UNIVERSITA' DEGLI STUDI DI MILANO
Dipartimento di Scienze dell'Informazione

Dottorato di Ricerca in Informatica
XVII Ciclo

Violetta Lonati

Pattern statistics in rational models

Relatori: Prof. A. Bertoni, Prof. M. Goldwurm

Anno Accademico 2003-2004

Contents

Introduction

Motivation

Probability on pattern occurrences in a random sequence of letters (generally called text) has been widely studied and has applications in many areas of bio-informatics, code theory and data compression, pattern matching, design and analysis of algorithms, games. Different aspects have been considered: the length of the longest matching, the moments and the distributions of the waiting times for first time occurrences of patterns, the distances between occurrences of patterns.

Here we focus on the frequency of occurrences of a repeated pattern in a random sequence of letters. If we assume to know the probabilistic model (and its parameters) that generates the text, the central question is: *how many occurrences of a given pattern shall we expect in such a random sequence?* Below, we shall refer to this problem as the *frequency problem*.

Among the motivations for the study of this problem, one should mention code synchronization [32], approximated pattern matching [40, 57, 26] and models for database system in wireless communications [1]. However, the most recent applications are in molecular biology. Because of many important sequencing projects, biologists have now large sets of DNA sequences from many different organisms and they need quantitative tools and statistical methods to help them in analyzing sequences. Identifying words that show relevant deviations between their observed frequency and their frequency predicted by a given model is an important way to extract information from DNA sequences. Among the problems that may benefit from results on words count we quote gene recognition: it is known that motif in DNA sequences have statistical properties that are different in coding and non-coding regions and most techniques for the recognition of the genes (coding regions) rely on such a difference. Another biological problem related to the frequency problem is the search of patterns that are significantly over or under-represented in experimentally observed DNA sequences. When the frequency of a short pattern is either too high or too low, the pattern often turns out to have some biological function. Thus, identifying deviant short motifs might point out unknown biological information [27, 47].

The frequency problem

The frequency problem can be studied under different assumptions concerning the source that generates the text, or the pattern to search for through the text. The simplest model represents a memoryless source: here the text is a sequence of letters chosen independently, according to a fixed probability distribution. Such model is referred to as the Bernoulli model; if in particular all letters are assigned the same probability, the model is said to be symmetric. Another classical model, more general than the previous one, is defined by Markov processes [41, 54], where the probability to generate the next letter depends on a fixed number of previous occurrences. Other models considered in the literature are called dynamical sources and describe non-Markovian processes, characterized by unbounded dependency on past history [58].

The choice of pattern can lead to different settings, too. String matching is the basic pattern matching problem; here, one counts the occurrences of a given string as a factor in the text. One can also search for a finite set of strings and count the occurrences of all of them. Moreover one may be interested in occurrences of the pattern as a subsequence of the text; in this case the letters no longer need to be consecutive. A generalization of all these problems is attained when the pattern is defined by a general regular expression, thus including infinite sets of words.

When a pattern is searched for through a text, various constraints can be imposed on the counting of overlapping occurrences; occurrences are considered valid if they satisfy these constraints. In the overlapping model, any occurrences is valid and two overlapping patterns both contribute to the count. However, in some cases this assumption is not correct. For instance, biologists remark that when an enzyme has bound to one occurrence of a pattern in a DNA sequence, other enzymes cannot bind to the same portion of DNA. Thus, two overlapping sequences cannot be considered valid simultaneously: one only counts the first occurrence and another occurrence is valid if it does not overlap on the left with any other valid occurrence. As variants, one may count overlapping occurrences of different patterns, or one may set a minimal distance between valid occurrences.

Several authors contributed to the study of the frequency problem, generally considering the Bernoulli or the Markov models to generate the random text [21, 32, 33, 34, 26, 44, 50, 49, 45, 10]. The most important recent contributions belong to Guibas and Odlyzko who in a series of seminal papers [32, 33, 34] laid the foundations for the analysis of the symmetric Bernoulli case.

The results have then been extended to the Markovian model, first by Li [44], who considered the problem with no pattern occurrences, and, more recently, by Régnier and Szpankovski. In [50], using a method that treats uniformly both the Bernoulli and the Markov models, they established that the number of occurrences of a string is asymptotically normal, under a primitivity hypothesis of the stochastic model. They also obtained large deviations results.

A recent improvement is due to Nicodème, Salvy e Flajolet, that in [45] extended all the previous results considering a text generated by a Markov source and counting the occurrences of a pattern defined by an unrestricted regular expression. Their results hold under a primitivity hypothesis on the stochastic matrix defining the Markov process.

Finally, non-Markovian models have been considered by Bourdon and Vallée. In [10] they assumed the text was generated by dynamical sources and they considered generalized pattern, entailing classical and patterns with "don't-care-symbols".

Our contribution

In this thesis, we study pattern occurrences in a new framework, introducing a stochastic model defined via rational formal series in non-commuting variables (or, equivalently, by weighted automata). More precisely, given a formal series $r : \{a,b\}^* \to \mathbb{R}_+$, for every integer n (satisfying $(r,x) \neq 0$ for some $x \in \{a,b\}^n$) we consider the probability space of all words in $\{a,b\}^n$ equipped with the probability measure given by

$$\mathrm{P}_n\{\omega\} = \frac{(r,\omega)}{\sum_{x\in\{a,b\}^n}(r,x)} \qquad (\omega \in \{a,b\}^n).$$

Then, we define the random variable $Y_n : \{a,b\}^n \to \{0,1,\ldots,n\}$ such that $Y_n(\omega)$ equals the number of occurrences of a in the word ω of length n. The *rational symbol frequency* (RSF) problem concerns the study of the distribution properties of the sequence $\{Y_n\}_n$, assuming that the series r defining the model is rational.

This setting generalizes the frequency problem studied in [45], which in fact turns out to be a special case of the RSF problem. Indeed, we prove that the question of studying the number of occurrences of a regular pattern in a text generated by a Markovian source can always be translated into the RSF problem for a suitable rational series over two non-commuting variables, while the converse does not hold. In this sense, the rational stochastic model properly extends the Markovian models.

Our goals are estimating the moments of the random variable Y_n and determining local and central limit distributions of the sequence $\{Y_n\}_n$ as n tends to infinity. We first assume that the transition matrix associated with the series defining the model is primitive. Then:

- We prove that the mean and the variance are asymptotically linear, that is there exist two constants β and γ such that $\mathbb{E}(Y_n) = \beta n + \mathrm{O}(1)$ and $\mathbb{V}ar(Y_n) = \gamma n + \mathrm{O}(1)$; we provide precise expressions for β and γ and we prove that they are strictly positive (except for degenerate cases).

- We show that a central limit theorem holds; the limit distribution approximates a Gaussian behaviour and we explicitly determine the approximation error.

- We provide a condition that guarantees the existence of a Gaussian local limit theorem; to state this condition, we introduce a notion of symbol periodicity for weighted automata which extends the classical periodicity theory of Perron–Frobenius for non-negative matrices.

- As an application of the previous analysis, we obtain an asymptotic estimation of the growth of the coefficients for a subclass of rational formal series in two commuting variables.

The results are then extended, dropping the primitive hypothesis usually assumed in the literature. In particular:

- We study bicomponent models, defined by weighted automaton with two strongly connected components, obtaining in many cases limit distributions quite different from the Gaussian one.

- We present a general approach to deal with arbitrary non-primitive models. Again, we start from the decomposition of the weighted automaton defining the model into strongly connected components, in order to detect the elements that mainly determine the limit distribution. In the most relevant cases we establish the limit distribution, that is characterized by a unimodal density function defined by polynomials over adjacent intervals.

Overview

The first part of the thesis presents the *tools of the trade*, consisting of preliminary notions, basic properties but also more advanced results concerning formal series and languages, non-negative matrices, limit theorems in probability theory. Moreover, we introduce the notion and the properties of symbol periodicity for non-negative matrices (see Section 2.4) and we prove a criterion for the local convergence of a general sequence of random variables (see Section 3.7).

Chapter 1 is an introduction to rational formal series and their relation to languages. After defining formal series in non-commuting variables having coefficients in a semiring, we present the algebraic and topological structure of the set of formal series. Then we introduce the classes of rational and recognizable series, focusing on the equality between such classes, due to the Schützenberger Representation Theorem. We also consider weighted automata associated with rational series, and define their counting matrices. Moreover, we extend the definitions to formal series in partially commuting variables, by introducing the notion of trace monoid. Finally we take into consideration the problem of estimating the maximum coefficient of a rational series, briefly illustrating some results known in the literature.

In **Chapter 2** we deal with matrices with coefficients in a positive semiring. On the one hand we recall the Perron–Frobenius Theory for matrices with coefficients in $\mathbb{R}_+$; on the other hand we introduce the notion of symbol periodicity for matrices with polynomial entries. While the former is a well-known subject (the main result [25] dates back to 1908), the definition of symbol periodicity has been introduced recently [7]. However, such a notion and its properties are included in this chapter, since in a certain sense they extend the Perron–Frobenius Theory.

Chapter 3 concerns probability theory and in particular central and local limit theorems for sequences of random variables. First, we recall some basic notions and present some typical examples of probability distributions. Then we consider a sequence of Bernoulli trials; in particular we focus on DeMoivre–Laplace limit theorems and their extensions to partial sums of more general sequences of random variables. We also present Markov processes as a generalization of the Bernoulli scheme. Finally, we take into consideration arbitrary sequences of discrete random variables, without assuming any condition of independence: we present the "quasi-power" theorem and we prove a criterion to establish local limit properties holding for such sequences.

In the second part of the thesis we carry out the analysis of pattern statistics in rational models, using the tools presented in Part I.

In **Chapter 4** we start off the discussion, formally defining the rational model and the rational symbol frequency (RSF) problem. In order to compare this problem with those previously dealt with in the literature, we show how our model can be viewed as a proper extension of the Markovian model as far as counting the occurrences of a regular set in a random text is concerned. Then we analyze the primitive case, assuming that the transition matrix associated with the series defining the rational model is primitive. We obtain asymptotic estimates for the mean value and the variance of the statistics in exam, showing that they have asymptotic linear behaviour; we also prove that they converge in distribution to a normal random variable; finally we establish a local limit theorem which turns out to be related to the notion of symbol periodicity introduced in Chapter 2. As an application of the previous analysis, we obtain an asymptotic estimation of the growth of the coefficients for rational formal series in commuting variables.

