Counting l-letter subwords in compositions
Toufik Mansour, Basel O. Sirhan

To cite this version:

HAL Id: hal-00961119
https://inria.hal.science/hal-00961119
Submitted on 20 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Counting $\ell$-letter subwords in compositions

Toufik Mansour and Basel O. Sirhan

Department of Mathematics, University of Haifa, 31905 Haifa, Israel.
toufik@math.haifa.ac.il, bosirhan@yahoo.com

received April 25, 2006, revised October 20, 2006, accepted November 2, 2006.

Let $\mathbb{N}$ be the set of all positive integers and let $A$ be any ordered subset of $\mathbb{N}$. Recently, Heubach and Mansour enumerated the number of compositions of $n$ with $m$ parts in $A$ that contain the subword $\tau$ exactly $r$ times, where $\tau \in \{111, 112, 221, 123\}$. Our aims are (1) to generalize the above results, i.e., to enumerate the number of compositions of $n$ with $m$ parts in $A$ that contain an $\ell$-letter subword, and (2) to analyze the number of compositions of $n$ with $m$ parts that avoid an $\ell$-letter pattern, for given $\ell$. We use tools such as asymptotic analysis of generating functions leading to Gaussian asymptotic.

Keywords: Carlitz compositions, compositions, subwords, generating functions, Gaussian asymptotic

1 Introduction

A composition $\pi = \pi_1 \pi_2 \ldots \pi_m$ of $n \in \mathbb{N}$ is an ordered collection of one or more positive integers whose sum is $n$. The number of summands, namely $m$, is called the number of parts of the composition. Let $A = \{a_1, a_2, \ldots, a_d\}$ or $A = \{a_1, a_2, a_3, \ldots\}$, where $a_1 < a_2 < \cdots$ are positive integers. We will refer to such a set as an ordered subset of $\mathbb{N}$. We will look at compositions of $n$ with parts in $A$, i.e., compositions whose parts are restricted to be from a set $A \subseteq \mathbb{N}$.

The problem of counting the number of compositions (words), which contain a set of given number of strings as substrings, is a classical problem in combinatorics. This problem can be studied using the transfer matrix method (see [12, Section 4.7] and [6]). In particular, it is a well-known fact that the generating function of such compositions is always rational [12, Theorem 4.7.2]. For example, following [12, Example 4.7.5], one can find that the generating function for the number of compositions with parts in $\{1, 2, 3\}$, where neither 11 nor 23 appear as two consecutive parts, is given by $1 + x - 2x^3 - x^4 - x^5 + x^6$. We note that the term pattern in [12, Section 4.7] and [6] is used to denote an exact string rather than its type with respect to order isomorphism. For example, the pattern 11 is the actual string 11, whereas in our setting an occurrence of the subword pattern 11 is any substring aa. Another example, the pattern rise, non-rise as defined in [6] includes the subword patterns 121, 122, 132, 231 as defined in this paper.

However, we show that each of the subword patterns 121, 122, 132 and 231, is avoided by a different number of compositions of $n$.

In this paper, we analyze, in several cases, different asymptotic parameters for the number of compositions of $n$ with $m$ parts in $\mathbb{N}$, which do not contain a given subword pattern (see below for the precise definition). In order to do that, first we need to find explicit formulas for the generating functions for the
number of compositions of $n$ with $m$ parts in $\mathbb{N}$, which does not contain given subword patterns. While the transfer matrix method expresses these generating functions in terms of determinants, we present here another technique which leads to compact formulas.

Let $\pi = \pi_1 \pi_2 \ldots \pi_m$ be any composition of $n$ with $m$ parts in $A$ and $\sigma = \sigma_1 \sigma_2 \ldots \sigma_m$ be any word of length $m'$, where $m \geq m'$. An occurrence of $\sigma$ in $\pi$ is a subword $\pi_i \ldots \pi_{i+m'-1}$ which is order-isomorphic to $\sigma$, i.e., $\pi_{i-1} + a < \pi_{i+b}$ if and only if $\sigma_a < \sigma_b$ for all $1 \leq a < b \leq m'$. We denote the number of occurrences of $\sigma$ in $\pi$ by $\sigma(\pi)$, and in such a context $\sigma$ is usually called a pattern of length $m'$ (or $m'$-letter pattern). For example, the number of occurrences of the pattern 112 in the composition $\pi = 1123113$ of 14 with 8 parts is 3, namely, 112, 223 and 113; thus $112(\pi) = 3$. If $r = 0$, we say that $\pi$ avoids $\sigma$.

Our aim is to count the number of compositions of $n$ with $m$ parts in $A$ which contain an $\ell$-letter pattern $\sigma$ exactly $r$ times, and to analyze the number of compositions of $n$ with $m$ parts in $A$ which avoid an $\ell$-letter pattern $\sigma$.

We denote the set of compositions of $n$ with parts in $A$ that avoid a pattern $\tau$ (resp. with $m$ parts) by $C^A_\tau(n)$ (resp. $C^A_\tau(n, m)$), and the set of compositions of $n$ with parts in $A$ that contain a pattern $\tau$ exactly $r$ times (resp. with $m$ parts) by $C^A_\tau(r, n)$ (resp. $C^A_\tau(r, n, m)$). Denote the generating function for the sequences $\#C^A_\tau(n, m)$ and $\#C^A_\tau(r, n, m)$ by $C^A_\tau(x, y)$ and $C^A_\tau(x, y, z)$, in other words,

$$C^A_\tau(x, y) = \sum_{n \geq 0} \sum_{m \geq 0} \#C^A_\tau(n, m)x^ny^m, \quad C^A_\tau(x, y, z) = \sum_{r \geq 0} \sum_{n \geq 0} \sum_{m \geq 0} \#C^A_\tau(r, n, m)x^ny^rz^r.$$

