Counting words by number of occurrences of some patterns

Zhicheng Gao* Andrew MacFie† Daniel Panario‡

School of Mathematics and Statistics
Carleton University, Ottawa, Canada

Submitted: Dec 1, 2010; Accepted: Jun 28, 2011; Published: Jul 15, 2011
Mathematics Subject Classification: 05A05

Abstract

We give asymptotic expressions for the number of words containing a given number of occurrences of a pattern for two families of patterns with two parameters each. One is the family of classical patterns in the form $22\cdots212\cdots22$ and the other is a family of partially ordered patterns. The asymptotic expressions are in terms of the number of solutions to an equation, and for one subfamily this quantity is the number of integer partitions into qth order binomial coefficients.

Keywords: classical pattern, occurrence, asymptotics, word, partially ordered pattern

This paper is dedicated to the memory of Philippe Flajolet (1948–2011).

1 Introduction

Let $[k] = \{1, 2, ..., k\}$ be a totally ordered alphabet on k letters. A *k-ary word of length n* is an element of $[k]^n$. Given a word $w = w_1 \cdots w_n \in [k]^n$, the *reduction* of w, denoted $\eta(w)$, is the word obtained by replacing the ith smallest letters of w with i's, for all i. For example, $\eta(46632) = 34421$.

Given words σ of length n and $\tau = \eta(\tau)$ of length l (τ is called the *pattern*), an *occurrence* of τ in σ is a sequence of indices $1 \le i_1 < \cdots < i_l \le n$ and the corresponding letter subsequence $\sigma_{i_1} \cdots \sigma_{i_l}$ that satisfy certain conditions related to τ; *classical patterns*

*zgao@math.carleton.ca
†amacfie@connect.carleton.ca
‡daniel@math.carleton.ca

require $\eta(\sigma_{i_1}\cdots\sigma_{i_l}) = \tau$, and *subword patterns* are classical patterns that require $i_1 + l - 1 = i_l$. If there are no occurrences of τ in σ, then σ is said to *avoid* the pattern τ.

As in [12], a partially ordered pattern (POP) is one in which not all letters are comparable. The letters in a POP are from a partially ordered alphabet; letters shown with the same number of primes are comparable to each other (e.g. $1''$ and $2''$), while letters shown without primes are comparable to all letters of the alphabet. An occurrence of a classical POP in a word σ is a distinguished subsequence of entries of σ such that the relative order of two entries in the subsequence need be the same as that of the corresponding letters in the pattern only if the corresponding letters in the pattern are comparable; e.g. the classical POP $1'1''2$ is found in the word 42213 three times as $4\underline{22}1\underline{3}$, $4\underline{2}2\underline{13}$ and $42\underline{213}$ (the subsequences of length three in which the third letter is larger than the first two). In this paper, the patterns are all classical POPs, although not all have noncomparable letters.

Let $\mathcal{W}_n^{[k]}(\tau;r) = \{\sigma \in [k]^n \mid \sigma \text{ contains exactly } r \text{ occurrences of } \tau\}$. For a classical POP τ of length l and greatest entry m, it is easy to see that there are maps φ such that $|\mathcal{W}_n^{[k]}(\varphi(\tau);r)| = |\mathcal{W}_n^{[k]}(\tau;r)|$ for all n, k, r. One is left-right reversal, $r : \tau_i \mapsto \tau_{l+1-i}$ (primes move with the entries), and another is complement, or "vertical reflection", $c : \tau_i \mapsto m + 1 - \tau_i$ (primes do not move). For example, if $\tau = 1'1''2$, then $r(\tau) = 21''1'$ and $c(\tau) = 2'2''1$. We get one more equivalent pattern by composing complement and reversal ($c \circ r = r \circ c$). The patterns studied in this paper are therefore representatives of equivalence classes of patterns for which the same results hold.

The number of k-ary words of length n avoiding a given classical pattern has been studied for a number of different patterns [2, 3, 8, 11, 14, 15, 16]. Specifically, exact results for the avoidance of a number of classical patterns with at most 2 distinct letters were found in [5]. As far as we know nothing has been studied for the general case of counting words with r occurrences of a classical pattern. For subword patterns some results are known for r occurrences [6, 7]. For POP-based enumeration for words and other objects, see [4, 10, 12, 13].

Notation 1. *If j is a letter, we use j^p to represent*

$$\underbrace{jj\cdots j}_{p \text{ copies of } j}.$$

For example, $2^3 1 3^2 = 222133$.