In **Chapter 5** we improve the analysis of the RSF problem, dropping the primitivity hypothesis. More precisely, we consider bicomponent rational models, defined by rational series corresponding to weighted automata with two primitive components. Two special examples are of particular interest: they occur when the formal series defining the model is, respectively, the sum or the product of two primitive formal series. The main results concern the asymptotic evaluation of mean value and variance and the limit distribution of our statistics. We obtain in many cases limit distributions quite different from the Gaussian one.

Finally, in **Chapter 6** we present a general approach to the analysis of arbitrary rational models, based on the decomposition of the weighted automaton defining the model into strongly connected components. We introduce the notion of main chains and we show that they mainly determine the limit behaviour of our statistics. In the most significant cases, we explicitly establish the limit distribution, that is characterized by a unimodal density function defined by polynomials over adjacent intervals. We finally provide a natural method to determine the limit distribution in the general case.

Part I

Tools of the trade

Chapter 1

Rational formal series

This chapter is an introduction to rational formal series in non-commuting variables having coefficients in a semiring. Formal series have long been in use in all branches of mathematics; they are fundamental especially in enumeration and combinatorics. In particular, the class of rational series has many remarkable properties and plays a role that in some sense corresponds to the role of regular languages in language theory.

The chapter is organized as follows. We first recall the definitions of monoid and semiring, presenting typical examples. In Section 1.2 we introduce the notion of formal series over a free monoid and the first relations with languages, provided by the support of a series and the characteristic series of a language. Then we present the algebraic and topological structure of the set of formal series. The class of rational series is defined in Section 1.4, together with its relation to the class of regular languages. In Section 1.5 we consider recognizable series and the corresponding weighted automata. We focus on the equality between recognizable and rational series, due to the Schützenberger Representation Theorem. As a special consequence, we state an important result in language theory, namely the Kleene Theorem. Afterwards, in 1.6, we define the counting matrices associated with a rational series, which will be basic tools for the analysis of pattern statistic in rational models we develop in Part II. In Section 1.7 we extend the definitions to formal series in partially commuting variables, by introducing the notion of trace monoid. Finally, in the last section we take into consideration the problem of estimating the maximum coefficient of a rational series and we briefly illustrate some results known in the literature. A specific result on this topic will be proven in Chapter 4 for the case of rational series in commuting variables and non-negative coefficients.

1.1 Monoids and semirings

In this section we recall the basic definitions of monoid and semiring and we present some classical examples.

A *monoid* is defined as a set $\mathcal{M}$ equipped with an operation "$\cdot$" called *product*. The product must be associative and have an identity element $1_{\mathcal{M}}$. Thus, a monoid is a group-like object that in general fails to be a group because elements need not have an inverse within the object.

In language theory, the most important monoid is given by the family of words over a finite alphabet. Formally, an *alphabet* is a finite set of symbols $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_\ell\}$. Such symbols are also called *letters*. Any finite sequences $\omega = \omega_1\omega_2\cdots\omega_n$ of symbols $\omega_i \in \Sigma$ is called *word* or *string*. The sequence composed by no letters is called the *empty word* and it is denoted by ϵ. The set of all words over Σ is indicated by Σ^*. The *concatenation* of two words $\omega = \omega_1\omega_2\cdots\omega_n$ and

$v = v_1 v_2 \cdots v_n$ is defined as the word $\omega_1 \cdots \omega_n v_1 \cdots v_n$. Clearly, the concatenation is associative and its identity element is the empty word. Thus, Σ^* is a monoid. Since for any monoid $\mathcal{M}$ and any function $f : \Sigma \longrightarrow \mathcal{M}$ there exists a unique monoid morphism $\overline{f} : \Sigma^* \longrightarrow \mathcal{M}$ extending f, Σ^* is referred to as the *free monoid generated by* Σ.

The *length* of a word ω is the number of its letters and it is denoted by $|\omega|$. Moreover, for any letter $\sigma \in \Sigma$, we use $|\omega|_\sigma$ to denote the number of occurrences of σ in ω. Clearly $|\epsilon| = |\epsilon|_\sigma = 0$ for every $\sigma \in \Sigma$. A *language* over Σ is simply a subset of Σ^*. An infinite language may be specified by means of a generating system or using a recognition device: a generating system, namely a *grammar*, defines a scheme to generate all words of the language; a recognizer, in most cases a *finite automaton*, provides an algorithm that halts with the answer 'yes' for words in the language and halts with the answer 'no' otherwise. We shall not linger on this topic here, the interested reader may refer to [36].

A *semiring* is a set $\mathcal{S}$ equipped with two binary operations "$\cdot$" and "$+$", respectively called *product* and *sum*, such that: $< \mathcal{S}, + >$ is a commutative monoid with neutral element $0_\mathcal{S}$; $< \mathcal{S}, \cdot >$ is a monoid with neutral element $1_\mathcal{S}$; the product is distributive with respect to the sum; finally for every $a \in \mathcal{S}$ one has $a \cdot 0_\mathcal{S} = 0_\mathcal{S} \cdot a = 0_\mathcal{S}$. In general, a semiring fails to be a ring because elements need not have an inverse with respect to the addition, in other term the subtraction is not defined. Thus the last condition, which in the case of rings follows by the previous axioms, must be explicitly required.

We usually assume that the product is commutative and in this case the semiring is said to be *commutative*, too. Moreover, we say that the semiring $\mathcal{S}$ is *positive*, if $x + y = 0_\mathcal{S}$ implies $x = y = 0_\mathcal{S}$ and $x \cdot y = 0_\mathcal{S}$ implies $x = 0_\mathcal{S}$ or $y = 0_\mathcal{S}$ for any pairs of elements $x, y \in \mathcal{S}$. In this case we also write $x > 0_\mathcal{S}$ whenever $x \neq 0_\mathcal{S}$.

As an example of semiring one can clearly consider all usual numeric rings as $\mathbb{C}, \mathbb{R}, \mathbb{Q}, \mathbb{Z}$. The set of positive integers $\mathbb{N}$ and the set of positive reals $\mathbb{R}_+$ are not ring, but they are positive semirings. The *boolean semiring* $\mathbb{B}$ is defined over the set $\{0, 1\}$ by setting $1 + 1 = 1$.

Another interesting example is given by the *tropical semiring* $\mathbb{T}$ whose support is the set $\mathbb{N} \cup \{\infty\}$ and whose operations are the *min* for the addition and the usual sum for the multiplication. Clearly the neutral element with respect to the addition is 0, while the identity element with respect to the multiplication is ∞. Sometimes in the literature $\mathbb{N}$ is replaced by $\mathbb{R}$; in this case someone also gives the definition using *max* instead of *min*.

A semiring can be built in a natural way from the power set $2^\mathcal{M}$ of a monoid $\mathcal{M}$: for every pair of subsets $A, B \subseteq \mathcal{M}$, we set $A + B = A \cup B$ and $A \cdot B = \{ab \mid a \in A, b \in B\}$. Then the identity element with respect to the sum is the empty set, while the identity element with respect to the product is the set $\{1_\mathcal{M}\}$.

If $\mathcal{S}$ is a semiring, then also $\mathcal{S}[x]$ and $\mathcal{S}^{Q \times Q}$ are semirings, where $\mathcal{S}[x]$ denotes of polynomials in the variable x and coefficients in $\mathcal{S}$, while $\mathcal{S}^{Q \times Q}$ indicates the set of matrices with entries in $\mathcal{S}$ and indices in a finite set Q. Such matrices are considered in Chapter 2, assuming $\mathcal{S}$ to be positive.

1.2 Formal series

From now on, let Σ be a finite alphabet and $\mathcal{S}$ a semiring.

Definition 1.1 *A* formal series *over* Σ *is a mapping* $r : \Sigma^* \longrightarrow \mathcal{S}$ *which associates each word* $\omega \in \Sigma^*$ *with the element* $(r, \omega) \in \mathcal{S}$*, called the* coefficient *of* ω *in* r*. The series* r *is usually written as a formal sum*

$$r = \sum_{\omega \in \Sigma^*} (r, \omega)\, \omega$$

and the collection of all formal series is denoted $\mathcal{S}\langle\langle \Sigma^* \rangle\rangle$.

This terminology reflects the intuitive ideas connected with power series. Indeed, if $\mathcal{S}$ is $\mathbb{R}$ or $\mathbb{C}$ and Σ reduces to a unique element z, then we obtain the usual definition of power series studied in classical analysis. We use the adjective 'formal' to indicate that the study of convergence properties is not our main interest, as is in classical analysis.

Each series $r \in \mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ identifies a language in a natural way, namely the language

$$\mathrm{Supp}(r) = \{\omega \in \Sigma^* | (r, \omega) \neq 0_{\mathcal{S}}\}$$

that is called the *support* of r. The subset of $\mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ consisting of all series with a finite support is denoted by $\mathcal{S}\langle\Sigma^*\rangle$ and its elements are called *polynomials*. To emphasize the set of generators $\Sigma = \{\sigma_1, \sigma_2, \cdots, \sigma_\ell\}$ of the monoid Σ^*, we may also denote the family of all formal series by $\mathcal{S}\langle\langle\sigma_1, \cdots, \sigma_\ell\rangle\rangle$ and the set of polynomials over Σ by $\mathcal{S}\langle\sigma_1, \cdots, \sigma_\ell\rangle$.

On the other hand, given a language $L \subseteq \Sigma^*$, there are many series r such that $\mathrm{Supp}(r) = L$. However we can uniquely define the *characteristic series* $\chi_L \in \mathcal{S}\langle\langle\Sigma^*\rangle\rangle$, by setting

$$(\chi_L, \omega) = \begin{cases} 1_{\mathcal{S}} & \text{if } \omega \in L \\ 0_{\mathcal{S}} & \text{otherwise} \end{cases}$$

for each $\omega \in \Sigma^*$. In other terms

$$\chi_L = \sum_{\omega \in L} \omega \ .$$

Clearly $\mathrm{Supp}(\chi_L) = L$ while in general $\chi_{Supp(r)} \neq r$. A series r is said to be *non-ambiguous* if $(r, \omega) \in \{0, 1\}$ for every $\omega \in \Sigma^*$ and this is equivalent to assuming $\chi_{Supp(r)} = r$. Of course, all series in $\mathbb{B}\langle\langle\Sigma^*\rangle\rangle$ are non-ambiguous.

