

# Algebraic aspects of some Riordan arrays related to binary words avoiding a pattern

D. Merlini and R. Sprugnoli  
Dipartimento di Sistemi e Informatica  
viale Morgagni 65, 50134, Firenze, Italia  
*merlini@dsi.unifi.it, renzo.sprugnoli@unifi.it*

July 26, 2010

## Abstract

We consider some Riordan arrays related to binary words avoiding a pattern  $\mathfrak{p}$ , which can be easily studied by means of an *A-matrix* rather than their *A-sequence*. Both concepts allow us to define every element as a linear combination of other elements in the array; the *A-sequence* is unique and corresponds to a linear dependence from the previous row. The *A-matrix* is not unique and corresponds to a linear dependence from several previous rows. However, for the problems considered in the present paper, we show that the *A-matrix* approach is more convenient. We provide explicit algebraic generating functions for these Riordan arrays and obtain many statistics on the corresponding languages. We thus obtain a deeper insight of the languages  $\mathfrak{L}^{[\mathfrak{p}]}$  of binary words avoiding  $\mathfrak{p}$  having a number of zeroes less or equal to the number of ones.

## 1 Introduction

The problem of determining the appearance of a certain *pattern* in long sequences of observations is interesting in many scientific situations (see the motivations in [6] for a general discussion). For example in the study of a genomic sequence it could be interesting to detect the occurrences of a particular pattern over the alphabet  $\{A, G, C, T\}$  (see, e.g., [13, 18]) and in computer network security, certain attacks can be detected by some well defined sequences of events (see, e.g., [1, 10]). These kinds of applications require to determine which patterns occur with high probability and which are very unlikely to arise, and this probabilistic phenomena often can be reduced to an enumerative problem over the corresponding language.

The notion of a pattern can be formalized in several ways and in this paper we consider *factor patterns*, that is, patterns whose letters must appear in an exact order and contiguously in the sequence under observation. There are two interesting categories of problems related to patterns. The first one consists in determining the probability that a random word contains (or excludes) a given pattern: this problem can be formulated as an enumeration problem and consists in counting the words in which the pattern occurs independently of the number of occurrences. The second one consists in determining the number of occurrences of a pattern in a random text: this problem involves the enumeration of the words according to the number

of occurrences of the pattern. For what concerns the first category of problems, if we fix an arbitrary pattern  $\mathbf{p} = p_0 \dots p_{h-1}$  of length  $h$  and consider the language  $\mathfrak{L}^{[\mathbf{p}]}$  of words containing at least one occurrence of  $\mathbf{p}$  as a factor, then by using automata theory we can prove that  $\mathfrak{L}^{[\mathbf{p}]}$  is a regular language. In fact, there exists a deterministic finite state automaton with  $h + 1$  states that recognizes  $\mathfrak{L}^{[\mathbf{p}]}$ ; in particular, at each stage, the states memorize the largest prefix of the pattern  $\mathbf{p}$  just recognized. The corresponding automaton is known as the Knuth-Morris-Pratt automaton (see, e.g., [9]). In the late 1950's, Chomsky and Schützenberger [3] proved that the generating function of the language of all the words accepted by a deterministic finite state automaton is a rational function and proved that this function can be determined as the solution of a linear system of equations which reflect the automaton transitions (see also [5, 14]). In particular, this result provides a generating function in determinant form and the relation between this rational form and the structure of the pattern is not evident. An explicit construction due to Guibas and Odlyzko [8] nicely circumvents this problem and it based on the notion of an *autocorrelation vector*. For a given pattern  $\mathbf{p}$ , this vector of bits  $c = (c_0, \dots, c_{h-1})$  can be defined in terms of Iverson's bracket notation (for a predicate  $P$ , the expression  $\llbracket P \rrbracket$  has value 1 if  $P$  is true and 0 otherwise) as follows:

$$c_i = \llbracket p_0 p_1 \dots p_{h-1-i} = p_i p_{i+1} \dots p_{h-1} \rrbracket.$$

In other words, the bit  $c_i$  is determined by shifting  $\mathbf{p}$  right by  $i$  positions and setting  $c_i = 1$  if and only if the remaining letters match the original. For example, when  $\mathbf{p} = 110011$  the autocorrelation vector is  $c = (1, 0, 0, 0, 1, 1)$ , as illustrated in Table 2.1. The polynomial  $C(z) = \sum_{j=0}^{h-1} c_j z^j$  is called the *autocorrelation polynomial*. The result of Guibas and Odlyzko states that the generating function of the words not containing the pattern  $\mathbf{p} = p_0 \dots p_{h-1}$  as a factor is:

$$S(z) = \frac{C(z)}{z^h + (1 - mz)C(z)}$$

where  $m$  is the alphabet cardinality. Moreover, the generating function of the words containing at least once the pattern is:

$$L(z) = \frac{1}{1 - mz} - S(z) = \frac{C(z)}{(1 - mz)(z^h + (1 - mz)C(z))}.$$

More precisely, the coefficients  $[z^n]S(z)$  and  $[z^n]L(z)$  count the number of words of length  $n$  not containing the pattern  $\mathbf{p}$  or containing the pattern at least once, respectively. For what concerns the second category of problems, when the alphabet has cardinality  $m = 2$ , Flajolet, Kirschenhofer and Tichy [4] have determined the following bivariate generating function

$$F(z, u) = \frac{1 - (C(z) - 1)(u - 1)}{1 - 2z - (u - 1)(z^h + (1 - 2z)(C(z) - 1))}$$

where the coefficient  $[z^n u^k]F(z, u)$  counts the number of words of length  $n$  over  $\mathfrak{L}^{[\mathbf{p}]}$  having  $k$  occurrences of the pattern  $\mathbf{p}$ ; as before,  $h$  is the length of the pattern and  $C(z)$  is the autocorrelation polynomial.

Recently, Baccherini, Merlini and Sprugnoli [2] studied the relation between binary words excluding a pattern and proper Riordan arrays. In particular, they proved necessary and sufficient conditions under which the number of words, counted with respect to the number of zeroes

and ones, is related to proper Riordan arrays. This approach is related to the first category of problems mentioned before and, as we will discuss later, allows us to find many properties of the corresponding languages. Moreover, the problem is interesting in the context of the Riordan array theory because the matrices which arise are naturally defined by recurrence relations following the characterization given in Merlini, Rogers, Sprugnoli and Verri [12]. Some history is necessary at this point.

The concept of a *Riordan array* was introduced in 1991 by Shapiro, Getu, Woan and Woodson [15], with the aim of defining a class of infinite lower triangular arrays with properties analogous to those of the Pascal triangle. This concept was successively studied by Sprugnoli [16] in the context of the computation of combinatorial sums. In these papers, Riordan arrays correspond to matrices  $D = (d_{n,k})_{n,k \in \mathbb{N}}$  where each element  $d_{n,k}$  is described by a linear combination of the elements in the previous row, starting from the previous column. The coefficients of this linear combination are independent of  $n$  and  $k$  and constitute a specific sequence called the  $A$ -sequence of the Riordan array. Later, several new characterizations of Riordan arrays were given in [12]: the main result in that paper shows that a lower triangular array  $(d_{n,k})_{n,k \in \mathbb{N}}$  is Riordan whenever its generic element  $d_{n+1,k+1}$  linearly depends on the elements  $d_{r,s}$  lying in a well-defined, but large zone of the array. The coefficients of this dependence constitute the so-called  $A$ -matrix and are illustrated in Figure 1.1. There is no difference between Riordan arrays defined in either ways: the  $A$ -sequence is a particular case of  $A$ -matrix and, given a Riordan array defined by an  $A$ -matrix, this corresponds to a well defined  $A$ -sequence.

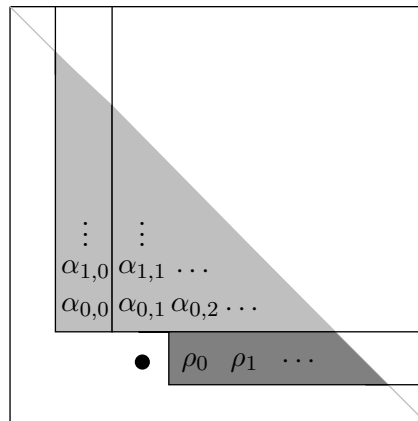


Figure 1.1: The zones which  $d_{n+1,k+1}$  can depend on.

However, there are some examples in which a Riordan array can be easily studied by means of the  $A$ -matrix while the  $A$ -sequence is very complex. This is just the case for the problem studied in [2]. From a combinatorial point of view, this means that it is very challenging to find a construction allowing to obtain objects of size  $n + 1$  from objects of size  $n$ . Instead, the existence of a simple  $A$ -matrix corresponds to a possible construction from objects of different sizes less than  $n + 1$ .

