quad + \frac{1}{4} |n|_{\mathtt{10}} |n|_{\mathtt{1100}} + \frac{1}{16} |n|_{\mathtt{10}} |n|_{\mathtt{1110}} + 4 |n|_{\mathtt{10000}} + |n|_{\mathtt{10010}} + |n|_{\mathtt{10100}} \\ &\quad + \frac{1}{4} |n|_{\mathtt{10110}} + |n|_{\mathtt{11000}} + \frac{1}{4} |n|_{\mathtt{11010}} + \frac{1}{4} |n|_{\mathtt{11100}} + \frac{1}{16} |n|_{\mathtt{11110}} .\end{aligned}$$

Moreover, Howard [24] found an expression for $\vartheta_p(2,n)$ for general primes p; see also [25, 38]. We also refer to Spearman and Williams [35, Theorem 1]. They reproved the formulas above by expressing $\vartheta_2(j,n)/\vartheta_2(0,n)$ as a sum of nonoverlapping subwords of the binary expansion of n. We note that the factors that are counted in the expressions for $\vartheta_2(j,n)$ always start with the digit $\mathtt{0}$ (read from right to left) and end with the digit $\mathtt{1}$. That is, the words w occurring in these expressions belong to the set W_j defined in Theorem 0, for some $j \ge 1$. By this theorem we can always require the condition $w \in W_j$, while Proposition 2.1 ensures uniqueness of an expression for $\vartheta_2(j,n)$ as above.

We refrained from listing formulas for $j \ge 5$ for the obvious reason: P_5 contains 69 monomials, P_6 already 174.

Remark 1. As we noted before, the statement of the Theorem 0 formulated by Rowland can already be found implicitly in Barat and Grabner [5]. That is, their method of proof can be adapted to show the theorem. More precisely, in the course of proving Theorem 5 in that paper, they proved that $\vartheta_p(j,n)/\vartheta_p(0,n)$ is a sum of products of block-additive functions. Here a function $f:\mathbb{N}\rightarrow\mathbb{C}$ is called *ℓ-block-additive* in base p, if there is a function $F:\{\mathtt{0},\ldots,\mathtt{q}\}^\ell\rightarrow\mathbb{C}$ satisfying $F(0,\ldots,0)=0$ such that for the base-p expansion $n=\sum_{i\geq 0}\varepsilon_i p^i$ we have

$$f(n)=\sum_{i\geq 0}F(\varepsilon_{i+\ell-1},\cdots,\varepsilon_i).$$

These functions were first defined by Cateland in his thesis [8]. We note that ℓ-block-additive functions are precisely the complex linear combinations of factor counting functions $|\cdot|_w$, where w contains a nonzero letter and the length $|w|$ is bounded by ℓ. It follows from [5, (3.3), (3.4)] that the ℓ-block-additive functions occurring in the representation of $\vartheta_p(j,n)/\vartheta_p(0,n)$ take only those factors $(w_{\nu-1}\cdots w_0)\in\{\mathtt{0},\ldots,\mathtt{q}\}^\nu$ into account such that $w_{\nu-1}\neq \mathtt{0}$ and $w_0\neq\mathtt{q}$. Moreover, enhancing the induction hypothesis in the proof of [5, Theorem 5], it can be shown that only ℓ-block-additive functions, where $1\leq\ell\leq j$, appear, and that the occurring products of block-additive functions have length $\leq j$.

Rowland [31] used an approach very similar to Barat and Grabner's [5] (see also Spearman and Williams [35]) in order to obtain Theorem 0. More precisely, it follows from the proof of this theorem that the monomials $X_{w^{(1)}}\cdots X_{w^{(\ell)}}$ occurring in the polynomial P_j satisfy

$$\left|w^{(1)}\right|+\cdots\left|w^{(\ell)}\right|-\ell\leq j. \tag{1.7}$$

For example, if $p=2$ and $j=2$, only the monomials 1, $X_{\mathtt{10}}$, $X_{\mathtt{10}}^2$, $X_{\mathtt{100}}$ and $X_{\mathtt{110}}$ can occur. Based on (1.7) we will derive in Corollary 2.7 an upper bound for the number of monomials in P_j.

We note that we always write words from right to left, since our interest in them stems from base-p expansions of an integer. Correspondingly, to name a consequence of this convention, a prefix of a word starts with the rightmost letter.

2. Results

2.1. Computing the coefficients of P_j.

Let p be a prime number throughout this section. For brevity of notation, we omit the index p whenever there is no risk of confusion. As in Theorem 0, let

$$W_j = \left\{w \in \{\mathtt{0}, \dots, \mathtt{q}\}^\nu : 2 \le \nu \le j+1, w_{\nu-1} \ne \mathtt{0}, w_0 \ne \mathtt{q}\right\},$$

moreover we define the set of *admissible* words,

$$W = \bigcup_{j \ge 1} W_j.$$

In order to get meaningful statements on the coefficients of P_j, we have to show that the polynomial P_j is well-defined, i.e., uniquely determined. Note that it is not clear a priori that there is only one polynomial P_j representing $\vartheta_p(j,n)/\vartheta_p(0,n)$ as in (1.3): the values inserted into this polynomial are not independent of each other, therefore we can not use Lagrange interpolation directly for establishing uniqueness. For example, we have $|n|_{\mathtt{10}} \ge |n|_{\mathtt{100}}$ for all n, so that not all tuples $(n_w)_{w \in W_j}$ of nonnegative integers can occur as family $(|n|_w)_{w \in W_j}$ of block counts of a nonnegative integer n. Moreover, for the polynomial to be unique it is necessary that the blocks we are counting satisfy some restrictions, since there are obvious identities such as $|n|_{\mathtt{1}} = |n|_{\mathtt{01}} + |n|_{\mathtt{11}}$. We will show that the restriction $w_{\nu-1} \ne \mathtt{0}, w_0 \ne \mathtt{q}$ leads to a unique polynomial P_j after all.

Proposition 2.1. *There is at most one polynomial P_j in the variables X_w, where $w \in W$, such that*

$$\frac{\vartheta_p(j,n)}{\vartheta_p(0,n)} = P_j\left((|n|_w)_{w \in W}\right)$$

for all $n \ge 0$.

In order to prepare for the main theorem, we define generating functions of the values $\vartheta_p(j,n)$, which occupy a central position in the statements of the main results.

$$T_n(x) \coloneqq \sum_{j \ge 0} \vartheta_p(j,n) x^j = \sum_{0 \le t \le n} x^{\nu_p\binom{n}{t}}. \tag{2.1}$$

Obviously, $T_n(x)$ is a polynomial of degree $\max_{0 \le t \le n} \nu_p\binom{n}{t}$, which is sequence A119387 in Sloane's OEIS for the case $p=2$. The recurrence (1.5) for ϑ_p translates to the generating functions $T_n(x)$ as follows:

$$\begin{aligned} T_a(x) &= a+1, \\ T_{pn+a}(x) &= (a+1)T_n(x) + (p-a-1)x^{s+1}T_{n-1}(x), \end{aligned} \tag{2.2}$$

for $n \ge 1$ and $0 \le a < p$, where $s = \nu_p(n)$. We note the special case

$$T_{cp^t-1}(x) = T_{(c-1)\mathtt{q}^t} = cp^t, \qquad 1 \le c < p, t \ge 0,$$

which we will use often.

Remark 2. Using the recurrence (2.2), one can show by induction that

$$\deg T_n(x) = \lambda - \nu_p(m+1)$$

for $n \geq 1$, where $\lambda \geq 0$ and $m \in \{0, \ldots, p^\lambda - 1\}$ are chosen such that $n = cp^\lambda + m$ for some $c \in \{1, \ldots, p-1\}$.

Let us compute some polynomials T_n for $p = 2$. From the recurrence (2.2), we obtain

$$\begin{aligned}
T_0(x) &= 1, & T_1(x) &= 2,\\
T_2(x) &= 2 + x, & T_3(x) &= 4,\\
T_4(x) &= 2 + x + 2x^2, & T_5(x) &= 4 + 2x,\\
T_6(x) &= 4 + 2x + x^2, & T_7(x) &= 8,\\
T_8(x) &= 2 + x + 2x^2 + 4x^3, & T_9(x) &= 4 + 2x + 4x^2.
\end{aligned}$$

Note that $T_n(1) = n + 1$, since the n-th row of Pascal's triangle contains $n + 1$ entries. Moreover, we define normalized generating functions $\overline{T}_n$:

$$\overline{T}_n(x) = \frac{1}{\vartheta_p(0,n)} T_n(x).$$

By definition, we have $[x^0]\,\overline{T}_n(x) = 1$. We are extending these notations to finite words v in $\{\mathtt{0}, \ldots, \mathtt{q}\}$ via the base-p expansion: if $(v)_p = n$, we set $T_v \coloneqq T_n$ and $\overline{T}_v \coloneqq \overline{T}_n$. Based on the polynomials $\overline{T}_n(x)$, we shall define the rational functions r_w occurring in the main theorem. In order to do so, we define the *left truncation* w_L and the *right truncation* w_R on the set $W \cup \{\varepsilon\}$, as follows. For $w \in W$, $r \geq 1$ $s \geq 0$, and digits $c \neq \mathtt{0}$ and $a \neq \mathtt{q}$, let

$$\begin{aligned}
&\varepsilon_L = \varepsilon, \quad (c\mathtt{0}^r)_L = \varepsilon, \quad (c\mathtt{0}^s a)_L = \varepsilon, \quad (c\mathtt{0}^s w)_L = w;\\
&\varepsilon_R = \varepsilon, \quad (\mathtt{q}^r a)_R = \varepsilon, \quad (c\mathtt{q}^s a)_R = \varepsilon, \quad (w\mathtt{q}^s a)_R = w.
\end{aligned}$$

In other words, for $w \in W$ the word w_L is the longest proper prefix u of w (read from right to left) such that $u \in W \cup \{\varepsilon\}$. Analogously, w_R is the longest proper suffix u of w such that $u \in W \cup \{\varepsilon\}$. Note that we have $(w_L)_R = (w_R)_L$ for all $w \in W \cup \{\varepsilon\}$; we write w_{LR} for the common value. In what follows, we write $\overline{T}_w \equiv \overline{T}_w(x)$ as a shorthand. The following proposition, a telescoping product, is the first out of two pillars on which the main theorem rests.

Proposition 2.2. *Let $v \in W \cup \{\varepsilon\}$. Then we have the identity*

$$\overline{T}_v = \prod_{w \in W} \left(\frac{\overline{T}_w \overline{T}_{w_{LR}}}{\overline{T}_{w_R} \overline{T}_{w_L}} \right)^{|v|_w}. \tag{2.3}$$

We note that we do not use the explicit definition of $\overline{T}_w$ in the proof of this proposition. We only need the property $\overline{T}_w(0) = 1$, so that we may take quotients, and the property $\overline{T}_\varepsilon = 1$. In other words, we will show that the product reduces to the fraction $\overline{T}_v/\overline{T}_\varepsilon$ by cancelling identical factors. The following example clarifies this point.

