Discrete Mathematics and Theoretical Computer Science DMTCS vol. **8**, 2006, 285–298

Counting ℓ-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 τ exactly r times, where $\tau \in \{111, 112, 221, 123\}$. Out 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 ℓ-letter subword, and (2) to analyze the number of compositions of n with m parts that avoid an ℓ-letter pattern, for given ℓ. 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 $\frac{1+x}{1-2x^3-x^4-x^2+x^5+x^6}$. 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 $\pi = \pi_1\pi_2\dots\pi_m$ be any composition of n with m parts in A and $\sigma = \sigma_1\sigma_2\dots\sigma_{m'}$ be any word of length m', where $m \geq m'$. An *occurrence* of σ in π is a *subword* $\pi_i\dots\pi_{i+m'-1}$ which is order-isomorphic to σ, i.e., $\pi_{i-1+a} < \pi_{i-1+b}$ if and only if $\sigma_a < \sigma_b$ for all $1 \leq a < b \leq m'$. We denote the number of occurrences of σ in π by $\sigma(\pi)$, and in such a context σ 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 $\pi = 11223113$ of 14 with 8 parts is 3, namely, 112, 223 and 113; thus $112(\pi) = 3$. If $r = 0$, we say that π *avoids* σ. Our aim is to count the number of compositions of n with m parts in A which contain an ℓ-letter pattern σ exactly r times, and to analyze the number of compositions of n with m parts in A which avoid an ℓ-letter pattern σ.

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

$$C_\tau^A(x,y) = \sum_{n\geq 0}\sum_{m\geq 0}\#C_\tau^A(n,m)x^ny^m, \quad C_\tau^A(x,y,z) = \sum_{r\geq 0}\sum_{n\geq 0}\sum_{m\geq 0}\#C_{\tau;r}^A(n,m)x^ny^mz^r.$$

Recently, Heubach and Mansour [8] found that

$$\begin{aligned}
C_{111}^A(x,y,z) &= \frac{1}{1-\sum_{j=1}^d \frac{x^{a_j}yz(1+(1-z)x^{a_j}y)}{1+x^{a_j}y(1+x^{a_j}y)(1-z)}}, & C_{112}^A(x,y,z) &= \frac{1}{1-\sum_{j=1}^d x^{a_j}y\prod_{i=1}^{j-1}(1-(1-z)x^{2a_i}y^2)},\\
C_{221}^A(x,y,z) &= \frac{1}{1-\sum_{j=1}^d x^{a_j}y\prod_{i=j+1}^{d}(1-(1-z)x^{2a_i}y^2)}, & C_{123}^A(x,y,z) &= \frac{1}{1-t^p(A)-\sum_{p=3}^d\sum_{j=0}^{p-3}\binom{p-3}{j}t^{p+j}(A)(z-1)^{p-2}},
\end{aligned} \tag{1.1}$$

where $t^p(A) = \sum_{1\leq i_1<\dots<i_p\leq d} y^p\prod_{j=1}^p x^{a_{i_j}}$. Our first goal is to generalize these results to obtain explicit formula for the generating function $C_\tau^A(x,y,z)$, where τ is an ℓ-letter pattern. In particular, we find explicit formulas for the generating functions $C_\tau^A(x,y,z)$, where $\tau \in \{212, 121, 132, 213\}$, see Section 2. We note that our generating function $C_\tau^A(x,y,z)$ for $x = 1$ and $A = \{1,2,\dots,k\}$ gives the generating function for the number k-ary words that avoid τ, 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 $\underbrace{11\dots1}_k$, for all $k \geq 2$.

Recently, several authors analyzed different asymptotic parameters of patterns in 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 polyominoes 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 ℓ-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 $C_\tau^A(x,y,z)$ for several interesting cases of ℓ-letter patterns τ. To do that we need the following notation. Denote the generating function for all compositions $\pi \in C_{\tau;r}^A(n,m)$ with $\pi_1\ldots\pi_s = \alpha_1\ldots\alpha_s$ by $C_\tau^A(\alpha_1\ldots\alpha_s|x,y,z)$.