The aim of this paper is to re-consider from an algebraic point of view some cases of the problem studied in [2]. More precisely, in this paper we study the languages  $\mathfrak{L}^{[p]} \subset \{0, 1\}^*$  of binary words  $w$  avoiding a given pattern  $\mathfrak{p}$  such that  $|w_0| \leq |w_1|$  for any  $w \in \mathfrak{L}^{[p]}$ , where  $|w_0|$  and

$|w_1|$  correspond to the number of zeroes and ones in the word  $w$ , respectively. We prove new results in Theorem 2.3 and Corollaries 2.5, 2.7 and 2.8. They show that, when the pattern has a particular shape (is a *Riordan pattern* according to Definition 2.2), then the array  $R_{n,k}^{[\mathfrak{p}]}$ , counting the number of words avoiding  $\mathfrak{p}$  and having  $n$  bits one and  $n - k$  bits zero, corresponds to a Riordan array which can be *explicitly* defined in terms of its generating functions. Definition 2.2 is equivalent to the conditions on  $\mathfrak{p}$  given in [2, Theorem 5.1 (c)], however it makes explicit the shape of the autocorrelation polynomial of the pattern. This allows us to find the generating functions defining the Riordan array simply in terms of the autocorrelation polynomial. If  $\mathfrak{p}$  is a Riordan pattern, the conjugate pattern  $\bar{\mathfrak{p}}$  also is Riordan and therefore matrix  $R_{n,k}^{[\bar{\mathfrak{p}}]}$  also can be explicitly defined in terms of its generating functions.

These results improve Theorems 6.1 and 6.2 in [2], which give only *implicit* equations for the generating functions defining the Riordan arrays  $R_{n,k}^{[\mathfrak{p}]}$  and  $R_{n,k}^{[\bar{\mathfrak{p}}]}$ . These facts allow us to give a deeper insight into the corresponding languages; in particular, we prove that Riordan patterns correspond to a simple  $A$ -matrix (see Corollary 2.4 and Figure 2.2), while the corresponding  $A$ -sequence is (in general) very complicated. In other words,  $R_{n,k}^{[\mathfrak{p}]}$  satisfies a quite simple recurrence relation whose coefficients are determined by the autocorrelation vector of  $\mathfrak{p}$  (see Corollary 2.4 and Figure 2.2). These results are proved in Section 2. In Section 3 we illustrate some applications of the theory studied in the first part of the paper. In particular, we study in detail the Riordan patterns which correspond to recurrence relations in which  $R_{n+1,k+1}^{[\mathfrak{p}]}$  linearly depends on rows  $n + 1, n$  and  $n - 1$ . The results of this taxonomy are illustrated in Tables 3.9, 3.10, 3.11 and 3.12. In Theorem 3.3 we also examine some families of Riordan patterns defined in terms of a parameter  $j$ . Each of these families has a common autocorrelation polynomial and the resulting generating functions can be determined explicitly as functions of the parameter  $j$ .

An important property of Riordan arrays concerns the computation of combinatorial sums. In fact, as we will illustrate in the next section, every combinatorial sum  $\sum_{k=0}^n d_{n,k} f_k$  involving a Riordan array  $D = (d_{n,k})_{n,k \in \mathbb{N}}$  can be computed by extracting the coefficient from a particular univariate power series (see Formula (2.6) below). As particular cases, one can compute the row sums  $\sum_{k=0}^n d_{n,k}$  and the weighted row sums  $\sum_{k=0}^n k d_{n,k}$  but the formula is very general and can be applied to any sequence  $f_k$  provided the corresponding generating function is known. In Section 3.1 we show some examples of application of Formula (2.6). In particular, for the pattern  $\mathfrak{p} = 101$ , we find the average number of zero bits among the words of length  $n$  in  $\mathfrak{L}^{[\mathfrak{p}]}$ . As a second example, we consider the pattern  $\mathfrak{p} = 11100$ , and find the number of words of length  $n$  in  $\mathfrak{L}^{[\mathfrak{p}]}$  such that each bit zero can assume two different configurations. Each word in  $\mathfrak{L}^{[\mathfrak{p}]}$  can be naturally represented as a path starting at the origin in the integer lattice by associating a rise step to the bit 1 and a fall step to the bit 0. Therefore, we can give a nice combinatorial interpretation to the previous statistic, imagining to have the fall step of two different colours. This kind of statistic can be computed for any Riordan pattern  $\mathfrak{p}$ , assuming more generally to have  $\gamma > 0$  colours for the rise step. Many other statistics on the languages  $\mathfrak{L}^{[\mathfrak{p}]}$  can be found by using Formula (2.6) and the results proved in this paper.

In conclusion, the result examined in this paper are not only interesting in the theory of Riordan arrays but can also be used to find many enumerative properties characterizing the languages  $\mathfrak{L}^{[\mathfrak{p}]}$  avoiding a Riordan pattern. We wish also to point out that, as for the original problem treated by Guibas and Odlyzko [8], the structure of the pattern is transparent in our results, since all the generating functions defining the Riordan arrays are based on the

1	1	0	0	1	1	Tails			
1	1	0	0	1	1				1
	1	1	0	0	1	1			0
		1	1	0	0	1	1		0
			1	1	0	0	1	1	0
				1	1	0	0	1	1
					1	1	0	0	1
						1	0	0	1
							1	1	1

Table 2.1: The autocorrelation vector for  $\mathbf{p} = 110011$ .

autocorrelation polynomial associated to the pattern.

## 2 Riordan arrays for Riordan patterns

In this paper we are interested in studying binary words excluding a given pattern  $\mathbf{p} = p_0 \dots p_{h-1} \in \{0, 1\}^h$ . In particular, if  $F_{n,k}^{[\mathbf{p}]}$  denotes the number of words excluding the pattern and having  $n$  bits 1 and  $k$  bits 0, then by using the results in Baccherini, Merlini, Sprugnoli [2] we have the following rational bivariate generating function:

$$F^{[\mathbf{p}]}(x, y) = \sum_{n,k \geq 0} F_{n,k}^{[\mathbf{p}]} x^n y^k = \frac{C^{[\mathbf{p}]}(x, y)}{(1-x-y)C^{[\mathbf{p}]}(x, y) + x^{n_1^{\mathbf{p}}} y^{n_0^{\mathbf{p}}}}, \quad (2.1)$$

where  $n_1^{[\mathbf{p}]}$  and  $n_0^{[\mathbf{p}]}$  correspond to the number of ones and zeroes in the pattern and  $C^{[\mathbf{p}]}(x, y)$  is the autocorrelation polynomial with coefficients given by the autocorrelation vector (see also [7, 8, 14]). For example, when  $\mathbf{p} = 110011$  the autocorrelation vector is  $c = (1, 0, 0, 0, 1, 1)$ , as illustrated in Table 2.1, and  $C^{[\mathbf{p}]}(x, y) = 1 + x^2 y^2 + x^3 y^2$ , that is, we mark with  $x^j y^i$  the tails of the pattern with  $j$  bits 1,  $i$  bits 0 and  $c_{j+i} = 1$ . Therefore, in this case we have:

$$F^{[\mathbf{p}]}(x, y) = \frac{1 + x^2 y^2 + x^3 y^2}{(1-x-y)(1 + x^2 y^2 + x^3 y^2) + x^4 y^2}.$$

$n/k$	0	1	2	3	4	5	6	7
0	1	1	1	1	1	1	1	1
1	1	2	3	4	5	6	7	8
2	1	3	6	10	15	21	28	36
3	1	4	10	20	35	56	84	120
4	1	5	14	33	67	122	205	324
5	1	6	19	50	114	232	432	750
6	1	7	25	72	181	404	822	1552
7	1	8	32	100	273	660	1451	2952

Table 2.2: The matrix  $\mathcal{F}^{[\mathbf{p}]}$  for  $\mathbf{p} = 110011$

$n/k$	0	1	2	3	4	5	6	7
0	1							
1	2	1						
2	6	3	1					
3	20	10	4	1				
4	67	33	14	5	1			
5	232	114	50	19	6	1		
6	822	404	181	72	25	7	1	
7	2952	1451	660	272	100	32	8	1

Table 2.3: The triangle  $\mathcal{R}^{[\mathbf{p}]}$  for  $\mathbf{p} = 110011$

$n/k$	0	1	2	3	4	5	6	7
0	1							
1	2	1						
2	6	3	1					
3	20	10	4	1				
4	67	35	15	5	1			
5	232	122	56	21	6	1		
6	822	432	205	84	28	7	1	
7	2952	1552	750	324	120	36	8	1

Table 2.4: The triangle  $\mathcal{R}^{[\bar{\mathbf{p}}]}$  for  $\bar{\mathbf{p}} = 001100$