Example. Let $p = 2$ and $v = \texttt{100100}$. Then we have

$$\frac{\overline{T}_v}{\overline{T}_\varepsilon} = \left(\frac{\overline{T}_{\texttt{10}}\overline{T}_\varepsilon}{\overline{T}_\varepsilon\overline{T}_\varepsilon}\right)^2 \left(\frac{\overline{T}_{\texttt{100}}\overline{T}_\varepsilon}{\overline{T}_{\texttt{10}}\overline{T}_\varepsilon}\right)^2 \left(\frac{\overline{T}_{\texttt{10010}}\overline{T}_\varepsilon}{\overline{T}_{\texttt{100}}\overline{T}_{\texttt{10}}}\right) \left(\frac{\overline{T}_{\texttt{100100}}\overline{T}_{\texttt{10}}}{\overline{T}_{\texttt{10010}}\overline{T}_{\texttt{100}}}\right).$$

For each admissible word w we can finally define the rational generating function

$$r_w(x) \coloneqq \frac{\overline{T}_w(x)\overline{T}_{w_{LR}}(x)}{\overline{T}_{w_R}(x)\overline{T}_{w_L}(x)}.$$

Now that we know r_w, our main theorem can be stated completely explicitly.

Theorem 2.3. *Let $w^{(1)}, \ldots, w^{(\ell)}$ be admissible words and $k_1, \ldots, k_\ell$ positive integers. Assume that c_j is the coefficient of the monomial*

$$X_{w^{(1)}}^{k_1} \cdots X_{w^{(\ell)}}^{k_\ell}$$

in the polynomial P_j. Then

$$\sum_{j\ge 0} c_j x^j = \frac{1}{k_1!}\big(\log r_{w^{(1)}}(x)\big)^{k_1} \cdots \frac{1}{k_\ell!}\big(\log r_{w^{(\ell)}}(x)\big)^{k_\ell}.$$

We list the first few rational functions r_w for the case $p = 2$:

$$\begin{aligned} r_{\texttt{10}}(x) &= 1 + \tfrac{1}{2}x, & r_{\texttt{100}}(x) &= 1 + \frac{x^2}{1 + x/2}, \\ r_{\texttt{110}}(x) &= 1 + \frac{\frac{1}{4}x^2}{1 + x/2}, & r_{\texttt{1000}}(x) &= 1 + \frac{2x^3}{1 + x/2 + x^2}, \\ r_{\texttt{1010}}(x) &= 1 + \frac{\frac{1}{2}x^3}{(1 + x/2)^2}, & r_{\texttt{1100}}(x) &= 1 + \frac{\frac{1}{2}x^3}{(1 + x/2 + x^2)(1 + x/2 + x^2/4)}. \end{aligned}$$

As a straightforward application of Theorem 2.3 we obtain the corollary from the introduction, which we restate here.

Corollary 2.4. *Let $p = 2$. The coefficient of $X_{\texttt{10}}$ in the polynomial P_j equals $[x^j]\log(1 + x/2)$. In particular,*

$$\sum_{j\ge 0} c_j = \log(3/2).$$

Proof. In this simple case all we need is $r_{\texttt{10}}(x) = \overline{T}_2(x) = 1 + \frac{x}{2}$, which does not have a singularity or a zero in the closed unit disc. $\square$

We continued the computation of the rational functions r_w and performed analogous experiments for the prime numbers $3, 5, 7$ in order to obtain a conjecture on the structure of r_w. The statement of the following proposition is the result of these experiments and constitutes the second main ingredient in the proof of our theorem. The proof can be found at the end of this paper.

Proposition 2.5. *Let p be a prime and assume that $w = w_{\nu-1} \cdots w_0 \in W$. The rational function $r_w(x)$ satisfies*

$$r_w(x) = 1 + \frac{\alpha x^{\nu-1}}{\overline{T}_{w_L}(x)\overline{T}_{w_R}(x)},$$

where

$$\alpha = p^{\nu-2}\frac{w_{\nu-1}}{w_{\nu-1}+1} \cdot \frac{p-w_0-1}{w_0+1} \prod_{2\le d\le p} d^{-2|w'|_{d-1}}, \tag{2.4}$$

and $w' = w_{\nu-2} \cdots w_1$.

Remark 3. Consider the special case $w = ca$ of this proposition. We obtain $\alpha = \frac{c}{c+1}\frac{p-a-1}{a+1}$, which gives the formula $\overline{T}_{ca}(x) = r_{ca}(x) = 1 + \frac{c}{c+1}\frac{p-a-1}{a+1}x$ (compare to (3.4)). By Theorem 2.3 we obtain the coefficient of X_{ca} in the polynomial P_1 by extracting the coefficient

$$[x^1] \log\left(1 + \frac{c}{c+1}\frac{p-a-1}{a+1}x\right) = \frac{c}{c+1}\frac{p-a-1}{a+1},$$

which is consistent with (1.6).

The proof of Theorem 2.3 is a combination of Propositions 2.2 and 2.5, and consists of a series of identities.

Proof of Theorem 2.3. By Proposition (2.2), by the definition $[x^j]\overline{T}_n(x) = \vartheta_p(j,n)/\vartheta_p(0,n)$, and by Theorem 0, we have

$$[x^j] \prod_{w\in W} r_w(x)^{|n|_w} = P_j\left(\left(|n|_w\right)_{w\in W_j}\right)$$

for all $n \in \mathbb{N}$. Proposition 2.5 implies that words $w \in W \setminus W_j$ do not contribute to the left hand side, since $|w| \ge j+2$ for these words and therefore $r_w(x) = 1 + \mathcal{O}\big(x^{j+1}\big)$. Let us reveal how the polynomial structure emerges in the left hand side. The idea is to apply an exp-log decomposition on (2.3). This is legitimate, as the constant term of $\overline{T}_n(x)$ and therefore of $r_w(x)$ is 1, compare (2.1). We have the identities

$$[x^j] \prod_{w\in W} r_w(x)^{|n|_w} = [x^j] \prod_{w\in W_j} r_w(x)^{|n|_w}$$

$$= [x^j] \prod_{w \in W_j} \exp\bigl(|n|_w \log r_w(x)\bigr)$$

$$= [x^j] \prod_{w \in W_j} \sum_{k \geq 0} |n|_w^k \frac{\bigl(\log r_w(x)\bigr)^k}{k!}$$

$$= \sum_{\substack{k_w \geq 0 \\ w \in W_j}} \left([x^j] \prod_{w \in W_j} \frac{\bigl(\log r_w(x)\bigr)^{k_w}}{k_w!} \right) \prod_{w \in W_j} |n|_w^{k_w},$$

where the last step is justified since there are only finitely many summands contributing to the j-th coefficient. (This is the case by the condition $r_w(0) = 1$, which implies $\log r_w(x) = \mathcal{O}(x)$ for $x \to 0$).

The right hand side is a polynomial in $|n|_w$ for $w \in W$, and by the uniqueness result (Proposition 2.1) the theorem is proved. □

Note that the argument given in the proof also gives a new proof of existence of the polynomials P_j.

Remark 4. By Proposition 2.5 we can determine exactly for which j a given monomial occurs first. Since $\overline{T}_w(0) = 1$ for all admissible words w, we have $r_w(x) = 1 + \alpha x^k + \mathcal{O}\bigl(x^{k+1}\bigr)$, where α is given by (2.4) and $k = |w| - 1$, therefore $\log r_w(x) = \alpha x^k + \mathcal{O}(x^{k+1})$ for some $\alpha \neq 0$. By Theorem 2.3 the monomial X_w occurs first in the polynomial P_j, where $j = |w| - 1$. More generally, the monomial $X_{w^{(1)}} \cdots X_{w^{(\ell)}}$ (repetitions allowed) occurs first in P_j, where $j = \bigl|w^{(1)}\bigr| + \cdots + \bigl|w^{(\ell)}\bigr| - \ell$. That is, the lower bound for the first occurrence of a monomial given by (1.7) is sharp.

We note that this observation is not sufficient to determine the number of terms in P_j; in the generating function appearing in Theorem 2.3 some higher coefficients may vanish. This is for example the case for $w = \texttt{110}$. We have

$$\log r_{\texttt{110}}(x) = \log \left(\frac{1 - (x/2)^3}{1 - (x/2)^2} \right) = \sum_{i \geq 1} \frac{x^{2i}}{i4^i} - \sum_{i \geq 1} \frac{x^{3i}}{i8^i},$$

and consequently the monomial $X_{\texttt{110}}$ does not occur in P_j for $j = 6\ell \pm 1$, where $\ell \geq 1$. It is however true that each nontrivial monomial occurs in infinitely many P_j.

Corollary 2.6. *Each monomial $X_{w^{(1)}}^{k_1} \cdots X_{w^{(\ell)}}^{k_\ell}$ except for the constant term 1 occurs in infinitely many P_j.*

Proof. By Theorem 2.3 the claim is equivalent to the statement that the power series $\prod_{i=1}^{\ell} \bigl(\log r_{w^{(i)}}(x)\bigr)^{k_i}$ is not a polynomial. We will analyse the possible singularities, which will contradict a polynomial behaviour.

Assume that ρ_i is the radius of convergence of the power series $\log r_{w^{(i)}}(x)$ and choose $j \in \{1, \dots, \ell\}$ such that $\rho_j = \min_{1\le i\le \ell} \rho_i$, moreover let x_j be a singularity of $\log r_{w^{(j)}}(x)$ on the circle $\{x : |x| = \rho_j\}$. By Proposition 2.5 we have $0 < \rho_j < \infty$, and that the power series $\log r_{w^{(i)}}(x)$ does not have a zero apart from $x = 0$. Therefore the singularities cannot cancel, which implies that x_j is a singularity of $\big(\log r_{w^{(1)}}(x)\big)^{k_1} \cdots \big(\log r_{w^{(\ell)}}(x)\big)^{k_\ell}$. Consequently, this expression is not a polynomial. □

Moreover, we want to derive an asymptotic estimate of the number of terms in P_j, using Proposition 2.5.

Corollary 2.7. *The number of terms N_j in the polynomial P_j satisfies the bound*

$$N_j \le \left[x^j\right] \frac{1}{1-x} \exp\left(\sum_{k\ge 1} \frac{1}{k} \frac{(p-1)^2 x^k}{1-px^k}\right).$$

Asymptotically, for $j \to \infty$, this upper bound is

$$\frac{e^{\mu(\sigma-1/2)}}{2p\mu^{1/4}\sqrt{\pi}} \frac{e^{2\sqrt{\mu j}} p^j}{j^{3/4}} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right),$$

with the constants $\mu = \frac{(p-1)^2}{p}$ and $\sigma = \sum_{k\ge 2} \frac{1}{k} \frac{1}{p^{k-1}-1}$. Moreover, we have

$$N_j = \Theta\big(p^j e^{2\sqrt{\mu j}} j^{-3/4}\big).$$

The same estimates are true for the number N'_j of terms in the polynomials P'_j.

Proof. The terms in P_j are built from the variables in W_j, see (1.2). In $W = \bigcup_{j\ge 1} W_j$ there are $p^{k-1}(p-1)^2$ many words w of *weight* $|w|-1$ equal to k, for $k \ge 2$. The corresponding generating function is $\mathcal{W}(x) = (p-1)^2 \frac{x}{1-px}$.

First, we want to determine the number of monomials having total weight j. These are the monomials that, by (1.7), may appear in P_j, but cannot appear in P_{j-1}. We obtain therefore the maximal number of "new" monomials in P_j.

A monomial is nothing else but a multiset of variables in W. Thus, by the multiset construction (see [17, page 27]) we obtain the *exp*-part of the generating function in the corollary. Finally, the factor $\frac{1}{1-x}$ stems from the fact that also monomials from $P_0, \dots, P_{j-1}$ are allowed in P_j.