1.3 The semiring of formal series

Let us now introduce some operations between formal series. If $r, s \in \mathcal{S}\langle\langle\Sigma^*\rangle\rangle$, then their *sum* $r + s$ and their *Cauchy product* [1] $r \cdot s$ are defined by setting

$$(r + s, \omega) = (r, \omega) + (s, \omega) \ ,$$

$$(r \cdot s, \omega) = \sum_{xy = \omega} (r, x) \cdot (s, y) \ ,$$

for every $\omega \in \Sigma^*$. The set $\mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ constitutes a semiring with respect to the previous binary operations and $\mathcal{S}\langle\Sigma^*\rangle$ is one of its subsemiring. The neutral element of the sum is given by the series 0 such that $(0, \omega) = 0_{\mathcal{S}}$ for every $\omega \in \Sigma^*$, while the neutral element of the product is given by the series 1 such that $(1, \omega) = 1_{\mathcal{S}}$ if $\omega = \epsilon$ and $(1, \omega) = 0_{\mathcal{S}}$ otherwise.

Furthermore, we can define two external operations of $\mathcal{S}$ on $\mathcal{S}\langle\langle\Sigma^*\rangle\rangle$, one acting on the left, the other on the right, by setting

$$(a \cdot r, \omega) = a \cdot (r, \omega) \ , \qquad (r \cdot a, \omega) = (r, \omega) \cdot a \ ,$$

for each $a \in \mathcal{S}$, $r \in \mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ and $\omega \in \Sigma^*$. With respect to these external products, $\mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ is a $\mathcal{S}$-module, that is the external products are compatible with the internal operations of $\mathcal{S}$ and their neutral elements.

Notice that we use the same symbol "$\cdot$" to indicate both the Cauchy product and the external products. In general, the dot will be omitted.

[1] This definition is well-set not only for the monoid Σ^*, but also for a general monoid $\mathcal{M}$, provided that each $\omega \in \mathcal{M}$ has only finitely many factorization $\omega = xy$.

When applied to the series ϵ, the external products give the same result. Thus, for each $a \in \mathcal{S}$ we can define a new series $a \cdot \epsilon = \epsilon \cdot a$, so obtaining a natural injection of $\mathcal{S}$ into $\mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ as a subsemiring. In particular the series corresponding to $1_\mathcal{S}$ and $0_\mathcal{S}$ equal the neutral elements for the product and the sum of series and they coincide with the series 1 and 0 defined above. Similarly, there is a natural injection of Σ^* into $\mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ as a submonoid: with each $\omega \in \Sigma^*$ we associate the series, still denoted by ω, such that $(\omega, \omega') = 1_\mathcal{S}$ if $\omega' = \omega$ and $(\omega, \omega') = 0_\mathcal{S}$ otherwise. In particular, the series determined by the empty word ϵ coincides with the series 1 defined above. We call *monomials* the series $a\omega$, for $a \in \mathcal{S}$ and $\omega \in \Sigma^*$. Note that $a\omega = \omega a$, moreover all coefficients of $a\omega$ are null except the coefficient of ω which equals a.

The semiring $\mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ can be equipped with a topological structure. Indeed, it turns out to be an ultrametric space with respect to the following *distance* [2]. If r and $s \in \mathcal{S}\langle\langle\Sigma^*\rangle\rangle$, we set

$$d(r,s) = \begin{cases} 0 & \text{if } r = s \\ 2^{-k} & \text{otherwise} \end{cases}$$

where $k = min\{|\omega| \mid \omega \in \Sigma^*, (r, \omega) \neq (s, \omega)\}$, $|\omega|$ indicating the length of the word ω.

Given a sequence $\{r_i\}_{i \in I}$ of formal series, we say it is *summable* if there exist a series r such that, for all $\varepsilon > 0$, one can find a finite subset I' of I such that every finite $J \supseteq I$, satisfies the equality

$$d\left(\sum_{j \in J} r_j, r\right) \leq \varepsilon \ .$$

The series r is called the *sum* of the family $\{r_i\}$ and is unique.

1.4 The class of rational series

A formal series $r \in \mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ is called *proper* or *quasi-regular* if the coefficient of ϵ (i.e. the *constant term* of r) vanishes. This kind of series has the desirable property that the family $\{r^i\}_{i \geq 0}$ is summable. The sum of this family is denoted by r^*

$$r^* = \sum_{i \geq 0} r^i$$

and it is called the *star* of r. Moreover, r^+ denotes the sum of the positive powers of r, that is

$$r^+ = \sum_{i > 0} r^i \ .$$

By the definition we easily get

$$r^* = 1 + r^+ \ , \qquad r^+ = rr^* = r^* + r \ .$$

If $\mathcal{S}$ is a ring, then the series $-r$ is defined, and r^* is just the inverse of $1 - r$ with respect to the Cauchy product. Indeed, $r^*(1 - r) = r^* - r^*r = r^* - r^+ = 1$.

The sum, the Cauchy product and the star operation are called *rational operations*. Given a subset E of $\mathcal{S}\langle\langle\Sigma^*\rangle\rangle$, we call *rational closure* of E the minimum subset of $\mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ containing E and closed and closed under rational operations.

[2] This definition is well-set not only for the monoid Σ^*, but also for a general monoid $\mathcal{M}$, provided that it admits a *length function* $|\cdot|$ satisfying $|\omega_1\omega_2| = |\omega_1| \cdot |\omega_2|$.

Definition 1.2 *A series r is said to be $\mathcal{S}$-*rational *if it is in the rational closure of the set of polynomials $\mathcal{S}\langle\Sigma^*\rangle$.*

The class of rational series is usually indicated as $\mathcal{S}^{Rat}\langle\langle\Sigma^*\rangle\rangle$, or $\mathcal{S}^{Rat}\langle\langle\sigma_1, \sigma_2, \ldots, \sigma_\ell\rangle\rangle$, if one wants to emphasize the set of generators $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_\ell\}$.

If Σ reduces to a unique element z, then one can prove that any rational series in $\mathbb{R}\langle\langle z\rangle\rangle$ converges to a rational function $p(z)/q(z)$, where p and q are polynomials and $q(z) \neq 0$. This justifies the use of the term "rational".

The rational operations on the semiring of formal series are related to the following operations among languages:

- *union* of two languages
$$L_1 \cup L_2 = \{\omega \mid \omega \in L_1, \text{or } \omega \in L_2\} \ ,$$
- *product* or *concatenation* of two languages
$$L_1 \cdot L_2 = \{\omega_1\omega_2 \mid \omega_1 \in L_1, \omega_2 \in L_2\} \ ,$$
- *Kleene closure* of a language, defined by the union of all its nonnegative powers
$$L^* = \bigcup_{j\geq 0} L^j \ .$$

Such operations are called *rational* and determine a well-known class of languages over a finite alphabet.

Definition 1.3 *A language $L \subseteq \Sigma^*$ is said to be regular if it belongs to the subset of 2^{Σ^*} containing the finite languages and closed under the rational operations.*

Notice that this definition means that a rational language can be obtained from atomic languages $\{\epsilon\}$ and $\{\sigma\}$, where $\sigma \in \Sigma$, by a finite number of applications of rational operations. The formula expressing how a specific language is obtained from atomic languages by regular operations is called *regular expression.*

Now, as a straightforward consequence of the definitions of support and characteristic series, one can prove the following result, which is the reason why regular languages are often called *rational.*

Theorem 1.4 *A language $L \subseteq \Sigma^*$ is regular if and only if it is the support of a rational series in $\mathbb{B}\langle\langle\Sigma^*\rangle\rangle$.*

Finally, note that if $\mathcal{H} : \mathcal{S} \to \mathcal{S}'$ is a semiring morphism, then it can be extended to a semiring morphism $\mathcal{H} : \mathcal{S}\langle\langle\Sigma^*\rangle\rangle \to \mathcal{S}'\langle\langle\Sigma^*\rangle\rangle$, by setting for each series r with coefficients in $\mathcal{S}$

$$\mathcal{H}(r) = \sum_{\omega\in\Sigma^*} \mathcal{H}(r, \omega)\ \omega \ .$$

As a consequence, if $r \in \mathcal{S}^{Rat}\langle\langle\Sigma^*\rangle\rangle$, then $\mathcal{H}(r) \in \mathcal{S}'^{Rat}\langle\langle\Sigma^*\rangle\rangle$. In particular, if $\mathcal{S}$ is a positive semiring, we can consider the morphism from $\mathcal{S}$ to $\mathbb{B}$ that associates $0_\mathcal{S}$ with $0_\mathbb{B}$ and any other $a \in \mathcal{S}$ with $1_\mathbb{B}$, so obtaining the following result.

Corollary 1.5 *Let $\mathcal{S}$ be a positive semiring and $r \in \mathcal{S}^{Rat}\langle\langle\Sigma^*\rangle\rangle$. Then the support of r is a regular language.*

1.5 Recognizable series and weighted automata

Definition 1.6 *A series $r \in \mathcal{S}\langle\langle\Sigma^*\rangle\rangle$ is called* $\mathcal{S}$-recognizable *if there exists a non-empty finite set Q, two vectors $\xi, \eta \in \mathcal{S}^Q$ and a monoid morphism $\mu : \Sigma^* \longrightarrow \mathcal{S}^{Q\times Q}$ such that*

$$(r, \omega) = \xi_T \mu(\omega)\eta \tag{1.1}$$

for each word $\omega \in \Sigma^*$. The triple (ξ, μ, η) is called a linear representation *of r and the cardinality of the set Q is said to be its* size. *(Remark that the morphism μ is uniquely determined by its restriction $\mu_{|\Sigma}$).*

Using the terminology of automata theory, the elements of Q are often referred as *states* and each $p \in Q$ such that $\xi_p \neq 0$ (resp. $\eta_p \neq 0$) is said to be an *initial* (resp. *final state*). Moreover we can define a *transition map* $\delta : Q \times \Sigma \to 2^Q$ by setting $\delta(p, \sigma) = \{q \in Q \mid \mu(\sigma)_{pq} \neq 0\}$ for every state p. Thus, defining a recognizable series is equivalent to defining a nondeterministic finite automaton where transitions, initial and final states are equipped with weights in $\mathcal{S}$. In the sequel we use the expression *weighted automaton over $\mathcal{S}$* to refer to this kind of automata. Notice that if $\mathcal{S}$ is the boolean semiring $\mathbb{B}$, we obtain the usual definition of finite state automaton.