2.1 *The pattern* $\tau = 11\ldots1$

Let $\tau = 11\ldots1$ be a pattern of length ℓ. By the definitions we have that

$$C_\tau^A(x,y,z) = 1 + \sum_{a\in A} C_\tau^A(a|x,y,z), \tag{2.1}$$

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

$$C_\tau^A(\underbrace{aa\ldots a}_{s}|x,y,z) = x^{as}y^s + \sum_{b\neq a\in A} C_\tau^A(\underbrace{aa\ldots a}_{s}b|x,y,z) + C_\tau^A(\underbrace{aa\ldots a}_{s+1}|x,y,z).$$

Observe that if $\pi \in C_{\tau;r}^A(n;m)$ 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 τ in π can involve one of the first s letters of π. Therefore, $\pi \in C_{\tau;r}^A(n;m)$ if and only if $(b,\pi_{s+2},\ldots,\pi_m) \in C_{\tau;r}^A(n-sa;m-s)$. Hence,

$$C_\tau^A(\underbrace{aa\ldots a}_{s}|x,y,z) = x^{as}y^s + x^{sa}y^s \sum_{b\neq a\in A} C_\tau^A(b|x,y,z) + C_\tau^A(\underbrace{aa\ldots a}_{s+1}|x,y,z),$$

equivalently, see (2.1),

$$C_\tau^A(\underbrace{aa\ldots a}_{s}|x,y,z) = x^{sa}y^s(C_\tau^A(x,y,z) - C_\tau^A(a|x,y,z)) + C_\tau^A(\underbrace{aa\ldots a}_{s+1}|x,y,z), \tag{2.2}$$

Moreover,

$$C_\tau^A(\underbrace{aa\ldots a}_{\ell}|x,y,z) = x^{a\ell}y^\ell z + \sum_{b\neq a\in A} C_\tau^A(\underbrace{aa\ldots a}_{\ell}b|x,y,z) + C_\tau^A(\underbrace{aa\ldots a}_{\ell+1}|x,y,z).$$

Observe that if $\pi \in C_{\tau;r}^A(n;m)$ 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 τ in π can involve one of the first ℓ letters of π. Therefore, $\pi \in C_{\tau;r}^A(n;m)$ if and only if $(b,\pi_{\ell+2},\ldots,\pi_m) \in C_{\tau;r-1}^A(n-\ell a;m-\ell)$. In the case of $b=a$ we get that $\pi \in C_{\tau;r}^A(n;m)$ if and only if $(\pi_2,\ldots,\pi_m) \in C_{\tau;r-1}^A(n-a;m-1)$. Hence,

$$C_\tau^A(\underbrace{aa\ldots a}_{\ell}|x,y,z) = x^{a\ell}y^\ell z + x^{a\ell}y^\ell z \sum_{b\neq a\in A} C_\tau^A(b|x,y,z) + x^a yz C_\tau^A(\underbrace{aa\ldots a}_{\ell}|x,y,z),$$

which is equivalent to, using (2.1),

$$C_\tau^A(\underbrace{aa\dots a}_{\ell}|x,y,z) = \frac{x^{a\ell}y^\ell z}{1-x^a yz}(C_\tau^A(x,y,z) - C_\tau^A(a|x,y,z)). \tag{2.3}$$

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

$$C_\tau^A(\underbrace{aa\dots a}_{s}|x,y,z) = \left(x^{as}y^s + \dots + x^{a(\ell-1)}y^{\ell-1} + \frac{x^{a\ell}y^\ell z}{1-x^a yz}\right)(C_\tau^A(x,y,z) - C_\tau^A(a|x,y,z)),$$

for all $s = 1, 2, \dots \ell$. Hence, putting $s = 1$ and using Equation 2.1 we get

$$C_\tau^A(x,y,z) = 1 + \sum_{a\in A} C_\tau^A(a|x,y,z),$$

with

$$C_\tau^A(a|x,y,z) = \frac{x^a y + x^{2a}y^2 + \cdots + x^{a(\ell-1)}y^{\ell-1} + \frac{x^{a\ell}y^\ell z}{1-x^a yz}}{1 + x^a y + x^{2a}y^2 + \cdots + x^{a(\ell-1)}y^{\ell-1} + \frac{x^{a\ell}y^\ell z}{1-x^a yz}} C_\tau^A(x,y,z),$$

which gives the following result.

Theorem 2.1 *Let A be any ordered subset of $\mathbb{N}$ and let $\tau = 11\dots1$ be a pattern of length ℓ. Then*

$$C_\tau^A(x,y,z) = \frac{1}{1 - \sum_{a\in A} \frac{x^a y + x^{2a}y^2 + \cdots + x^{a(\ell-1)}y^{\ell-1} + x^{a\ell}y^\ell z/(1-x^a yz)}{1 + x^a y + x^{2a}y^2 + \cdots + x^{a(\ell-1)}y^{\ell-1} + x^{a\ell}y^\ell z/(1-x^a yz)}}.$$