For the asymptotic result, we first need to find the dominant singularity, i.e., the one closest to the origin. Note that the possible singularities are at $\omega_k^\ell p^{-1/k}$, for $\ell = 0, \dots, k-1$, where $\omega_k = \exp(2\pi i/k)$ is a k-th root of unity. As $p \ge 2$, the dominant one is found at $1/p$ for $k = 1$. Thus, we may decompose our generating function into

$$\exp\left(\frac{(p-1)^2x}{1-xp}\right)S(x),$$

where $S(x)$ is the generating function of the remaining factors. The crucial observation is that $S(x)$ is analytic for $|x| < 1/\sqrt{p}$, hence, for $|x| < 1/p$. This is a well-known type of functions for which a complete asymptotic expansion is known. Using Wright's result from [39, Theorem 2] we get the final result. The constants are coming from $S(1/p)$. The last statement follows from Proposition 2.5 and the asymptotic statement, since all monomials of weight j actually appear in P_j with a nonzero coefficient, and their number is a positive portion of the asymptotic main term. □

This type of functions was already intensively considered in the literature. It appears in the enumeration of permutations. The analysis builds on a saddle point method, see [17, Example VIII.7, p. 562]. Wright [39] derived the asymptotics for the general form of an exponential singularity we encounter here, extending the work of Perron [30].

Remark 5. We note that for the upper bound in Corollary 2.7 we do not need Proposition 2.5, but it suffices to use Rowland's paper, see (1.7). The lower bound however uses Proposition 2.5, which implies that all monomials of weight j do occur in the polynomial P_j.

For the prime $p = 2$, we implemented the method of finding the coefficients of P_j by Theorem 2.3 in the Sage Mathematics Software System [37]. In particular, we retrieve the formulas for $\vartheta_2(2,n), \dots, \vartheta_2(4,n)$ obtained by Howard [23], Spearman and Williams [35] and Rowland [31] before. Computing $P_0, \dots, P_{11}$ took less than five minutes using our implementation, which is a significant improvement over Rowland's algorithm [31].

We compare the actual number of nonzero coefficients in P_j (first line of numbers) with the upper bound from Corollary 2.7 (second line).

P_0	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}
1	1	4	11	29	69	174	413	995	2364	5581	13082
1	2	5	12	30	72	176	420	1005	2378	5611	13144

From this numerical evidence it seems reasonable to conjecture that the upper bound given in Corollary 2.7 gives in fact the asymptotic main term of the number N_j of nonzero coefficients of P_j. However, the exact behaviour of the integers N_j seems to be difficult to grasp, and remains an open problem at the moment.

2.2. **Asymptotic behaviour of coefficients of a given monomial.** In this chapter we study the different asymptotic behaviours exhibited by a sequence $(c_j)_{j\ge 0}$ of coefficients of a monomial. More precisely, we restrict ourselves to $p=2$ and monomials X_w for $w\in W$. The following lemma explains how the coefficients of the logarithm of a rational function behave asymptotically. We will apply it repeatedly in the subsequent discussion.

Lemma 2.8 (Coefficient asymptotics of $\log\circ\mathrm{rat}$)**.** *Let $r(x)$ be a rational function defined at 0 such that $r(0)=1$. Choose $L\ge 0$, $\varepsilon_0,\ldots,\varepsilon_{L-1}\in\mathbb{Z}\setminus\{0\}$ and pairwise different $\xi_0,\ldots,\xi_{L-1}\in\mathbb{C}\setminus\{0\}$ in such a way that*

$$r(x)=(1-\xi_0 x)^{\varepsilon_0}\cdots(1-\xi_{L-1}x)^{\varepsilon_{L-1}}.$$

(Note that this decomposition is unique up to the order of the factors.) Then

$$[x^n]\log r(x) = -\frac{1}{n}\sum_{0\le i<L}\varepsilon_i\xi_i^n \tag{2.5}$$

for $n\ge 1$. In particular, assume without loss of generality that $\xi_0,\ldots,\xi_{m-1}$, for some $1\le m\le L$, have maximal absolute value among the ξ_i, and $M=|\xi_0|$. Then

$$[x^n]\log r(x) = -\frac{1}{n}\sum_{0\le i<m}\varepsilon_i\xi_i^n + \mathcal{O}\big((M-\varepsilon)^n\big)$$

for some $\varepsilon>0$. If moreover $m=1$, we have for all $k\ge 1$

$$[x^n]\big(\log r(x)\big)^k = k(-\varepsilon_0)^k\big(\log n\big)^{k-1}\frac{\xi_0^n}{n}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right). \tag{2.6}$$

Proof. The first two statements follow immediately from the identity

$$[x^n]\log\left(\frac{1}{1-x}\right) = [x^n]\sum_{n\ge 1}\frac{x^n}{n} = \frac{1}{n}.$$

The asymptotic statements can be proved using standard results from singularity analysis (see Flajolet and Sedgewick [17]). We begin with the case $m=1$. First of all, the location of the dominant singularity (the one closest to the origin) is responsible for the exponential growth of the coefficients. Next note that the function $\log r(x)$ is singular if the rational function is either singular, or takes the value 0. If we assume that $\varepsilon_0>0$, the dominant singularity comes from the zero $1/\xi_0$ of the numerator of $r(x)$, and the exponential growth of the n-th coefficient is given by ξ_0^n. More precisely, a Taylor expansion of $r(x)$ at $x=r$ shows that

$$\log\left(r(x)\right) = \log\left(h(x)(x-r)^{d_r}\right) = -d_r\log\left(\frac{1}{1-x/r}\right)+\log(h(x)),$$

where $\log(h(x))$ is analytic for $|x| \leq |r| + \varepsilon$. If $\varepsilon_0 < 0$, we simply swap numerator and denominator of $r(x)$ and adjust the sign. If $m > 1$ one deals separately with the different singularities.

If higher powers of the logarithm are considered we have to deal with Cauchy products. In this case one can elementarily show the appearance of the $\left(\log n\right)^{k-1}$ terms by partial summation combined with $\sum_{k=1}^{n} \frac{1}{k} = \log n + \mathcal{O}\left(1\right)$. For more details we refer to [17, Chapter VI]. □

Examples. Let $p = 2$ and consider $\log\left(r_{110}(x)\right) = \log\left(\frac{1+x/2+x^2/4}{1+x/2}\right)$. Here, the numerator has the two roots $2e^{2\pi i/3}$ and $2\,\mathrm{e}^{-2\pi i/3}$, whereas the denominator has the root -2. In this case all roots lie on the same circle $|x| = 2$, and therefore cancellations take place (compare Remark 4). By (2.5) we obtain

$$[x^n] \log r_{110}(x) = \frac{2^{-n}}{n}\left((-1)^n - e^{2\pi i n/3} - e^{-2\pi i n/3}\right).$$

In this special case we have equality, as no other roots are involved. Since the radius of convergence is larger than 1, we can obtain the infinite sum of coefficients c_j of X_{110} by inserting 1 into the generating function:

$$\sum_{j\geq 0} c_j = \sum_{j\geq 0} \left[x^j\right] \log r_{110}(x) = \lim_{j\to\infty} \left[x^j\right] \frac{\log r_{110}(x)}{1-x}$$
$$= \log r_{110}(1) = \log(7/6).$$

Now we consider the generating function $\frac{1}{2}\left(\log(1+x/2)\right)^2$ corresponding to the coefficients c_j of X_{10}^2. In this case we have, by (2.6),

$$c_j = \frac{(-1)^j \log j}{j \cdot 2^j}\left(1 + \mathcal{O}(1/j)\right).$$

In this simple case an exact form of the coefficients can be obtained from (2.5), using the Cauchy product of

$$\log r_{10}(x) = \sum_{j\geq 1} \frac{(-1)^j}{j \cdot 2^j} x^j$$

with itself:

$$c_j = \left[x^j\right] \frac{1}{2}\left(\log r_{10}(x)\right)^2 = \frac{(-1)^j}{2^{j+1}} \sum_{\substack{i_1,i_2\geq 1\\ i_1+i_2=j}} \frac{1}{i_1 i_2}.$$

Moreover, similarly as in the first example we have

$$\sum_{j\geq 0} c_j = \frac{1}{2}\left(\log(3/2)\right)^2.$$

Let us now consider special classes of monomials, whose generating function has a large radius of convergence and can be evaluated at $x = 1$.

Corollary 2.9. *Consider the words $w = \mathtt{1}^s\mathtt{0}$ or $w = \mathtt{1}^{4s+1}\mathtt{00}$ for $s \geq 1$. For fixed word w and an integer $k \geq 0$ let c_j be the coefficient of the corresponding monomial X_w^k. Then the radius of convergence of $\sum_{j\geq 0} c_j x^j$ is greater than 1 (more precisely, equal to 2 for the first family of values). Thus,*

$$\sum_{j\geq 0} c_j = \frac{1}{k!}\big(\log r_w(1)\big)^k.$$

Proof. By the main theorem the considered generating function is given by $\frac{1}{k!}\log\big(r_w(x)\big)^k$. Let us start with the first family of words. We need to analyse the rational function $r_w(x) = \frac{T_{\mathtt{1}^s\mathtt{0}}(x)}{T_{\mathtt{1}^{s-1}\mathtt{0}}(x)}$, as our plan is to apply Lemma 2.8. It is not difficult to show (see also (3.2)) that

$$T_{\mathtt{1}^s\mathtt{0}}(x) = \frac{1-(x/2)^{s+1}}{1-x/2}.$$

Thus, $r_w(x) = \frac{1-(x/2)^{s+1}}{1-(x/2)^s}$, and we see that all roots of the numerator and the denominator are located on the circle $|x| = 2$.

For the second family of words, we get

$$T_{\mathtt{1}^r\mathtt{00}}(x) = \frac{q_{r+1}(x/2)}{q_r(x/2)}\cdot\frac{1-(x/2)^r}{1-(x/2)^{r+1}}, \quad \text{with} \quad q_r(t) = 4t^{r+1} + t^r - 4t^2 - 1.$$

Hence, we are interested in the roots of the polynomials $q_r(x)$. By Rouche's Theorem there are exactly 2 roots inside the disc $|t| < 2^{-1}(1+2^{-r+2})$. These two are very close to $\pm i/2$. In particular, by Newton's method starting with $i/2$, we get after one iteration the very good approximation

$$\frac{i}{2} + \left(\frac{i}{2}\right)^r\left(\frac{1}{2} - \frac{i}{4}\right) + O\left(\frac{1}{2^{2r}}\right).$$

Therefore, the roots of $q_r(t)$ are in absolute value greater than $1/2$ for $r \equiv 1,2 \mod 4$ and less than $1/2$ for $r \equiv 0,3 \mod 4$. In particular, for $r \equiv 1 \mod 4$ we have that the roots of $q_{r+1}(x/2)$ and $q_r(x/2)$ are both in absolute value greater than 1. Thus, the radius of convergence is larger than 1, and it is legitimate to insert 1. □

By Lemma 2.8 the sequence of coefficients $(c_j)_{j\geq 0}$ for a given word w can exhibit different kinds of behaviours, corresponding to the position of the zeros and singularities of $r_w(x)$. Because of the construction of $r_w(x)$, there is a convergence–divergence dichotomy, which we summarize in the following corollary.