Recently, Heubach and Mansour [8] found that

$$C^A_{111}(x, y, z) = \frac{1}{1 - x - y - \sum_{d \geq 1} x^dy(1+(-1)z)x^dy), \quad C^A_{112}(x, y, z) = \frac{1}{1 - \sum_{d \geq 1} x^dy\prod_{i=1}^{d-1} (1-(1-z)x^ay^2)},$$

$$C^A_{221}(x, y, z) = \frac{1}{1 - \sum_{d \geq 1} x^dy\prod_{i=1}^{d-1} (1-(1-z)x^ay^2)}, \quad C^A_{222}(x, y, z) = \frac{1}{1 - \sum_{d \geq 1} x^dy\prod_{i=1}^{d-1} (1-(1-z)x^ay^2)},$$

where $t^p(A) = \sum_{1 \leq i_1 < \ldots < i_p \leq d} y^p \prod_{j=1}^{p} x^{a_i}$. Our first goal is to generalize these results to obtain explicit formula for the generating function $C^A_\tau(x, y, z)$, where $\tau$ is an $\ell$-letter pattern. In particular, we find explicit formulas for the generating functions $C^A_\tau(x, y, z)$, where $\tau \in \{121, 122, 131, 132\}$, see Section 2. We note that our generating function $C^A_\tau(x, y, z)$ for $x = 1$ and $A = \{1, 2, \ldots, k\}$ gives the generating function for the number $k$-ary words that avoid $\tau$, thus this presents a generalization of Burstein and Mansour’s results, see [2].

A Carlitz composition of $n$, introduced in [3], is a composition of $n$ in which no adjacent parts are the same, namely, one that avoids the pattern 11. More generally, a $k$-Carlitz composition of $n$ is a composition of $n$ in which no $k$ consecutive parts are the same, that is, a composition that avoids the pattern $11\ldots1_k$ for all $k \geq 2$.

Recently, several authors analyzed different asymptotic parameters of patterns of combinatorial objects such as trees [4] and compositions [10, 11]. For instance, in [10], Knopfmacher and Prodinger found asymptotic values for the number of Carlitz compositions of $n$ and the mean number of parts, when all the Carlitz compositions of $n$ are considered as equiprobable. In [11], Louchard and Prodinger considered compositions as particular polynomials and obtained a stochastic description of their behavior for large $n$. They analyzed the number of parts in the set of Carlitz compositions of $n$. Our second goal is to generalize this approach to compositions that avoid a more general $\ell$-letter pattern. In particular, we analyze different
asymptotic parameters of the number of \( k \)-Carlitz compositions of \( n \) with \( m \) parts in \( \mathbb{N} \), see Section 3.

We use tools such as asymptotic analysis of generating functions (based on their singularities) leading to Gaussian asymptotic (see [1], [5, Theorem IX.8], and [9]).

## 2 Enumeration of compositions

In this section we propose an explicit formula for the generating function \( C^A_\tau(x, y, z) \) for several interesting cases of \( \ell \)-letter patterns \( \tau \). To do that we need the following notation. Denote the generating function for all compositions \( \pi \in C^A_{\tau, r}(n, m) \) with \( \pi_1 \ldots \pi_s = \alpha_1 \ldots \alpha_s \) by \( C^A_\tau(\alpha_1 \ldots \alpha_s|x, y, z) \).

### 2.1 The pattern \( \tau = 11 \ldots 1 \)

Let \( \tau = 11 \ldots 1 \) be a pattern of length \( \ell \). By the definitions we have that

\[
C^A_\tau(x, y, z) = 1 + \sum_{a \in A} C^A_\tau(a|x, y, z),
\]

and for all \( s = 1, 2, \ldots, \ell - 1, \)

\[
C^A_\tau(aa \ldots a|x, y, z) = x^{as} y^s + \sum_{b \neq a \in A} C^A_\tau(a b|x, y, z) + C^A_\tau(aa \ldots a|x, y, z).
\]

Observe that if \( \pi \in C^A_{\tau, r}(n; m) \) is such that \( \pi_1 \ldots \pi_s \pi_{s+1} = aa \ldots ab \) where \( b \neq a \) and \( 1 \leq s \leq \ell - 1 \), then no occurrence of the generalized pattern \( \tau \) in \( \pi \) can involve one of the first \( s \) letters of \( \pi \). Therefore, \( \pi \in C^A_{\tau, r}(n; m) \) if and only if \( (b, \pi_{s+2}, \ldots, \pi_m) \in C^A_{\tau, r}(n-sa; m-s) \). Hence,

\[
C^A_\tau(aa \ldots a|x, y, z) = x^{as} y^s + \sum_{b \neq a \in A} C^A_\tau(b|x, y, z) + C^A_\tau(aa \ldots a|x, y, z),
\]

equivalently, see (2.1).

\[
C^A_\tau(aa \ldots a|x, y, z) = x^{as} y^s(C^A_\tau(x, y, z) - C^A_\tau(a|x, y, z)) + C^A_\tau(aa \ldots a|x, y, z),
\]

Moreover,

\[
C^A_\tau(aa \ldots a|x, y, z) = x^{as} y^sz + \sum_{b \neq a \in A} C^A_\tau(a b|x, y, z) + C^A_\tau(aa \ldots a|x, y, z).
\]

Observe that if \( \pi \in C^A_{\tau, r}(n; m) \) is such that \( \pi_1 \ldots \pi_{\ell} \pi_{\ell+1} = aa \ldots ab \) where \( b \neq a \), then exactly one occurrence of the pattern \( \tau \) in \( \pi \) can involve one of the first \( \ell \) letters of \( \pi \). Therefore, \( \pi \in C^A_{\tau, r}(n; m) \) if and only if \( (b, \pi_{\ell+2}, \ldots, \pi_m) \in C^A_{\tau, r-1}(n-\ell a; m-\ell) \). In the case of \( b = a \) we get that \( \pi \in C^A_{\tau, r}(n; m) \) if and only if \( (\pi_2, \ldots, \pi_m) \in C^A_{\tau, r-1}(n-a; m-1) \). Hence,

\[
C^A_\tau(aa \ldots a|x, y, z) = x^{as} y^sz + x^{as} y^s(z + yz) C^A_\tau(a a \ldots a|x, y, z),
\]
which is equivalent to, using (2.1),

\[ C^A_{\tau}(aa \ldots a|x, y, z) = \frac{x^{a\ell} y^\ell z}{1 - x^a y z} (C^A_{\tau}(x, y, z) - C^A_{\tau}(a|x, y, z)). \]  