Using the standard approach of automata theory, we can also represent a recognizable series (or, more precisely, each of its linear representation) by a *state diagram*. This is a graph consisting of:

- a node for every state in Q, equipped with the weights ξ_p and η_p;
- an oriented edge from state p to state q with label $a \in \Sigma$ and weight $\mu(a)_{pq}$, whenever $\mu(a)_{pq} \neq 0$.

Example 1.7 Consider the series $r \in \mathbb{B}\langle\langle a, b\rangle\rangle$ defined by setting

$$(r, w) = \begin{cases} 1 & \text{if } |w|_a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

for every word $w \in \{a, b\}^*$. Such a series admits the following linear representation

$$\xi_T = (10)\ , \quad \mu(a) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\ , \quad \mu(b) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\ , \quad \eta = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

corresponding to the automaton represented in Fig. 1.1.

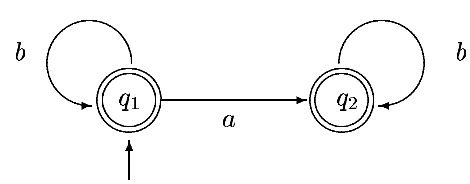


Figure 1.1: State diagram associated with the recognizable series of Example 1.7. The double circles denote the final states, while the entering arrow denotes the initial state.

Now, consider a computation path, that is a sequence of edges in the state diagram of the form

$$\ell = p_0 \xrightarrow{x_1} p_1 \xrightarrow{x_2} p_2 \cdots p_{n-1} \xrightarrow{x_n} p_n.$$

We say that $\omega = x_1 x_2 \cdots x_n$ is the *label* of ℓ, $|\ell| = n$ is its *length* and we denote by $|\ell|_a =$ the number of occurrences of a in ω. Moreover if p_0 is an initial state and p_n is a final one we say that ℓ is an *accepting computation path* for ω. We also call *weight* of ℓ the value

$$w(\ell) = \mu(x_1)_{p_0 p_1} \cdot \mu(x_2)_{p_1 p_2} \cdots \mu(x_n)_{p_{n-1} p_n} \in \mathcal{S} \ .$$

Thus, the series r satisfies

$$(r, \omega) = \sum_{p,q \in Q} \ \sum_{\ell : p \xrightarrow{\omega} q} \xi(p) \cdot w(\ell) \cdot \eta(q) \ .$$

Furthermore, let P be the $(Q \times Q)$-matrix with entries in $\mathcal{S}\langle\langle \Sigma^* \rangle\rangle$ defined by

$$P = \sum_{\sigma \in \Sigma} \mu(\sigma) \sigma \ .$$

It is easy to see that the series ${P^n}_{pq}$ associates each word ω with the sum of weights of all paths of length n starting in p, ending in q and labelled by ω. In other terms we have

$$\xi_T P^n \eta = \sum_{\omega \in \Sigma_n} (r, \omega) \ \omega \ ,$$

where Σ^n denotes the set of words of length n in Σ^*. Since $r = \sum_{n \geq 0} \sum_{\omega \in \Sigma_n} (r, \omega) \ \omega$, we also have

$$r = \sum_{n \geq 0} \xi_T P^n \eta \ .$$

Observe that if $\mathcal{S}$ is the tropical semiring, then the operation denoted by "$\cdot$" is interpreted as the sum, while the operation denoted by "$+$" is interpreted as the min. Hence in this case weights are summed along paths, while the coefficient of a word is determined by the minimum weight among all paths. Furthermore, if $\mathcal{S} = \mathbb{N}$ is the traditional semiring of non-negative integers and ξ, μ, η take on values only in $\{0, 1\}$, then the coefficient of ω in r is the number of its accepting computation paths. In this case, the weighted automaton is nothing more than a (nondeterministic) finite automaton and we say that a word is *accepted* by the automaton if it admits at least one accepting computation path; we also call *language accepted* by the automaton the set of all accepted words.

In general, we remark that the total sum of weights may vanish and hence there could exist some accepting computation paths for a word ω having null coefficient in r. Anyway, if $\mathcal{S}$ is a positive semiring, this cannot happen and hence we obtain what follows.

Theorem 1.8 *A language $L \subseteq \Sigma^*$ is accepted by a nondeterministic finite automaton if and only if its characteristic series is $\mathcal{S}$-recognizable for every positive semiring $\mathcal{S}$.*

We conclude this section stating two fundamental results. The first one was established in 1961 and its proof can be found in [5, Section 1.6] or [52, Theorem 2.3].

Theorem 1.9 (Schützenberger Representation Theorem) *Let $\mathcal{S}$ be a semiring. Then a series $r \in \mathcal{S}\langle\langle \Sigma^* \rangle\rangle$ is $\mathcal{S}$-rational if and only if r is $\mathcal{S}$-recognizable.*

As a consequence, a rational formal series $r \in \mathcal{S}^{Rat}\langle\langle\Sigma^*\rangle\rangle$ may be equivalently defined via a linear representation or a weighted automaton over $\mathcal{S}$.

The previous theorem, together with Theorems 1.4 and 1.8, yields another important result, which was actually found a few years earlier (1956) and can be proved without reference to the formal series, see for instance [36, Theorem 3.10].

Theorem 1.10 (Kleene's Theorem) *A language is regular if and only if it is accepted by a finite automaton.*

1.6 Counting matrices associated with rational series

We now introduce a natural notion of counting matrices associated with a given linear representation (ξ, μ, η) over a positive semiring $\mathcal{S}$. Such matrices will be basic tools for the analysis of pattern statistics in rational models we develop in Part II.

First consider the matrix $M \in \mathcal{S}^{Q \times Q}$ defined by setting

$$M_{pq} = \sum_{\sigma \in \Sigma} \mu(\sigma)_{pq}$$

for any $p, q \in Q$. This matrix is related to the paths of the state diagram. Indeed, for every positive integer n and for every pairs of states $p, q \in Q$, the entry ${M^n}_{pq}$ sums up the weights of all paths of length n starting in p and ending in q:

$${M^n}_{pq} = \sum_{\ell : p \to q, \ |\ell| = n} w(\ell) \ .$$

If $\mathcal{S} = \mathbb{N}$ and ξ, μ, η take on values only in $\{0, 1\}$, then ${M^n}_{pq}$ is exactly the number of paths of length n in the state diagram. For this reason, we name M the *counting matrix* of the linear representation (ξ, μ, η).

Also, for any symbol $a \in \Sigma$ and for any pair of states $p, q \in Q$, we set

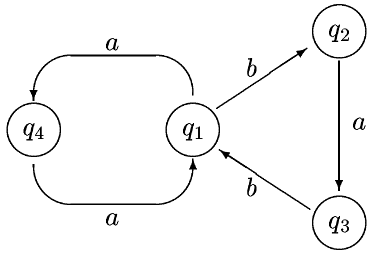
$$M_a(x)_{pq} = \mu(a)_{pq} \ x + \sum_{\sigma \neq a} \mu(\sigma)_{pq} \ ,$$

where x is a variable. Then, it is easy to verify that

$${(M_a(x))^n}_{pq} = \sum_{\ell : p \to q} \mu(\ell) \ x^{|\ell|_a} \ .$$

In other terms, ${(M_a(x))^n}_{pq}$ is the sum of monomials (according to the usual definition) like sx^k, where s is the sum of weights of paths from p to q that have k occurrences of the letter a. $M_a(x)$ can be again interpreted by assuming that $\mathcal{S} = \mathbb{N}$ and ξ, μ, η take on values only in $\{0, 1\}$. Indeed, in this case the coefficient of x^k in $M_a(x)^n$ equals the number of paths from p to q that have k occurrences of the letter a. For this reason $M_a(x)$ is called the *a-counting matrix* of the linear representation.

Example 1.11 Fig. 1.2 represents the state diagram of a linear representation over $\mathbb{B}$ and the corresponding a-counting matrix.



$$M = \begin{pmatrix} 0 & 1 & 0 & x \\ 0 & 0 & x & 0 \\ 1 & 0 & 0 & 0 \\ x & 0 & 0 & 0 \end{pmatrix}$$

Figure 1.2: Example of state diagram and a-counting matrix.

1.7 Trace monoids and languages

Up to now, we considered words over a non-commuting alphabet, that is we always assumed that the order of letters into words was fixed. Here we generalize the definition of alphabet, to include also the existence of pairs of commuting letters; words on such alphabet are called *traces*. In particular, the commutative case occurs when all letters commute. We just note that traces have been studied [19, 20, 12] as a formal model to describe concurrent processes: each letter represents a process and when two processes are independent (i.e. they can be run in any order), then the corresponding letters commute.

Given an alphabet $\Sigma = \{\sigma_1, \cdots \sigma_\ell\}$, let $I \subseteq \Sigma \times \Sigma$ be an irreflexive and symmetric relation. Then (Σ, I) is called *concurrent alphabet*, I is called *independence relation* and two symbols a and b of Σ such that $(a, b) \in I$ are said to be *independent*. The relation I can be interpreted as follows: if $(a, b) \in I$, then their order may be switch and ab is equivalent to ba. Thus, I is also called *commutativity relation*.