In order to study the binary words avoiding a pattern in terms of Riordan arrays, we consider the array  $\mathcal{R}^{[\mathbf{p}]} = (R_{n,k}^{[\mathbf{p}]})$  given by the lower triangular part of the array  $\mathcal{F}^{[\mathbf{p}]} = (F_{n,k}^{[\mathbf{p}]})$ , that is,  $R_{n,k}^{[\mathbf{p}]} = F_{n,n-k}^{[\mathbf{p}]}$  with  $k \leq n$ . More precisely,  $R_{n,k}^{[\mathbf{p}]}$  counts the number of words avoiding  $\mathbf{p}$  and having length  $2n - k$ ,  $n$  bits one and  $n - k$  bits zero. Given a pattern  $\mathbf{p} = p_0 \dots p_{h-1} \in \{0, 1\}^h$ , let  $\bar{\mathbf{p}} = \bar{p}_0 \dots \bar{p}_{h-1}$  be the conjugate pattern with  $\bar{p}_i = 1 - p_i, \forall i = 0, \dots, h - 1$ . We obviously have  $R_{n,k}^{[\bar{\mathbf{p}}]} = F_{n,n-k}^{[\bar{\mathbf{p}}]} = F_{n-k,n}^{[\mathbf{p}]}$ , therefore, the matrices  $\mathcal{R}^{[\mathbf{p}]}$  and  $\mathcal{R}^{[\bar{\mathbf{p}}]}$  represent the lower and upper triangular part of the array  $\mathcal{F}^{[\mathbf{p}]}$ , respectively. Moreover, we have  $R_{n,0}^{[\mathbf{p}]} = R_{n,0}^{[\bar{\mathbf{p}}]} = F_{n,n}^{[\mathbf{p}]}$ ,  $\forall n \in \mathbb{N}$ , that is, columns zero of  $\mathcal{R}^{[\mathbf{p}]}$  and  $\mathcal{R}^{[\bar{\mathbf{p}}]}$  correspond to the main diagonal of  $\mathcal{F}^{[\mathbf{p}]}$ . Tables 2.2, 2.3 and 2.4 illustrate some rows for the matrices  $\mathcal{F}^{[\mathbf{p}]}$ ,  $\mathcal{R}^{[\mathbf{p}]}$  and  $\mathcal{R}^{[\bar{\mathbf{p}}]}$  when  $\mathbf{p} = 110011$ .

We briefly recall that a Riordan array is an infinite lower triangular array  $(d_{n,k})_{n,k \in \mathbb{N}}$ , defined by a pair of formal power series  $(d(t), h(t))$ , such that  $d(0) \neq 0, h(0) = 0, h'(0) \neq 0$  and the generic element  $d_{n,k}$  is the  $n$ -th coefficient in the series  $d(t)h(t)^k$ , i.e.:

$$d_{n,k} = [t^n]d(t)h(t)^k, \quad n, k \geq 0.$$

From this definition we have  $d_{n,k} = 0$  for  $k > n$ . An alternative definition is in terms of the so-called  $A$ -sequence and  $Z$ -sequence, with generating functions  $A(t)$  and  $Z(t)$  satisfying the relations:

$$h(t) = tA(h(t)), \quad d(t) = \frac{d_0}{1 - tZ(h(t))} \quad \text{with} \quad d_0 = d(0).$$

Another characterization states that a lower triangular array  $(d_{n,k})_{n,k \in \mathbb{N}}$  is Riordan if and only if there exists an array  $(\alpha_{i,j})_{i,j \in \mathbb{N}}$ , with  $\alpha_{0,0} \neq 0$ , and a sequence  $(\rho_j)_{j \in \mathbb{N}}$  such that:

$$d_{n+1,k+1} = \sum_{i \geq 0} \sum_{j \geq 0} \alpha_{i,j} d_{n-i,k+j} + \sum_{j \geq 0} \rho_j d_{n+1,k+j+2}. \quad (2.2)$$

Figure 1.1 gives a graphical representation of this kind of dependence. Matrix  $(\alpha_{i,j})_{i,j \in \mathbb{N}}$  is called the  $A$ -matrix of the Riordan array. If  $P^{[0]}(t), P^{[1]}(t), P^{[2]}(t), \dots$  denote the generating functions of rows  $0, 1, 2, \dots$  in the  $A$ -matrix, i.e.:

$$P^{[i]}(t) = \alpha_{i,0} + \alpha_{i,1}t + \alpha_{i,2}t^2 + \alpha_{i,3}t^3 + \dots$$

and  $Q(t)$  is the generating function for the sequence  $(\rho_j)_{j \in \mathbb{N}}$ , then we have:

$$\frac{h(t)}{t} = \sum_{i \geq 0} t^i P^{[i]}(h(t)) + \frac{h(t)^2}{t} Q(h(t)), \quad (2.3)$$

$$A(t) = \sum_{i \geq 0} t^i A(t)^{-i} P^{[i]}(t) + tA(t)Q(t). \quad (2.4)$$

The generic element  $d_{n+1,k+1}$  often only depends on the two previous rows and sometimes on the elements of its own row. In this case, the functional equation (2.4) reduces to a second degree equation in  $A(t)$  and, as a result, we give an explicit expression for the generating function of the  $A$ -sequence.

**Theorem 2.1** *Let  $(d_{n,k})_{n,k \in \mathbb{N}}$  be a Riordan array whose generic element  $d_{n+1,k+1}$  only depends on the two previous rows and, possibly, on its own row. If  $P^{[0]}(t), P^{[1]}(t)$  and  $Q(t)$  are the generating functions for the coefficients of this dependence, then we have:*

$$A(t) = \frac{P^{[0]}(t) + \sqrt{P^{[0]}(t)^2 + 4tP^{[1]}(t)(1-tQ(t))}}{2(1-tQ(t))}. \quad (2.5)$$

Examples of applications of this theorem will be shown in the next section.

Another important property of Riordan array concerns the computation of combinatorial sums. In particular we have the following result (see Sprugnoli [16]):

$$\sum_{k=0}^n d_{n,k} f_k = [t^n] d(t) f(h(t)) \quad (2.6)$$

that is, every combinatorial sum involving a Riordan array can be computed by extracting the coefficient of  $t^n$  from the generating function  $d(t)f(h(t))$  where  $f(t)$  is the generating function of the sequence  $(f_k)_{k \in \mathbb{N}}$ . The theory of Riordan arrays and the proofs of their properties can be found in [11, 12].

Coming back to our original problem, Baccherini, Merlini, Sprugnoli in [2, Theorem 5.1] have proved necessary and sufficient conditions under which the matrices  $\mathcal{R}^{[p]}$  and  $\mathcal{R}^{[\bar{p}]}$  are Riordan arrays. In this paper, we are interested to examine in detail the case when both  $\mathcal{R}^{[p]}$  and  $\mathcal{R}^{[\bar{p}]}$  are Riordan arrays. To this purpose, we introduce the following definition:

**Definition 2.2 (Riordan pattern)** We say that  $\mathbf{p} = p_0 \dots p_{h-1}$  is a Riordan pattern if and only if

$$C^{[\mathbf{p}]}(x, y) = C^{[\mathbf{p}]}(y, x) = \sum_{i=0}^{\lfloor (h-1)/2 \rfloor} c_{2i} x^i y^i, \quad \text{and} \quad |n_1^{[\mathbf{p}]} - n_0^{[\mathbf{p}]}| \in \{0, 1\}.$$

As already observed in the Introduction, this definition is equivalent to the condition on  $\mathbf{p}$  given in [2, Theorem 5.1 (c)]. However it makes explicit the shape of the autocorrelation polynomial of the pattern and this allows us to find the generating functions defining the Riordan array just in terms of the autocorrelation polynomial of the pattern. We prove the following result, which improves the results in [2, Theorems 6.1 and 6.2]:

**Theorem 2.3** The matrices  $\mathcal{R}^{[\mathbf{p}]}$  and  $\mathcal{R}^{[\bar{\mathbf{p}}]}$  are both Riordan arrays  $\mathcal{R}^{[\mathbf{p}]} = (d^{[\mathbf{p}]}(t), h^{[\mathbf{p}]}(t))$  and  $\mathcal{R}^{[\bar{\mathbf{p}}]} = (d^{[\bar{\mathbf{p}}]}(t), h^{[\bar{\mathbf{p}}]}(t))$  if and only if  $\mathbf{p}$  is a Riordan pattern. Moreover we have:

$$d^{[\mathbf{p}]}(t) = d^{[\bar{\mathbf{p}}]}(t) = [x^0] F \left( x, \frac{t}{x} \right) = \frac{1}{2\pi i} \oint F \left( x, \frac{t}{x} \right) \frac{dx}{x}$$

and

$$h^{[\mathbf{p}]}(t) = \frac{1 - \sum_{i=0}^{n_1^{\mathbf{p}}-1} \alpha_{i,1} t^{i+1} - \sqrt{(1 - \sum_{i=0}^{n_1^{\mathbf{p}}-1} \alpha_{i,1} t^{i+1})^2 - 4 \sum_{i=0}^{n_1^{\mathbf{p}}-1} \alpha_{i,0} t^{i+1} (\sum_{i=0}^{n_1^{\mathbf{p}}-1} \alpha_{i,2} t^{i+1} + 1)}}{2(\sum_{i=0}^{n_1^{\mathbf{p}}-1} \alpha_{i,2} t^{i+1} + 1)}$$