In [9], Flajolet et al. studied in detail some properties of the random variable $X(\tau')$, the number of occurrences of a *hidden word* τ' in a random k-ary word of length n, including the mean and variance of its distribution. A hidden word is simply a word that must be found as an exact subsequence of another word, e.g. 132 is found in 1432 once as $\underline{1}4\underline{32}$. For a classical pattern τ, and fixed k, let $T(\tau) = \{w \in [k]^{|\tau|} \mid \eta(w) = \tau\}$. In a random word $\sigma \in [k]^n$, if $Y(\tau)$ is the number of occurrences of τ as a classical pattern and $X(\tau')$ is the number of occurrences of τ' as a hidden word, then

$$Y(\tau) = \sum_{\tau' \in T(\tau)} X(\tau'). \tag{1}$$

We note that the distribution of $X(\tau')$ is the same for all $\tau' \in T(\tau)$. Thus (1) implies $\mathbf{E}Y(\tau) = |T(\tau)|\,\mathbf{E}X(\tau)$. However, the random variables $\{X(j^p) \mid 1 \le j \le k\}$ are not asymptotically independent since, for example

$$\begin{aligned} P\left[Y\left(1^p\right) = \binom{i}{p}, Y\left(2^p\right) = \binom{j}{p}\right] &= \binom{n}{i}\binom{n-i}{j}\left(1/k\right)^i\left(1/k\right)^j\left(1-2/k\right)^{n-i-j} \\ &\not\approx \binom{n}{i}\left(1/k\right)^i\left(1-1/k\right)^{n-i}\binom{n}{j}\left(1/k\right)^j\left(1-1/k\right)^{n-j} \\ &= P\left[Y\left(1^p\right) = \binom{i}{p}\right] \cdot P\left[Y\left(2^p\right) = \binom{j}{p}\right], \end{aligned}$$

which means that known results for hidden words are not directly transformed into results for classical POPs.

The structure of the paper is as follows. In Section 2 we find a recursion for $|\mathcal{W}_n^{[k]}(\tau; r)|$ where $\tau = 1'1'' \cdots 1^{(p)}2^q$ and obtain an asymptotic expression. In Section 2.1 we simplify the asymptotic expression for the case of $\tau = 12^q$ and establish a connection to integer partitions. In Section 3 we derive a recursion for $|\mathcal{W}_n^{[k]}(\tau; r)|$ where $\tau = 2^p12^q$ and also obtain an asymptotic expression. We conclude in Section 4, mentioning possible extensions to this work.

2 The pattern $1'1'' \cdots 1^{(p)}2^q$

For $p, q \ge 1$, we let $1'2_{p,q}$ represent the partially ordered pattern $1'1'' \cdots 1^{(p)}2^q$, where $1^{(p)}$ means 1 with p primes. An occurrence of $1'2_{p,q}$ is formed by a subsequence $\phi = (\phi_1, \ldots, \phi_p, \phi_{p+1}, \ldots, \phi_{p+q})$ where

$$\phi_i < \phi_{p+1} = \phi_{p+2} = \cdots = \phi_{p+q}, \quad 1 \le i \le p.$$

Let $f_r(n,k) = |\mathcal{W}_n^{[k]}(1'2_{p,q}; r)|$, and let $F_{r,k}(x) = \sum_{n \ge 0} f_r(n,k)x^n$.

Notation 2. *We use*

$$[x]_n = x(x-1)\cdots(x-(n-1))$$

to denote the nth falling factorial of x, and

$$[x]^n = x(x+1)\cdots(x+(n-1))$$

to denote the nth rising factorial of x.

Notation 3. *For a proposition S, the notation $[S]$ stands for 1 if S is true, 0 otherwise.*

Notation 4. *We say that*

$$f(x) = O\left((1-x)^{-a}\right),$$

where $a > 0$, if

$$f(x) = \frac{p(x)}{(1-x)^a}$$

for some polynomial $p(x)$.

Theorem 1. *For $k \geq 1$, $F_{r,k}(x)$ is a rational function of the form*

$$\frac{\mathfrak{p}_{r,k}(x)}{(1-x)^{\alpha_{r,k}}},$$

where $\mathfrak{p}_{r,k}(x)$ is either 0 or a polynomial such that $\mathfrak{p}_{r,k}(1) \neq 0$, and $\alpha_{r,k} > 0$.

Proof. We begin by deriving a recursion for $f_r(n,k)$. We comment that this extends the work in [5], that deals with avoidance of classical patterns with at most two distinct letters. For the initial values we have, for $0 \leq n < p+q$,

$$f_r(n,k) = [r{=}0]\, k^n.$$

For the general case $n \geq p+q$, we recursively count $\sigma \in \mathcal{W}_n^{[k]}(1'2_{p,q}; r)$ by first counting σ such that at least one of the first p letters is k. By the principle of inclusion-exclusion, the number of such σ is

$$\sum_{m=1}^{p} N_m (-1)^{m+1},$$

where N_m is the sum, over all m-subsets of the first p positions, of the number of words σ with k's in the positions given by the subset. The quantity N_m is given by

$$N_m = \binom{p}{m} f_r(n-m,k),$$

since inserting m copies of k into any of the first p positions of words from the set $\mathcal{W}_{n-m}^{[k]}(1'2_{p,q}; r)$ is reversible and does not affect the number of occurrences of $1'2_{p,q}$.