(2.3)

Using (2.2) and (2.3) and induction on \( s \) we obtain that

\[ C^A_{\tau}(aa \ldots a|x, y, z) = \left( x^{a\ell} y^s + \ldots + x^{a(\ell-1)} y^{\ell-1} + \frac{x^{a\ell} y^z}{1 - x^a y z} \right) (C^A_{\tau}(x, y, z) - C^A_{\tau}(a|x, y, z)), \]

for all \( s = 1, 2, \ldots, \ell \). Hence, putting \( s = 1 \) and using Equation (2.1) we get

\[ C^A_{\tau}(x, y, z) = 1 + \sum_{a \in A} C^A_{\tau}(a|x, y, z), \]

with

\[ C^A_{\tau}(a|x, y, z) = \frac{x^{a} y + x^{2a} y^2 + \ldots + x^{a(\ell-1)} y^{\ell-1} + x^{a\ell} y^z}{1 - x^a y + x^{2a} y^2 + \ldots + x^{a(\ell-1)} y^{\ell-1} + x^{a\ell} y^z/(1 - x^a y z)} C^A_{\tau}(x, y, z), \]

which gives the following result.

**Theorem 2.1** Let \( A \) be any ordered subset of \( \mathbb{N} \) and let \( \tau = 11 \ldots 1 \) be a pattern of length \( \ell \). Then

\[ C^A_{\tau}(x, y, z) = \frac{1}{1 - \sum_{a \in A} \frac{x^a y + x^{2a} y^2 + \ldots + x^{a(\ell-1)} y^{\ell-1} + x^{a\ell} y^z}{1 - x^a y + x^{2a} y^2 + \ldots + x^{a(\ell-1)} y^{\ell-1} + x^{a\ell} y^z/(1 - x^a y z)}}. \]

For instance, Theorem 2.1 for \( z = 0 \) and \( A = \mathbb{N} \) gives that the generating function for the number of compositions of \( n \) with \( m \) parts in \( \mathbb{N} \) that avoid \( \tau = 11 \ldots 1 \) (\( \ell \)-letter pattern) is given by

\[ C^N_{\tau}(x, y) = C^N_{\tau}(x, y, 0) = \frac{1}{1 - \sum_{j \geq 1} \frac{x^{j} y + x^{2j} y^2 + \ldots + x^{j(\ell-1)} y^{\ell-1}}{1 + x^{j} y + x^{2j} y^2 + \ldots + x^{j(\ell-1)} y^{\ell-1}}} \]

In particular, if \( \ell = 3 \) then we obtain [8, Theorem 2.1], which is given for easy reference also in (1.1).

### 2.2 The pattern \( \tau = 11 \ldots 12 \)

Let \( \tau = 11 \ldots 12 \) be a pattern of length \( \ell \) and let \( A = \{a_1, \ldots, a_d\} \) be any ordered subset of \( \mathbb{N} \). Define \( d^A_{r,r'}(n; m) \) to be the number of compositions \( \pi \in C^A_{r,r'}(n; m) \) such that the string concatenation \( \pi(a_d + 1) (a_d + 1 > a_d) \) contains \( \tau \) exactly \( r \) times, and denote the corresponding generating function by \( D^A_{\tau}(x, y, z) = \sum_{n,m,r \geq 0} d^A_{r,r'}(n; m)x^ny^rz^r \). Let us find a recurrence for \( C^A_{\tau}(x, y, z) \).

Let \( \pi \in C^A_{\tau}(n; m) \) be such that \( \pi \) contains exactly \( s \) copies of the letter \( a_d \). If \( s = 0 \), then \( \pi \in C^A_{r,r'}(n; m) \) where \( A' = \{a_1, \ldots, a_{d-1}\} \). If \( s \geq 1 \), then \( \pi = \pi^{(1)} a_d \pi^{(2)} \), where \( \pi^{(1)} \in C^A_{\tau}(n_1; m_1) \) with \( \tau(\pi^{(1)} a_d) = r_1 \) (number occurrences of \( \tau \) in \( \pi^{(1)} a_d \)), \( \pi^{(2)} \in C^A_{\tau}(n_2; m_2) \), \( n_1 + n_2 = n - a_d \), \( m_1 + m_2 + 1 = m \), and \( r_1 + r_2 = r \). In terms of generating functions, the above translates into

\[ C^A_{\tau}(x, y, z) = C^{A'}_{\tau}(x, y, z) + x^{a_d} y D^{A'}_{\tau}(x, y, z) C^A_{\tau}(x, y, z), \]
or, equivalently,

\[ C^\ell_A(x, y, z) = \frac{C^\ell_A(x, y, z)}{1 - x^a_d y D^\ell_A(x, y, z)}. \]  

(2.4)

Let us now find the recurrence for \( D^\ell_A(x, y, z) \). Let \( \pi \in C^A(n; m) \) be such that the string concatenation \( \pi(a_d+1) \) contains \( \tau \) exactly \( r \) times, and has exactly \( s \) letters \( a_d \). Then \( \pi = \pi^{(1)} a_d \pi^{(2)} \ldots a_d \pi^{(s)} a_d \pi^{(s+1)} \) for some \( \pi^{(i)} \in C^A(n_i; m_i) \), \( 1 \leq i \leq s + 1 \), where \( n_1 + \cdots + n_{s+1} = n - sa_d, m_1 + \cdots + m_{s+1} = m - s \), and

\[ \tau^{(1)} a_d + \cdots + \tau^{(s)} a_d + \tau^{(s+1)} (a_d + 1) \] and \( \delta(\pi) \) ends on \( \ell - 1 \) \( a_d \)'s = \( r \),

where \( \delta(\xi) = 1 \) if \( \xi \) holds and \( \delta(\xi) = 0 \) otherwise. Translating to generating functions, we obtain

\[
\begin{align*}
D^\ell_A(x, y, z) &= \sum_{s=0}^{\ell-2} x^{a_d s} y^s (D^\ell_A(x, y, z))^{s+1} \\
&+ \sum_{s=\ell-1}^{\infty} x^{a_d s} y^s \left( (D^\ell_A(x, y, z))^{s+1} - (D^\ell_A(x, y, z))^{(s+1)-(\ell-1)} + z(D^\ell_A(x, y, z))^{(s+1)-(\ell-1)} \right),
\end{align*}
\]

which, after summing over \( s \), yields

\[
D^\ell_A(x, y, z) = \frac{(1 - x^{(\ell-1)a_d} y^{\ell-1}(1 - z)) D^\ell_A(x, y, z)}{1 - x^{a_d} y D^\ell_A(x, y, z)}. \]  

(2.5)

These two recurrences, together with \( D^\ell_A(x, y, z) = C^\ell_A(x, y, z) = 1 \) and induction, yield the following theorem.