It is easy to see that I induces an equivalence relation over Σ^*: for any pairs of words $u, v \in \Sigma^*$, if a and b are symbols in Σ such that $(a, b) \in I$, then we set $uabv =_I ubav$. The equivalence classes are called *traces* and they are denoted $[\omega]$, for $\omega \in \Sigma^*$. Observe that all words in the same equivalence class have the same length and the same numbers of occurrences of each letter σ. Hence the values $|[\omega]|$ and $|[\omega]|_\sigma$ are well-defines. The relation $=_I$ is also invariant with respect to the concatenation product. Thus, we can define the following operation between traces:

$$[\omega_1] \cdot [\omega_2] = [\omega_1 \omega_2].$$

The set of all traces is a monoid with respect to the previous operation and it is usually denoted $\mathcal{M}(\Sigma, I)$. Given any monoid $\mathcal{M}$ and any function $f : \Sigma \longrightarrow A$ such that $f(a) \cdot f(b) = f(b) \cdot f(a)$ for each pair $(a, b) \in I$, there exits a unique monoid morphism $\overline{f} : \mathcal{M}(\Sigma, I) \longrightarrow A$ that extends f. For this reason, $\mathcal{M}(\Sigma, I)$ is termed *partially commutative free monoid generated by* (Σ, I). More briefly, we abbreviate it to *trace monoid generated by* Σ.

Special cases occur when I is the empty or the total irreflexive relation. If $I = \emptyset$, then the trace monoid generated by Σ is just the free monoid Σ^*. Similarly, if $(a, b) \in I$ for every $a \neq b$ in Σ, then we get the free commutative monoid $\Sigma^\otimes$ with generators in Σ. In this case, any element $\sigma_1^{i_1} \cdots \sigma_\ell^{i_\ell}$ of $\Sigma^\otimes$ is represented in the form $\underline{\sigma}^{\underline{i}}$, where $\underline{\sigma} = (\sigma_1, \ldots, \sigma_\ell)$ and $\underline{i} = (i_1, \ldots, i_\ell) \in \mathbb{N}^\ell$. The relationship between Σ^* and $\Sigma^\otimes$ is given by the *canonical monoid morphism* $\mathcal{F} : \Sigma^* \to \Sigma^\otimes$ defined, for each word $\omega \in \Sigma^*$, by $\mathcal{F}(\omega) = \underline{\sigma}^{\underline{i}}$, where $i_j = |\omega|_{\sigma_j}$ for every j.

The definition of formal power series can be given for a general trace monoid $\mathcal{M}(\Sigma, I)$ instead of Σ^*. In particular one can generalize the notions of rational and recognizable series. Most of the properties illustrated above still hold, but an important exception occurs as far as Schützenberger Representation Theorem is concerned. Indeed, one can prove that it does not hold when Σ^* is replaced by any other monoid.

Let us now introduce some notations relating to formal series in commutative variables. If $\mathcal{M} = \Sigma^{\otimes}$, then we use $\mathcal{S}[[\sigma_1, \cdots, \sigma_\ell]]$ to denote the family of all formal series in commutative variables, $\mathcal{S}^{Rat}[[\sigma_1, \cdots, \sigma_\ell]]$ to denote the set of $\mathcal{S}$-rational series over $\Sigma^{\otimes}$ and $\mathcal{S}[\sigma_1, \cdots, \sigma_\ell]$ to denote the set of polynomials. Observe that in this case, the notion of polynomial corresponds to the usual ones. The canonical morphism $\mathcal{F} : \Sigma^* \to \Sigma^{\otimes}$ extends to the semiring of formal series:

$$\begin{array}{ccc} \mathcal{S}\langle\langle \sigma_1, \cdots, \sigma_\ell \rangle\rangle & \longrightarrow & \mathcal{S}[[\sigma_1, \cdots, \sigma_\ell]] \\ r & \longmapsto & \mathcal{F}(r) \end{array}$$

where the commutative series $\mathcal{F}(r)$ is defined by setting

$$\left(\mathcal{F}(r), \underline{\sigma}^{\underline{i}}\right) = \sum_{\substack{|x|_{\sigma_j} = i_j \\ j=1,2,\ldots,\ell}} (r, x)$$

for every $\underline{\sigma}^{\underline{i}} \in \Sigma^{\otimes}$. The series $\mathcal{F}(r)$ is called the *commutative image* of r. Clearly, the extended map is a semiring morphism and therefore it preserves the rational operations. Thus, the commutative image $\mathcal{F}(r)$ of any $r \in \mathcal{S}^{Rat}\langle\langle \sigma_1, \cdots, \sigma_\ell \rangle\rangle$ is in $\mathcal{S}^{Rat}[[\sigma_1, \cdots, \sigma_\ell]]$.

We conclude this section by introducing trace languages over partially commutative free monoids. We define *trace language* any subset of $\mathcal{M}(\Sigma, I)$. Note that each language $L \subseteq \Sigma^*$ defines in a natural way the trace language

$$[L] = \{[\omega] \mid \omega \in L\}.$$

A trace language $T \subseteq \mathcal{M}(\Sigma, I)$ is said to be *rational* if there exists a regular language $L \subseteq \Sigma^*$ such that

$$T = \{[\omega] \mid \omega \in L\}.$$

An algebraic characterization of the rational trace language is based on the usual rational operations: the class of rational trace languages over $\mathcal{M}(\Sigma, I)$ coincides with the smallest class of subsets of $\mathcal{M}(\Sigma, I)$ containing all finite sets and closed with respect to the operations of union, product and closure.

Another natural class of languages is given by the set of all trace languages recognized by finite state automata over the trace monoid $\mathcal{M}(\Sigma, I)$, which can be defined extending the notion of finite automaton over Σ^*. It turns out that a trace language T is recognizable if and only if the language

$$\text{lin } T = \{\omega \in \Sigma^* \mid [\omega] \in T\}$$

is regular. It is easy to verify that every recognizable trace language is rational. Nevertheless the classes of rational and recognizable trace languages do not coincide unless the independence relation is empty, as Example 1.12 shows. In other word an analogous to Kleene's Theorem does not hold for rational trace languages.

Example 1.12 Let $\mathcal{M} = \{a, b\}^{\otimes}$ be the trace monoid generated by the independence alphabet $\{a, b\}$, where $(a, b) \in I$ and consider the regular language $L = (ab)^*$. Then, the trace language

$$[L] = \{[\omega] \in \mathcal{M} \mid |\omega|_a = |\omega|_b\}$$

is regular, but it is not recognizable, since

$$\text{lin } [L] = \{\omega \in \{a, b\}^* \mid |\omega|_a = |\omega|_b\}$$

is not a regular language. □

1.8 Growth of coefficients

Assume we are given a semiring $\mathcal{S}$ whose underlying set is contained in the reals, e.g., $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{R}_+$, etc. Then consider a formal series r over a trace monoid $\mathcal{M}$ with coefficients in $\mathcal{S}$ and denote by (r, ω) the corresponding coefficient of the element $\omega \in \mathcal{M}$. The evaluation of the growth of coefficients of r is a problem of interest in language theory, especially relating to the ambiguity of formal grammar (or finite automata) generating (recognizing, resp.) the support of the series. More formally, let us give the following definition.

Definition 1.13 *Given a monoid $\mathcal{M}$ and a semiring $\mathcal{S} \subseteq \mathbb{R}$, the* maximum function *associated with the series $r \in \mathcal{S}\langle\langle \mathcal{M} \rangle\rangle$ is defined as*

$$g_r(n) = \max_{|\omega|=n} |(r, \omega)|$$

where $|(r,x)|$ denotes the absolute value of (r,x), while $|\omega|$ is the length of ω.

For rational formal series over a free monoid with integer coefficients, the growth of the coefficients was investigated in [53] (see also [51]), where it is proved that for such a series r either there exists $k \in \mathbb{N}$ such that $g_r(n) = O(n^k)$ or $|(r, \omega_j)| \geq 2^{|\omega_j|}$ for a sequence of words $\{\omega_j\}$ of increasing length. In the first case, the series is the sum of products of at most $k+1$ characteristic series of regular languages over the free monoid (see also [5, Corollary 2.11.]).

Different results are obtained in the algebraic case (for precise definitions see [52]); a wide literature has been devoted to this problem (see for instance [42, 61, 64]). In particular, it has been recently proved that there are context-free grammars that have a logarithmic degree of ambiguity [63] and this implies the existence of algebraic formal series with logarithmic maximum function.

Similarly, the result cannot be extended to rational series over partially commutative free monoids. Given a trace language L and the trace language $T = [L]$, we call *ambiguity of a trace* $t \in T$ and *ambiguity of T of degree n* (with respect to L) the integers

$$Amb_L(t) = \#\{\omega \in t \cap L\} \qquad \text{and} \qquad Amb_L(n) = \max_{t \in T, |t|=n} Amb_L(t) \ .$$

From an example given in [62] one can prove that there exists a regular trace language, defined over the monoid generated by the independence alphabet $(\{a,b,c,d\}, \{(a,b),(b,a), (b,c),(c,b), (c,d),(d,c), (a,d),(d,a)\})$, that has a logarithmic ambiguity degree. Again, this implies the existence of rational series over trace monoids having logarithmic maximum function.

Referring to the growth of coefficients in a series, another case can be found in the literature, which is related to the tropical semiring $\mathbb{T}$. In [55], Imre Simon proves that for all $\mathbb{T}$-rational series r over the free monoid $\{a,b\}^*$, there exists an integer k such that $(r,w) = O(|w|^{1/k})$ holds for all $w \in \{a,b\}^*$. Moreover it is proven that for each positive integer k, there exist a $\mathbb{T}$-rational series r_k such that $g_{r_k}(n) = \Theta(n^{1/k})$. Thus, the hierarchy is strict though it is not proven that all series have an asymptotic growth of this kind.

As far as we know, the general problem of characterizing the order of magnitude of $g_r(n)$ for series in commutative variables is still open. In Section 4.5 we prove a result concerning the rational case with two commuting variables and coefficients in $\mathbb{R}_+$. For each $k \in \mathbb{N}$, we provide a class of $\mathbb{R}_+$-rational formal series in commutative variables whose maximum functions satisfy $g_r(n) = \Theta(n^{k-(1/2)}\lambda^n)$ for some positive real λ. This result somehow generalizes the following well-known property: given a rational fraction $p(x)/q(x)$ where $p(x)$ and $q(x)$ are two polynomials with coefficients in the field of real numbers (with $q(0) \neq 0$), the coefficient of the term x^n in its Taylor expansion is asymptotically equivalent to a linear combination of expressions of the form $n^{k-1}\lambda^n$, where λ is a root of $q(x)$ and k its multiplicity, cf. [35, Theorem 6.8] or [52, Lemma II.9.7].