Now we count the σ's that have no k's in their first p positions. Let b be the number of k's in σ. If $b \leq q-1$, then there are not enough k's to be part of a pattern, so there are

$$\sum_{b=0}^{q-1} \binom{n-p}{b} f_r(n-b, k-1),$$

words of this kind. But if $b \geq q$ then there will be at least one occurrence of the pattern, and we count in the following manner: We use the position vector $\boldsymbol{a}_b = (a_1, \ldots, a_{b-(q-1)})$ to denote the positions in σ of the 1st through $(b-(q-1))$th copies of k (the positions of the last $q-1$ copies of k do not affect the number of occurrences of the pattern), and we let $A = a_{b-(q-1)}$. The number of occurrences of $1'2_{p,q}$ that the k's of σ are part of is seen to be

$$\bar{a} = \sum_{i=1}^{b-(q-1)} \binom{b-i}{q-1} \binom{a_i - i}{p}. \tag{2}$$

Once $\boldsymbol{a}_b$ is known, the number of ways of placing the remaining $q-1$ copies of k is $\binom{n-A}{q-1}$. Thus we have, for $n \geq p+q, k \geq 1$,

$$f_r(n,k) = \sum_{m=1}^{p} \binom{p}{m} f_r(n-m,k)(-1)^{m+1} + \sum_{b=0}^{q-1} \binom{n-p}{b} f_r(n-b,k-1)$$
$$+ \sum_{q \le b \le n-p} \sum_{\substack{\boldsymbol{a}_b \\ A \le n-q+1 \\ 1 \le \bar{a} \le r}} \binom{n-A}{q-1} f_{r-\bar{a}}(n-b,k-1), \tag{3}$$

where $\bar{a}$ depends on $\boldsymbol{a}_b$ and is given in (2).

After multiplying (3) by x^n and summing, we have

$$F_{r,k}(x) = \frac{1}{(1-x)^p} \left(\sum_{b=0}^{q-1} \sum_{i=0}^{b} \lambda_{b,i} x^{i+b} \frac{d^i}{dx^i} F_{r,k-1}(x) \right.$$
$$\left. + \sum_{b \ge q} \sum_{\substack{\boldsymbol{a}_b \\ 1 \le \bar{a} \le r}} \sum_{i=0}^{q-1} \lambda_{b,i,A} x^{i+b} \frac{d^i}{dx^i} F_{r-\bar{a},k-1}(x) + P(x) \right), \tag{4}$$

for rational λ's, where $P(x)$ is the polynomial

$$P(x) = \sum_{m=0}^{p} \binom{p}{m} (-1)^m \sum_{n=0}^{p+q-m-1} f_r(n,k) x^{n+m} - \sum_{b=0}^{q-1} \sum_{n=0}^{p+q-b-1} \binom{n+b-p}{b} f_r(n,k-1) x^{n+b}$$
$$- \sum_{b \ge q} \sum_{\substack{\boldsymbol{a}_b \\ 1 \le \bar{a} \le r}} \sum_{n=0}^{A+q-1-b} \binom{n+b-A}{q-1} f_{r-\bar{a}}(n,k-1) x^{n+b}.$$

We observe that $F_{r,k}(x)$ can never be a nonzero polynomial. Indeed, from its combinatorial definition we have that

$$\sigma \in \mathcal{W}_n^{[k]}(1'2_{p,q};r) \text{ implies } k\sigma \in \mathcal{W}_{n+1}^{[k]}(1'2_{p,q};r).$$

Hence, the theorem can now be proved by induction on k, starting with $k=1$ from the initial values:

$$F_{r,0}(x) = [r{=}0], \qquad F_{r,1}(x) = [r{=}0] \frac{1}{1-x}. \qquad \square$$

For $j \ge 1$, we let c_j be the number of solutions $\boldsymbol{a}_b = (a_1, a_2, \ldots, a_{b-(q-1)})$, $1 \le a_1 < \cdots < a_{b-(q-1)}$, to

$$j = \sum_{i=1}^{b-(q-1)} \binom{b-i}{q-1} \binom{a_i - i}{p},$$

for any b. We take $c_0 = 1$, and we let $C(x) = \sum_{j \ge 0} c_j x^j$.

Corollary 1. *The function $F_{r,k}(x)$ has the following asymptotic form:*

$$F_{r,k}(x) = D_{r,k}\,(1-x)^{-\alpha_k} + O\left((1-x)^{-\alpha_k+1}\right), \quad k \geq 1$$

where

$$\alpha_k = (k-1)(q+p-1)+1, \quad D_k(x) = \sum_{r\geq 0} D_{r,k}x^r = C^{k-1}(x)\prod_{i=1}^{k-1}\binom{\alpha_i+q-2}{q-1}.$$

Proof. We proceed by induction on k. For the base case

$$F_{r,1}(x) = [r{=}0]\,\frac{1}{1-x},$$

we have $\alpha_k = 1 = (1-1)(q+p-1)+1$, and $D_1(x) = 1 = C^{1-1}(x)\prod_{i=1}^{1-1}\binom{\alpha_i+q-2}{q-1}$.