where  $\delta_{i,j}$  is the Kronecker delta,

$$\begin{aligned} \sum_{i=0}^{n_1^{\mathbf{p}}-1} \alpha_{i,0} t^i &= \sum_{i=0}^{n_1^{\mathbf{p}}-1} c_{2i} t^i - \delta_{-1, n_0^{\mathbf{p}} - n_1^{\mathbf{p}}} t^{n_1^{\mathbf{p}}-1}, \\ \sum_{i=0}^{n_1^{\mathbf{p}}-1} \alpha_{i,1} t^i &= - \sum_{i=0}^{n_1^{\mathbf{p}}-1} c_{2(i+1)} t^i - \delta_{0, n_0^{\mathbf{p}} - n_1^{\mathbf{p}}} t^{n_1^{\mathbf{p}}-1}, \\ \sum_{i=0}^{n_1^{\mathbf{p}}-1} \alpha_{i,2} t^i &= \sum_{i=0}^{n_1^{\mathbf{p}}-1} c_{2(i+1)} t^i - \delta_{1, n_0^{\mathbf{p}} - n_1^{\mathbf{p}}} t^{n_1^{\mathbf{p}}-1}, \end{aligned}$$

and the coefficients  $c_i$  are given by the autocorrelation vector of  $\mathbf{p}$ . An analogous formula holds for  $h^{[\bar{\mathbf{p}}]}(t)$ .

**Proof:** The first part of the proof is analogous to that of [2, Theorem 3.1] and consists in extracting the coefficients

$$[x^{n+1} y^{k+1}] \left( (1-x-y) C^{[\mathbf{p}]}(x, y) + x^{n_1^{\mathbf{p}}} y^{n_0^{\mathbf{p}}} \right) F^{[\mathbf{p}]}(x, y) = [x^{n+1} y^{k+1}] C^{[\mathbf{p}]}(x, y)$$

and then putting  $R_{n,k}^{[\mathbf{p}]} = F_{n,n-k}^{[\mathbf{p}]} = [x^n y^{n-k}] F^{[\mathbf{p}]}(x, y)$ . We thus obtain:

$$\begin{aligned} R_{n+1,k+1}^{[\mathbf{p}]} &= R_{n,k}^{[\mathbf{p}]} + R_{n+1,k+2}^{[\mathbf{p}]} - R_{n+1-n_1^{\mathbf{p}}, k+1+n_0^{\mathbf{p}}-n_1^{\mathbf{p}}}^{[\mathbf{p}]} \\ &\quad - \sum_{i \geq 1} c_{2i} \left( R_{n+1-i, k+1}^{[\mathbf{p}]} - R_{n-i, k}^{[\mathbf{p}]} - R_{n+1-i, k+2}^{[\mathbf{p}]} \right). \end{aligned}$$

This recurrence relation follows the schema of formula (2.2) and the formula for  $h^{[p]}(t)$  can be found, after some computations, from equation (2.3) with  $P^{[i]}(t) = \alpha_{i,0} + \alpha_{i,1}t + \alpha_{i,2}t^2$  and  $Q(t) = 1$ . The coefficients  $\alpha_{i,0}, \alpha_{i,1}$  and  $\alpha_{i,2}$  with  $0 \leq i \leq n_1^p - 1$  are defined as in the statement of the theorem in terms of the autocorrelation vector of  $\mathbf{p}$ . For what concerns  $d^{[p]}(t)$ , we simply use the Cauchy formula for finding the main diagonal of matrix  $\mathcal{F}^{[p]}$  (see, e.g., [17, Chap. 6, p. 182]).  $\blacksquare$

For example, the pattern  $\mathbf{p} = 110011$  introduced at the beginning of Section 2 is not a Riordan pattern and, in particular, by following the classification of [2] we have that  $\mathcal{R}^{[\bar{\mathbf{p}}]}$  is a Riordan array while  $\mathcal{R}^{[\mathbf{p}]}$  is not.

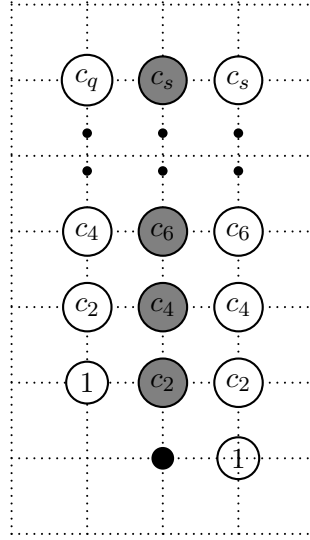


Figure 2.2: The  $A$ -matrix corresponding to a Riordan pattern; the coefficients in the gray circles are negative,  $s = 2n_1^p$ ,  $q = 2(n_1^p - 1)$ . Moreover, we have the contribution of  $-R_{n+1-n_1^p, k+1+n_0^p-n_1^p}^{[p]}$ .

As a direct consequence of the previous theorem we have:

**Corollary 2.4** *Let  $\mathbf{p}$  be a Riordan pattern. Then the Riordan array  $\mathcal{R}^{[\mathbf{p}]}$  is characterized by the  $A$ -matrix defined by the following relation:*

$$R_{n+1, k+1}^{[p]} = R_{n, k}^{[p]} + R_{n+1, k+2}^{[p]} - R_{n+1-n_1^p, k+1+n_0^p-n_1^p}^{[p]} - \sum_{i \geq 1} c_{2i} \left( R_{n+1-i, k+1}^{[p]} - R_{n-i, k}^{[p]} - R_{n+1-i, k+2}^{[p]} \right),$$

where the coefficients  $c_i$  are given by the autocorrelation vector of  $\mathbf{p}$ .

Figure 2.2, gives a graphical representation of the  $A$ -matrix: in particular,  $s = 2n_1^p$ ,  $q = 2(n_1^p - 1)$  and the coefficients in the gray circles have to be taken as negative while the coefficients in the white circles have to be taken as positive. Moreover we have to consider the contribution of  $-R_{n+1-n_1^p, k+1+n_0^p-n_1^p}^{[p]}$ .

By specializing Theorem 2.3 to the cases  $|n_1^p - n_0^p| \in \{0, 1\}$  and setting  $C^{[p]}(t) = C^{[p]}(\sqrt{t}, \sqrt{t}) = \sum_{i \geq 0} c_{2i} t^i$ , we have the following corollaries:

**Corollary 2.5** *Let  $\mathbf{p}$  be a Riordan pattern with  $n_1^{\mathbf{p}} - n_0^{\mathbf{p}} = 1$ . Then we have:*

$$d^{[\mathbf{p}]}(t) = \frac{C^{[\mathbf{p}]}(t)}{\sqrt{C^{[\mathbf{p}]}(t)^2 - 4tC^{[\mathbf{p}]}(t)(C^{[\mathbf{p}]}(t) - t^{n_0^{\mathbf{p}}})}},$$

$$h^{[\mathbf{p}]}(t) = \frac{C^{[\mathbf{p}]}(t) - \sqrt{C^{[\mathbf{p}]}(t)^2 - 4tC^{[\mathbf{p}]}(t)(C^{[\mathbf{p}]}(t) - t^{n_0^{\mathbf{p}}})}}{2C^{[\mathbf{p}]}(t)}.$$

**Proof:** From Theorem 2.3 we have

$$d^{[\mathbf{p}]}(t) = [x^0]F^{[\mathbf{p}]} \left( x, \frac{t}{x} \right) = \frac{1}{2\pi i} \oint F^{[\mathbf{p}]} \left( x, \frac{t}{x} \right) \frac{dx}{x}$$

and when  $n_1^{\mathbf{p}} - n_0^{\mathbf{p}} = 1$  we obtain:

$$\frac{1}{x}F^{[\mathbf{p}]} \left( x, \frac{t}{x} \right) = \frac{-C^{[\mathbf{p}]}(t)}{x^2(C^{[\mathbf{p}]}(t) - t^{n_0^{\mathbf{p}}}) - xC^{[\mathbf{p}]}(t) + tC^{[\mathbf{p}]}(t)}.$$

In order to compute the integral, it is necessary to find the singularities  $x(t)$  such that  $x(t) \rightarrow 0$  with  $t \rightarrow 0$  and apply the Residue theorem. We have two singularities ( $x_1(t)$  corresponds to the plus and  $x_2(t)$  to the minus sign):

$$x_{1,2}(t) = \frac{C^{[\mathbf{p}]}(t) \pm \sqrt{C^{[\mathbf{p}]}(t)^2 - 4tC^{[\mathbf{p}]}(t)(C^{[\mathbf{p}]}(t) - t^{n_0^{\mathbf{p}}})}}{2(C^{[\mathbf{p}]}(t) - t^{n_0^{\mathbf{p}}})}$$

and

$$\frac{1}{x}F^{[\mathbf{p}]} \left( x, \frac{t}{x} \right) = \frac{-C^{[\mathbf{p}]}(t)}{(C^{[\mathbf{p}]}(t) - t^{n_0^{\mathbf{p}}})(x - x_1(t))(x - x_2(t))}.$$