Corollary 2.10. *Let $w \in W$ and write $r_w(x) = (1-\xi_0 x)^{\varepsilon_0}\cdots(1-\xi_{L-1}x)^{\varepsilon_{L-1}}$ with pairwise different, nonzero $\xi_i \in \mathbb{C}$ and nonzero $\varepsilon_i \in \mathbb{Z}$, such that $|\xi_0| \geq \cdots \geq |\xi_{L-1}|$.*

(a) If $|\xi_0| \le 1$, the sequence c_w converges, moreover we have the convergent series

$$\sum_{j\ge 0} c_j = \log r_w(1).$$

(b) If $|\xi_0| > 1$, the sequence c_w diverges. If moreover $1/\xi_0$ is the only dominant singularity, then ξ_0 is a real number in $(-\infty, -1]$, and we have $c_w(j) \sim -\varepsilon_0 \xi_0^j / j$.

Proof. The case $|\xi_0| < 1$ is clear, since the function $\log r_w(x)$ has no singularity in the closed unit circle in this case. For the case $|\xi_0| = 1$ we note that $\xi_i \ne 1$ for all i, since T_v has only positive coefficients. Since the sum $\sum_{j\ge 1} \xi^j / j$ converges for all j on the unit circle such that $j \ne 1$, the sum $\sum_{j\ge 1} c_j$ converges by (2.5). Abel's limit theorem finishes the proof for this case. Finally, case (b) follows from Lemma 2.8 and the positivity of coefficients of T_v. □

In the following, let $p = 2$. We have seen (Corollaries 2.4 and 2.9) that case (a) occurs for $w = \mathtt{1}^s\mathtt{0}$, where $s \ge 1$.

Case (b) appears for $w = \mathtt{1010}$ (dominant singularity at $x_0 \sim -0.86408$). In this case the singularity is coming from the log, as $r_w(x_0) = 0$. Thus log becomes singular. This is also called a supercritical composition scheme, as the outer function is responsible for the singularity.

This case also appears for $w = \mathtt{10100}$ (dominant singularity again at $x_0 \sim -0.86408$). In this case however, the denominator of r_w is zero at x_0, thus the singularity is coming from a simple pole. This is also called a subcritical composition scheme, as the inner function is responsible for the singularity.

By approximate computation of the roots of $\overline{T}_v$ using GNU Octave [12] we determined all words of length at most 10 for which case (a) occurs. Besides for the words of the form $\mathtt{1}^s\mathtt{0}$ or $\mathtt{1}^{4s+1}\mathtt{00}$, this also seems to be the case for the words $\mathtt{1}^s\mathtt{01}^t\mathtt{0}$, where $s \ge 1$ and $t \ge 2$. Here is the list of remaining words $w \in W$ of length at most 10, not falling into one of these three classes, for which this case occurs too.

10011110, 101101110, 101110110, 101111010,
101111100, 111011010, 1011011110, 1011101110,
1011110110, 1101101110, 1101110110, 1101111010,
1101111100, 1111011010.

We leave the classification of the words $w \in W$ for which the sum $\sum_{j\ge 0} c_j$ converges as an open problem.

2.3. **A simplified recurrence for $\vartheta_p(j,n)$.** Rarefying $\vartheta_p(j,n)$ in the first coordinate by the factor $p-1$, and shifting j by $s_p(n)$ many places, the recurrence (1.5) is transformed into a simpler form: the term ν_p disappears, instead the maximal shift occurring in the first coordinate is $2p-1$. We pass to the details. Define, for $k,n \ge 0$,

$$\tilde{\vartheta}_p(k,n) = \begin{cases} \vartheta_p\left(\frac{k-s_p(n)}{p-1}, n\right), & k \ge s_p(n) \text{ and } p-1 \mid k-s_p(n); \\ 0, & \text{otherwise.} \end{cases}$$

Setting for simplicity $\tilde{\vartheta}_p(k,n) = 0$ if $k < 0$ or $n < 0$, we obtain the following recurrence relation for $k,n \ge 0$, where we use the Kronecker delta, which is defined by $\delta_{i,i} = 1$, and $\delta_{i,j} = 0$ for $i \ne j$.

$$\begin{aligned} \tilde{\vartheta}_p(0,n) &= \delta_{0,n}, \quad n \ge 0; \\ \tilde{\vartheta}_p(k,0) &= \delta_{k,0}, \quad k \ge 0, \end{aligned}$$

and for $n \ge 0$ and $0 \le a < p$,

$$\tilde{\vartheta}_p(k, pn+a) = (a+1)\tilde{\vartheta}_p(k-a,n) + (p-a-1)\tilde{\vartheta}_p(k-p-a,n-1).$$

The proof of this new recurrence is straightforward and uses the identity

$$s_p(n+1) - s_p(n) = 1 - (p-1)\nu_p(n+1), \tag{2.7}$$

which follows immediately by writing n in base p and counting the number of times the digit $\mathtt{q}$ occurs at the lowest digits of n, and also the recurrence

$$s_p(pn+a) = s_p(n) + a \qquad (0 \le a < p).$$

In Tables 1–3 we list some coefficients of $\tilde{\vartheta}_p(k,n)$ for $p = 2,3,5$, respectively.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	1																	
1		2	2		2				2								2	
2			1	4	1	4	4		1	4	4		4				1	4
3					2	2	2	8	2	2	4	8	2	8	8		2	2
4							1		4	4	1	4	5	4	4	16	4	4
5											2		2	2	2		8	8
6															1			

TABLE 1. Some coefficients of $\tilde{\vartheta}_2(k,n)$. The variable k corresponds to the row number in this table.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	1																	
1		2		2						2								
2			3		4		3				4		4					
3				2		6		6		2		6		8		6		
4					1		4		9		4		5		12		12	
5								2		6		6		4		8		18
6											3		4		3		4	
7														2		2		
8																	1	

Table 2. Some coefficients of $\tilde{\vartheta}_3(k,n)$.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	1																	
1		2				2												
2			3				4				3							
3				4				6				6				4		
4					5				8				9				8	
5						4				10				12				12
6							3				8				15			
7								2				6				12		
8									1				4				9	
9														2				6

Table 3. Some coefficients of $\tilde{\vartheta}_5(k,n)$.

We want to derive a product representation for $\tilde{\vartheta}_p(j,n)$. In order to do so, we note the well-known fact due to Legendre stating that

$$\nu_p(n!) = \frac{n - s_p(n)}{p-1}, \tag{2.8}$$

for prime p. This can be proved easily by summing the identity (2.7). Applying (2.8) three times, we obtain

$$\nu_p\binom{n}{t} = \frac{s_p(n-t) + s_p(t) - s_p(n)}{p-1}. \tag{2.9}$$

We note that, by Kummer's theorem [28], the left hand side of (2.9) is the number of borrows occurring in the subtraction $n-t$. Let us define the bivariate generating function $\widetilde{T}(x,z) \coloneqq \sum_{k,n\geq 0} \tilde{\vartheta}_p(k,n)x^k z^n$. We will prove that $\widetilde{T}$ can be written compactly as an infinite product. By the definition of $\tilde{\vartheta}$, the binomial coefficient $\binom{n}{t}$ contributes to $k = s_p(n) + (p-1)\nu_p\binom{n}{t}$. Thus, we obtain by (2.9)

$$\begin{aligned}\widetilde{T}(x,z) &= \sum_{n\geq 0} z^n \sum_{t=0}^{n} x^{s_p(n)+(p-1)\nu_p\binom{n}{t}} = \sum_{n\geq 0} z^n \sum_{t=0}^{n} x^{s_p(t)+s_p(n-t)} \\ &= \left(\sum_{n\geq 0} z^n x^{s_p(n)}\right)^2 = \prod_{i\geq 0}\left(1 + xz^{p^i} + x^2 z^{2p^i} + \cdots + x^{p-1}z^{(p-1)p^i}\right)^2,\end{aligned}$$

where the last equality holds due the uniqueness of the base-p expansion of an integer n. This product representation should be compared to [7, Equations (3.3), (3.12)]. Since Carlitz does not use the transformation in the first coordinate, his product takes a more complicated form. For $p=2$ we have the special case

$$\sum_{k,n\geq 0} \tilde{\vartheta}_2(k,n)x^k z^n = \prod_{i\geq 0}\left(1 + xz^{2^i}\right)^2.$$

We note that this product representation can be used for an alternative proof of Carlitz' recurrence (1.4).

We finish this section with a remark on divisibility in *columns* of Pascal's triangle.

2.4. **Divisibility in columns of Pascal's triangle.** In the recent paper [11] by Drmota, Kauers, and the first author, we deal with a conjecture by Cusick (private communication, 2012, 2015) stating that

$$c_t \coloneqq \operatorname{dens}\{m \geq 0 : s_2(m+t) \geq s_2(m)\} > 1/2,$$

for all $t \geq 0$. Here $\operatorname{dens} A$ denotes the asymptotic density of a set $A \subseteq \mathbb{N}$, which exists in this case. By (2.9) this corresponds to a problem on divisibility in columns of Pascal's triangle: if we define $\rho_2(j,t) = \operatorname{dens}\bigl\{m \geq 0 : \nu_2\binom{m+t}{m} = j\bigr\}$[2], the conjecture states that

$$\sum_{j\leq s_2(t)} \rho_2(j,t) > 1/2.$$

We gave [11, Theorem 1] a partial answer, solving the conjecture for almost all t in the sense of asymptotic density. More precisely, we proved that for

[2] In [11], we use the notations $\delta(j,t) = \operatorname{dens}\bigl\{m \geq 0 : s_2(m+t) - s_2(m) = j\bigr\}$ for all $j \in \mathbb{Z}$, and $b_{2^j} = \operatorname{dens}\bigl\{m : 2^j \nmid \binom{m+t}{m}\bigr\}$. We have $\rho_2(j,t) = \delta(s_2(t)-j,t)$ for all $j \geq 0$ and $b_{2^j}(t) = \rho_2(0,t) + \cdots + \rho_2(j-1,t)$ for $j \geq 1$.

all $\varepsilon > 0$,

$$\bigl|\{t \le T : 1/2 < c_t < 1/2 + \varepsilon\}\bigr| = T + O(T/\log T).$$

The full statement of Cusick's conjecture is however still an open problem. We also want to note the recent work by Emme and Hubert [13] (preprint), which continues earlier work by Emme and Prikhod'ko [14] (preprint). They proved that for almost all $X \in \{0,1\}^{\mathbb{N}}$ with respect to the balanced Bernoulli measure the values

$$\operatorname{dens}\bigl\{n \in \mathbb{N} : s_2(n + a_X(k)) - s_2(n) \le x\sqrt{k/2}\bigr\}$$

converge pointwise to the standard normal distribution as $k \to \infty$, where $a_X(k) = \sum_{0 \le j < k} X_j 2^j$.