For the inductive step, we assume Corollary 1 holds for all $k, 1 \leq k < K$. Theorem 1 allows us to turn (4) (where $\lambda_{q-1,q-1} = \lambda_{b,q-1,A} = \frac{1}{(q-1)!}$) into the following asymptotic relation:

$$F_{r,K}(x) = \frac{1}{(1-x)^p}\left(\sum_{0\leq j\leq r} c_j \frac{1}{(q-1)!}\frac{d^{q-1}}{dx^{q-1}}F_{r-j,K-1}(x) + P(x)\right) + O\left((1-x)^{-\alpha_K+1}\right). \tag{5}$$

By the inductive hypothesis, the terms in the sum on j dominate $P(x)$ unless they are 0. However, we show that if the sum in (5) is 0, then $P(x)$ is 0 as follows: Since $F_{0,K-1}(x)$ is nonzero and $c_0 = 1$, if the sum is 0, then $r > 0$. In this case,

$$P(x) = -\sum_{q\geq b}\sum_{\substack{\boldsymbol{a}_b \\ 1\leq \bar{a}\leq r}}\sum_{n=0}^{A+q-1-b}\binom{n+b-A}{q-1} f_{r-\bar{a}}(n, K-1)x^{n+b}.$$

Let us assume the sum is 0, and pick a $j, 1 \leq j \leq r$. If $c_j = 0$, then there are no $f_{r-j}(n, K-1)$ terms in $P(x)$. If $c_j \neq 0$, then $F_{r-j,K-1}(x) = 0$, in which case all $f_{r-j}(n, K-1)$ terms in $P(x)$ are 0. This shows that $P(x) = 0$.

This means that we have

$$F_{r,K}(x) = \frac{1}{(1-x)^p}\left(\sum_{0\leq j\leq r} c_j \frac{1}{(q-1)!}\frac{d^{q-1}}{dx^{q-1}}F_{r-j,K-1}(x)\right) + O\left((1-x)^{-\alpha_K+1}\right).$$

By the inductive hypothesis:

$$
\begin{aligned}
F_{r,K}(x) &= \frac{1}{(1-x)^p}\left(\sum_{j=0}^{r}\frac{c_j}{(q-1)!}D_{r-j,K-1}[\alpha_{K-1}]^{q-1}(1-x)^{-\alpha_{K-1}-(q-1)}\right) \\
&\qquad\qquad\qquad\qquad\qquad + O\left((1-x)^{-\alpha_K+1}\right) \\
&= \left(\binom{\alpha_{K-1}+q-2}{q-1}\sum_{j=0}^{r}c_jD_{r-j,K-1}\right)(1-x)^{-((K-2+1)(q+p-1)+1)} \\
&\qquad\qquad\qquad\qquad\qquad + O\left((1-x)^{-\alpha_K+1}\right) \\
&= \left(\binom{\alpha_{K-1}+q-2}{q-1}\sum_{j=0}^{r}c_jD_{r-j,K-1}\right)(1-x)^{-\alpha_K} + O\left((1-x)^{-\alpha_K+1}\right).
\end{aligned}
$$

This means that

$$D_K(x) = \binom{\alpha_{K-1}+q-2}{q-1}C(x)D_{K-1}(x),$$

which, along with the inductive hypothesis gives

$$D_K(x) = C^{K-1}(x)\prod_{i=1}^{K-1}\binom{\alpha_i+q-2}{q-1}.$$

So the theorem is proved. □

Corollary 2. *We have that as $n \to \infty$*

$$f_r(n,k) = \frac{D_{r,k}}{((k-1)(q+p-1))!}n^{(k-1)(q+p-1)} + O\left(n^{(k-1)(q+p-1)-1}\right).$$

Proof. We note that

$$[x^n](1-x)^{-a} = \binom{n+a-1}{a-1} = n^{a-1} + O\left(n^{a-2}\right)$$

and the result is seen directly. □

2.1 The pattern 12^q

In this subsection p is set to 1 and we look at the pattern $1'2_{1,q}$, which is the (classical) pattern $12^q = 122\cdots2$. This is a particular case of interest for which we can produce more precise estimates. For q fixed, let $\tilde{f}_r(n,k) = |\mathcal{W}_n^{[k]}(12^q;r)|$, and $\tilde{F}_{r,k}(x) = \sum_{n\geq 0}\tilde{f}_r(n,k)x^n$.

If instead of using $\boldsymbol{a}_b$ for the positions of the k's we use it for the spacing in between them, we can get an expression for $C(x)$, and a simpler asymptotic expression for $\tilde{f}_r(n,k)$. Thus we now let $\boldsymbol{a}_b = (a_1, a_2, \ldots, a_{b-(q-1)})$ where a_1 is the number of entries before the first k, minus 1 (there is at least one non-k entry at the beginning of σ), and, for $i > 1$,

a_i is the number of non-k entries between the $(i-1)$th and ith k. Thus for all i, $a_i \geq 0$. We let $\bar{a}$ be the number of occurrences of 12^q that the k's are part of. It can be seen that

$$\bar{a} = \binom{b}{q} + \sum_{i=1}^{b-(q-1)} a_i \binom{b-i+1}{q} = \binom{b}{q} + \sum_{i=q}^{b} a_{b-i+1} \binom{i}{q}$$

where the $\binom{b}{q}$ is correcting for the 1 subtracted from a_1.