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 $\tau = 11\dots1$ (ℓ-letter pattern) is given by

$$C_\tau^{\mathbb{N}}(x,y) = C_\tau^{\mathbb{N}}(x,y,0) = \frac{1}{1 - \sum_{j\geq1} \frac{x^j y + x^{2j}y^2 + \cdots + x^{j(\ell-1)}y^{\ell-1}}{1 + x^j y + x^{2j}y^2 + \cdots + x^{j(\ell-1)}y^{\ell-1}}}.$$

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 $\tau = 11\dots12$

Let $\tau = 11\dots12$ be a pattern of length ℓ and let $A = \{a_1, \dots, a_d\}$ be any ordered subset of $\mathbb{N}$. Define $d_{\tau;r}^A(n;m)$ to be the number of compositions $\pi \in C_{\tau;r}^A(n;m)$ such that the string concatenation $\pi(a_d+1)$ ($a_d + 1 > a_d$) contains τ exactly r times, and denote the corresponding generating function by $D_\tau^A(x,y,z) = \sum_{n,m,r\geq0} d_{\tau;r}^A(n;m)x^n y^m z^r$. Let us find a recurrence for $C_\tau^A(x,y,z)$.

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

$$C_\tau^A(x,y,z) = C_\tau^{A'}(x,y,z) + x^{a_d}yD_\tau^{A'}(x,y,z)C_\tau^A(x,y,z),$$

or, equivalently,

$$C_\tau^A(x,y,z) = \frac{C_\tau^{A'}(x,y,z)}{1 - x^{a_d} y D_\tau^{A'}(x,y,z)}. \tag{2.4}$$

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

$$\tau(\pi^{(1)} a_d) + \cdots + \tau(\pi^{(s)} a_d) + \tau(\pi^{(s+1)}(a_d+1)) + \delta(\pi \text{ ends on } l-1\ a_d\text{'s}) = r,$$

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

$$\begin{aligned} D_\tau^A(x,y,z) &= \sum_{s=0}^{\ell-2} x^{a_d s} y^s (D_\tau^{A'}(x,y,z))^{s+1} \\ &+ \sum_{s=\ell-1}^{\infty} x^{a_d s} y^s \Big((D_\tau^{A'}(x,y,z))^{s+1} - (D_\tau^{A'}(x,y,z))^{(s+1)-(\ell-1)} + z (D_\tau^{A'}(x,y,z))^{(s+1)-(l-1)} \Big), \end{aligned}$$

which, after summing over s, yields

$$D_\tau^A(x,y,z) = \frac{(1 - x^{(\ell-1)a_d} y^{\ell-1}(1-z)) D_\tau^{A'}(x,y,z)}{1 - x^{a_d} y D_\tau^{A'}(x,y,z)}. \tag{2.5}$$

These two recurrences, together with $D_\tau^\varnothing(x,y,z) = C_\tau^\varnothing(x,y,z) = 1$ and induction, yield the following theorem.

Theorem 2.2 *Let A be any ordered subset of $\mathbb{N}$ and let $\tau = 11\dots12$ be a pattern of length ℓ. Then*

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

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 ℓ-letter pattern $\tau = 11\dots12$ is given by

$$C_\tau^{\mathbb{N}}(x,y) = C_\tau^{\mathbb{N}}(x,y,0) = \frac{1}{1 - \sum_{j\ge1} x^j y \prod_{i=1}^{j-1} (1 - x^{(\ell-1)i} y^{\ell-1})}.$$

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 $\tau = 22\dots21$ be a pattern of length ℓ. Then*

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

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 ℓ-letter pattern $\tau = 22\dots21$ is given by

$$C_\tau^{\mathbb{N}}(x,y) = C_\tau^{\mathbb{N}}(x,y,0) = \frac{1}{1 - \sum_{j\ge1} x^j y \prod_{i\ge j+1} (1 - x^{(\ell-1)i} y^{\ell-1})}.$$

In particular, if $\ell = 3$ then we get [8, Theorem 2.2], which is listed in (1.1).

2.3 The subword $\tau = 211\dots112$

Let $A = \{a_1, \dots, a_d\}$ be any ordered subset of $\mathbb{N}$, and let $\tau = 211\dots112$ be a pattern of length ℓ. We define $d^A_{\tau;r}(n;m)$ 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 τ exactly r times, and denote the corresponding generating function by $D^A_\tau(x,y,z)$. Let $\pi \in C^A_{\tau;r}(n;m)$ such that π contains s occurrences of the letter a_d. For $s = 0$, the generating function for the number of such compositions π is given by $C^{A'}_\tau(x,y,z)$ where $A' = \{a_1, \dots, a_{d-1}\}$, 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 $\pi = \pi^{(0)} a_d \pi^{(1)} a_d \dots a_d \pi^{(s-1)} a_d \pi^{(s)}$, where all $\pi^{(i)} \in C^{A'}(n_i; m_i)$, $\sum_{i=0}^s n_i = n - sa_d$, $\sum_{i=0}^s m_i = m - s$, and $\tau(\pi) = \tau(\pi^{(0)}) + \sum_{i=1}^{s-1} \tau(a_d \pi^{(1)} a_d) + \tau(\pi^{(s)})$. Hence, if we sum over all $s \geq 0$, we get