**Theorem 2.2** Let \( A \) be any ordered subset of \( \mathbb{N} \) and let \( \tau = 11 \ldots 12 \) be a pattern of length \( \ell \). Then

\[ C^\ell_A(x, y, z) = \frac{1}{1 - \sum_{j=1}^{d} x^{a_j} y \prod_{i=1}^{j-1} (1 - x^{(\ell-1)a_i} y^{\ell-1}(1 - z))}. \]

For example, Theorem 2.2 for \( z = 0 \) and \( A = \mathbb{N} \) gives that the generating function for the number of compositions of \( n \) with \( m \) parts in \( \mathbb{N} \) which avoid the \( \ell \)-letter pattern \( \tau = 11 \ldots 12 \) is given by

\[ C^\ell_A(x, y) = C^\ell_A(x, y, 0) = \frac{1}{1 - \sum_{j \geq 1} x^{j} y \prod_{i=1}^{j-1} (1 - x^{(\ell-1)} y^{\ell-1})}. \]

In particular, if \( \ell = 3 \) then we get [8] Theorem 2.2], which is also listed in (1,1).

Similar arguments as in the proof of Theorem 2.2, replacing \( a_d \) by \( a_1 \), lead us to the following result.

**Theorem 2.3** Let \( A \) be any ordered subset of \( \mathbb{N} \) and let \( \tau = 22 \ldots 21 \) be a pattern of length \( \ell \). Then

\[ C^\ell_A(x, y, z) = \frac{1}{1 - \sum_{j=1}^{d} x^{a_j} y \prod_{i=j+1}^{d} (1 - x^{(\ell-1)a_i} y^{\ell-1}(1 - z))}. \]

For example, Theorem 2.3 for \( z = 0 \) and \( A = \mathbb{N} \) gives that the generating function for the number of compositions of \( n \) with \( m \) parts in \( \mathbb{N} \) which avoid the \( \ell \)-letter pattern \( \tau = 22 \ldots 21 \) is given by

\[ C^\ell_A(x, y) = C^\ell_A(x, y, 0) = \frac{1}{1 - \sum_{j \geq 1} x^{j} y \prod_{i=j+1}^{d} (1 - x^{(\ell-1)} y^{\ell-1})}. \]

In particular, if \( \ell = 3 \) then we get [8] Theorem 2.2], which is also listed in (1,1).
2.3 The subword $\tau = 211 \ldots 112$

Let $A = \{a_1, \ldots, a_d\}$ be any ordered subset of $\mathbb{N}$, and let $\tau = 211 \ldots 112$ be a pattern of length $\ell$. We define $d^{A}_{\tau}(n; m)$ to be the number of compositions $\pi' \in C^{A}(n; m)$ such that the string concatenation $(a_d + 1)\pi'(a_d + 1)$ contains $\tau$ exactly $r$ times, and denote the corresponding generating function by $D^{A}_{\tau}(x, y, z)$. Let $\pi \in C^{A}_{\tau}(n; m)$ such that $\pi$ contains $s$ occurrences of the letter $a_d$. For $s = 0$, the generating function for the number of such compositions $\pi$ is given by $C^{A}_{\tau}(x, y, z)$ where $A' = \{a_1, \ldots, a_{d-1}\}$, and for $s \geq 1$, by $x^{sa_d}y^s(C^{A'}_{\tau}(x, y, z))^2(D^{A'}_{\tau}(x, y, z))^{s-1}$; since in that case $\pi = \pi(0)_{a_d\pi(1)}a_d \ldots a_d\pi(s-1)_{a_d\pi(s)}$, where all $\pi(i) \in C^{A'}(n_i; m_i)$, $\sum_{i=0}^{s}n_i = n - sa_d$, $\sum_{i=0}^{d}m_i = m - s$, and $\tau(\pi) = \tau(\pi(0)) + \sum_{i=1}^{s-1} \tau(a_d\pi(i)_{a_d}) + \tau(\pi(s))$. Hence, if we sum over all $s \geq 0$, we get

$$C^{A}_{\tau}(x, y, z) = C^{A'}_{\tau}(x, y, z) + \frac{x^{sa_d}y^s(C^{A'}_{\tau}(x, y, z))^2}{1 - x^{sa_d}y D^{A'}_{\tau}(x, y, z)}.$$  

On the other hand, the string concatenation $(a_d + 1)\pi'(a_d + 1)$, with $\pi'$ as above, contains an occurrence of $\tau$ involving the two letters $a_d + 1$ if and only if $\pi'$ is a constant string of length $\ell - 2$, otherwise, $\tau((a_d + 1)\pi'(a_d + 1)) = \tau(\pi')$ (that is, the number occurrences of $\tau$ in the string concatenation $(a_d + 1)\pi'(a_d + 1)$ is the same as the number occurrences of $\tau$ in $\pi'$). Translating to generating functions, we obtain

$$D^{A}_{\tau}(x, y, z) = \sum_{i=1}^{d} x^{(\ell-2)a_i}y^{\ell - 2}z + C^{A}_{\tau}(x, y, z) - \sum_{i=1}^{d} x^{(\ell-2)a_i}y \tau(z).$$

Therefore, using the initial conditions $C^{A}_{\tau}(x, y, z) = D^{A}_{\tau}(x, y, z) = 1$ and induction on $d$, we get the following theorem.