The definition of c_j is now the (finite) number of solutions $\boldsymbol{a}_b$ (for any b, $b \geq q$) to

$$j - \binom{b}{q} = \sum_{i=q}^{b} a_i \binom{i}{q},$$

and $c_0 = 1$. Thus for $p = 1$ we can define $C(x)$ as

$$\sum_{j \geq 0} c_j x^j = C(x) = 1 + \sum_{b \geq q} x^{\binom{b}{q}} \prod_{i=q}^{b} \frac{1}{1 - x^{\binom{i}{q}}}, \tag{6}$$

since $[x^0]C(x) = 1$ and for $j \geq 1$,

$$\begin{aligned} [x^j]C(x) &= \sum_{b \geq q} [x^{j-\binom{b}{q}}] \prod_{i=q}^{b} \frac{1}{1 - x^{\binom{i}{q}}} \\ &= c_j. \end{aligned}$$

Remark 1. *We have that* $C(x) = \prod_{i \geq q} \frac{1}{1-x^{\binom{i}{q}}}$ *is the ordinary generating function for the number of partitions of* n *into* q*th order binomial coefficients. This can be easily seen since the terms of* $C(x)$ *in (6) correspond to such a partition either being empty, or having largest part* $\binom{b}{q}$.

The new expression for $C(x)$ allows us to supply the following computational recursion for c_n:

$$c_0 = 1, \qquad c_n = \frac{1}{n} \sum_{j=1}^{n} \sum_{\substack{\binom{i}{q} | j \\ i \geq q}} \binom{i}{q} c_{n-j}, \quad n \geq 1.$$

The sequences c_n for $q = 1, 2$ and 3 are found as EIS A000041, EIS A007294 and EIS A068980, respectively, in [17]. We note that $C(x)$ is also the ordinary generating function for the number of partitions of n with non-negative q-th differences [1].

Now if we let

$$\tilde{D}_{r,k} = \frac{D_{r,k}}{(qk-q)!},$$

then we have

$$\tilde{D}_k(x) = \sum_{r \geq 0} \tilde{D}_{r,k} x^r = \frac{1}{(qk-q)!} C^{k-1}(x) \prod_{i=1}^{k-1} \binom{iq-1}{q-1} = \frac{C^{k-1}(x)}{(q!)^{k-1}(k-1)!}.$$

$r \backslash k$	1	2	3	4
0	1.0	0.50	0.13	0.021
1	0	0.50	0.25	0.063
2	0	0.50	0.38	0.13
3	0	1.0	0.75	0.27
4	0	1.0	1.1	0.50
5	0	1.0	1.5	0.81
6	0	2.0	2.5	1.4
7	0	2.0	3.5	2.3
8	0	2.0	4.5	3.4
9	0	3.0	6.5	5.2
10	0	3.5	8.8	7.7
11	0	3.5	11	11
12	0	5.0	15	16

Table 1: Rounded values of $\tilde{D}_{r,k}$ for the pattern 122.

By Corollary 1 we have, for $k \geq 1$,

$$\tilde{F}_{r,k}(x) = (qk-q)!\tilde{D}_{r,k}(1-x)^{-qk+q-1} + O\left((1-x)^{-qk+q}\right).$$

In addition, from Corollary 2 we have that as $n \to \infty$

$$\tilde{f}_r(n,k) = \tilde{D}_{r,k} n^{qk-q} + O\left(n^{qk-q-1}\right).$$

The magnitudes and growth of some initial values of $\tilde{D}_{r,k}$ are provided in Table 1 for the pattern 122.

3 The pattern 2^p12^q

We now turn to the pattern $2\cdots 212\cdots 2 = 2^p12^q$. An occurrence of 2^p12^q is formed by a subsequence $\phi = (\phi_1, \phi_2, \ldots, \phi_{p+q+1})$ where

$$\phi_{p+1} < \phi_1 = \cdots = \phi_p = \phi_{p+2} = \cdots = \phi_{p+q+1}.$$

If we set $p = 0$, we have the pattern 12^q from Section 2.1 and the following recursion is valid for this case. However, the asymptotic results derived in this section only apply for $p, q \geq 1$. We let $h_r(n,k) = |\mathcal{W}_n^{[k]}(2^p12^q; r)|$ and $H_{r,k}(x) = \sum_{n\geq 0} h_r(n,k)x^n$.

Theorem 2. *For $k \geq 1$, $H_{r,k}(x)$ is a rational function of the form*

$$\frac{\mathfrak{q}_{r,k}(x)}{(1-x)^{\alpha_{r,k}}},$$

where $\mathfrak{q}_{r,k}(x)$ is either 0 or a polynomial such that $\mathfrak{q}_{r,k}(1) \neq 0$, and $\alpha_{r,k} > 0$.

Proof. We derive a recursion for $h_r(n,k)$, in a different manner from Section 2, but again with reference to [5]. Given a word $\sigma \in \mathcal{W}_n^{[k]}(2^p12^q;r)$, we again let b represent the number of letters k in σ.

This time we immediately split into two cases: whether or not the b letters k are part of an occurrence of 2^p12^q.

In the case that they are not, our counting depends on b. If $b \le p+q-1$, their positions do not matter, so there are $\binom{n}{b}h_r(n-b,k-1)$ such words σ. If $b \ge p+q$, then the pth k from the left through to the qth k from the right must be consecutive in σ, and there are $\binom{n-b+p+q-1}{p+q-1}h_r(n-b,k-1)$ such words. This comes from the following procedure: Let the number of k's between the pth k from the left and the qth k from the right (inclusive) be

$$m = b-(p-1)-(q-1).$$

Let us say we are given $n-m+1$ slots in which we place the $(p-1)$ and $(q-1)$ letters k and one extra k. Then the extra k is replaced with all m copies of k. The remaining slots are filled with a $(k-1)$-ary word of length $n-b$ with r occurrences of the pattern, giving

$$\binom{n-m+1}{(p-1)+(q-1)+1}h_r(n-b,k-1) = \binom{n-b+p+q-1}{p+q-1}h_r(n-b,k-1)$$

words.