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

On the other hand, the string concatenation $(a_d+1)\pi'(a_d+1)$, with π' as above, contains an occurrence of τ involving the two letters a_d+1 if and only if π' 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 τ in the string concatenation $(a_d+1)\pi'(a_d+1)$ is the same as the number occurrences of τ in π'). Translating to generating functions, we obtain

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

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

Theorem 2.4 *Let A be any ordered subset of $\mathbb{N}$ and let $\tau = 211\dots112$ be a pattern of length ℓ. For all $\ell \geq 3$,*

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

For example, Theorem 2.4 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 ℓ-letter pattern $\tau = 21\dots12$ is given by

$$C^{\mathbb{N}}_\tau(x,y) = C^{\mathbb{N}}_\tau(x,y,0) = \frac{1}{1 - \sum_{j\geq1} \left(\frac{x^j y}{1 + x^j y^{\ell-1} \sum_{i=1}^{j-1} x^{(\ell-2)i}} \right)}.$$

In particular, for $\ell = 3$ we have that $C^{\mathbb{N}}_{212}(x,1) = \frac{1}{1 - \sum_{j\geq1} \left(\frac{x^j(1-x)}{1-x+x^j(x-x^j)} \right)}$, 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, 58, 114, 222, 434, 846, 1655, 3230, 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 $\tau = 122\dots221$ be a pattern of length ℓ. Then*

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

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 ℓ-letter pattern $\tau = 12\ldots21$ is given by

$$C_\tau^{\mathbb{N}}(x, y) = C_\tau^{\mathbb{N}}(x, y, 0) = \frac{1}{1 - \sum_{j\geq1}\left(\frac{x^j y}{1+x^j y^{\ell-1}\sum_{i\geq j+1} x^{(\ell-2)i}}\right)}.$$

In particular, for $\ell = 3$ we have that $C_{121}^{\mathbb{N}}(x, 1) = \frac{1}{1-\sum_{j\geq1}\left(\frac{x^j(1-x)}{1-x+x^{2j+1}}\right)}$, 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 $\tau = p\tau' p$, where τ' ia a pattern with $\ell - 2$ letters in $\{1, 2, \ldots, p-1\}$. Then*

$$C_\tau^A(x, y, z) = \frac{1}{1 - \sum_{j=1}^{d} \frac{x^{a_j} y}{1+x^{a_j} y^{\ell-1}(1-z)\sum_{(a_{i_1},\ldots,a_{i_{\ell-2}})\in B_j} x^{a_{i_1}+\cdots+a_{i_{\ell-2}}}}},$$

where B_j is the set of all the compositions $\beta = a_{i_1}\ldots a_{i_{\ell-2}}$ with parts in $\{a_1, \ldots, a_j\}$ such that β is order-isomorphic to τ'.

Proof: Let $\pi \in C^A(n; m)$. The generating function for the number of compositions π which do not contain the letter a_d and contain τ exactly r times is given by $C_\tau^{A'}(x, y, z)$ 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 τ in π satisfies $\tau(\pi) = \tau(\pi^{(1)}) + \tau(a_d\pi^{(2)})$, so the generating function for the number of such compositions π is given by $x^{a_d} y C_\tau^{A'}(x, y, z) D_\tau^A(x, y, z)$, where $D_\tau^A(x, y, z)$ is the generating function for the number of compositions $\pi \in C^A(n; m)$ such that $a_d\pi$ contains τ exactly r times. Therefore,

$$C_\tau^A(x, y, z) = C_\tau^{A'}(x, y, z) + x^{a_d} y C_\tau^{A'}(x, y, z) D_\tau^A(x, y, z).$$