Chapter 2

Non-negative matrices

In this chapter we deal with matrices with coefficients in a positive semiring. Our aim is to provide both classical and new tools that we shall apply in Part II. Thus, on the one hand we present the *Perron–Frobenius Theory* for matrices with coefficients in $\mathbb{R}_+$; on the other hand we introduce the notion of *symbol periodicity* for matrices with polynomial entries. Notice that, while the former is a well-known subject (the main result [25] dates back to 1908), the definition of symbol periodicity is very recent. Indeed, we introduced it in 2001 [7] to prove a local limit theorem in pattern statistics (see Section 4.4.2). However, such a notion and its properties are included in this chapter, since in a certain sense they extend the Perron–Frobenius Theory; moreover, strong analogies occur (compare for instance Theorem 2.12 and Proposition 2.25).

The chapter is organized as follows. We begin recalling the basic definitions and notations concerning matrices. In Section 2.2 we illustrate the general structure of a matrix with coefficients in a positive semiring and we give the definitions of primitive, irreducible and periodic matrices. In Section 2.3 we summarize the main results of Perron–Frobenius Theory for matrices with coefficients in $\mathbb{R}_+$. All proofs are omitted and can be found for instance in [54]. Section 2.4 is devoted to the notion of symbol periodicity for matrices with polynomial entries: after the definitions, we present some properties and in particular we focus on the eigenvalues of such matrices. Finally, we give some general notation on matrix functions in Section 2.5.

In Part II we shall apply these results to matrices associated with weighted automata (or linear representations). For this reason, this kind of matrices are used in this chapter as leading examples.

2.1 Basics on matrices

Given a semiring $\mathcal{S}$ and a finite set Q, consider the set $\mathcal{S}^{Q\times Q}$ of all *square matrices* $M : Q \times Q \to \mathcal{S}$. The elements in Q are referred to as *indices*; we write $M = (M_{pq})_{(p,q)\in Q\times Q}$ or simply $M = (M_{pq})$ and we say that that $M_{pq} \in \mathcal{S}$ is the *pq-component* (or *entry*) of M. Clearly, $\mathcal{S}^{Q\times Q}$ is a semiring with respect to the usual operations of sum and product between matrices. The neutral elements are the null matrix 0, whose entries are all null, and the *diagonal* matrix I, defined by setting $I_{qq} = 1$ for any index q and $I_{pq} = 0$ for every pairs of indices $p \neq q$. We also use $\mathcal{S}^Q$ to denote the set of *(column) vectors* with indices in Q and components in $\mathcal{S}$. Given a vector $v = (v_q)_{q\in Q}$, we indicate the corresponding *row vector* by v_T. For any index q, the row q of a matrix M is the vector $(M_{pq})_{q\in Q}$, while the column q of M is the vector $(M_{pq})_{p\in Q}$. Moreover we call the *transpose* of M the matrix M_T whose *pq*-entry is given by M_{qp}. To avoid the use of brackets, we use the expression ${M^n}_{pq}$ to denote the *pq*-entry of the matrix M^n.

If the semiring $\mathcal{S}$ is included in $\mathbb{R}$, then we say that M is *non-negative* (and we write $M \geq 0$)

whenever its entries are all greater than or equal to 0. If all entries are positive, we say that the matrix is *positive* too and we write $M > 0$. In general, if $\mathcal{S}$ is a positive semiring, then all matrices are said to be non-negative and M is positive if its components are all different from zero. A trivial remark: $M \neq 0$ and $M \geq 0$ do not imply $M > 0$!

One says that $\lambda \in \mathcal{S}$ is an *eigenvalue* of M if there exists a vector v such that $Mv = \lambda v$. In this case v is a *(right) eigenvector* of M related to λ. Then, λ is also an eigenvalue of M_T, that is there exists a vector u such that $\lambda u = M_T u = (u_T M)_T$. The vector u is also called *left eigenvalue* of M related to λ.

If in particular $\mathcal{S}$ is a field (that is each $a \in \mathcal{S}$ admits an opposite $-a \in \mathcal{S}$ such that $a + (-a) = 0$ and moreover the product is commutative), then we can define the *determinant* of a matrix M to be the value in $\mathcal{S}$

$$\mathrm{Det}(M) = \sum_{\rho} (-1)^{\sigma(\rho)} \cdot \prod_{q \in Q} M_{q\rho(q)} \ ,$$

where ρ is a permutation of the indices and $\sigma(\rho)$ is the number of inversions within ρ. One can prove that, for each pair of matrices A and B, one has $\mathrm{Det}(A \cdot B) = \mathrm{Det}(A) \cdot \mathrm{Det}(B)$. The determinant of the matrix $M - wI$ is a polynomial in the variable w and coefficients in $\mathcal{S}$. It is called the *characteristic polynomial* of M and its degree equals the cardinality of Q. It turns out that its roots are the eigenvalues of the matrix M.

For the sake of simplicity, let us now assume $Q = \{1, 2, \cdots m\}$. Then, for every pair of indices p, q we define *minor* the determinant m_{pq} of the matrix obtained by M deleting row p and column q. It turns out that, for any $p \in Q$, the determinant of $M = (M_{pq})$ can be written as

$$\mathrm{Det}(M) = \sum_{q \in Q} (-1)^{p+q} M_{pq} \cdot m_{pq} \ .$$

The matrix whose entries are $(-1)^{p+q} m_{pq}$ is called the *adjoint matrix* of M and is denoted by $\mathrm{Adj}(M)$. A matrix M admits an *inverse* M^{-1} such that $M \cdot M^{-1} = M^{-1} \cdot M = I$ if and only if the determinant of M is not null. Such an inverse is defined as

$$M^{-1} = \frac{\mathrm{Adj}(M)}{\mathrm{Det}(M)} \ .$$

2.2 Decomposition of matrices over a positive semiring

From now on, we take into consideration matrices with coefficients in a positive semiring $\mathcal{S}$ and we refer to such matrices as *nonnegative* ones. In particular, in this section we study the general structure of a non-negative matrix. To provide a clearer exposition, we shall represent matrices by means of graphs. Notice that we defined a similar correspondence between matrices and graphs in Section 1.5, when considering weighted automata.

Given a matrix $M \in \mathcal{S}^{Q \times Q}$, the *incidence graph* of M is a directed graph where the set of vertices is Q and an edge from index p to index q is drawn if $M_{pq} \neq 0$. Now, $\ell : q_0, \to q_1 \to \cdots \to q_n$, $n \geq 1$ is a *path* in the incidence graph if $q_i \in Q$ for each $i = 0, 1, \ldots, n$ and $M_{q_{i-1}, q_i} \neq 0$ for each $i = 1, 2, \ldots, n$ (if $q_0 = q_n$ we say that ℓ is a q_0*-cycle*). In other terms, ${M^n}_{pq} \neq 0$ if and only if there exists at least one path from p to q of length n. In this case, using the standard terms of graph theory, we say that p *leads to* q. This defines a partial order between indices.

If p leads to q and q leads to p, then we say that p and q *communicate* and we write $p \leftrightarrow q$. Now, consider the reflexive closure of $\leftrightarrow$. This defines an equivalence relation and its equivalence classes are called *strongly connected components* or *irreducible components* of the graph. For the sake of brevity we simply call them *components* of the matrix M.

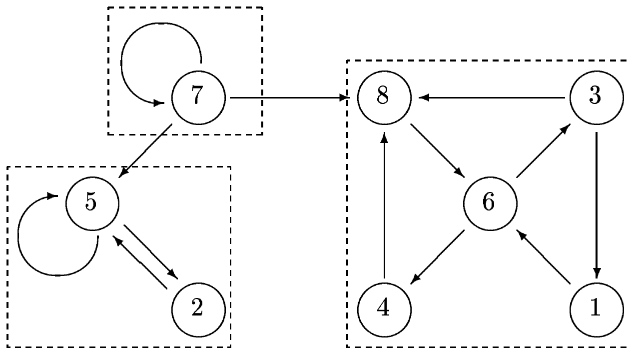


Figure 2.1: Incidence graph of the matrix of Example 2.1; the dashed boxes denote the irreducible components.

Example 2.1 Consider the following square matrix with index in $Q = \{1, 2, \cdots, 8\}$ and coefficients in $\mathbb{B}$

$M =$

	1	2	3	4	5	6	7	8
1						1		
2					1			
3	1							1
4								1
5		1			1			
6			1	1				
7					1		1	1
8						1		

Note that the null entries are omitted, while the indices for both rows and columns are shown. Fig. 2.1 illustrates the incidence graph of M and its components $\{1, 3, 4, 6, 8\}$, $\{2, 5\}$ and $\{7\}$. □

Definition 2.2 *If a non-negative matrix admits exactly one strongly connected component, it is said to be* irreducible.

Observe that if M is an irreducible matrix, then all indices communicate. Therefore, for each pair of indices p, q there exists an integer h such that ${M^h}_{pq} > 0$.

Definition 2.3 *Given a non-negative matrix M, for any index q, we call* period *of q the greatest common divisor (GCD) of the positive integers h such that ${M^h}_{qq} \neq 0$, with the convention that $GCD(\emptyset) = +\infty$.*

In other terms, to compute the period of an index q, one has to consider all q-cycles and take the greatest common divisor of their lengths. If $p \leftrightarrow q$, then it turns out that q and p have the same period; hence the *period of a component* is also well-defined. This yields the following

Definition 2.4 *The* period *of an irreducible matrix M is the common period of its indices. If such period is greater than one, then M is said to be* periodic.

We now introduce the definition of primitive matrices.

Definition 2.5 *A nonnegative matrix M is called* primitive *if there exists a positive integer h such that $M^h > 0$, which implies $M^n > 0$ for every $n \geq h$.*

Clearly, if a matrix is primitive, then it is also irreducible. More precisely, one can prove the following result.

Theorem 2.6 *A matrix M is primitive if and only if M is irreducible and has period 1.*

It should be emphasized that the definitions of irreducibility and primitivity, as the definition of strongly connected components of a matrix, are referred only to the positions of its non-null entries, whose exact values are not relevant. As an example consider a weighted automaton over the alphabet $\{a, b\}$, let M be its counting matrix and $M(x)$ its a-counting matrix. Then M and $M(x)$ have the same components; moreover M is irreducible (resp. primitive) if and only if $M(x)$ does.