Finally, for the case in which the k's in σ are involved in at least one occurrence of the 2^p12^q pattern, we need only know the positions of the k's within the subword α between the pth k from the left in σ and the qth k from the right, exclusive. If α contains at least one non-k letter, the k's are part of an occurrence of 2^p12^q in σ. Let $\boldsymbol{a}_b = (a_1, a_2, \ldots, a_{b-p-q+1})$ be the spacing in between the k's in α, where a_1 is the number of non-k entries in α before its first k, $a_{b-p-q+1}$ is the number of non-k entries in α after its last k, and for $2 \le i \le b-p-q$, a_i is the number of non-k entries between the ith and $(i+1)$th k in α (in the particular case $\boldsymbol{a}_b = (a_1)$, a_1 is the length of α). It can be seen that in σ, the k's are part of

$$\bar{a} = \sum_{i=1}^{b-p-q+1} a_i \binom{p+i-1}{p}\binom{b-p+1-i}{q}$$

occurrences of 2^p12^q.

To see that the number of σ with a given $\boldsymbol{a}_b$ is $\binom{n-|\alpha|-1}{p+q-1}h_{r-\bar{a}}(n-b,k-1)$, we note the following. Let $|\alpha| = \|\boldsymbol{a}_b\|_1 + b-p-q$ be the length of α, where $\|\boldsymbol{a}_b\|_1$ is the sum of entries in $\boldsymbol{a}_b$. Given $n-|\alpha|$ slots there are $\binom{n-|\alpha|-1}{p+q-1}$ ways to place $p+q$ letters k such that the pth and $(p+1)th$ k are adjacent. For each of these ways, we insert $|\alpha|$ slots between the pth and $(p+1)$th k for α, place k's in the inserted slots according to $\boldsymbol{a}_b$ and fill the rest with some $\sigma' \in \mathcal{W}_{n-b}^{[k]}(2^p12^q;r-\bar{a})$.

Putting all the pieces together, we get for $n \geq 0$ and $k \geq 1$

$$h_r(n,k) = \sum_{b=0}^{p+q-1} \binom{n}{b} h_r(n-b,k-1) + \sum_{b=p+q}^{n} \binom{n-b+p+q-1}{p+q-1} h_r(n-b,k-1)$$
$$+ \sum_{b \geq p+q} \sum_{\substack{\boldsymbol{a}_b \succeq 0 \\ 1 \leq \bar{a} \leq r \\ b+\|\boldsymbol{a}_b\|_1 \leq n}} \binom{n-|\alpha|-1}{p+q-1} h_{r-\bar{a}}(n-b,k-1), \tag{7}$$

We also have $h_r(n,0) = [n{=}r{=}0]$. We take $h_r(n,k) = 0$ for negative n. Similarly to what we have seen in Theorem 1, when we multiply (7) by x^n and sum on $n \geq 0$ we have

$$H_{r,k}(x) = \sum_{b=0}^{p+q-1} \sum_{i=0}^{b} \lambda_{i,b} x^{i+b} \frac{d^i}{dx^i} H_{r,k-1}(x) + \frac{x^{p+q}}{1-x} \sum_{i=0}^{p+q-1} \lambda_i x^i \frac{d^i}{dx^i} H_{r,k-1}(x)$$
$$+ \sum_{b \geq p+q} \sum_{\substack{\boldsymbol{a}_b \succeq 0 \\ 1 \leq \bar{a} \leq r}} \left(\sum_{i=0}^{p+q-1} \lambda_{|\alpha|,i,b} x^{i+b} \frac{d^i}{dx^i} H_{r-\bar{a},k-1}(x) - P(x) \right), \tag{8}$$

where $P(x)$ is the polynomial $\sum_{n=0}^{\|a\|_1-1} \binom{n+b-|\alpha|-1}{p+q-1} h_{r-\bar{a}}(n,k-1) x^{n+b}$. Now using the initial value

$$H_{r,1}(x) = [r{=}0] \, \frac{1}{1-x},$$

and the fact that $H_{r,k}(x)$ cannot be a nonzero polynomial, the theorem can be seen directly by induction. □

We let c_j be the number of solutions $\boldsymbol{a}_b$ to

$$j = \sum_{i=1}^{b-p-q+1} a_i \binom{p+i-1}{p} \binom{b-p+1-i}{q}.$$

We observe that c_j is only guaranteed to be finite if we require $p, q \geq 1$; because of this, the asymptotics that follow assume these conditions. We remark that this means the following results are not a generalization of Section 2.1.