Therefore, we have:

$$d^{[\mathbf{p}]}(t) = \lim_{x \rightarrow x_2(t)} \frac{1}{x}F^{[\mathbf{p}]} \left( x, \frac{t}{x} \right) (x - x_2(t)) = \frac{C^{[\mathbf{p}]}(t)}{(C^{[\mathbf{p}]}(t) - t^{n_0^{\mathbf{p}}})(x_1(t) - x_2(t))}$$

and, after some simplification, we obtain the formula in the statement. The expression for  $h^{[\mathbf{p}]}(t)$  follows directly from Theorem 2.3, by specializing the Kronecker deltas. ■

For example, when  $\mathbf{p} = 11100$  we have the Riordan array defined by the following functions:

$$d^{[\mathbf{p}]}(t) = \frac{1}{\sqrt{1 - 4t + 4t^3}}, \quad h^{[\mathbf{p}]}(t) = \frac{1 - \sqrt{1 - 4t + 4t^3}}{2}$$

and illustrated in Table 2.5. As another example, the triangle in Table 2.6 corresponds to  $\mathbf{p} = 101$  and to the functions:

$$d^{[\mathbf{p}]}(t) = \frac{1+t}{\sqrt{1-2t-3t^2}}, \quad h^{[\mathbf{p}]}(t) = \frac{1+t - \sqrt{1-2t-3t^2}}{2(1+t)}.$$

We observe that in Tables 2.5 and 2.6 every element in column 0, except the first one, is twice the value in column 1. In fact we have the following result:

$n/k$	0	1	2	3	4	5	6	7
0	1							
1	2	1						
2	6	3	1					
3	18	9	4	1				
4	58	29	13	5	1			
5	192	96	44	18	6	1		
6	650	325	151	64	24	7	1	
7	2232	1116	524	288	90	31	8	1

Table 2.5: The triangle for  $\mathfrak{p} = 11100$

$n/k$	0	1	2	3	4	5	6	7
0	1							
1	2	1						
2	4	2	1					
3	10	5	2	1				
4	26	13	6	2	1			
5	70	35	16	7	2	1		
6	192	96	45	19	8	2	1	
7	534	267	126	56	22	9	2	1

Table 2.6: The triangle for  $\mathfrak{p} = 101$

**Theorem 2.6** *Let  $\mathfrak{p}$  be a Riordan pattern with  $n_1^{\mathfrak{p}} - n_0^{\mathfrak{p}} = 1$ . Then the Riordan array  $\mathcal{R}^{[\mathfrak{p}]}$  satisfies the following relation:*

$$R_{n+1,0}^{[\mathfrak{p}]} = 2R_{n+1,1}^{[\mathfrak{p}]}.$$

**Proof:** The relation can be found by observing that by Corollary 2.5 we have

$$\frac{d^{[\mathfrak{p}]}(t) - 1}{h^{[\mathfrak{p}]}(t)} = 2.$$

■

The proof of Corollaries 2.7 and 2.8 can be done similarly to Corollary 2.5.

**Corollary 2.7** *Let  $\mathfrak{p}$  be a Riordan pattern with  $n_1^{\mathfrak{p}} - n_0^{\mathfrak{p}} = 0$ . Then we have:*

$$d^{[\mathfrak{p}]}(t) = \frac{C^{[\mathfrak{p}]}(t)}{\sqrt{(C^{[\mathfrak{p}]}(t) + t^{n_0^{\mathfrak{p}}})^2 - 4tC^{[\mathfrak{p}]}(t)^2}},$$

$$h^{[\mathfrak{p}]}(t) = \frac{C^{[\mathfrak{p}]}(t) + t^{n_0^{\mathfrak{p}}} - \sqrt{(C^{[\mathfrak{p}]}(t) + t^{n_0^{\mathfrak{p}}})^2 - 4tC^{[\mathfrak{p}]}(t)^2}}{2C^{[\mathfrak{p}]}(t)}.$$

**Corollary 2.8** *Let  $\mathbf{p}$  be a Riordan pattern with  $n_0^{\mathbf{p}} - n_1^{\mathbf{p}} = 1$ . Then we have:*

$$d^{[\mathbf{p}]}(t) = \frac{C^{[\mathbf{p}]}(t)}{\sqrt{C^{[\mathbf{p}]}(t)^2 - 4tC^{[\mathbf{p}]}(t)(C^{[\mathbf{p}]}(t) - t^{n_1^{\mathbf{p}}})}},$$

$$h^{[\mathbf{p}]}(t) = \frac{C^{[\mathbf{p}]}(t) - \sqrt{C^{[\mathbf{p}]}(t)^2 - 4tC^{[\mathbf{p}]}(t)(C^{[\mathbf{p}]}(t) - t^{n_1^{\mathbf{p}}})}}{2(C^{[\mathbf{p}]}(t) - t^{n_1^{\mathbf{p}}})}.$$

$n/k$	0	1	2	3	4	5	6	7
0	1							
1	2	1						
2	6	3	1					
3	19	10	4	1				
4	65	33	15	5	1			
5	228	116	53	21	6	1		
6	814	414	193	80	28	7	1	
7	2947	1497	707	304	115	36	8	1

Table 2.7: The triangle for  $\mathbf{p} = 010101$

For the sake of completeness, the triangle in Table 2.7 corresponds to the pattern  $\mathbf{p} = 010101$  and to the functions:

$$d^{[\mathbf{p}]}(t) = \frac{1 + t + t^2}{\sqrt{1 - 2t - 5t^2 - 8t^3 - 5t^4 - 2t^5 + t^6}},$$

$$h^{[\mathbf{p}]}(t) = \frac{1 + t + t^2 + t^3 - \sqrt{1 - 2t - 5t^2 - 8t^3 - 5t^4 - 2t^5 + t^6}}{2(1 + t + t^2)}.$$

Finally, for  $\mathbf{p} = 00011$  we have the triangle illustrated in Table 2.8 and the functions:

$n/k$	0	1	2	3	4	5	6	7
0	1							
1	2	1						
2	6	3	1					
3	18	10	4	1				
4	58	32	15	5	1			
5	192	106	52	21	6	1		
6	650	357	180	79	28	7	1	
7	2232	1222	624	288	114	36	8	1

Table 2.8: The triangle for  $\mathbf{p} = 00011$

$$d^{[\mathbf{p}]}(t) = \frac{1}{\sqrt{1 - 4t + 4t^3}}, \quad h^{[\mathbf{p}]}(t) = \frac{1 - \sqrt{1 - 4t + 4t^3}}{2(1 - t^2)}.$$

We wish to conclude this section by observing that by Corollaries 2.5, 2.7 and 2.8 if  $\mathbf{p}$  is a Riordan pattern with  $n_1^{\mathbf{p}} - n_0^{\mathbf{p}} = 0$  than  $h^{[\mathbf{p}]}(t) = h^{[\bar{\mathbf{p}}]}(t)$  while

$$\frac{h^{[\mathbf{p}]}(t)}{h^{[\bar{\mathbf{p}}]}(t)} = \frac{C^{[\mathbf{p}]}(t) - t^{n_0^{\mathbf{p}}}}{C^{[\bar{\mathbf{p}}]}(t)} \quad \text{if } n_1^{\mathbf{p}} - n_0^{\mathbf{p}} = 1$$

and

$$\frac{h^{[\mathbf{p}]}(t)}{h^{[\bar{\mathbf{p}}]}(t)} = \frac{C^{[\mathbf{p}]}(t)}{C^{[\bar{\mathbf{p}}]}(t) - t^{n_1^{\mathbf{p}}}} \quad \text{if } n_0^{\mathbf{p}} - n_1^{\mathbf{p}} = 1.$$