Theorem 2.4 Let $A$ be any ordered subset of $\mathbb{N}$ and let $\tau = 211 \ldots 112$ be a pattern of length $\ell$. For all $\ell \geq 3$,

$$C^{A}_{\tau}(x, y, z) = \frac{1}{1 - \sum_{j=1}^{d} x^{a_j}y^j \sum_{i=1}^{j} x^{(\ell - 2)a_i} \tau^i}.$$  

For example, Theorem 2.4 for $z = 0$ and $A = \mathbb{N}$ gives that the generating function for the number of compositions of $n \in \mathbb{N}$ with $m$ parts in $\mathbb{N}$ which avoid the $\ell$-letter pattern $\tau = 211 \ldots 12$ is given by

$$C^{A}_{\tau}(n, y, 0) = \frac{1}{1 - \sum_{j \geq 1} x^{j}y^{j-1} \sum_{i=1}^{j} x^{(\ell - 2)a_i}}.$$  

In particular, for $\ell = 3$ we have that $C^{A}_{212}(x, 1) = \frac{1}{1 - \sum_{j \geq 1} (\frac{x^{j-1}(1-x)}{1-x^{j+1}})^2}$, and the sequence for the number of compositions of $n$ with parts in $\mathbb{N}$ that avoid 212 for $n = 0$ to $n = 20$ is given by 1, 1, 2, 4, 8, 15, 30, 54, 114, 222, 434, 846, 1655, 3238, 6310, 12322, 24067, 46997, 91791, 179262, 350106. Note that the first time the pattern 212 can occur is for $n = 5$, as the composition 212.

Similar arguments as in the proof of Theorem 2.4 replacing $a_d$ by $a_1$, lead us to the following result.

Theorem 2.5 Let $A$ be any ordered subset of $\mathbb{N}$ and let $\tau = 212 \ldots 221$ be a pattern of length $\ell$. Then

$$C^{A}_{\tau}(x, y, z) = \frac{1}{1 - \sum_{j=1}^{d} x^{a_j}y^j \sum_{i=1}^{j} x^{(\ell - 2)a_i} \tau^i}.$$
Counting \(\ell\)-letter subwords in compositions

For example, Theorem 2.5 for \(z = 0\) and \(A = \mathbb{N}\) gives that the generating function for the number of compositions of \(n\) with \(m\) parts in \(\mathbb{N}\) which avoid the \(\ell\)-letter pattern \(\tau = 12 \ldots 21\) is given by

\[
C^N_\tau(x, y) = C^N_\tau(x, y, 0) = \frac{1}{1 - \sum_{j \geq 1} \frac{x^{i+j}}{1 + xjy^{i+j}}}. 
\]

In particular, for \(\ell = 3\) we have that \(C^{121}_\tau(x, 1) = \frac{1}{1 - \sum_{j \geq 1} \frac{x^{j+1}}{1 + xjy^{j+1}}}\), and the sequence for the number of compositions of \(n\) with parts in \(\mathbb{N}\) that avoid 121 for \(n = 0\) to \(n = 20\) is given by 1, 1, 2, 4, 7, 13, 24, 44, 82, 153, 284, 528, 981, 1820, 3378, 6270, 11638, 21608, 40121, 74494, 138317. Note that the first time the pattern 121 can occur is for \(n = 4\), as the composition 121. We remark that in [8] the two patterns 212 and 121 were not treated individually, but as part of a peak and valley.

2.4 General patterns

In this section we consider several general cases of patterns. We start by the following result which generalizes Theorem 2.4.

**Theorem 2.6** Let \(A\) be any ordered subset of \(\mathbb{N}\), and let \(\tau = pr \ell p\), where \(\tau'\) is a pattern with \(\ell - 2\) letters in \(\{1, 2, \ldots, p-1\}\). Then

\[
C^A_\tau(x, y, z) = \frac{1}{1 - \sum_{j=1}^{d} x^{n_j} y^{l_j-1} (1-z) \sum_{\{a_1, \ldots, a_{l_j-2}\} \in B_j} x^{a_1} y^{a_2} \cdots y^{a_{l_j-2}}} 
\]

where \(B_j\) is the set of all the compositions \(\beta = a_1 \ldots a_{l_j-2}\) with parts in \(\{a_1, \ldots, a_j\}\) such that \(\beta\) is order-isomorphic to \(\tau'\).

**Proof:** Let \(\pi \in C^A_n(m)\). The generating function for the number of compositions \(\pi\) which do not contain the letter \(a_d\) and contain \(\tau\) exactly \(r\) times is given by \(C^A_\tau(x, y, z)\) where \(A' = \{a_1, \ldots, a_{d-1}\}\). Now assume that \(\pi = \pi(1)a_d\pi(2)\) such that \(\pi(1)\) does not contain the part \(a_d\). Then the number occurrences of \(\pi\) in \(\pi\) satisfies \(\tau(\pi) = \tau(\pi(1)) + \tau(a_d\pi(2))\), so the generating function for the number of such compositions \(\pi\) is given by \(x^{a_d} y C^A_\tau(x, y, z) D^A_\tau(x, y, z)\), where \(D^A_\tau(x, y, z)\) is the generating function for the number of compositions \(\pi \in C^A_n(m)\) such that \(a_d\pi\) contains \(\tau\) exactly \(r\) times. Therefore,

\[
C^A_\tau(x, y, z) = C^A_\tau(x, y, z) + x^{a_d} y C^A_\tau(x, y, z) D^A_\tau(x, y, z). 
\]

On the other hand, let \(a_d \pi \in C^A_n(m + a_d, m + 1)\). If \(\pi\) does not contain the letter \(a_d\), then the generating function for the number of such \(\pi\) is given by \(C^A_n(x, y, z)\). Otherwise, let \(i\) be the position of the longest letter \(a_d\) and let \(\pi_i\) be the left prefix of \(\pi\) of length \(i\), then the generating function for these compositions is given by

\[
x^{a_d} y \left( C^A_\tau(x, y, z) - y^{l-2} \sum_{\{a_1, \ldots, a_{l-2}\} \in B_d} x^{a_1} y^{a_2} \cdots y^{a_{l-2}} \right) D^A_\tau(x, y, z) 
\]

if \((a_d, \pi_i)\) is not order-isomorphic to \(\tau\), or by

\[
x^{a_d} y^{l-1} \sum_{\{a_1, \ldots, a_{l-2}\} \in B_d} x^{a_1} y^{a_2} \cdots y^{a_{l-2}} D_\tau(x, y, z) 
\]
if \((a_d, \pi_i)\) is order-isomorphic to \(\tau\). Therefore,

\[
D^A_{\tau}(x, y, z) = C^A_{\tau}(x, y, z) + x^{a_d}y \left(C^A_{\tau}(x, y, z) - y^{d-2}\sum_{(a_1, \ldots, a_{d-2}) \in B_d} x^{a_1 + \cdots + a_{d-2}} \right) D^A_{\tau}(x, y, z) + x^{a_d}y^{d-1}\sum_{(a_1, \ldots, a_{d-2}) \in B_d} x^{a_1 + \cdots + a_{d-2}} D_{\tau}(x, y, z).
\]