Let $t = p+q-1$ and consider (8). Since $\lambda_t = \lambda_{t,t} = \lambda_{|\alpha|,t,b} = \frac{1}{t!}$, setting x factors to 1 and keeping only the highest derivative for each $H_{j,k}(x)$, $0 \leq j \leq r$, since the other derivatives are of smaller order, yields, for $k \geq 2$

$$H_{r,k}(x) = \frac{1}{t!} \left(\frac{1}{1-x} \frac{d^t}{dx^t} H_{r,k-1}(x) + \sum_{j=1}^{r} c_j \frac{d^t}{dx^t} H_{r-j,k-1}(x) \right) + O\left((1-x)^{-\alpha_{r,k}+1} \right). \tag{9}$$

Corollary 3. *The function $H_{r,k}(x)$ has the following asymptotic form:*

$$H_{r,k}(x) = D_{r,k}(1-x)^{-\alpha_{r,k}} + O\left((1-x)^{-\alpha_{r,k}+1}\right),$$

where

$$\alpha_{r,k} = \begin{cases} k(t+1)-t-1 & \text{if } r>0, \\ k(t+1)-t & \text{if } r=0, \end{cases}$$

and

$$D_{r,1} = [r{=}0], \qquad D_{0,k} = \frac{((k-1)(t+1)-1)!}{((t+1)!)^{k-1}(k-2)!}, \quad k \geq 2,$$

and for $r > 0$ and $k \geq 2$

$$D_{r,k} = \sum_{i=2}^{k} y_i x_{i+1} x_{i+2} \cdots x_k, \tag{10}$$

where

$$x_i = \frac{1}{t!}[\alpha_{r,i-1}]^t, \qquad y_i = c_r D_{0,i-1}[\alpha_{0,i-1}]^t \frac{1}{t!}.$$

Proof. We first prove the value of $\alpha_{r,k}$. We begin with the case $r=0$, using induction on k. The base case is $k=1$, for which we have

$$H_{0,1}(x) = \frac{1}{1-x},$$

giving $\alpha_{0,1} = 1(t+1)-t = 1$.

We assume the theorem holds for $k, 1 \leq k < K$. From (9) we have

$$H_{0,K}(x) = \frac{1}{t!}\frac{1}{1-x}\frac{d^t}{dx^t}H_{r,K-1}(x) + O\left((1-x)^{-\alpha_{0,K}+1}\right).$$

By the inductive hypothesis:

$$\begin{aligned} H_{0,K}(x) &= \frac{1}{t!}\frac{1}{1-x}D_{0,K-1}[\alpha_{0,K-1}]^t(1-x)^{-\alpha_{0,K-1}-t} + O\left((1-x)^{-\alpha_{0,K}+1}\right) \\ &= \frac{1}{t!}D_{0,K-1}[\alpha_{0,K-1}]^t(1-x)^{-((K-1)(t+1)-t)-t-1} + O\left((1-x)^{-\alpha_{0,K}+1}\right) \\ &= \frac{1}{t!}D_{0,K-1}[\alpha_{0,K-1}]^t(1-x)^{-K(t+1)+t} + O\left((1-x)^{-\alpha_{0,K}+1}\right). \end{aligned}$$

Now we consider the case $r > 0$. For the base case $k=1$, $H_{r,1}(x) = 0$, and we take $\alpha_{r,1} = 0 = 1(t+1)-t-1$.

For the inductive step we assume the theorem holds for $k, 1 \leq k < K$. We have

$$H_{r,K}(x) = \frac{1}{t!}\left(\frac{1}{1-x}\frac{d^t}{dx^t}H_{r,K-1}(x) + \sum_{j=1}^{r} c_j \frac{d^t}{dx^t}H_{r-j,K-1}(x)\right) + O\left((1-x)^{-\alpha_{r,K}+1}\right).$$

Using the inductive hypothesis,

$$
\begin{aligned}
H_{r,K}(x) = \frac{1}{t!}\Bigg(& \frac{1}{1-x} D_{r,K-1}[\alpha_{r,K-1}]^t (1-x)^{-(K-2)(t+1)-t} \\
& + \sum_{j=1}^{r-1} c_j D_{r-j,K-1}[\alpha_{r-j,K-1}]^t (1-x)^{-(K-2)(t+1)-t} \\
& + c_r D_{0,K-1}[\alpha_{0,K-1}]^t (1-x)^{-(K-1)(t+1)} \Bigg) + O\left((1-x)^{-\alpha_{r,K}+1}\right).
\end{aligned}
$$

The dominating terms are the first and the third, as long as they are nonzero. The third term is nonzero because we know c_r, $D_{0,K-1}$, and $\alpha_{0,K-1}$ are positive, so we can say:

$$
\begin{aligned}
H_{r,K}(x) &= \frac{1}{t!}\left(D_{r,K-1}[\alpha_{r,K-1}]^t (1-x)^{-(K-2)(t+1)-t-1} \right. \\
&\qquad \left. + c_r D_{0,K-1}[\alpha_{0,K-1}]^t (1-x)^{-(K-1)(t+1)}\right) + O\left((1-x)^{-\alpha_{r,K}+1}\right) \\
&= \frac{1}{t!}\left(D_{r,K-1}[\alpha_{r,K-1}]^t + c_r D_{0,K-1}[\alpha_{0,K-1}]^t\right)(1-x)^{-(K(t+1)-t-1)} \\
&\qquad\qquad + O\left((1-x)^{-\alpha_{r,K}+1}\right).
\end{aligned}
$$

Next we consider the value of $D_{r,k}$. From what we have shown so far, it is clear that

$$
D_{0,1} = 1 \quad D_{0,k} = \frac{1}{t!}[\alpha_{0,k-1}]^t D_{0,k-1}, \quad k \geq 2.
$$

This gives for $k \geq 2$

$$
D_{0,k} = \frac{((k-1)(t+1)-1)!}{((t+1)!)^{k-1}(k-2)!}.
$$

For $r > 0$, we have the first-order linear recursion

$$
D_{r,1} = 0 \quad D_{r,k} = \frac{1}{t!}\left(D_{r,k-1}[\alpha_{r,k-1}]^t + c_r D_{0,k-1}[\alpha_{0,k-1}]^t\right), \quad k \geq 2.
$$

Its solution is given in (10). □

The following corollary has a proof similar to the one in Corollary 2.