### 3 Applications

Let us consider a first application of the theory presented in Section 2. As we pointed out in the Introduction, the Riordan arrays studied in this paper are characterized by a complex  $A$ -sequence. In fact, the generating function  $A(t)$  of this sequence can be found from the relation  $h(t) = tA(h(t))$ , and, due to Theorem 2.3, we can expect, in general, a complex solution. Fortunately, for Riordan patterns, we have always a simple  $A$ -matrix, as we proved in Corollary 2.4. For example, for the Riordan array in Table 2.8, corresponding to  $\mathbf{p} = 00011$ , we find:

$$A(t) = \frac{1 + \sqrt{1 + 4t^4 - 4t^3}}{2(1-t)} = 1 + t + t^2 + t^4 + t^5 + 2t^7 + t^8 - t^9 + 5t^{10} - t^{11} - 4t^{12} + \\ + 16t^{13} - 14t^{14} - 8t^{15} + 57t^{16} - 83t^{17} + 15t^{18} + 197t^{19} + O(t^{20}),$$

while for the  $A$ -matrix we have:  $R_{n+1,k+1}^{[\mathbf{p}]} = R_{n,k}^{[\mathbf{p}]} + R_{n+1,k+2}^{[\mathbf{p}]} - R_{n-1,k+2}^{[\mathbf{p}]}$ . This example corresponds to an  $A$ -matrix in which  $R_{n+1,k+1}^{[\mathbf{p}]}$  linearly depends on rows  $n+1, n$  and  $n-1$  (the same and the two preceding rows); due to Theorem 2.1, these cases can be treated with algebraic elementary tools. In the following theorems, we give a taxonomy for the Riordan pattern corresponding to dependence of these types. As shown in Figure 2.2, in order to have a dependence limited to the two preceding rows, we need to have  $c_{2i} = 0$  for all  $i \geq 3$  while  $c_4$  can be zero or one; in fact, if  $c_4 = 1$ , it may be eliminated from the position  $(n-2, k)$  by the term  $-R_{n+1-n_1^{\mathbf{p}}, k+1+n_0^{\mathbf{p}}-n_1^{\mathbf{p}}}^{[\mathbf{p}]}$ . On the other hand, when  $c_4 = 0$ , we have to take under control the same term  $-R_{n+1-n_1^{\mathbf{p}}, k+1+n_0^{\mathbf{p}}-n_1^{\mathbf{p}}}^{[\mathbf{p}]}$  by imposing that  $n-1 \leq n+1-n_1^{\mathbf{p}} \leq n+1$  and  $k \leq k+1+n_0^{\mathbf{p}} \leq k+2$ . These facts prove the following theorems:

**Theorem 3.1** *Let  $\mathbf{p}$  be a Riordan pattern with  $C^{[\mathbf{p}]}(t) = C^{[\mathbf{p}]}(\sqrt{t}, \sqrt{t}) = 1 + c_2t + t^2$ ; then, in order to have a dependence limited to the two preceding rows, the pattern needs to satisfy the conditions:*

$$n_1^{\mathbf{p}} = 3, \quad n_0^{\mathbf{p}} = 2.$$

*Therefore, we have exactly the following possible Riordan patterns:  $\mathbf{p} = 11001, 10101, 10011$ .*

**Theorem 3.2** *Let  $\mathbf{p}$  be a Riordan pattern with  $C^{[\mathbf{p}]}(t) = C^{[\mathbf{p}]}(\sqrt{t}, \sqrt{t}) = 1 + c_2t$ ; then, in order to have a dependence limited to the two preceding rows, the pattern needs to satisfy the conditions:*

$$n_1^{\mathbf{p}} \leq 2, \quad n_1^{\mathbf{p}} - 1 \leq n_0^{\mathbf{p}} \leq n_1^{\mathbf{p}} + 1.$$

*Therefore, we have exactly the following possible Riordan patterns:*

$\mathbf{p}$	$C^{[\mathbf{p}]}(t)$	$d^{[\mathbf{p}]}(t)$	$h^{[\mathbf{p}]}(t)$	$A^{[\mathbf{p}]}(t)$
110, 011	1	$\frac{1}{1-2t}$	$t$	1
100, 001	1	$\frac{1}{1-2t}$	$\frac{t}{1-t}$	$1+t$
010	$1+t$	$\frac{1+t}{\sqrt{1-2t-3t^2}}$	$\frac{1+t-\sqrt{1-2t-3t^2}}{2}$	$\frac{1-t+\sqrt{1+2t-3t^2}}{2(1-t)}$
101	$1+t$	$\frac{1+t}{\sqrt{1-2t-3t^2}}$	$\frac{1+t-\sqrt{1-2t-3t^2}}{2(1+t)}$	$\frac{1-t+t^2}{1-t}$

Table 3.9: The Riordan arrays for the patterns of length 3 in Theorem 3.1.

$\mathbf{p}$	1100, 0011	1010, 0101
$C^{[\mathbf{p}]}(t)$	1	$1+t$
$d^{[\mathbf{p}]}(t)$	$\frac{1}{\sqrt{1-4t+2t^2+t^4}}$	$\frac{1+t}{\sqrt{1-2t-5t^2-2t^3+t^4}}$
$h^{[\mathbf{p}]}(t)$	$\frac{1+t^2-\sqrt{1-4t+2t^2+t^4}}{2}$	$\frac{1+t+t^2-\sqrt{1-2t-5t^2-2t^3+t^4}}{2(1+t)}$
$A^{[\mathbf{p}]}(t)$	$\frac{1+\sqrt{1-4t^2+4t^3}}{2(1-t)}$	$\frac{1-t+t^2+\sqrt{1+2t-5t^2+2t^3+t^4}}{2(1-t)}$

Table 3.10: The Riordan arrays for the patterns of length 4 in Theorem 3.1.

- $n_1^{\mathbf{p}} = 0, n_0^{\mathbf{p}} = 0, 1 : \mathbf{p} = \epsilon, 0;$
- $n_1^{\mathbf{p}} = 1, n_0^{\mathbf{p}} = 0, 1, 2 : \mathbf{p} = 1, 01, 10, 001, 010, 100;$
- $n_1^{\mathbf{p}} = 2, n_0^{\mathbf{p}} = 1, 2 : \mathbf{p} = 110, 101, 011, 0011, 0101, 1010, 1100;$
- $n_1^{\mathbf{p}} = 2, n_0^{\mathbf{p}} = 3 : \mathbf{p} = 11000, 10100, 00101, 00011.$

The patterns in the previous taxonomy, with length less than or equal to 2, correspond to trivial Riordan arrays. In particular, when  $\mathbf{p} = 0$  we have  $d^{[\mathbf{p}]}(t) = 1$  and  $h^{[\mathbf{p}]}(t) = t$ ; when  $\mathbf{p} = 1$  we have  $d^{[\mathbf{p}]}(t) = 1$  and  $h^{[\mathbf{p}]}(t) = 0$ , and, finally, when  $\mathbf{p} = 01, 10$  we have  $d^{[\mathbf{p}]}(t) = 1/(1-t)$  and  $h^{[\mathbf{p}]}(t) = t$ . The remaining patterns are more interesting and the corresponding Riordan arrays are given in Tables 3.9, 3.10 and 3.11. In these tables, for each pattern, we give the autocorrelation polynomial  $C^{[\mathbf{p}]}(t)$  and the functions  $d^{[\mathbf{p}]}(t), h^{[\mathbf{p}]}(t)$  and  $A^{[\mathbf{p}]}(t)$ . These functions are computed by using the results of Section 2. Moreover, in Table 3.12 we give, for each of the previous patterns, the recurrence relation defining the  $A$ -matrix.

We observe that in Tables 3.9 and 3.10 we find both  $\mathbf{p}$  and  $\bar{\mathbf{p}}$ . This is not true for the patterns in Table 3.11; in fact, the conjugate patterns correspond to a dependence on rows  $n+1, n, n-1$  and  $n-2$ , as illustrated in Table 3.13. In these cases, the function  $A^{[\mathbf{p}]}(t)$  has a very complicated expression or cannot be found explicitly.

$\mathfrak{p}$	10100, 00101 11000, 00011	11001, 10011	10101
$C^{[\mathfrak{p}]}(t)$	1	$1 + t^2$	$1 + t + t^2$
$P^{[\mathfrak{p}]}(t)$	$\sqrt{1 - 4t + 4t^3}$	$\sqrt{1 - 4t + 2t^2 - 4t^3 + t^4}$	$\sqrt{1 - 2t - 5t^2 - 6t^3 - 3t^4}$
$d^{[\mathfrak{p}]}(t)$	$\frac{1}{P^{[\mathfrak{p}]}(t)}$	$\frac{1 + t^2}{P^{[\mathfrak{p}]}(t)}$	$\frac{1 + t + t^2}{P^{[\mathfrak{p}]}(t)}$
$h^{[\mathfrak{p}]}(t)$	$\frac{1 - P^{[\mathfrak{p}]}(t)}{2(1 - t^2)}$	$\frac{1 + t^2 - P^{[\mathfrak{p}]}(t)}{2(1 + t^2)}$	$\frac{1 + t + t^2 - P^{[\mathfrak{p}]}(t)}{2(1 + t + t^2)}$
$A^{[\mathfrak{p}]}(t)$	$\frac{1 + P^{[\mathfrak{p}]}(t)}{2(1 - t)}$	$\frac{1 + P^{[\mathfrak{p}]}(t)}{2(1 - t)}$	$\frac{1 - t + t^2 + P^{[\mathfrak{p}]}(t)}{2(1 - t)}$

Table 3.11: The Riordan arrays for the patterns of length 5 in Theorems 3.1 and Theorems 3.2.