Hence, from the above two equations, we obtain

\[
C^A_{\tau}(x, y, z) = \frac{1 + (1 - z)x^{a_d}y^{d-1}\sum_{(a_1, \ldots, a_{d-2}) \in B_d} x^{a_1 + \cdots + a_{d-2}} C^A_{\tau}(x, y, z) - x^{a_d}yC^A_{\tau}(x, y, z)}{1 + (1 - z)x^{a_d}y^{d-1}\sum_{(a_1, \ldots, a_{d-2}) \in B_d} x^{a_1 + \cdots + a_{d-2}} - x^{a_d}yC^A_{\tau}(x, y, z)},
\]

so, by induction on \(d\) with the initial condition \(C^{(a_1, \ldots, a_{p-1})}_{\tau}(x, y, z) = \frac{1}{1 - y^{p-1}}\) we get the desired result.

Now, let \(\tau = 1\tau'1\) be a pattern, where \(\tau'\) is a pattern with parts in \(\{2, 3, \ldots\}\). Then similar arguments as in the proof of Theorem \ref{thm:order-isomorphic} with replacing \(a_d\) by \(a_1\) lead us to the following result.

**Theorem 2.7** Let \(A\) be any ordered subset of \(\mathbb{N}\), and let \(\tau = 1\tau'1\), where \(\tau'\) is a pattern with \(\ell - 2\) letters in \(\{2, 3, \ldots\}\). Then

\[
C^A_{\tau}(x, y, z) = \frac{1}{1 - \sum_{j=1}^{d} x^{a_j}y^{\ell-1}(1-z)\sum_{(a_1, \ldots, a_{\ell-2}) \in B_j} x^{a_1 + \cdots + a_{\ell-2}}},
\]

where \(B_j\) is the set of all the compositions \(\beta = a_1 \ldots a_{\ell-2}\) with parts in \(\{a_j, \ldots, a_d\}\) such that \(\beta\) is order-isomorphic to \(\tau'\).

Note that the above theorem generalizes Theorem \ref{thm:order-isomorphic}.

Let \(\tau = pr'(p + 1)\) be a pattern such that \(\tau'\) is a pattern with letters in \(\{1, 2, \ldots, p-1\}\). This case is treated in a similar manner as the case of \(\tau = pr'p\). As a result, we obtain the following theorem.

**Theorem 2.8** Let \(A\) be any ordered subset of \(\mathbb{N}\), and let \(\tau = pr'(p + 1)\), where \(\tau'\) is a pattern with \(\ell - 2\) letters in \(\{1, 2, \ldots, p-1\}\). Then

\[
C^A_{\tau}(x, y, z) = \frac{1}{1 - \sum_{i=1}^{d} x^{a_i}y^{p-1}(1 + x^{a_i}y^{\ell-1}(z - 1)\sum_{(a_1, \ldots, a_{\ell-2}) \in B_{j-1}} x^{a_1 + \cdots + a_{\ell-2}})},
\]

where \(B_{j-1}\) is the set of all the compositions \(\beta = a_1 \ldots a_{\ell-2}\) with parts in \(\{a_1, \ldots, a_{j-1}\}\) such that \(\beta\) is order-isomorphic to \(\tau'\).

For instance, Theorem \ref{thm:order-isomorphic} for \(z = 0, y = 1, A = N\) and \(\tau = 213\) gives

\[
C^A_{213}(x, 1) = \frac{1}{1 - \sum_{i \geq 1} (x^i \prod_{j=1}^{i-1} (1 - x^j(x - x^j))/(1 - x))},
\]
and the sequence for the number of compositions of \( n \) with parts in \( \mathbb{N} \) that avoid 213 for \( n = 0 \) to \( n = 20 \) is given by 1, 1, 2, 4, 8, 16, 31, 61, 119, 232, 452, 881, 1716, 3342, 6508, 12674, 24681, 48062, 93591, 182251, 354900. Note that the first time the pattern 213 can occur is for \( n = 6 \), as the composition 213.

Now, let \( \tau = 1 \tau'2 \) be a composition such that \( \tau' \) is a pattern with letters in \( \{3,4,\ldots\} \). This case is treated in a similar manner as the case of \( \tau = p \tau' p \). As a result, we obtain the following theorem.

**Theorem 2.9** Let \( A \) be any ordered subset of \( \mathbb{N} \), and let \( \tau = 1 \tau'2 \), where \( \tau' \) is a pattern with \( \ell - 2 \) letters in \( \{3,4,\ldots\} \). Then

\[
C_A^\ell(x, y, z) = \frac{1}{1 - \sum_{i=1}^d \left( x^{a_i} y \prod_{j=i+1}^d (1 + x^{a_j}) y^{\ell-1} (z - 1) \sum_{(a_{i_1}, \ldots, a_{i_{\ell-2}}) \in B_{j+1}} x^{a_{i_1} + \ldots + a_{i_{\ell-2}}} \right)},
\]

where \( B_{j+1} \) is the set of all the compositions \( \beta = a_{i_1} \ldots a_{i_{\ell-2}} \) with parts in \( \{a_{j+1}, \ldots, a_d\} \) such that \( \beta \) is order-isomorphic to \( \tau' \).