On the other hand, let $a_d\pi \in C^A(n + a_d; m + 1)$. If π does not contain the letter a_d, then the generating function for the number of such π is given by $C_\tau^{A'}(x, y, z)$. Otherwise, let i be the position of the leftmost letter a_d and let $\pi|_i$ be the left prefix of π of length i, then the generating function for these compositions is given by

$$x^{a_d} y \left(C_\tau^{A'}(x, y, z) - y^{\ell-2} \sum_{(a_{i_1},\ldots,a_{i_{\ell-2}})\in B_d} x^{a_{i_1}+\cdots+a_{i_{\ell-2}}} \right) D_\tau^A(x, y, z)$$

if $(a_d, \pi|_i)$ is not order-isomorphic to τ, or by

$$x^{a_d} z y^{\ell-1} \sum_{(a_{i_1},\ldots,a_{i_{\ell-2}})\in B_d} x^{a_{i_1}+\cdots+a_{i_{\ell-2}}} D_\tau(x, y, z)$$

if $(a_d,\pi|_i)$ is order-isomorphic to τ. Therefore,

$$D_\tau^A(x,y,z)=C_\tau^{A'}(x,y,z)+x^{a_d}y\left(C_\tau^{A'}(x,y,z)-y^{\ell-2}\sum\nolimits_{(a_{i_1},\dots,a_{i_{\ell-2}})\in B_d}x^{a_{i_1}+\cdots+a_{i_{\ell-2}}}\right)D_\tau^A(x,y,z)$$
$$+x^{a_d}zy^{\ell-1}\sum\nolimits_{(a_{i_1},\dots,a_{i_{\ell-2}})\in B_d}x^{a_{i_1}+\cdots+a_{i_{\ell-2}}}D_\tau(x,y,z).$$

Hence, from the above two equations, we obtain

$$C_\tau^A(x,y,z)=\frac{\left(1+(1-z)x^{a_d}y^{\ell-1}\sum_{(a_{i_1},\dots,a_{i_{\ell-2}})\in B_d}x^{a_{i_1}+\cdots+a_{i_{\ell-2}}}\right)C_\tau^{A'}(x,y,z)}{1+(1-z)x^{a_d}y^{\ell-1}\sum_{(a_{i_1},\dots,a_{i_{\ell-2}})\in B_d}x^{a_{i_1}+\cdots+a_{i_{\ell-2}}}-x^{a_d}yC_\tau^{A'}(x,y,z)},$$

so, by induction on d with the initial condition $C_\tau^{\{a_1,\dots,a_{p-1}\}}(x,y,z)=\frac{1}{1-y\sum_{j=1}^{p-1}x^{a_j}}$ we get the desired result. □

Now, let $\tau=1\tau'1$ be a pattern, where τ' is a pattern with parts in $\{2,3,\dots\}$. Then similar arguments as in the proof of Theorem 2.6 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 τ' is a pattern with $\ell-2$ letters in $\{2,3,\dots\}$. Then*

$$C_\tau^A(x,y,z)=\frac{1}{1-\sum\limits_{j=1}^{d}\frac{x^{a_j}y}{1+x^{a_j}y^{\ell-1}(1-z)\sum_{(a_{i_1},\dots,a_{i_{\ell-2}})\in B_j}x^{a_{i_1}+\cdots+a_{i_{\ell-2}}}}},$$

where B_j is the set of all the compositions $\beta=a_{i_1}\dots a_{i_{\ell-2}}$ with parts in $\{a_j,\dots,a_d\}$ such that β is order-isomorphic to τ'.

Note that the above theorem generalizes Theorem 2.5.

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

$$C_\tau^A(x,y,z)=\frac{1}{1-\sum_{i=1}^{d}\left(x^{a_i}y\prod_{j=1}^{i-1}(1+x^{a_j}y^{\ell-1}(z-1)\sum_{(a_{i_1},\dots,a_{i_{\ell-2}})\in B_{j-1}}x^{a_{i_1}+\dots+a_{i_{\ell-2}}}\right)},$$

where B_{j-1} is the set of all the compositions $\beta=a_{i_1}\dots a_{i_{\ell-2}}$ with parts in $\{a_1,\dots,a_{j-1}\}$ such that β is order-isomorphic to τ'.

For instance, Theorem 2.8 for $z=0$, $y=1$, $A=\mathbb{N}$ and $\tau=213$ gives

$$C_{213}^{\mathbb{N}}(x,1)=\frac{1}{1-\sum_{i\geq1}\left(x^i\prod_{j=1}^{i-1}(1-x^j(x-x^j)/(1-x)\right)},$$

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 τ' 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 τ' is a pattern with $\ell - 2$ letters in $\{3, 4, \ldots\}$. Then*

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

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 β is order-isomorphic to τ'.