$\mathfrak{p}$	$R_{n+1,k+1}^{[\mathfrak{p}]}$
110, 011	$R_{n+1,k+2}^{[\mathfrak{p}]} + R_{n,k}^{[\mathfrak{p}]} - R_{n-1,k}^{[\mathfrak{p}]}$
100, 001	$R_{n+1,k+2}^{[\mathfrak{p}]} + R_{n,k}^{[\mathfrak{p}]} - R_{n,k+2}^{[\mathfrak{p}]}$
010	$R_{n+1,k+2}^{[\mathfrak{p}]} + R_{n,k}^{[\mathfrak{p}]} - R_{n,k+1}^{[\mathfrak{p}]} + R_{n-1,k}^{[\mathfrak{p}]}$
101	$R_{n+1,k+2}^{[\mathfrak{p}]} + R_{n,k+2}^{[\mathfrak{p}]} - R_{n,k+1}^{[\mathfrak{p}]} + R_{n,k}^{[\mathfrak{p}]}$
1100, 0011	$R_{n+1,k+2}^{[\mathfrak{p}]} + R_{n,k}^{[\mathfrak{p}]} - R_{n-1,k+1}^{[\mathfrak{p}]}$
1010, 0101	$R_{n+1,k+2}^{[\mathfrak{p}]} + R_{n,k+2}^{[\mathfrak{p}]} - R_{n,k+1}^{[\mathfrak{p}]} + R_{n,k}^{[\mathfrak{p}]} - R_{n-1,k+1}^{[\mathfrak{p}]} + R_{n-1,k}^{[\mathfrak{p}]}$
10100, 00101 11000, 00011	$R_{n+1,k+2}^{[\mathfrak{p}]} + R_{n,k}^{[\mathfrak{p}]} - R_{n-1,k+2}^{[\mathfrak{p}]}$
11001, 10011	$R_{n+1,k+2}^{[\mathfrak{p}]} + R_{n,k+2}^{[\mathfrak{p}]} + R_{n-1,k+2}^{[\mathfrak{p}]} - R_{n-1,k+1}^{[\mathfrak{p}]}$
10101	$R_{n+1,k+2}^{[\mathfrak{p}]} + R_{n,k+2}^{[\mathfrak{p}]} - R_{n,k+1}^{[\mathfrak{p}]} + R_{n,k}^{[\mathfrak{p}]} + R_{n-1,k+2}^{[\mathfrak{p}]} - R_{n-1,k+1}^{[\mathfrak{p}]} + R_{n-1,k}^{[\mathfrak{p}]}$

Table 3.12: The A-matrices for the patterns of Theorems 3.1 and 3.2.

$\mathbf{p}$	01011, 11010 00111, 11100	00110, 01100	01010
$P^{[\mathbf{p}]}(t)$	$\sqrt{1-4t+4t^3}$	$\sqrt{1-4t+2t^2-4t^3+t^4}$	$\sqrt{1-2t-5t^2-6t^3-3t^4}$
$d^{[\mathbf{p}]}(t)$	$\frac{1}{P^{[\mathbf{p}]}(t)}$	$\frac{1+t^2}{P^{[\mathbf{p}]}(t)}$	$\frac{1+t+t^2}{P^{[\mathbf{p}]}(t)}$
$h^{[\mathbf{p}]}(t)$	$\frac{1-P^{[\mathbf{p}]}(t)}{2}$	$\frac{1+t^2-P^{[\mathbf{p}]}(t)}{2}$	$\frac{1+t+t^2-P^{[\mathbf{p}]}(t)}{2(1+t)}$
$R_{n+1,k+1}^{[\mathbf{p}]}$	$R_{n+1,k+2}^{[\mathbf{p}]} + R_{n,k}^{[\mathbf{p}]} - R_{n-2,k}^{[\mathbf{p}]}$	$R_{n+1,k+2}^{[\mathbf{p}]} + R_{n,k}^{[\mathbf{p}]} - R_{n-1,k+1}^{[\mathbf{p}]} + R_{n-2,k}^{[\mathbf{p}]}$	$R_{n+1,k+2}^{[\mathbf{p}]} + R_{n,k+2}^{[\mathbf{p}]} - R_{n,k+1}^{[\mathbf{p}]} + R_{n,k}^{[\mathbf{p}]} - R_{n-1,k+1}^{[\mathbf{p}]} + R_{n-1,k}^{[\mathbf{p}]} + R_{n-2,k}^{[\mathbf{p}]}$

Table 3.13: The Riordan arrays for the conjugate patterns of Table 3.11.

A second application of the general theory concerns some special classes of patterns. These patterns are characterized by an autocorrelation polynomial which can be easily computed; actually, by using the results of Section 2 we can prove the following results:

**Theorem 3.3** For  $\mathbf{p} = 1^{j+1}0^j$  we have the Riordan array:

$$d^{[\mathbf{p}]}(t) = \frac{1}{\sqrt{1-4t+4t^{j+1}}}, \quad h^{[\mathbf{p}]}(t) = \frac{1-\sqrt{1-4t+4t^{j+1}}}{2};$$

for  $\mathbf{p} = 0^{j+1}1^j$  we have the Riordan array:

$$d^{[\mathbf{p}]}(t) = \frac{1}{\sqrt{1-4t+4t^{j+1}}}, \quad h^{[\mathbf{p}]}(t) = \frac{1-\sqrt{1-4t+4t^{j+1}}}{2(1-t^j)};$$

for  $\mathbf{p} = 1^j0^j$  and  $\mathbf{p} = 0^j1^j$  we have the Riordan array:

$$d^{[\mathbf{p}]}(t) = \frac{1}{\sqrt{1-4t+2t^j+t^{2j}}}, \quad h^{[\mathbf{p}]}(t) = \frac{1+t^j-\sqrt{1-4t+2t^j+t^{2j}}}{2};$$

for  $\mathbf{p} = (10)^j1$  we have the Riordan array:

$$d^{[\mathbf{p}]}(t) = \frac{\sum_{i=0}^j t^i}{\sqrt{1-2\sum_{i=1}^j t^i-3\left(\sum_{i=1}^j t^i\right)^2}}, \quad h^{[\mathbf{p}]}(t) = \frac{\sum_{i=0}^j t^i - \sqrt{1-2\sum_{i=1}^j t^i-3\left(\sum_{i=1}^j t^i\right)^2}}{2\sum_{i=0}^j t^i};$$

for  $\mathbf{p} = (01)^j0$  we have the Riordan array:

$$d^{[\mathbf{p}]}(t) = \frac{\sum_{i=0}^j t^i}{\sqrt{1-2\sum_{i=1}^j t^i-3\left(\sum_{i=1}^j t^i\right)^2}}, \quad h^{[\mathbf{p}]}(t) = \frac{\sum_{i=0}^j t^i - \sqrt{1-2\sum_{i=1}^j t^i-3\left(\sum_{i=1}^j t^i\right)^2}}{2\sum_{i=0}^{j-1} t^i}.$$

### 3.1 Some statistics for the patterns 101 and 11100

We conclude this paper with a third application, by observing that by Formula (2.6) and Corollaries 2.5, 2.7 and 2.8 we can compute many statistics on the languages considered in the present paper. For example we have for  $f_k = 1$  and  $f_k = k$  :

$$S_n^{[p]} = \sum_{k=0}^n R_{n,k}^{[p]} = [t^n] \frac{d^{[p]}(t)}{1 - h^{[p]}(t)},$$

$$W_n^{[p]} = \sum_{k=0}^n k R_{n,k}^{[p]} = [t^n] \frac{d^{[p]}(t) h^{[p]}(t)}{(1 - h^{[p]}(t))^2}.$$

In some cases these coefficients can be extracted as exact formulas; in other cases, we can compute an asymptotic approximation. For example, we can study the case of the pattern  $\mathfrak{p} = 101$ . In particular, we are interested in finding the average number of zero bits in all the words avoiding  $\mathfrak{p}$ , having  $n$  bits one and  $n - k$  bits zero. This is obtained by computing the following quantity:

$$m_n^{[p]} = \frac{\sum_{k=0}^n (n - k) R_{n,k}^{[p]}}{\sum_{k=0}^n R_{n,k}^{[p]}} = n - \frac{W_n^{[p]}}{S_n^{[p]}}.$$