Surprisingly, the "column densities" $\rho_2(j,t)$ can be expressed by the same polynomial P_j as the "row counts" $\vartheta_2(j,n)$ (see [11, Sections 3.2 and 3.3]). We have $\rho_2(0,t) = 2^{-|t|_1}$ and, for example,

$$\rho_2(1,t)/\rho_2(0,t) = \frac{1}{2}\,|t|_{\mathtt{01}}\,,$$

$$\rho_2(2,t)/\rho_2(0,t) = -\frac{1}{8}\,|t|_{\mathtt{01}} + \frac{1}{8}\,|t|_{\mathtt{01}}^2 + |t|_{\mathtt{011}} + \frac{1}{4}\,|t|_{\mathtt{001}}\,.$$

In general, if we denote by $\overline{w}$ the Boolean complement of the word $w \in W$, these expressions are obtained by inserting the value $|t|_{\overline{w}}$ for the variable X_w in P_j (compare to (1.3)):

$$t \mapsto \bigl(|t|_{\overline{w}}\bigr)_{w \in W_j} \mapsto P_j\Bigl(\bigl(|t|_{\overline{w}}\bigr)_{w \in W_j}\Bigr) = \frac{\rho_2(j,t)}{\rho_2(0,t)}.$$

3. Proofs

Proof of Proposition 2.1. Assume that P_j and $\widetilde{P}_j$ are two polynomials in the variables X_w ($w \in W$), representing $\vartheta(j,n)/\vartheta(0,n)$, and let R be the maximal degree with which a variable X_w occurs in P_j or $\widetilde{P}_j$. Moreover, let ℓ be such that $\ell+1$ is the maximal length of a word w such that the variable X_w occurs in one of the polynomials. The strategy is to compute the coefficients of a polynomial starting from its values. For a multivariate polynomial in M variables, where the degree of each variable is bounded by R, this can be done by evaluating the polynomial at each tuple in $\{0,\ldots,R\}^M$, and applying recursively the fact that a univariate polynomial q is determined by $\deg q + 1$ of its values. We adapt this strategy, taking the dependence between the variables into account.

On the set W_ℓ we have a partial order $\preceq$ defined by $v \preceq w$ if and only if v is a factor of w. For convenience, we extend this order to a total order on W_ℓ and denote it the same symbol $\preceq$. Let $w_0,\ldots,w_{M-1}$ be the increasing enumeration of W_ℓ (where $M = |W_\ell|$). We will work with certain "test integers", defined as follows. For a vector $a = (a_m)_{m<M}$ in $\{0,\ldots,R\}^M$ let $n(a)$ be the integer whose binary expansion is given by the concatenation $v_{M-1}\cdots v_0$, where

$$v_m = \left(w_m\mathtt{q}^\ell\mathtt{0}^\ell\right)^{a_m}\left(\mathtt{q}^\ell\mathtt{0}^\ell\right)^{R-a_m}.$$

The idea behind this is that $\mathtt{q}^\ell\mathtt{0}^\ell$ acts as a "separator" in the sense that admissible factors of $n(a)$ of length $\le \ell+1$ are contained completely in one of the building blocks $w_m\mathtt{q}^\ell\mathtt{0}^\ell$ or $\mathtt{q}^\ell\mathtt{0}^\ell$. (At this point the restrictions $w_{\nu-1}\neq\mathtt{0}$, $w_0\neq\mathtt{q}$ for a word $w_{\nu-1}\cdots w_0\in W$ come into play.) By varying the values a_m we can therefore vary the factor count $|\cdot|_{w_m}$ without changing $|\cdot|_{w_{m'}}$ for $m'>m$. For simplicity, we rename the variables X_{w_m} to X_m. We prove the following statement by induction on s.

Claim. *Assume that s is an integer, $0\le s\le M$. For all $a_0,\ldots,a_{M-1}$, $k_0,\ldots,k_{s-1}\in\{0,\ldots,R\}$ we have*

$$\left[X_0^{k_0}\cdots X_{s-1}^{k_{s-1}}\right]\left(P_j-\widetilde{P}_j\right)\left(X_0,\ldots,X_{s-1},|n(a)|_{w_s},\ldots,|n(a)|_{w_{M-1}}\right)=0.$$

The case $s=0$ follows from the assumption that P_j and $\widetilde{P}_j$ yield the same value for all assignments $X_w=|n|_w$, where $n\ge 0$. The case $s=M$ is the desired statement that $P_j=\widetilde{P}_j$, by the fact that the degree of each variable in P_j and $\widetilde{P}_j$ is bounded by R. Assume therefore that the statement holds for some $s<M$ and let $a_0,\ldots,a_{M-1},k_0,\ldots,k_{s-1}\in\{0,\ldots,R\}$. We define polynomials $Q(X_s)$ and $\widetilde{Q}(X_s)$ in one variable, of degree at most R, by

$$Q(X_s)=\left[X_0^{k_0}\cdots X_{s-1}^{k_{s-1}}\right]P_j\left(X_0,\ldots,X_s,|n(a)|_{w_{s+1}},\ldots,|n(a)|_{w_{M-1}}\right),$$

analogously $\widetilde{Q}$. By the definition of the total order $\preceq$ we have

$$\left|n\left(a^{(r)}\right)\right|_{w_m}=\left|n(a)\right|_{w_m}$$

for $0\le r\le R$ and $m>s$, where

$$a_\ell^{(r)}=\begin{cases}a_\ell, & \ell\neq s;\\ r, & \ell=s.\end{cases}$$

By applying the induction hypothesis for $a^{(0)},\ldots,a^{(R)}$, we obtain the equality $Q(N)=\widetilde{Q}(N)$ for the $R+1$ values $\left|n\left(a^{(0)}\right)\right|_{w_s},\ldots,\left|n\left(a^{(R)}\right)\right|_{w_s}$ of N, therefore

$$\begin{aligned}0&=\left[X_s^{k_s}\right](Q-\widetilde{Q})(X_s)\\&=\left[X_0^{k_0}\cdots X_m^{k_s}\right]\left(P_j-\widetilde{P}_j\right)\left(X_0,\ldots,X_s,|n(a)|_{w_{s+1}},\ldots,|n(a)|_{w_{M-1}}\right).\end{aligned}$$

This proves that $P_j=\widetilde{P}_j$. □

Proof of Proposition 2.2. Let $v \in W \cup \{\varepsilon\}$. The proof is by induction on the length of v, the case $v = \varepsilon$ being trivial. Moreover, for the words $c\mathtt{0}^s a$, where $c \in \{\mathtt{1}, \ldots, \mathtt{q}\}$, $s \geq 0$ and $a \in \{\mathtt{0}, \ldots, \mathtt{q}-\mathtt{1}\}$, we obtain

$$\prod_{w \in W} \left(\frac{\overline{T}_w \overline{T}_{w_{LR}}}{\overline{T}_{w_R} \overline{T}_{w_L}} \right)^{|v|_w} = \frac{\overline{T}_{c\mathtt{0}^s a}}{\overline{T}_{c\mathtt{0}^s}} \cdot \frac{\overline{T}_{c\mathtt{0}^s}}{\overline{T}_{c\mathtt{0}^{s-1}}} \cdots \frac{\overline{T}_{c\mathtt{0}}}{\overline{T}_{\varepsilon}} = \overline{T}_{c\mathtt{0}^s a}.$$

Suppose that the statement holds for some $v' \in W$. It is sufficient to show that it is also true for $v = a\mathtt{0}^s v'$, where $a \in \{\mathtt{1}, \ldots, \mathtt{q}\}$ and $s \geq 0$.

Since words in W do not end with the letter $\mathtt{0}$ (read from right to left), an admissible factor of v is either a factor of v' or a suffix of v. This implies that the product corresponding to v is obtained from the product corresponding to v', multiplying by $\overline{T}_w \overline{T}_{w_{LR}} / (\overline{T}_{w_R} \overline{T}_{w_L})$ for each suffix w of v such that $w \in W$. This product of suffixes equals

$$\prod_{\substack{w \text{ suffix of } v \\ w \in W}} \frac{\overline{T}_w \overline{T}_{w_{LR}}}{\overline{T}_{w_R} \overline{T}_{w_L}} = \prod_{\substack{w \text{ suffix of } v \\ w \in W}} \frac{\overline{T}_w}{\overline{T}_{w_R}} \prod_{\substack{w \text{ suffix of } v' \\ w \in W}} \frac{\overline{T}_{w_R}}{\overline{T}_w} = \frac{\overline{T}_v}{\overline{T}_{v'}}.$$

This shows the desired form and together with the induction hypothesis it yields the claim. □

Finally, we prove Proposition 2.5 by a somewhat tedious case distinction.

Proof of Proposition 2.5. Assume that $w = w_{\nu-1} \cdots w_0 \in W$. The statement we want to prove is equivalent to

$$\overline{T}_w \overline{T}_{w_{LR}} - \overline{T}_{w_L} \overline{T}_{w_R} = \alpha x^{\nu-1}, \tag{3.1}$$

where

$$\alpha = p^{\nu-2} \frac{w_{\nu-1}}{w_{\nu-1}+1} \frac{p - w_0 - 1}{w_0 + 1} \prod_{2 \leq d \leq p} d^{-2|w'|_{d-1}},$$

and w' is obtained from w by omitting the left- and rightmost digits. We want to prove the statement by induction on the *right depth* of $w \in W$. This is the number of right truncations needed to map w to a *base case*, which are words v such that $v_{LR} = \varepsilon$. Among the base cases there are words v satisfying $v_L = \varepsilon$. These are exactly the words of the form $c\mathtt{0}^t a$, for $c \neq \mathtt{0}$, $t \geq 0$ and $a \neq \mathtt{q}$. Each remaining base case falls into exactly one of the following classes, where $c \in \{\mathtt{1}, \ldots, \mathtt{q}\}$ and $a \in \{\mathtt{0}, \ldots, \mathtt{q}-\mathtt{1}\}$.

$v = c\mathtt{q}^s a$	with $s \geq 1$;
$v = cb\mathtt{q}^s a$	with $b \notin \{\mathtt{0}, \mathtt{q}\}$ and $s \geq 0$;
$v = c\mathtt{0}^t b\mathtt{q}^s a$	with $t \geq 1$, $b \notin \{\mathtt{0}, \mathtt{q}\}$ and $s \geq 0$;
$v = c\mathtt{0}^t \mathtt{q}^s a$	with $t \geq 1$ and $s \geq 1$.