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

$$C_{132}^{\mathbb{N}}(x, 1) = \frac{1}{1 - \sum_{i \geq 1} \left(x^i \prod_{j \geq i+1} (1 - x^{2j+1}/(1-x) \right)},$$

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 ℓ-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 ℓ-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 τ be any given pattern, the function $C_\tau^{\mathbb{N}}(z, w)$ is a bivariate analytic function at $(0, 0)$ and has nonnegative coefficients there. The asymptotic behavior of $C_\tau^{\mathbb{N}}(n)$, the number of compositions of n with parts in $\mathbb{N}$ that avoid a pattern τ, is determined by the dominant pole of the function $C_\tau^{\mathbb{N}}(z, 1) = 1/h_\tau(z)$, i.e., the smallest positive z_τ^* root of $h_\tau(z)$ (see for example [13]). For instance,

$$\begin{array}{cccc} z_{111}^* = 0.5233508903, & z_{1111}^* = 0.5095953468, & z_{11111}^* = 0.5043484367, & z_{112}^* = 0.5534397072, \\ z_{1112}^* = 0.5203736252, & z_{11112}^* = 0.5089798475, & z_{221}^* = 0.5133872872, & z_{2221}^* = 0.5024555452, \\ z_{22221}^* = 0.5005424388, & z_{212}^* = 0.5120333765, & z_{2112}^* = 0.5056218556, & z_{21112}^* = 0.5027148656, \\ z_{121}^* = 0.5386079645, & z_{1221}^* = 0.5057043128, & z_{12221}^* = 0.5011623153, & z_{123}^* = 0.5132858388, \\ z_{132}^* = 0.5055555128, & z_{213}^* = 0.5220916260 & & \end{array} . \tag{3.1}$$

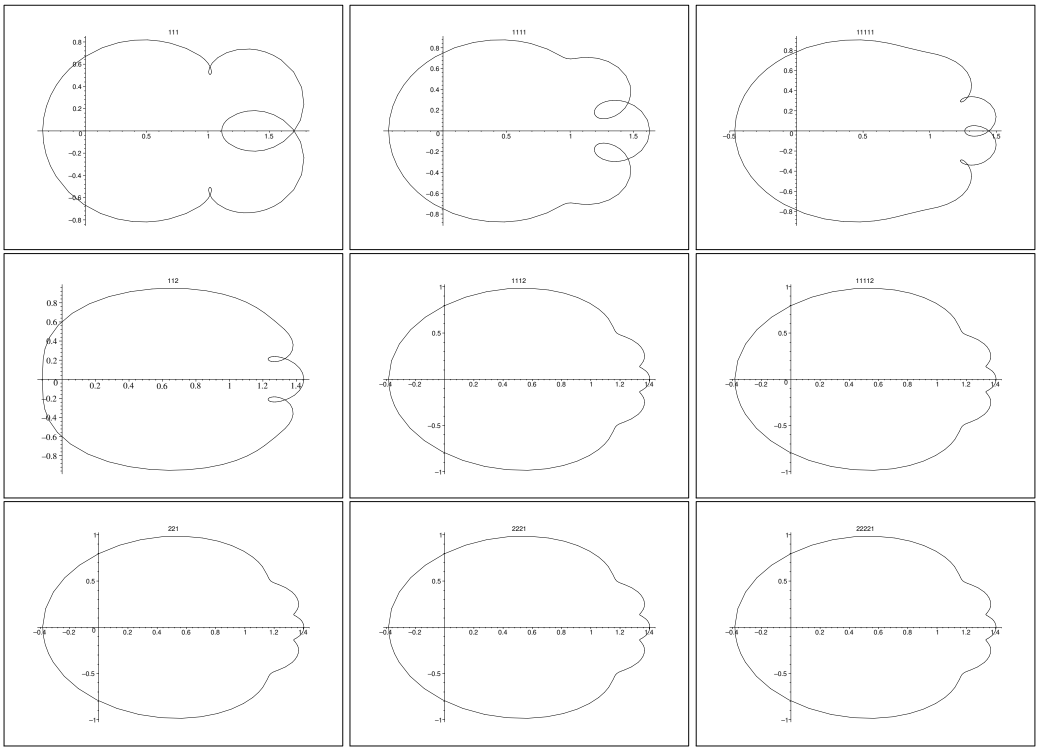

Fig. 1: Winding number for $h_\tau(z)$.