Example 2.7 Consider the matrix $M \in \mathbb{B}^{2\times 2}$ defined by setting $M_{11} = M_{22} = 0$ and $M_{12} = M_{21} = 1$. Such a matrix M is irreducible, but not primitive, having period 2. These properties can be easily deduced from the incidence graph of M drawn in Fig. 2.2 or observing that for every odd integer h we have $M^h = M$ while for every even h M^h is the identity matrix. □

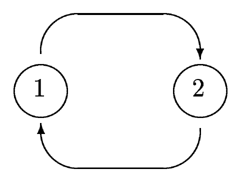


Figure 2.2: Incidence graph of the matrix of Example 2.7.

Let us now go back to a general matrix M and its components. The partial order defined in Q induces a partial order among such components: we say that the component C_1 *leads to* the component C_2 if there exist two indices $p \in C_1$ and $q \in C_2$ such that p leads to q. We also define *reduced graph* of M the directed acyclic graph where $C_1, C_2, \dots, C_s$ are the vertices and any pair (C_i, C_j) is an edge if and only if C_i leads to C_j.

As a consequence, any matrix $M \in \mathcal{S}^{Q\times Q}$ can be decomposed in a standard form, since one can always rearrange the set of indices Q, following the partial order among components. Thus, up to a permutation of indices, M can be written as a triangular block matrix of the form

$$M = \begin{pmatrix} M_1 & M_{12} & M_{13} & \cdots & M_{1s} \\ 0 & M_2 & M_{23} & \cdots & M_{2s} \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & M_s \end{pmatrix} \tag{2.1}$$

where each M_i corresponds to the strongly connected component C_i and every M_{ij} corresponds to the transitions from vertices of C_i to vertices of C_j in the incidence graph of M.

Example 2.8 Consider again the matrix of Example 2.1. Following the partial order among its components, we can rearrange the set of indices. Thus, up to a permutation, M can be written as

	7	2	5	1	3	4	6	8
7	1		1					1
2			1					
5		1	1					
1							1	
3				1				1
4								1
6					1	1		
8							1	

Remark that, since the order between indices is not total, neither the structure of the decomposition, nor the matrices M_i, M_{ij} are uniquely determined. For instance, when decomposing the matrix of Example 2.8, we may permutate the indices 2 and 5, belonging to the same class, thus obtaining a new matrix M_2. Or we may exchange the blocks M_2 and M_3 corresponding to the indices $\{2,5\}$ and $\{1,3,4,6,8\}$, respectively, thus determining new matrices M_{12}, M_{13}, M_{23}. Anyway, the number of components and the number of vanishing matrices M_{ij} are invariant.

2.3 The Perron–Frobenius Theory

When $\mathcal{S}$ is the semiring of positive real numbers a classical result is given by the following theorem (see [54, Theorem 1.1]).

Theorem 2.9 (Perron–Frobenius Theorem) *Let M be a primitive matrix with entries in $\mathbb{R}_+$. There exists an eigenvalue λ of M such that:*

1. *λ is real and positive;*
2. *with λ we can associate strictly positive left and right eigenvector;*
3. *$|\nu| < \lambda$ for every eigenvalue $\nu \neq \lambda$;*
4. *λ is a simple root of the characteristic polynomial of T;*
5. *the matrix $\mathrm{Adj}(\lambda I - M)$ is positive.*

In the sequel, we refer to the unique eigenvalue of maximum modulus of a primitive matrix as its *Perron–Frobenius eigenvalue.*

A first consequence of the previous theorem concerns the asymptotic growth of the entries of the n-th power of a primitive matrix M. More precisely, the following property holds [54, Theorem 1.2].

Proposition 2.10 *If M is a primitive matrix with entries in $\mathbb{R}_+$ and λ is its Perron–Frobenius eigenvalue, then*

$$M^n = \lambda^n \left(uv_{T} + C(n)\right) \qquad \text{for } n \to +\infty$$

where v and u are strictly positive left and right eigenvectors of M corresponding to the eigenvalue λ, normed so that $v_{T} u = 1$, while $C(n)$ is a real matrix such that each of its entries is $O(\varepsilon^n)$ for some $0 \leq \varepsilon < 1$.

A further application is given by the following proposition [54, Exercise 1.9], to be used in the next sections.

Proposition 2.11 *Let C be a complex matrix, set $|C| = (|C_{pq}|)$ and let γ be one of the eigenvalues of C. If M is a primitive matrix over $\mathbb{R}_+$ such that $|C_{pq}| \le M_{pq}$ for every p, q and if λ is its Perron–Frobenius eigenvalue, then $|\gamma| \le \lambda$. Moreover, if $|\gamma| = \lambda$, then necessarily $|C| = M$.*

If the matrix M is not primitive but it is irreducible, then the Perron-Frobenius theorem can be extended in the following sense [54, Theorems 1.5 and 1.7].

Theorem 2.12 (Perron-Frobenius Theorem for irreducible matrices) *Let M an irreducible matrix with period p. Then all the assertions of the Perron–Frobenius Theorem hold, except that* 3) *is replaced by the weaker statements: $|\nu| \le \lambda$ for any eigenvalue ν of M. Moreover there exist precisely p distinct eigenvalues with modulus λ, namely $\lambda \cdot z^k$, for $k = 0, 1, \ldots, p-1$, where z is the main p-th root of unity in $\mathbb{C}$; these eigenvalues are all simple roots of the characteristic polynomial of M.*

2.4 Symbol periodicity

In this section we introduce the notion of x-periodicity for matrices in the semiring $\mathcal{S}[x]$ of polynomials in the variable x with coefficients in $\mathcal{S}$ and focus more specifically on the case of irreducible matrices. We still assume $\mathcal{S}$ to be positive.

2.4.1 The definition of x-periodicity

Given a polynomial $F = \sum_k f_k x^k \in \mathcal{S}[x]$, we define the *$x$-period* of F as the integer $d(F) = \text{GCD}\{|h-k| \ : f_h \neq 0 \neq f_k\}$, where we assume $\text{GCD}(\{0\}) = \text{GCD}(\emptyset) = +\infty$. Observe that $d(F) = +\infty$ if and only if $F = 0$ or F is a monomial.

Now consider a matrix $M : Q \times Q \to \mathcal{S}[x]$. For any index $q \in Q$ and for each integer n we set $d(q,n) = d(M^n{}_{qq})$ and we define the x-period of q as the integer $d(q) = \text{GCD}\ \{d(q,n) \mid n \ge 0\}$, assuming that any non-zero element in $\mathbb{N} \cup \{+\infty\}$ divides $+\infty$. Notice that if M is the a-counting matrix of some linear representation, this definition implies that for every index q and for every pair of q-cycles $\mathcal{C}_1$ and $\mathcal{C}_2$ of equal length, $|\mathcal{C}_1|_a - |\mathcal{C}_2|_a$ is a multiple of $d(q)$. More precisely, $d(q)$ is the GCD of the differences of number of occurrences of a in all pairs of q-cycles of equal length.

Proposition 2.13 *If M is an irreducible matrix over $\mathcal{S}[x]$, then all indices have the same x-period.*

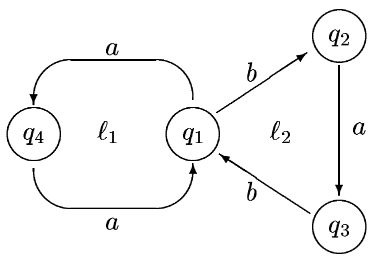
Proof. Consider an arbitrary pair of indices p, q. By symmetry, it suffices to prove that $d(p)$ divides $d(q)$, and this again can be proven by showing that $d(p)$ divides $d(q,n)$ for all $n \in \mathbb{N}$. As M is irreducible, there exist two integers s, t such that $M^s{}_{pq} \neq 0 \neq M^t{}_{qp}$. Then the polynomial $M^{s+t}{}_{pp} = \sum_r M^s{}_{pr} M^t{}_{rp} \neq 0$ and for some $k \in \mathbb{N}$ there exists a monomial in $M^{s+t}{}_{pp}$ with exponent k. Therefore, for every exponent h in $M^n{}_{qq}$, the integer $h+k$ appears as an exponent in $M^{n+s+t}{}_{pp}$. This proves that $d(p, n+s+t)$ divides $d(q,n)$ and since $d(p)$ divides $d(p, n+s+t)$, this establishes the result. □

Definition 2.14 *The x-*period *of an irreducible matrix over $\mathcal{S}[x]$ is the common x-period of its indices.*

Example 2.15 We compute the x-period of the a-counting matrix M over $\mathbb{B}[x]$ associated with the state diagram represented in Figure 2.3. Consider for instance state q_1 and let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two arbitrary q_1-cycles having the same length. Clearly they can be decomposed by using the simple q_1-cycles of the automaton, namely $\ell_1 = q_1 \xrightarrow{a} q_4 \xrightarrow{a} q_1$, $\ell_2 = q_1 \xrightarrow{b} q_2 \xrightarrow{a} q_3 \xrightarrow{b} q_1$. Hence, except for their order, $\mathcal{C}_1$ and $\mathcal{C}_2$ only differ in the number of cycles ℓ_1 and ℓ_2 they contain: for $k = 1, 2$, let $s_k \in \mathbb{Z}$ be the difference between the number of ℓ_k contained in $\mathcal{C}_1$ and the number of ℓ_k contained in $\mathcal{C}_2$. Then, necessarily, $s_1|\ell_1| + s_2|\ell_2| = 0$, that is $2s_1 + 3s_2 = 0$. This implies that $s_1 = 3n$ and $s_2 = -2n$ for some $n \in \mathbb{Z}$. Hence

$$|\mathcal{C}_1|_a - |\mathcal{C}_2|_a = 3n|\ell_1|_a - 2n|\ell_2|_a = 6n - 2n = 4n \ .$$

This proves that 4 is a divisor of the x-period of M. Moreover, both the q_1-cycles ${\ell_1}^3$ and ${\ell_2}^2$ have length 6 and the numbers of occurrences of a differ exactly by 4. Hence, in this case, the x-period of M is exactly 4. □



$$M = \begin{pmatrix} 0 & 1 & 0 & x \\ 0 & 0 & x & 0 \\ 1 & 0 & 0 & 0 \\ x & 0 & 0 & 0 \end{pmatrix}$$

Figure 2.3: State diagram and a-counting matrix of Example 2.15.