We begin with the following formulas, which can be proved from the recurrence (2.2) in a straightforward way, and which we will use throughout this proof. Assume that $w = \{\mathtt{0},\ldots,\mathtt{q}\}^*$, $s \ge 1$, $t \ge 0$, $c \in \{\mathtt{1},\ldots,\mathtt{q}\}$, and $a \in \{\mathtt{0},\ldots,\mathtt{q}-\mathtt{1}\}$. Then

$$\begin{aligned}
(3.2)\quad T_{w\mathtt{q}^s a}(x) &= p^s\Big((a+1)+(p-a-1)(p-1)A_1(s)\Big)T_w(x)\\
&\quad + (p-a-1)x^s T_{w(\mathtt{q}-\mathtt{1})}(x),\\
(3.3)\quad T_{wc\mathtt{0}^t a}(x) &= \frac{1}{p}\Big((p-a-1)(px)^{t+1}+(a+1)(p-1)A_2(t)\Big)T_{w(c-1)}(x)\\
&\quad + (a+1)T_{wc}(x),
\end{aligned}$$

where we set $A_1(s) = \sum_{1\le i<s}(x/p)^i$ and $A_2(t) = \sum_{1\le i\le t}(px)^i$. We note the following special case of (3.3):

$$T_{ca} = (c+1)(a+1) + c(p-a-1)x. \tag{3.4}$$

We proceed to evaluating $\overline{T}_w\overline{T}_{w_{LR}} - \overline{T}_{w_L}\overline{T}_{w_R}$ for the base cases, thus confirming (3.1) for these cases. If $w = ca$, $c \ne \mathtt{0}$, and $a \ne \mathtt{q}$, we have $\overline{T}_w\overline{T}_{w_{LR}} - \overline{T}_{w_L}\overline{T}_{w_R} = \overline{T}_w - 1 = \frac{c}{c+1}\frac{p-a-1}{a+1}x$ by (3.4). If $w = c\mathtt{q}^s a$, where $s \ge 1$, $c \ne \mathtt{0}$, and $a \ne \mathtt{q}$, we obtain by (3.2) and (3.4)

$$\begin{aligned}
\overline{T}_{c\mathtt{q}^s a}(x) &= 1 + \frac{p-a-1}{a+1}(p-1)A_1(s+1) + (x/p)^s\frac{c}{c+1}\frac{p-a-1}{a+1}x,\\
\overline{T}_{\mathtt{q}^s a}(x) &= 1 + \frac{p-a-1}{a+1}(p-1)A_1(s+1),
\end{aligned}$$

therefore

$$\overline{T}_w\overline{T}_{w_{LR}} - \overline{T}_{w_L}\overline{T}_{w_R} = \overline{T}_{c\mathtt{q}^s a}(x) - \overline{T}_{\mathtt{q}^s a}(x) = x^{s+1}p^{-s}\frac{c}{c+1}\frac{p-a-1}{a+1}.$$

If $w = c\mathtt{0}^t a$, where $t \ge 1$, $c \in \{\mathtt{1},\ldots,\mathtt{q}\}$, and $a \in \{\mathtt{0},\ldots,\mathtt{q}-1\}$, we obtain by (3.3)

$$\begin{aligned}
\overline{T}_{c\mathtt{0}^t a}(x) &= 1 + \frac{p-a-1}{a+1}\frac{c}{c+1}p^t x^{t+1} + \frac{p-1}{p}\frac{c}{c+1}A_2(t),\\
\overline{T}_{c\mathtt{0}^t}(x) &= 1 + \frac{p-1}{p}\frac{c}{c+1}A_2(t),
\end{aligned}$$

therefore

$$\overline{T}_w\overline{T}_{w_{LR}} - \overline{T}_{w_L}\overline{T}_{w_R} = p^t x^{t+1}\frac{p-a-1}{a+1}\frac{c}{c+1}.$$

Now let $w = cbq^s a$ for some $c \ne \mathtt{0}$, $b \in \{\mathtt{1},\ldots,\mathtt{q}-\mathtt{1}\}$, $s \ge 0$, and $a \ne \mathtt{q}$. The case $s = 0$ can be verified easily: after a short calculation we obtain the expected result

$$\overline{T}_{cba} - \overline{T}_{ba}\overline{T}_{cb} = \frac{c}{c+1}\frac{p}{(b+1)^2}\frac{p-a-1}{a+1}x^2.$$

Otherwise we get by (3.2):

$$\begin{aligned}(b+1)T_wT_{w_{LR}} - T_{w_L}T_{w_R} &= (b+1)T_{cb\mathtt{q}^sa} - T_{b\mathtt{q}^sa}T_{cb}\\ &= (b+1)\Big(p^s\big((a+1)+(p-a-1)(p-1)A_1(s)\big)T_{cb}\\ &\quad + (p-a-1)x^sT_{cb(\mathtt{q}-1)}\Big)\\ &\quad - \Big(p^s\big((a+1)+(p-a-1)(p-1)A_1(s)\big)T_b\\ &\quad + (p-a-1)x^sT_{b(\mathtt{q}-1)}\Big)T_{cb}\\ &= (p-a-1)x^s\Big((b+1)T_{cb(\mathtt{q}-1)} - T_{b(\mathtt{q}-1)}T_{cb}\Big).\end{aligned}$$

Using the case $s=0$, we obtain

$$\begin{aligned}\overline{T}_w\overline{T}_{w_{LR}} - \overline{T}_{w_L}\overline{T}_{w_R} &= \frac{p-a-1}{a+1}(p-1)p^{-s}x^s\big(\overline{T}_{cb(\mathtt{q}-1)} - \overline{T}_{b(\mathtt{q}-1)}\overline{T}_{cb}\big)\\ &= \frac{c}{c+1}\frac{1}{(b+1)^2}\frac{p-a-1}{a+1}p^{-s+1}x^{s+2}.\end{aligned}$$

Let $w = c\mathtt{0}^tb\mathtt{q}^sa$, where $c \neq \mathtt{0}$, $t \geq 1$, $b \notin \{\mathtt{0},\mathtt{q}\}$, $s \geq 0$, and $a \neq \mathtt{q}$. If $s=0$, we obtain by (3.4) and (3.3),

$$\begin{aligned}&(b+1)T_wT_{w_{LR}} - T_{w_L}T_{w_R} = (b+1)T_{c\mathtt{0}^tba} - T_{ba}T_{c\mathtt{0}^tb}\\ &\quad = (b+1)\Big((a+1)T_{c\mathtt{0}^tb} + (p-a-1)xT_{c\mathtt{0}^t(b-1)}\Big)\\ &\quad - \Big((b+1)(a+1) + b(p-a-1)x\Big)T_{c\mathtt{0}^tb}\\ &\quad = (b+1)(p-a-1)x\left(\frac{1}{p}\Big((p-b)(px)^{t+1} + b(p-1)A_2(t)\Big)T_{c-1} + bT_c\right)\\ &\quad - b(p-a-1)x\left(\frac{1}{p}\Big((p-b-1)(px)^{t+1}\right.\\ &\qquad \left. + (b+1)(p-1)A_2(t)\Big)T_{c-1} + (b+1)T_c\right)\\ &\quad = (p-a-1)p^{t+1}x^{t+2}c.\end{aligned}$$

Therefore we get in this case

$$\overline{T}_w\overline{T}_{w_{LR}} - \overline{T}_{w_L}\overline{T}_{w_R} = p^{t+1}x^{t+2}\frac{c}{c+1}\frac{1}{(b+1)^2}\frac{p-a-1}{a+1}.$$

If $s \geq 1$, we obtain, using (3.2)–(3.4),

$$\begin{aligned}&(b+1)T_wT_{w_{LR}} - T_{w_L}T_{w_R} = (b+1)T_{c\mathtt{0}^tb\mathtt{q}^sa} - T_{b\mathtt{q}^sa}T_{c\mathtt{0}^tb}\\ &\quad = (b+1)\bigg(p^s\Big((a+1)+(p-a-1)(p-1)A_1(s)\Big)T_{c\mathtt{0}^tb}\\ &\qquad + (p-a-1)x^sT_{c\mathtt{0}^tb(\mathtt{q}-1)}\bigg)\end{aligned}$$

$$
\begin{aligned}
&- \Bigl(p^s\Bigl((a+1)+(p-a-1)(p-1)A_1(s)\Bigr)T_b+(p-a-1)x^sT_{b(\mathtt{q}-1)}\Bigr)T_{c0^tb}\\
&=(p-a-1)x^s\bigl((b+1)T_{c0^tb(\mathtt{q}-1)}-T_{b(\mathtt{q}-1)}T_{c0^tb}\bigr)\\
&=(p-a-1)x^s\Bigl((b+1)\bigl((p-1)T_{c0^tb}+xT_{c0^t(b-1)}\bigr)\\
&\quad-\bigl((b+1)(p-1)+bx\bigr)T_{c0^tb}\Bigr)\\
&=(p-a-1)x^{s+1}\Bigl((b+1)T_{c0^t(b-1)}-bT_{c0^tb}\Bigr)\\
&=(p-a-1)(b+1)x^{s+1}\biggl(\frac{1}{p}\Bigl((p-b)(px)^{t+1}+b(p-1)A_2(t)\Bigr)T_{c-1}+bT_c\biggr)\\
&-(p-a-1)bx^{s+1}\biggl(\frac{1}{p}\Bigl((p-b-1)(px)^{t+1}\\
&\quad+(b+1)(p-1)A_2(t)\Bigr)T_{c-1}+(b+1)T_c\biggr)\\
&=(p-a-1)p^{t+1}x^{s+t+2}c,
\end{aligned}
$$

which yields the statement also for this case. We proceed with the case $w=c0^t\mathtt{q}^sa$, where $c\neq \mathtt{0}$, $t,s\geq 1$, and $a\neq \mathtt{q}$. In this case, we have

$$
\begin{aligned}
&T_wT_{w_{LR}}-T_{w_L}T_{w_R}=T_{c0^t\mathtt{q}^sa}-T_{\mathtt{q}^sa}T_{c0^t}\\
&\quad=\Bigl(p^s\Bigl((a+1)+(p-a-1)(p-1)A_1(s)\Bigr)T_{c0^t}+(p-a-1)x^sT_{c0^t(\mathtt{q}-1)}\Bigr)\\
&\quad-\Bigl(p^s\Bigl((a+1)+(p-a-1)(p-1)A_1(s)\Bigr)+(p-a-1)x^sT_{\mathtt{q}-1}\Bigr)T_{c0^t}\\
&\quad=(p-a-1)x^s\bigl(T_{c0^t(\mathtt{q}-1)}-(p-1)T_{c0^t}\bigr)\\
&\quad=(p-a-1)x^s\Bigl(\bigl((p-1)T_{c0^t}+x^{t+1}T_{(c-1)\mathtt{q}^t}\bigr)-(p-1)T_{c0^t}\Bigr)\\
&\quad=(p-a-1)p^tcx^{s+t+1},
\end{aligned}
$$

therefore

$$
\overline{T}_w\overline{T}_{w_{LR}}-\overline{T}_{w_L}\overline{T}_{w_R}=p^{t-s}x^{s+t+1}\frac{c}{c+1}\frac{p-a-1}{a+1}.
$$

Equation (3.1) therefore holds for the base cases. Assume that we have already established the statement for all $w\in W$ having right depth $\leq d-1$, where $d\geq 1$, and assume that $\widetilde{w}\in W$ has right depth equal to d. Then $\widetilde{w}$ is of (exactly) one of the following forms, which we have to treat one by one.

$$
\begin{aligned}
&(3.5)\qquad wb\mathtt{0}, && w\mathtt{0}\in W, b\in\{\mathtt{1},\ldots,\mathtt{q}-\mathtt{1}\};\\
&(3.6)\qquad wb\mathtt{0}^t, && w\mathtt{0}\in W, b\in\{\mathtt{1},\ldots,\mathtt{q}\}, t\geq 2;\\
&(3.7)\qquad w\mathtt{q}^sa, && w\in W, s\geq 1, a\in\{\mathtt{0},\ldots,\mathtt{q}-\mathtt{1}\};\\
&(3.8)\qquad wa, && w\in W, a\in\{\mathtt{1},\ldots,\mathtt{q}-\mathtt{1}\}.
\end{aligned}
$$

We will use the following auxiliary formulas. If $wb \in W$, where $b \neq \mathtt{0}$ and $(wb)_L \neq \varepsilon$, then

$$b(b+1)\Big(\overline{T}_{(wb)-\mathtt{1}}\overline{T}_{(wb)_L} - \overline{T}_{(wb)_L-\mathtt{1}}\overline{T}_{wb}\Big) = \frac{p}{p-1}\Big(\overline{T}_{w\mathtt{0}}\overline{T}_{(w\mathtt{0})_{LR}} - \overline{T}_{(w\mathtt{0})_L}\overline{T}_{(w\mathtt{0})_R}\Big). \tag{3.9}$$