To be sure that z_τ^*, for fixed τ, 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_τ^* in the domain $|z| < 0.6$. Thus, the singularity analysis gives that

$$C_\tau(n) \sim -\frac{1}{z_\tau^* \frac{\partial}{\partial z} h_\tau(z) \mid_{z=z_\tau^*}} \frac{1}{(z_\tau^*)^n} = \frac{d_\tau}{(z_\tau^*)^n}, \quad n \to \infty, \tag{3.2}$$

where

τ	d_τ	τ	d_τ	τ	d_τ
111	0.4993008432	1111	0.5062126253	11111	0.5061594376
112	0.6920054131	1112	0.5798584679	11112	0.5411295078
221	0.5453621593	2221	0.5118239345	22221	0.5035135167
212	0.5372331234	2112	0.5222046486	21112	0.5131763007
121	0.5838933009	1221	0.5224624915	12221	0.5065230389
123	0.5760960195	132	0.5353097161	213	0.6121896742

. (3.3)

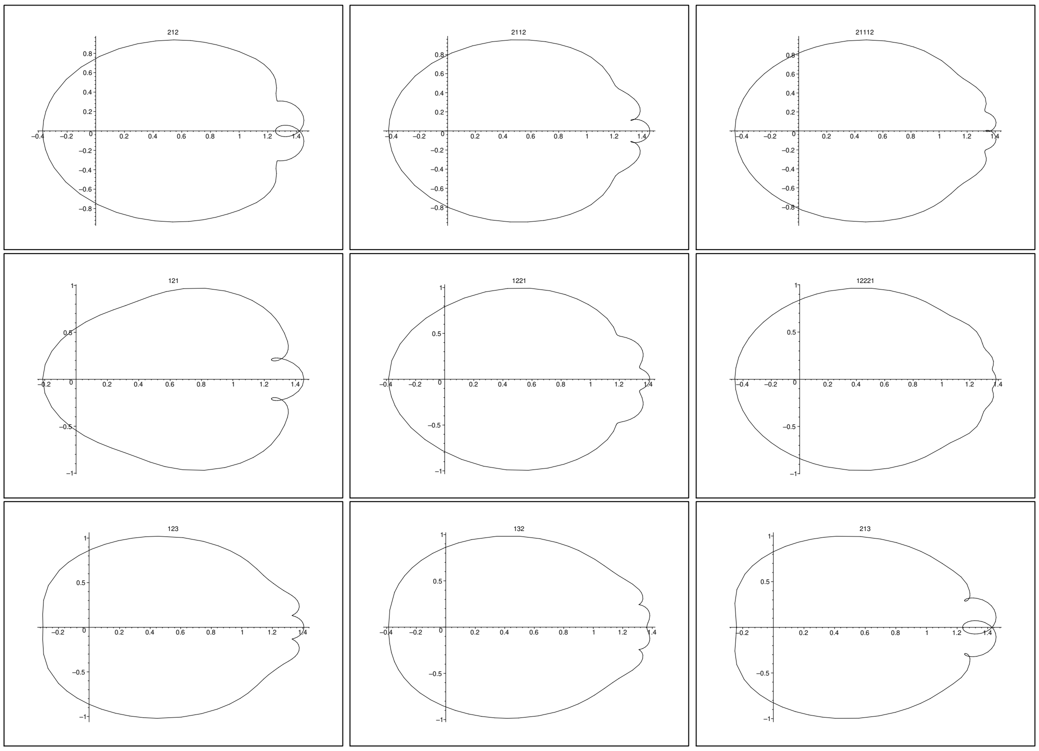

Fig. 2: Winding number for $h_\tau(z)$.

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 ℓ-letter pattern. Take $|z| \leq 0.6$ and $|w-1| \leq \epsilon_\tau$, where we assume that ϵ_τ 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 $h_{112}(z,w) = 1 - \sum_{j\geq 1} wz^j \prod_{i=1}^{j-1}(1-w^2z^{2i})$ 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 $\rho'(w)\mid_{w=1} = \partial_w \rho(w)\mid_{w=1}$. Differentiating the equation $h_{112}(\rho(w), w) = 0$ respect to w, we arrive at

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

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 τ is similar as the case of 112, where τ is any pattern given in Section 2.

In order to obtain an explicit Gaussian asymptotic for each case of τ, let (see [5, Section IX.5] and [1])

$$r_\tau = -\frac{\frac{\partial}{\partial w}h_\tau(z,w)}{\frac{\partial}{\partial z}h_\tau(z,w)},\ s_\tau = -\frac{r_\tau^2\frac{\partial^2}{\partial z^2}h_\tau(z,w)+2r_\tau\frac{\partial}{\partial z}\frac{\partial}{\partial w}h_\tau(z,w)+\frac{\partial}{\partial w}h_\tau(z,w)+\frac{\partial^2}{\partial w^2}h_\tau(z,w)}{\frac{\partial}{\partial z}h_\tau(z,w)},$$

then define $\mu_\tau = -\frac{r_\tau}{z^*_\tau}$ and $\sigma^2_\tau = \mu^2_\tau - \frac{s_\tau}{z^*_\tau}$. If we set $w = 1$ and $z = z^*_\tau$, then we derive the values of the mean $\mu_\tau = -\frac{r_\tau}{z^*_\tau}$ and the variance $\sigma^2_\tau = \mu^2_\tau - \frac{s_\tau}{z^*_\tau}$ of the number of parts in a composition of large given n that avoids a pattern τ, where τ is any pattern given in Section 2. For instance,