By applying the previous formulas we have:

$$S_n^{[p]} = [t^n] \frac{(1+t)(1-3t-\sqrt{1-2t-3t^2})}{2t(3t-1)},$$

$$W_n^{[p]} = [t^n] \frac{(1+t)(1-3t-(1-2t)\sqrt{1-2t-3t^2})}{2t^2(3t-1)}.$$

The singularity of minimal module of both functions is  $t = 1/3$ , therefore, by developing around this value, we obtain the following asymptotic expressions:

$$S_n^{[p]} = [t^n] \left( \frac{4}{3} \frac{\sqrt{3}}{\sqrt{1-3t}} - 2 + \frac{5\sqrt{3}\sqrt{1-3t}}{6} - \frac{3(1-3t)}{2} + O\left(\sqrt{(1-3t)^3}\right) \right) =$$

$$= \sqrt{3} \left( \frac{3}{4} \right)^{n-1} \binom{2n}{n} \left( 1 - \frac{5}{8(2n-1)} + O\left(\frac{1}{n^2}\right) \right),$$

$$W_n^{[p]} = [t^n] \left( \frac{4}{3} \frac{\sqrt{3}}{\sqrt{1-3t}} - 6 + \frac{29\sqrt{3}\sqrt{1-3t}}{6} - \frac{21(1-3t)}{2} + O\left(\sqrt{(1-3t)^3}\right) \right) =$$

$$= \sqrt{3} \left( \frac{3}{4} \right)^{n-1} \binom{2n}{n} \left( 1 - \frac{29}{8(2n-1)} + O\left(\frac{1}{n^2}\right) \right),$$

where we used the formulas:

$$[t^n](1-3t)^{-1/2} = \frac{(-1)^n}{4^n} \binom{2n}{n} (-3)^n, \quad [t^n](1-3t)^{1/2} = \frac{(-1)^{n-1}}{4^n(2n-1)} \binom{2n}{n} (-3)^n.$$

Finally, we obtain

$$m_n^{[p]} = n - \frac{1 - \frac{5}{8(2n-1)} + O\left(\frac{1}{n^2}\right)}{1 - \frac{29}{8(2n-1)} + O\left(\frac{1}{n^2}\right)} = n - 1 + \frac{3}{2n} + O\left(\frac{1}{n^2}\right).$$

For example, when  $n = 7$ , from Table 2.6 we obtain the exact value  $m_n^{[p]} = 6280/1017 = 6.175024582$  against an approximate value of 6.214285714, with a relative error of 0.64%.

The problems considered in this paper have a natural combinatorial interpretation in terms of lattice paths where a 1 corresponds to a rise step and a 0 to a fall step. These paths start from the origin and avoid the sub-path corresponding to the pattern. For example, let us consider the case  $\mathbf{p} = 11100$ , or, equivalently,  $\mathbf{p} = 01011, 11010, 00111$  (see Table 3.13). In the lattice path interpretation  $R_{n,k}^{[p]}$  counts the number of paths avoiding  $\mathbf{p}$ , having length  $2n - k$  and with  $n$  rise steps. Then, by using Formula (2.6) and the first case of Theorem 3.3 with  $j = 2$ , we can compute the following statistic:

$$\begin{aligned} \sum_{k=0}^n R_{n,k}^{[p]} 2^{n-k} &= 2^n [t^n] \frac{d^{[p]}(t)}{1 - \frac{1}{2}h^{[p]}(t)} = \\ &= 2^n [t^n] \frac{4}{\sqrt{1-4t+4t^3} \left(3 + \sqrt{1-4t+4t^3}\right)} = \\ &= 2^n [t^n] \left(1 + \frac{5}{2}t + \frac{31}{4}t^2 + \frac{189}{8}t^3 + \frac{1223}{16}t^4 + \frac{8117}{32}t^5 + O(t^6)\right) \end{aligned}$$

This statistic has a nice combinatorial interpretation: it counts the number of paths, avoiding the sub-path  $\mathbf{p} = 11100$ , of length between  $n$  and  $2n$ , having  $n$  rise steps and having the fall steps of two different colours. With the help of *Maple*, we can find an asymptotic approximation of the coefficients by observing that the polynomial  $1 - 4t + 4t^3$  has the following three real roots:

$$x_1 = \frac{2}{3}\sqrt{3}s \approx 0.83756, \quad x_2 = -\frac{1}{3}\sqrt{3}s - q \approx -1.10715, \quad x_3 = -\frac{1}{3}\sqrt{3}s + q \approx 0.26959$$

where

$$s = \cos\left(-\frac{1}{3}\arctan\left(\frac{\sqrt{3}\sqrt{37}}{9}\right) + \frac{\pi}{3}\right), \quad q = \sin\left(-\frac{1}{3}\arctan\left(\frac{\sqrt{3}\sqrt{37}}{9}\right) + \frac{\pi}{3}\right).$$

Therefore, by developing the above generating function around its singularity of minimal module  $x_3$ , we have the following asymptotic formula:

$$\begin{aligned} \sum_{k=0}^n R_{n,k}^{[p]} 2^{n-k} &= 2^n [t^n] \left( \frac{4\sqrt[4]{3}}{9\sqrt{\frac{(\sqrt{3}s-q)q}{s(\sqrt{3}s+3q)}}} \frac{1}{\sqrt{\left(1-\frac{t}{x_3}\right)}} - \frac{4}{9} + O\left(\sqrt{\left(1-\frac{t}{x_3}\right)}\right) \right) = \\ &= \frac{4\sqrt[4]{3}}{9\sqrt{\frac{(\sqrt{3}s-q)q}{s(\sqrt{3}s+3q)}}} \frac{1}{2^n} \binom{2n}{n} \left(\frac{1}{x_3}\right)^n \left(1 + O\left(\frac{1}{n}\right)\right) = \\ &= 1.45198 \binom{2n}{n} (1.85463)^n \left(1 + O\left(\frac{1}{n}\right)\right). \end{aligned}$$

For example, for  $n = 50$  the exact value is  $0.3793008365 \cdot 10^{43}$  against an approximate value of  $0.3791140086 \cdot 10^{43}$ , with a relative error of 0.05%.

A similar reasoning can be applied to any Riordan pattern  $\mathbf{p}$  for computing the sum

$$\sum_{k=0}^n R_{n,k}^{[\mathbf{p}]} \gamma^{n-k} = \gamma^n [t^n] \frac{d^{[\mathbf{p}]}(t)}{1 - \frac{1}{\gamma} h^{[\mathbf{p}]}(t)}, \quad \gamma \in \mathbb{N},$$

thus assuming to have  $\gamma$  different colours for the fall step, or, equivalently, for the bit zero. Other statistics on the languages on the alphabet  $\{0, 1\}$  avoiding a Riordan pattern can be found analogously by specializing the sequence  $f_k$  in Formula (2.6).

## References

- [1] A. Apostolico and M. Atallah. Compact recognizers of episode sequences. *Information and Computation*, 174(2):180–192, 2002.
- [2] D. Baccherini, D. Merlini, and R. Sprugnoli. Binary words excluding a pattern and proper Riordan arrays. *Discrete Mathematics*, 307:1021–1037, 2007.
- [3] N. Chomsky and M. P. Schützenberger. The algebraic theory of context-free languages. In *Computer programming and formal languages*, volume 65, pages 118–161, North Holland, 1963.
- [4] P. Flajolet, P. Kirschenhofer, and R. F. Tichy. Deviations from uniformity in random strings. *Probability Theory and Related Fields*, 80:139–150, 1988.
- [5] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, 2009.
- [6] P. Flajolet, W. Szpankowski, and B. Vallée. Hidden word statistics. *Journal of the ACM*, 53(1):147–183, 2006.
- [7] L. J. Guibas and M. Odlyzko. Long repetitive patterns in random sequences. *Zeitschrift für Wahrscheinlichkeitstheorie*, 53:241–262, 1980.
- [8] L. J. Guibas and M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *Journal of Combinatorial Theory, Series A*, 30:183–208, 1981.
- [9] D. E. Knuth, J. H. Morris, and V. R. Pratt. Fast pattern matching in strings. *SIAM journal on Computing*, 6:281–315, 1977.
- [10] S. Kumar and E. H. Spafford. A pattern matching model for misuse intrusion detection. In *Proceedings of the 17th National Computer Security Conference*, pages 11–21, 1994.
- [11] D. Merlini. I Riordan Array nell’Analisi degli Algoritmi. Tesi di Dottorato, Università degli Studi di Firenze, 1996.
- [12] D. Merlini, D. G. Rogers, R. Sprugnoli, and M. C. Verri. On some alternative characterizations of Riordan arrays. *Canadian Journal of Mathematics*, 49(2):301–320, 1997.
- [13] I. Rigoutsos, A. Floratos, L. Parida, Y. Gao, and D. Platt. The emergence of pattern discovery techniques in computational biology. *Metabolic Engineering*, 2:159–177, 2000.

- [14] R. Sedgewick and P. Flajolet. *An Introduction to the Analysis of Algorithms*. Addison-Wesley, Reading, MA, 1996.
- [15] L. W. Shapiro, S. Getu, W.-J. Woan, and L. Woodson. The Riordan group. *Discrete Applied Mathematics*, 34:229–239, 1991.
- [16] R. Sprugnoli. Riordan arrays and combinatorial sums. *Discrete Mathematics*, 132:267–290, 1994.
- [17] R. P. Stanley. *Enumerative Combinatorics*, volume 2. Cambridge University Press, Cambridge, 1999.
- [18] M. Waterman. *Introduction to Computational Biology*. Chapman & Hall, London, 1995.