If moreover $w = w_{\nu-1}\cdots w_r\mathtt{0}^r \in W$, where $r \geq 0$ is maximal, and $w_L \neq \varepsilon$ is satisfied, we have

$$x^{r+1}\Big(T_{w-\mathtt{1}}T_{w_L} - T_{w_L-\mathtt{1}}T_w\Big) = \frac{1}{p-1}\Big(T_{w\mathtt{0}}T_{(w\mathtt{0})_{LR}} - T_{(w\mathtt{0})_L}T_{(w\mathtt{0})_R}\Big). \tag{3.10}$$

Let us now prove these formulas. We handle the case $w_L = \varepsilon$ separately. Since $(wb)_L \neq \varepsilon$ by assumption, there exist $d \in \{\mathtt{1},\ldots,\mathtt{q}\}$, $c \in \{\mathtt{1},\ldots,\mathtt{q}-\mathtt{1}\}$ and $t \geq 0$ such that $w = d\mathtt{0}^t c$. We obtain by (3.4) and (3.3)

$$\begin{aligned}
&T_{(wb)-\mathtt{1}}T_{(wb)_L} - T_{(wb)_L-\mathtt{1}}T_{wb} = T_{d\mathtt{0}^tc(b-\mathtt{1})}T_{cb} - T_{c(b-\mathtt{1})}T_{d\mathtt{0}^tcb}\\
&\quad= \Big(bT_{d\mathtt{0}^tc} + (p-b)xT_{d\mathtt{0}^t(c-\mathtt{1})}\Big)\big((c+1)(b+1) + c(p-b-1)x\big)\\
&\quad- \big((c+1)b + c(p-b)x\big)\Big((b+1)T_{d\mathtt{0}^tc} + (p-b-1)xT_{d\mathtt{0}^t(c-\mathtt{1})}\Big)\\
&\quad= px\Big((c+1)T_{d\mathtt{0}^t(c-\mathtt{1})} - cT_{d\mathtt{0}^tc}\Big)\\
&\quad= px\Big((c+1)\Big(\frac{1}{p}\Big((p-c)(px)^{t+1} + c(p-1)A_2(t)\Big)T_{d-\mathtt{1}} + cT_d\Big)\\
&\quad- c\Big(\frac{1}{p}\Big((p-c-1)(px)^{t+1} + (c+1)(p-1)A_2(t)\Big)T_{d-\mathtt{1}} + (c+1)T_d\Big)\Big)\\
&\quad= p^{t+2}x^{t+2}d,
\end{aligned}$$

moreover

$$\begin{aligned}
&(c+1)T_{w\mathtt{0}}T_{(w\mathtt{0})_{LR}} - T_{(w\mathtt{0})_L}T_{(w\mathtt{0})_R} = (c+1)T_{d\mathtt{0}^tc\mathtt{0}} - T_{c\mathtt{0}}T_{d\mathtt{0}^tc}\\
&\quad= (c+1)\big(T_{d\mathtt{0}^tc} + (p-1)xT_{d\mathtt{0}^t(c-\mathtt{1})}\big) - \big((c+1) + c(p-1)x\big)T_{d\mathtt{0}^tc}\\
&\quad= (p-1)x\big((c+1)T_{d\mathtt{0}^t(c-\mathtt{1})} - cT_{d\mathtt{0}^tc}\big)\\
&\quad= (p-1)x\Bigg((c+1)\Big(\frac{1}{p}\Big((p-c)(px)^{t+1} + c(p-1)A_2(t)\Big)d + c(d+1)\Big)\\
&\quad- c\Big(\frac{1}{p}\Big((p-c-1)(px)^{t+1} + (c+1)(p-1)A_2(t)\Big)d + (c+1)(d+1)\Big)\Bigg)\\
&\quad= (p-1)p^{t+1}x^{t+2}d.
\end{aligned}$$

Passing from T to $\overline{T}$, we obtain the statement (3.9) for the case $w_L = \varepsilon$, $(wb)_L \neq \varepsilon$. If $w_L \neq \varepsilon$, we have $(wb)_L = w_Lb$, moreover r is also the number of zeros at the low digits of w_L. Therefore

$$T_{(wb)-\mathtt{1}}T_{(wb)_L} - T_{(wb)_L-\mathtt{1}}T_{wb}$$

$$\begin{aligned}
&= \bigl(bT_w + (p-b)x^{r+1}T_{w-\mathtt{1}}\bigr)\bigl((b+1)T_{w_L} + (p-b-1)x^{r+1}T_{w_L-\mathtt{1}}\bigr)\\
&- \bigl(bT_{w_L} + (p-b)x^{r+1}T_{w_L-\mathtt{1}}\bigr)\bigl((b+1)T_w + (p-b-1)x^{r+1}T_{w-\mathtt{1}}\bigr)\\
&= px^{r+1}\bigl(T_{w-\mathtt{1}}T_{w_L} - T_{w_L-\mathtt{1}}T_w\bigr),
\end{aligned}$$

and (3.9) and (3.10) follow easily using the instance $T_{w\mathtt{0}} = T_w + (p-1)x^{r+1}T_{w-\mathtt{1}}$ of the recurrence (2.2). We have to treat the cases (3.5)–(3.8). Assume that $\widetilde{w} = wb\mathtt{0}$, where $w\mathtt{0} \in W$ and $b \in \{\mathtt{1},\ldots,\mathtt{q}-\mathtt{1}\}$. Since $(wb)_L = \widetilde{w}_{RL} \neq \varepsilon$ (this holds since the right depth of $\widetilde{w}$ is not zero), we have $\widetilde{w}_L = (wb)_L\mathtt{0}$ and therefore

$$\begin{aligned}
T_{\widetilde{w}}T_{\widetilde{w}_{LR}} - T_{\widetilde{w}_L}T_{\widetilde{w}_R} &= T_{wb\mathtt{0}}T_{(wb)_L} - T_{(wb)_L\mathtt{0}}T_{wb}\\
&= \bigl(T_{wb} + (p-1)xT_{(wb)-\mathtt{1}}\bigr)T_{(wb)_L} - \bigl(T_{(wb)_L} + (p-1)xT_{(wb)_L-\mathtt{1}}\bigr)T_{wb}\\
&= (p-1)x\bigl(T_{(wb)-\mathtt{1}}T_{(wb)_L} - T_{(wb)_L-\mathtt{1}}T_{wb}\bigr).
\end{aligned}$$

By (3.9) we have

$$\begin{aligned}
\overline{T}_{\widetilde{w}}\overline{T}_{\widetilde{w}_{LR}} - \overline{T}_{\widetilde{w}_L}\overline{T}_{\widetilde{w}_R} &= (p-1)x\frac{b}{b+1}\bigl(\overline{T}_{(wb)-\mathtt{1}}\overline{T}_{(wb)_L} - \overline{T}_{(wb)_L-\mathtt{1}}\overline{T}_{wb}\bigr)\\
&= \frac{px}{(b+1)^2}\bigl(\overline{T}_{w\mathtt{0}}\overline{T}_{(w\mathtt{0})_{LR}} - \overline{T}_{(w\mathtt{0})_L}\overline{T}_{(w\mathtt{0})_R}\bigr).
\end{aligned}$$

Since the right depth of $w\mathtt{0}$ is smaller than d, we can apply the induction hypothesis and the case (3.5) is finished. Now we assume that $\widetilde{w} = wb\mathtt{0}^t$, where $w\mathtt{0} \in W$, $b \in \{\mathtt{1},\ldots,\mathtt{q}\}$, and $t \geq 2$. We first note that for a finite word $v \in \{\mathtt{0},\ldots,\mathtt{q}\}^*$ we have the identity $T_{vb\mathtt{0}^t} = T_{vb\mathtt{0}^{t-1}} + (p-1)x^tT_{vb\mathtt{0}^{t-1}-\mathtt{1}} = T_{vb\mathtt{0}^{t-1}} + (p-1)x^tp^{t-1}T_{v(b-\mathtt{1})}$, analogously for $t-1$ instead of t, therefore

$$T_{vb\mathtt{0}^t} = (1+px)T_{vb\mathtt{0}^{t-1}} - pxT_{vb\mathtt{0}^{t-2}}.$$

Moreover, we have $\widetilde{w}_L = (wb\mathtt{0})_L\mathtt{0}^{t-1} = w'b\mathtt{0}^t$ for some $w' \in \{\mathtt{0},\ldots,\mathtt{q}\}^*$. We may therefore calculate:

$$\begin{aligned}
T_{\widetilde{w}}T_{\widetilde{w}_{LR}} - T_{\widetilde{w}_L}T_{\widetilde{w}_R} &= \Bigl((1+px)T_{wb\mathtt{0}^{t-1}} - pxT_{wb\mathtt{0}^{t-2}}\Bigr)T_{w'b\mathtt{0}^{t-1}}\\
&- \Bigl((1+px)T_{w'b\mathtt{0}^{t-1}} - pxT_{w'b\mathtt{0}^{t-2}}\Bigr)T_{wb\mathtt{0}^{t-1}}\\
&= px\bigl(T_{wb\mathtt{0}^{t-1}}T_{w'b\mathtt{0}^{t-2}} - T_{(wb\mathtt{0}^{t-1})_L}T_{(wb\mathtt{0}^{t-1})_R}\bigr).
\end{aligned}$$

If $t > 2$ or $(wb)_L \neq \varepsilon$, we have $w'b\mathtt{0}^{t-2} = (wb\mathtt{0}^{t-1})_{LR}$, therefore

$$\overline{T}_{\widetilde{w}}\overline{T}_{\widetilde{w}_{LR}} - \overline{T}_{\widetilde{w}_L}\overline{T}_{\widetilde{w}_R} = px\bigl(\overline{T}_{wb\mathtt{0}^{t-1}}\overline{T}_{(wb\mathtt{0}^{t-1})_{LR}} - \overline{T}_{(wb\mathtt{0}^{t-1})_L}\overline{T}_{(wb\mathtt{0}^{t-1})_R}\bigr)$$

and we can use the induction hypothesis. Otherwise, we have $w = d\mathtt{0}^r$ for some $d \in \{\mathtt{1},\ldots,\mathtt{q}\}$ and $r \geq 0$, and we obtain

$$\begin{aligned}
\overline{T}_{\widetilde{w}}\overline{T}_{\widetilde{w}_{LR}} - \overline{T}_{\widetilde{w}_L}\overline{T}_{\widetilde{w}_R} &= \frac{1}{(d+1)(b+1)^2}\bigl(T_{\widetilde{w}}T_{\widetilde{w}_{LR}} - T_{\widetilde{w}_L}T_{\widetilde{w}_R}\bigr)\\
&= \frac{px}{(d+1)(b+1)^2}\bigl(T_{wb\mathtt{0}}T_b - T_{(wb\mathtt{0})_L}T_{(wb\mathtt{0})_R}\bigr)
\end{aligned}$$

$$\begin{aligned}&= px\big(\overline{T}_{wb\mathtt{0}}\overline{T}_{\varepsilon} - \overline{T}_{(wb\mathtt{0})_L}\overline{T}_{(wb\mathtt{0})_R}\big)\\ &= px\big(\overline{T}_{wb\mathtt{0}}\overline{T}_{(wb\mathtt{0})_{LR}} - \overline{T}_{(wb\mathtt{0})_L}\overline{T}_{(wb\mathtt{0})_R}\big),\end{aligned}$$