For instance, Theorem 2.9 for \( z = 0, y = 1, A = \mathbb{N} \) and \( \tau = 132 \) gives

\[
C_{132}^\ell(x, 1) = \frac{1}{1 - \sum_{i=1}^d \left( x^i \prod_{j=i+1}^d (1 - x^{j+1}/(1 - x)) \right)},
\]

and the sequence for the number of compositions of \( n \) with parts in \( \mathbb{N} \) that avoid 132 for \( n = 0 \) to \( n = 20 \) is given by 1, 1, 2, 4, 8, 16, 31, 61, 119, 232, 452, 880, 1712, 3331, 6479, 12601, 24505, 47653, 92664, 180187, 350372. Note that the first time the pattern 132 can occur is for \( n = 6 \), as the composition 132.

We remark that in [8] there two patterns 132 and 213 were not treated individually, but as part of a peak and valley.

### 3 Asymptotic distribution of the number of compositions that avoid \( \ell \)-letter pattern

We will now use methods from asymptotic analysis as described in [5] Theorem IX.8 (also, see [1]) to analyze different asymptotic parameters for the number of compositions of \( n \) with \( m \) parts in \( \mathbb{N} \) that avoid an \( \ell \)-letter pattern, for given \( \ell \geq 3 \). We look at the generating function as a complex function, and indicate this fact by using the variables \( z, w \) instead of the variables \( x, y \). Let \( \tau \) be any given pattern, the function \( C_{\tau}^\ell(z, w) \) is a bivariate analytic function at \((0, 0)\) and has nonnegative coefficients there. The asymptotic behavior of \( C_{\tau}^\ell(n) \), the number of compositions of \( n \) with parts in \( \mathbb{N} \) that avoid a pattern \( \tau \), is determined by the dominant pole of the function \( C_{\tau}^\ell(z, 1) = 1/h_{\tau}(z) \), i.e., the smallest positive \( z_{\tau}^* \) root of \( h_{\tau}(z) \) (see for example [13]). For instance,

\[
\begin{align*}
z_{111}^* &= 0.5233508903, & z_{111}^* &= 0.5005934689, & z_{111}^* &= 0.5043484367, & z_{112}^* &= 0.5534397072, \\
z_{1112}^* &= 0.5203736252, & z_{1112}^* &= 0.5089789475, & z_{221}^* &= 0.5133872872, & z_{221}^* &= 0.5024555452, \\
z_{2222}^* &= 0.5005424388, & z_{2221}^* &= 0.5120333765, & z_{212}^* &= 0.5056218556, & z_{2112}^* &= 0.5027148656, & z_{2112}^* &= 0.5027148656, \\
z_{12}^* &= 0.5386079645, & z_{122}^* &= 0.5057043128, & z_{1221}^* &= 0.5011623153, & z_{123}^* &= 0.5132858388, \\
z_{132}^* &= 0.5005555128, & z_{213}^* &= 0.5229016260
\end{align*}
\]

(3.1)

To be sure that \( z_{\tau}^* \), for fixed \( \tau \), is the dominant singularity, we use the argument principle [7]. Figures [1] and [2] present the curve \( h_{\tau}(z) \) with \( z = 0.6e^{it} \), where \( h_{\tau}(z) \) is converted into a series up to \( O(z^{30}) \). Thus, the winding number is 1 (this result can be achieved analytically, for example, Rouche’s theorem [7].
gives that the functions $h_{\tau}(z)$ and $z - z_{\tau}^*$ in the domain $|z| < 0.6$ have the same number of zeros, thus $h_{\tau}(z)$ has only one simple zero), so that the function $h_{\tau}(z)$ has only one root $z_{\tau}^*$ in the domain $|z| < 0.6$. Thus, the singularity analysis gives that

$$C_{\tau}(n) \sim -\frac{1}{z_{\tau}^*} \frac{d_{\tau}}{dz} h_{\tau}(z) |_{z = z_{\tau}^*} n = \frac{d_{\tau}}{(z_{\tau}^*)^n}, \quad n \to \infty,$$

(3.2)

where

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$d_{\tau}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>0.4993008432</td>
</tr>
<tr>
<td>112</td>
<td>0.6920054131</td>
</tr>
<tr>
<td>221</td>
<td>0.5453621593</td>
</tr>
<tr>
<td>212</td>
<td>0.5372311234</td>
</tr>
<tr>
<td>121</td>
<td>0.5838933009</td>
</tr>
<tr>
<td>123</td>
<td>0.5760960195</td>
</tr>
</tbody>
</table>

We are now ready for bivariate asymptotic by using Flajolet and Sedgewick’s book [5, Theorem IX.8]), which allows us to obtain results on the asymptotic distribution of the number of parts in the set of compositions (consider all compositions as equiprobable) of $n$ with $m$ parts in $\mathbb{N}$ that avoid an $\ell$-letter pattern. Take $|z| \leq 0.6$ and $|\tau - 1| \leq \epsilon_{\tau}$, where we assume that $\epsilon_{\tau}$ is a small positive real number.
Because of the form of the terms of the sum in denominator of $C_\tau(z, w) = \frac{1}{h_\tau(z, w)}$ that involve $wz^j$, the function $h_\tau(z, w)$ remain analytic there. Thus, there exists $\rho_\tau(w)$ a nonconstant analytic function for $w$ in a sufficiently small neighborhood of 1 (by Weierstrass preparation or implicit functions). The nondegeneracy conditions are easily verified by numerical computations.