τ	r_τ	s_τ	μ_τ	σ^2_τ
111	−0.2238894933	0.00918608541	0.4277999660	0.1654603703
1111	−0.2348403207	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	0.4796263956	0.2495846426
11112	−0.2499193682	−0.00448839272	0.4910201640	0.2499192110
221	−0.2498207882	−0.00668873965	0.4866127277	0.2498205897
2221	−0.2499939748	−0.00122768532	0.4975444638	0.2499938645
22221	−0.2499997094	−0.00027116650	0.4994575685	0.2499996080
212	−0.2498552052	−0.00601309945	0.4879666379	0.2498550090
2112	−0.2499683997	−0.00281050345	0.4943781542	0.2499682680
21112	−0.2499926336	−0.00135733359	0.4972851425	0.2499925199
121	−0.2485109840	−0.01917042125	0.4613949299	0.2484778134
1221	−0.2499674656	−0.00285171573	0.4942956967	0.2499673329
12221	−0.2499986526	−0.00058110024	0.4988376918	0.2499985479
123	−0.2491790209	−0.01283794488	0.4854585926	0.2606813427
132	−0.2584883173	0.00490192258	0.5112956159	0.2517270954
213	−0.2626317573	−0.01291499850	0.5030376743	0.2777839362.

(3.4)

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^*_τ which are given in (3.1). Therefore, by using [5, theorem IX.8] (see also [1]) we can state the following result.

Theorem 3.1 *The number of parts m in a composition of large given n that avoids τ is asymptotically Gaussian, i.e.,*

$$\frac{m-\mu_\tau n}{\sqrt{n}\sigma_\tau} \sim \mathcal{N}(0,1), \quad n\to\infty,$$

(with mean μ_τ and variance σ_τ), and

$$C_\tau^{\mathbb{N}}(n,m) \sim \frac{d_\tau}{(z_\tau^*)^n \cdot \sqrt{2\pi n} \cdot \sigma_\tau} e^{-(m-n\mu_\tau)^2/(2n\sigma_\tau^2)}, \quad n \to \infty, \quad m - n\mu_\tau = O(\sqrt{n}),$$

where τ, d_τ, z_τ^*, μ_τ and σ_τ^2 are given by (3.1), (3.3) and (3.4).

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

[1] E.A. Bender, Central and local limit theorems applied to asymptotic enumeration, *J. Combina. Theory, Ser. A* **15** (1973) 91-111.

[2] A. Burstein and T. Mansour, Counting occurrences of some subword patterns, *Discr. Math. Theor. Comp. Sci.* **6:1** (2003), 1–12.

[3] L. Carlitz, Restricted compositions, *The Fib. Quart.* **14** (1976) 254–264.

[4] P. Flajolet, X. Gourdon, and C. Martinez, Patterns in random binary search trees, *Rand. Struct. Alg.* **11** (1997) 223–244.

[5] P. Flajolet and R. Sedgewick, Analytic Combinatorics, online book draft available at http://algo.inria.fr/ flajolet/Publications/books.html. Accessed 16 August 2006.

[6] I.P. Goulden and D.M. Jackson, Combinatorial enumeration, John Wiley and Sons, New York, NY, 1983.

[7] P. Henrici, Applied and Computational Complex Analysis, Wiley, 1988.

[8] S. Heubach and T. Mansour, Enumeration of 3-letter patterns in compositions, *Integers Elect. J. Combin. Number Theory*, to appear.

[9] H.K. Hwang, On convergence rates in the central limit theorems for combinatorial structures, *Europ. J. Combin.* **19** (1998) 329–343.

[10] A. Knopfmacher and H. Prodinger, On Carlitz compositions, *Europ. J. Combin.* **19** (1998) 579–589.

[11] G. Louchard and H. Prodinger, Probabilistic analysis of Carlitz Compositions, *Discr. Math. Theor. Comp. Sci.* **5** (2002) 71-96.

[12] R. Stanley, Enumerative Combinatrics, vol. 1, Cambridge University Press, 1997.

[13] H. S. Wilf, Generatingfunctionology, 2nd Edition, Academic Press, Inc., San Diego, 1994.