so that we can apply the hypothesis also in this case. Assume that $\widetilde{w} = w\mathtt{q}^s a$, where $w = w_{\nu-1}\cdots w_r\mathtt{0}^r \in W$ and $r \geq 0$ is maximal, $s \geq 1$, and $a \in \{\mathtt{0},\ldots,\mathtt{q}-\mathtt{1}\}$. The right depth of $\widetilde{w}$ is at least one. Therefore $w_L \neq \varepsilon$, and we obtain, using (3.2) and (3.10),

$$\begin{aligned}&T_{\widetilde{w}}T_{\widetilde{w}_{LR}} - T_{\widetilde{w}_L}T_{\widetilde{w}_R}\\ &= \left(p^s\Big((a+1)+(p-a-1)(p-1)A_1(s)\Big)T_w + (p-a-1)x^sT_{w(\mathtt{q}-\mathtt{1})}\right)T_{w_L}\\ &- \left(p^s\Big((a+1)+(p-a-1)(p-1)A_1(s)\Big)T_{w_L} + (p-a-1)x^sT_{w_L(\mathtt{q}-\mathtt{1})}\right)T_w\\ &= (p-a-1)x^s\big(T_{w(\mathtt{q}-\mathtt{1})}T_{w_L} - T_{w_L(\mathtt{q}-\mathtt{1})}T_w\big)\\ &= (p-a-1)x^s\Big(\big((p-1)T_w + x^{r+1}T_{w-1}\big)T_{w_L}\\ &\quad - \big((p-1)T_{w_L} + x^{r+1}T_{w_L-1}\big)T_w\Big)\\ &= (p-a-1)x^{s+r+1}\big(T_{w-\mathtt{1}}T_{w_L} - T_{w_L-\mathtt{1}}T_w\big)\\ &= (p-a-1)\frac{1}{p-1}x^r\big(T_{w\mathtt{0}}T_{(w\mathtt{0})_{LR}} - T_{(w\mathtt{0})_L}T_{(w\mathtt{0})_R}\big),\end{aligned}$$

therefore

$$\overline{T}_{\widetilde{w}}\overline{T}_{\widetilde{w}_{LR}} - \overline{T}_{\widetilde{w}_L}\overline{T}_{\widetilde{w}_R} = \frac{p-a-1}{a+1}\frac{1}{p-1}p^{-r}x^r\big(\overline{T}_{w\mathtt{0}}\overline{T}_{(w\mathtt{0})_{LR}} - \overline{T}_{(w\mathtt{0})_L}\overline{T}_{(w\mathtt{0})_R}\big).$$

Now one of the two cases (3.5) or (3.6) is applicable. It remains to handle the fourth case. Assume that $\widetilde{w} = wa$, where $w = w_{\nu-1}\cdots w_r\mathtt{0}^r \in W$ and $r \geq 0$ is maximal, and $a \in \{\mathtt{1},\ldots,\mathtt{q}-\mathtt{1}\}$. As in the last case, we have $w_L \neq \varepsilon$, therefore we can use (3.10) and obtain

$$\begin{aligned}T_{\widetilde{w}}T_{\widetilde{w}_{LR}} - T_{\widetilde{w}_L}T_{\widetilde{w}_R} &= \big((a+1)T_w + (p-a-1)x^{r+1}T_{w-\mathtt{1}}\big)T_{w_L}\\ &- \big((a+1)T_{w_L} + (p-a-1)x^{r+1}T_{w_L-\mathtt{1}}\big)T_w\\ &= (p-a-1)x^{r+1}\big(T_{w-\mathtt{1}}T_{w_L} - T_{w_L-\mathtt{1}}T_w\big)\\ &= \frac{p-a-1}{p-1}\big(T_{w\mathtt{0}}T_{(w\mathtt{0})_{LR}} - T_{(w\mathtt{0})_L}T_{(w\mathtt{0})_R}\big),\end{aligned}$$

therefore

$$\overline{T}_{\widetilde{w}}\overline{T}_{\widetilde{w}_{LR}} - \overline{T}_{\widetilde{w}_L}\overline{T}_{\widetilde{w}_R} = \frac{p-a-1}{a+1}\frac{1}{p-1}\big(T_{w\mathtt{0}}T_{(w\mathtt{0})_{LR}} - T_{(w\mathtt{0})_L}T_{(w\mathtt{0})_R}\big).$$

As in the previous case, this expression can be treated with one of the cases (3.5) or (3.6). The proof is complete. □

References

[1] J.-P. Allouche and V. Berthé, *Triangle de Pascal, complexité et automates*, Bull. Belg. Math. Soc. Simon Stevin, 4 (1997), pp. 1–23. Journées Montoises (Mons, 1994).

[2] J.-P. Allouche and J. Shallit, *The ring of k-regular sequences*, Theoret. Comput. Sci., 98 (1992), pp. 163–197.

[3] ——, *Automatic sequences*, Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.

[4] T. Amdeberhan and R. P. Stanley, *Polynomial Coefficient Enumeration*, (2008). Preprint. arXiv:0811.3652.

[5] G. Barat and P. J. Grabner, *Distribution of binomial coefficients and digital functions*, J. London Math. Soc. (2), 64 (2001), pp. 523–547.

[6] D. Barbolosi and P. J. Grabner, *Distribution des coefficients multinomiaux et q-binomiaux modulo p*, Indag. Math. (N.S.), 7 (1996), pp. 129–135.

[7] L. Carlitz, *The number of binomial coefficients divisible by a fixed power of a prime*, Rend. Circ. Mat. Palermo (2), 16 (1967), pp. 299–320.

[8] E. Cateland, *Digital sequences and k-regular sequences*, thesis, Université Sciences et Technologies - Bordeaux I, June 1992.

[9] K. S. Davis and W. A. Webb, *Lucas' theorem for prime powers*, European J. Combin., 11 (1990), pp. 229–233.

[10] L. E. Dickson, *History of the theory of numbers. Vol. I: Divisibility and primality.*, 1919. Chapter IX: "Divisibility of factorials and multinomial coefficients".

[11] M. Drmota, M. Kauers, and L. Spiegelhofer, *On a Conjecture of Cusick Concerning the Sum of Digits of n and $n+t$*, SIAM J. Discrete Math., 30 (2016), pp. 621–649. arXiv:1509.08623.

[12] J. W. Eaton, D. Bateman, and S. Hauberg, *GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations*, CreateSpace Independent Publishing Platform, 2009. ISBN 1441413006.

[13] J. Emme and P. Hubert, *Central Limit Theorem for probability measures defined by sum-of-digits function in base 2*, (2016). Preprint. arXiv:1605.06297.

[14] J. Emme and A. Prikhodko, *On the asymptotic behaviour of the correlation measure of sum-of-digits function in base* 2, (2015). Preprint. arXiv:1504.01701.

[15] N. J. Fine, *Binomial coefficients modulo a prime*, Amer. Math. Monthly, 54 (1947), pp. 589–592.

[16] P. Flajolet, P. Grabner, P. Kirschenhofer, H. Prodinger, and R. F. Tichy, *Mellin transforms and asymptotics: digital sums*, Theoret. Comput. Sci., 123 (1994), pp. 291–314.

[17] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.

[18] R. D. Fray, *Congruence properties of ordinary and q-binomial coefficients*, Duke Math. J., 34 (1967), pp. 467–480.

[19] J. Glaisher, *On the residue of a binomial-theorem coefficient with respect to a prime modulus*, Quarterly Journal of Pure and Applied Mathematics, 30 (1899), pp. 150–156.

[20] A. Granville, *Zaphod Beeblebrox's brain and the fifty-ninth row of Pascal's triangle*, Amer. Math. Monthly, 99 (1992), pp. 318–331.

[21] ——, *Arithmetic properties of binomial coefficients. I. Binomial coefficients modulo prime powers*, vol. 20 of CMS Conf. Proc., Amer. Math. Soc., Providence, RI, 1997.

[22] F. T. Howard, *A combinatorial problem and congruences for the Rayleigh function*, Proc. Amer. Math. Soc., 26 (1970), pp. 574–578.

[23] ——, *The number of binomial coefficients divisible by a fixed power of* 2, Proc. Amer. Math. Soc., 29 (1971), pp. 236–242.

[24] ——, *Formulas for the number of binomial coefficients divisible by a fixed power of a prime*, Proc. Amer. Math. Soc., 37 (1973), pp. 358–362.

[25] J. G. Huard, B. K. Spearman, and K. S. Williams, *On Pascal's triangle modulo p^2*, Colloq. Math., 74 (1997), pp. 157–165.
[26] G. S. Kazandzidis, *Congruences on the binomial coefficients*, Bull. Soc. Math. Grèce (N.S.), 9 (1968), pp. 1–12.
[27] D. E. Knuth and H. S. Wilf, *The power of a prime that divides a generalized binomial coefficient*, J. Reine Angew. Math., 396 (1989), pp. 212–219.
[28] E. E. Kummer, *Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen*, J. Reine Angew. Math., 44 (1852), pp. 93–146.
[29] E. Lucas, *Sur les congruences des nombres eulériens et les coefficients différentiels des functions trigonométriques suivant un module premier*, Bull. Soc. Math. France, 6 (1878), pp. 49–54.
[30] O. Perron, *Über das infinitäre Verhalten der Koeffizienten einer gewissen Potenzreihe.*, Arch. der Math. u. Phys. (3), 22 (1914), pp. 329–340.
[31] E. Rowland, *The number of nonzero binomial coefficients modulo p^α*, J. Comb. Number Theory, 3 (2011), pp. 15–25.
[32] D. Singmaster, *Notes on binomial coefficients. I. A generalization of Lucas' congruence*, J. London Math. Soc. (2), 8 (1974), pp. 545–548.
[33] ——, *Notes on binomial coefficients. III. Any integer divides almost all binomial coefficients*, J. London Math. Soc. (2), 8 (1974), pp. 555–560.
[34] ——, *Divisibility of binomial and multinomial coefficients by primes and prime powers*, Fibonacci Assoc., Santa Clara, Calif., 1980.
[35] B. K. Spearman and K. S. Williams, *On a formula of Howard*, Bull. Hong Kong Math. Soc., 2 (1999), pp. 325–340. (Available on Spearman's website).
[36] K. B. Stolarsky, *Power and exponential sums of digital sums related to binomial coefficient parity*, SIAM J. Appl. Math., 32 (1977), pp. 717–730.
[37] The Sage Developers, *SageMath, the Sage Mathematics Software System (Version 6.9)*, 2015. `http://www.sagemath.org`.
[38] W. A. Webb, *The number of binomial coefficients in residue classes modulo p and p^2*, Colloq. Math., 60/61 (1990), pp. 275–280.
[39] E. M. Wright, *The Coefficients of a Certain Power Series*, J. London Math. Soc., S1-7 (1933), p. 256.

Institut Élie Cartan de Lorraine, Université de Lorraine, Vandœuvre-lès-Nancy, France
E-mail address: lukas.spiegelhofer@tuwien.ac.at

Institute of Discrete Mathematics and Geometry, Vienna University of Technology, Vienna, Austria
E-mail address: michael.wallner@tuwien.ac.at