Corollary 4. *We have that as $n \to \infty$*

$$
h_0(n,k) = \frac{D_{r,k}}{((k-1)(t+1))!} n^{(k-1)(t+1)} + O\left(n^{(k-1)(t+1)-1}\right),
$$

and for $r > 0$, as $n \to \infty$

$$
h_r(n,k) = \frac{D_{r,k}}{((k-1)(t+1)-1)!} n^{(k-1)(t+1)-1} + O\left(n^{(k-1)(t+1)-2}\right).
$$

4 Conclusion

In the previous sections we have shown how the asymptotic form of $|\mathcal{W}_n^{[k]}(\tau; r)|$ as $n \to \infty$ may be computed, for two families of patterns with two parameters each. Whether these expressions can be simplified further remains open, as well as the asymptotic form of the distribution of occurrences of a given classical pattern or classical POP. For a hidden word, the number of occurrences is asymptotically normal [9].

One area of possible further work is an extension of the recursions and generating functions in this paper to the following scenario. Consider a weight function w defined on letters, and additively on words, i.e. for a word σ of length n, $w(\sigma) = w(\sigma_1) + \cdots + w(\sigma_n)$. If $w(j) = 1$ for all j, then the weight of a word is its length; if $w(j) = j$ for all j, then the weight of a word is its order as an integer composition. Instead of counting k-ary words of length n as this paper does, consider counting k-ary words with weight m.

For example, the ordinary generating function for k-ary words with occurrences of 1^p marked by u and weight marked by x is

$$G_k(x, u) = \prod_{i=1}^{k} \sum_{n \geq 0} u^{\binom{n}{p}} \frac{x^{n\,w(i)}}{n!}.$$

Using a general weight function seems to complicate the analysis substantially, but it may be possible to treat some particular weight functions.

Acknowledgements

The authors wish to thank an anonymous referee for helpful comments.

References

[1] G.E. Andrews. MacMahon's Partition Analysis: II Fundamental Theorems. *Ann. Combin.*, 4(3/4):327–338, 2000.

[2] P. Brändén and T. Mansour. Finite automata and pattern avoidance in words. *J. Comb. Theory Ser. A*, 110(1):127–145, 2005.

[3] A. Burstein. *Enumeration of words with forbidden patterns.* PhD thesis, University of Pennsylvania, 1998.

[4] A. Burstein and S. Kitaev. Partially ordered patterns and their combinatorial interpretations. *Pure Math. and Appl. (PU.M.A.)*, 19(2-3):27–38, 2008.

[5] A. Burstein and T. Mansour. Words restricted by patterns with at most 2 distinct letters. *Electron. J. Combin.*, 9(2), 2002.

[6] A. Burstein and T. Mansour. Counting occurrences of some subword patterns. *Discrete Math. Theor. Comput. Sci.*, 6(1):1–11, 2003.

[7] A. Burstein and T. Mansour. Words restricted by 3-letter generalized multipermutation patterns. *Ann. Combin.*, 7(1):1–14, 2003.

[8] G. Firro and T. Mansour. Restricted k-ary words and functional equations. *Discrete Appl. Math.*, 157(4):602–616, 2009.

[9] P. Flajolet, W. Szpankowski, and B. Vallée. Hidden word statistics. *J. of the ACM*, 53(1):147–183, 2006.

[10] S. Heubach, S. Kitaev, and T. Mansour. Partially ordered patterns and compositions. *Pure Math. and Appl. (PU.M.A.)*, 17(1-2):1–12, 2007.

[11] S. Heubach and T. Mansour. *Combinatorics of Compositions and Words*. Chapman & Hall/CRC, 2010.

[12] S. Kitaev. Segmented partially ordered generalized patterns. *Theor. Comp. Sci.*, 349(3):420–428, 2005.

[13] S. Kitaev and T. Mansour. Partially ordered generalized patterns and k-ary words. *Ann. Combin.*, 7(2), 2003.

[14] T. Mansour. Restricted 132-avoiding k-ary words, Chebyshev polynomials, and continued fractions. *Adv. in Appl. Math.*, 36(2):175–193, 2006.

[15] L. Pudwell. *Enumeration Schemes for Pattern-Avoiding Words and Permutations*. PhD thesis, Rutgers University, 2008.

[16] A. Regev. Asymptotics of the number of k-words with an l-descent. *Electron. J. Combin.*, 5:R15, 1998.

[17] N.J.A. Sloane. The On-Line Encyclopedia of Integer Sequences, 2010. Published electronically at www.research.att.com/~njas/sequences/.