For instance, if $\tau = 112$ then the function $h_{112}(z, w) = 1 - \sum_{j \geq 1} wz^j \prod_{i=1}^{j-1} (1 - w^2 z^{2i})$ satisfies $\partial_z h_{112}(z_{112}^*, 1) = -5.014638 \ldots \neq 0$ and $\partial_w h_{112}(z_{112}^*, 1) = -1.2393392 \ldots \neq 0$ (in our calculations $h_{112}(z, w)$ is converted into a series up to $O(z^{30})$), which gives the nondegeneracy condition for $h_{112}(z, w)$. Thus there exists a nonconstant analytic function $\rho(w) = \rho_{112}(w)$ at $w = 1$ with $h_{112}(\rho(w), w) = 0$ and $\rho(1) = z_{112}^* = 0.5534397072$. To check the variability condition in the case $\tau = 112$, let us find $\rho'(w) |_{w=1} = \partial_w \rho(w) |_{w=1}$. Differentiating the equation $h_{112}(\rho(w), w) = 0$ respect to $w$, we arrive at

$$\rho'(w) = -\frac{\sum_{j \geq 1} \rho^j(w) \left(1 - 2 \sum_{a=1}^{j-1} \frac{w^2 \rho^{2a}(w)}{1 - w^2 \rho^{2a}(w)} \right) \prod_{i=1}^{j-1} (1 - w^2 \rho^{2i}(w))}{\sum_{j \geq 1} w^j \rho^{j-1}(w) \left(1 - 2 \sum_{a=1}^{j-1} \frac{w^2 \rho^{2a-1}(w)}{1 - w^2 \rho^{2a}(w)} \right) \prod_{i=1}^{j-1} (1 - w^2 \rho^{2i}(w))}.$$  

Substituting $w = 1$ with using the fact that $\rho(1) = z_{112}^*$, we get that $\rho'(1) = -0.24714426 \ldots$. Hence, the variability condition in the case $\tau = 112$ holds. Therefore, there results that [5, theorem IX.8]
applies. Hence, the number of compositions with \( m \) parts of large given \( n \) that avoid the pattern 112 is asymptotically Gaussian. We note that the study of Gaussian asymptotic of the number compositions with \( m \) parts of large given \( n \) that avoid a pattern \( \tau \) is similar as the case of 112, where \( \tau \) is any pattern given in Section \ref{section:patterns}.

In order to obtain an explicit Gaussian asymptotic for each case of \( \tau \), let (see \cite[Section IX.5]{5} and \cite{11})
\[
 r_\tau = -\frac{\partial}{\partial z} h_\tau(z, w), \quad s_\tau = -\frac{r_\tau^2 \partial^2}{\partial z^2} h_\tau(z, w) + 2r_\tau \frac{\partial}{\partial w} h_\tau(z, w) + \frac{\partial}{\partial w} h_\tau(z, w) + \frac{\partial^2}{\partial w^2} h_\tau(z, w),
\]
then define \( \mu_\tau = -\frac{r_\tau}{s_\tau} \) and \( \sigma_\tau^2 = \mu_\tau^2 - \frac{s_\tau}{s_\tau} \). If we set \( w = 1 \) and \( z = z_\tau^* \), then we derive the values of the mean \( \mu_\tau = -\frac{r_\tau}{s_\tau} \) and the variance \( \sigma_\tau^2 = \mu_\tau^2 - \frac{s_\tau}{s_\tau} \) of the number of parts in a composition of large given \( n \) that avoids a pattern \( \tau \), where \( \tau \) is any pattern given in Section \ref{section:patterns}. For instance,
\[
\begin{array}{cccc}
\tau & r_\tau & s_\tau & \mu_\tau & \sigma_\tau^2 \\
111 & -0.2238894933 & 0.00018608451 & 0.4277999660 & 0.165403703 \\
1111 & -0.2384043207 & 0.01266095475 & 0.4608368622 & 0.1875254991 \\
11111 & -0.2410563274 & 0.01178048689 & 0.4779559324 & 0.2050840396 \\
112 & -0.2471442624 & -0.02641380791 & 0.4465604097 & 0.2471428219 \\
1112 & -0.2495849262 & -0.01016974670 & 0.4796236956 & 0.2495846426 \\
11112 & -0.24999193682 & -0.0044839272 & 0.4910201640 & 0.2499192110 \\
221 & -0.24999207882 & -0.00068873965 & 0.4866127277 & 0.24998205897 \\
2221 & -0.2499939748 & -0.00122768532 & 0.4975444638 & 0.2499938645 \\
22221 & -0.2499997094 & -0.00027116650 & 0.4994575685 & 0.2499996080 \\
212 & -0.2498552052 & -0.00061309945 & 0.4879666379 & 0.2498550909 \\
2112 & -0.2499683997 & -0.00281050345 & 0.4943781542 & 0.2499682680 \\
21112 & -0.249926336 & -0.00135733359 & 0.4972851425 & 0.2499251999 \\
121 & -0.2485109840 & -0.01917042125 & 0.4613949299 & 0.2484778134 \\
1221 & -0.2499674656 & -0.00285171573 & 0.4942956967 & 0.2499673229 \\
12221 & -0.2499865526 & -0.000581110024 & 0.4988376918 & 0.2499985479 \\
123 & -0.2491790299 & -0.01283794488 & 0.4854585926 & 0.2606813427 \\
132 & -0.2584885173 & 0.00490192258 & 0.5112956159 & 0.2517279554 \\
213 & -0.2626317573 & -0.01294199850 & 0.5030376743 & 0.2777839362.
\end{array}
\]
For the above numerical calculations we converted \( h_\tau(z, w) \) into a series up to \( O(z^{30}) \) and used the values of \( z_\tau^* \) which are given in \eqref{equation:zスター}. Therefore, by using \cite[theorem IX.8]{5} (see also \cite{11}) we can state the following result.

**Theorem 3.1** The number of parts \( m \) in a composition of large given \( n \) that avoids \( \tau \) is asymptotically Gaussian, i.e.,
\[
\frac{m - \mu_\tau n}{\sqrt{n} \sigma_\tau} \sim N(0, 1), \quad n \to \infty,
\]
(with mean \( \mu_\tau \) and variance \( \sigma_\tau \), and
\[
C_\tau^n(n, m) \sim \frac{d_\tau}{(z_\tau^*)^n \sqrt{2\pi n} \sigma_\tau} e^{-(m - n\mu_\tau)^2/(2n\sigma_\tau^2)}, \quad n \to \infty, \quad m - n\mu_\tau = O(\sqrt{n}),
\]
where \( \tau, d_\tau, z_\tau^*, \mu_\tau \) and \( \sigma_\tau^2 \) are given by \eqref{equation:パターン}, \eqref{equation:zスター} and \eqref{equation:mean-variance}.
Acknowledgements

The authors express their appreciation to Silvia Heubach for her careful reading of the manuscript. Also, the authors wish to thank the referees for their valuable comments and suggestions.

References