In the particular case where the entries of the matrix are all linear in x, the matrix decomposes $M = Ax + B$, where A and B are matrices over $\mathcal{S}$; this clearly happens when M is the a-counting matrix of some linear representation. If further M is primitive, the following proposition holds.

Proposition 2.16 *Let A and B be matrices over $\mathcal{S}$ and set $M = Ax + B$. If M is primitive and $A \neq 0 \neq B$, then the x-period of M is finite.*

Proof. Let q be an arbitrary index and consider the finite family of pairs $\{(n_j, k_j)\}_{j \in J}$ such that $0 \leq k_j \leq n_j \leq m$ where m is the size of M and k_j appears as an exponent in ${M^{n_j}}_{qq}$. Notice that since M is irreducible J is not empty. Since every cycle can be decomposed into elementary cycles all of which of length at most m, the result is proved once we show that $d(q) = +\infty$ implies either $k_j = 0$ for all $j \in J$ or $k_j = n_j$ for all $j \in J$: in the first case we get $A = 0$ while in the second case we have $B = 0$.

Because of equality $M^{\prod_j n_j} = (M^{n_i})^{\prod_{j \neq i} n_j}$, the polynomial ${M^{\prod_j n_j}}_{qq}$ contains the exponent $k_i \prod_{j \neq i} n_j$ for each $i \in J$. Now, suppose by contradiction that $d(q)$ is not finite. This means that all exponents in ${M^{\prod_j n_j}}_{qq}$ are equal to a unique integer h such that $h = k_i \prod_{j \neq i} n_j$ for all $i \in J$. Hence, h must be a multiple of the least common multiple of all products $\prod_{j \neq i} n_j$. Now we have $\text{LCM}\{\prod_{j \neq i} n_j \mid i \in J\} \cdot \text{GCD}\{n_j \mid j \in J\} = \prod_j n_j$ and by the primitivity hypothesis $\text{GCD}\{n_j \mid j \in J\} = 1$ holds. Therefore h is a multiple of $\prod_j n_j$. Thus the conditions $k_j \leq n_j$ leave the only possibilities $k_j = 0$ for all $j \in J$ or $k_j = n_j$ for all $j \in J$. □

Observe that the previous theorem cannot be extended to the case when M is irreducible or when M is a matrix over $\mathcal{S}[x]$ that cannot be written as $Ax + B$ for some matrices A and B over $\mathcal{S}$.

Example 2.17 The matrix M with entries $M_{11} = M_{22} = 0$, $M_{12} = x$ and $M_{21} = 1$ is irreducible but it is not primitive since it has period 2. It is easy to see that the non-null entries of all its powers are monomials, thus M has infinite x-period. □

Example 2.18 Consider again Figure 2.3 and set $M_{2,3} = x^3$. Then we obtain a primitive matrix over $\mathbb{B}[x]$ that cannot be written as $Ax + B$ and does not have finite x-period. □

2.4.2 Properties of x-periodic matrices

Given a positive integer d, consider the cyclic group $C_d = \{1, g, g^2, \ldots, g^{d-1}\}$ of order d and the semiring $\mathcal{B}_d = \langle 2^{C_d}, +, \cdot \rangle$ (which is also called $\mathbb{B}$-algebra of the cyclic group) where 2^{C_d} denotes the family of all subsets of C_d and for every pair of subsets A, B of C_d we set $A + B = A \cup B$ and $A \cdot B = \{a \cdot b \mid a \in A, b \in B\}$; hence $\emptyset$ is the unit of the sum and $\{1\}$ is the unit of the product. Now, given a positive semiring $\mathcal{S}$, consider the map $\varphi_d : \mathcal{S}[x] \to \mathcal{B}_d$ that associates any polynomial $F = \sum_k f_k x^k \in \mathcal{S}[x]$ with the set $\{g^k \mid f_k \neq 0\} \in \mathcal{B}_d$. Note that since the semiring $\mathcal{S}$ is positive φ_d is a semiring morphism. Intuitively, φ_d associates F with the set of its exponents modulo the integer d. Of course φ_d extends to the semiring of $Q \times Q$-matrices over $\mathcal{S}[x]$ by setting $\varphi_d(T)_{pq} = \varphi_d(T_{pq})$, for every matrix $T : Q \times Q \to \mathcal{S}[x]$ and all $p, q \in Q$. Observe that, since φ_d is a morphism, ${\varphi_d(T)^n}_{pq} = \varphi_d(T^n)_{pq} = \varphi_d({T^n}_{pq})$.

Now, let $M : Q \times Q \to \mathcal{S}[x]$ be an irreducible matrix with finite x-period d. Simply by the definition of d and φ_d, we have that for each $n \in \mathbb{N}$ all non-empty entries $\varphi_d(M^n)_{pp}$ contains at most 1 element. The following results also concern the powers of $\varphi_d(M)$.

Proposition 2.19 *Let M be an irreducible matrix over $\mathcal{S}[x]$ with finite x-period d. Then, for each integer n and each pair of indices p and q, the set ${\varphi_d(M)^n}_{pq}$ contains at most one element; moreover, if $\varphi_d(M)_{qq} \neq \emptyset$, then ${\varphi_d(M)^n}_{qq} = (\varphi_d(M)_{qq})^n$.*

Proof. Let n be an arbitrary integer and p, q an arbitrary pair of indices. By the remarks above we may assume $p \neq q$ and ${M^n}_{pq} \neq 0$. M being irreducible, there exists an integer t such that ${M^t}_{qp} \neq 0$. Note that if B is a non–empty subset of C_d, then $|A \cdot B| \geq |A|$ holds for each $A \subseteq C_d$ and ${\varphi_d(M)^{n+t}}_{pp} \supseteq {\varphi_d(M)^n}_{pq} \cdot {\varphi_d(M)^t}_{qp}$. Therefore, since $|{\varphi_d(M)^{n+t}}_{pp}| \leq 1$, we have also $|{\varphi_d(M)^n}_{pq}| \leq 1$. The second statement is proved in a similar way reasoning by induction on n. □

Proposition 2.20 *Let M be an irreducible matrix over $\mathcal{S}[x]$ with finite x-period d. Then, for each integer n, all non-empty diagonal elements of $\varphi(M)^n$ are equal.*

Proof. Let n be an arbitrary integer and let p, q be an arbitrary pair of indices such that ${M^n}_{pp} \neq 0 \neq {M^n}_{qq}$. By the previous proposition, there exist h, k such that ${\varphi(M)^n}_{pp} = \{g^h\}$ and ${\varphi(M)^n}_{qq} = \{g^k\}$. If t is defined as in the previous proof, then the two elements ${\varphi(M)^t}_{qp} \cdot \{g^h\}$ and $\{g^k\} \cdot {\varphi(M)^t}_{qp}$ belong to ${\varphi(M)^{t+n}}_{qp}$; since this subset contains only one element they must be equal and this completes the proof. □

Proposition 2.21 *Let M be a primitive matrix over $\mathcal{S}[x]$ with finite x-period d. There exists an integer $0 \leq \gamma < d$ such that for each integer n and each index q, if ${M^n}_{qq} \neq 0$, then ${\varphi_d(M)^n}_{qq} = \{g^{\gamma n}\}$. Moreover, for each pair of indices p, q and for any integer n such that ${M^n}_{pq} \neq 0$, we have $\varphi_d\left({M^n}_{pq}\right) = \{g^{\gamma n + \delta_{pq}}\}$ for a suitable integer $0 \leq \delta_{pq} < d$ independent of n.*

Proof. Since M is primitive, there exists an integer t such that ${M^n}_{pq} \neq 0$ for every $n \geq t$ and for every pair of indices p and q. In particular, since $dt+1 > t$, we have $|\varphi_d({M^{dt+1}}_{qq})| = 1$ for each q and hence there exists $0 \leq \gamma < d$ such that ${\varphi_d(M)^{dt+1}}_{qq} = \{g^\gamma\}$. Observe that γ does not depend on q, by Proposition 2.20. Therefore, by Proposition 2.19, we have

$$\{g^{\gamma n}\} = {\varphi_d(M)^{dtn+n}}_{qq} \supseteq {\varphi_d(M)^{dtn}}_{qq} \cdot {\varphi_d(M)^n}_{qq} = \{1\} \cdot {\varphi_d(M)^n}_{qq}$$

which proves the first part of the statement.

Now, consider an arbitrary pair of indices p, q and let t be the smallest positive integer such that ${M^t}_{pq} \neq 0$ (the existence of such t is guaranteed by the primitivity of M). Then, for each integer n, we have

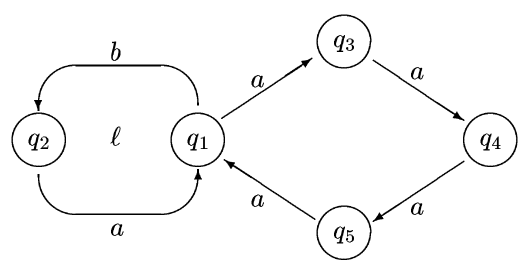
$${\varphi_d(M)^t}_{qp} \cdot {\varphi_d(M)^n}_{pq} \subseteq {\varphi_d(M)^{n+t}}_{qq} = \{g^{\gamma(n+t)}\} \; .$$

Moreover, by Proposition 2.19 we know that there exists an exponent k such that ${\varphi_d(M)^t}_{qp} = \{g^k\}$. This yields the result, by denoting by δ_{pq} the congruence class of $\gamma t - k$ modulo d. □

If M is the a-counting matrix of a linear representation, then the previous propositions can be interpreted by considering its state diagram. For any pair of states p, q, all paths of the same length starting in p and ending in q have the same number of occurrences of a modulo d. Secondly, if $\mathcal{C}_k$ is a q_k-cycle for $k = 1, 2$ and $\mathcal{C}_1$ and $\mathcal{C}_2$ have the same length, then they also have the same number of occurrences of a modulo d. Finally, if M is primitive, for each cycle ℓ we have $|\ell|_a = \gamma|\ell|$ modulo d for some integer γ and moreover, for every pair of states p, q there exists a constant δ_{pq} such that the number of a in any path of length n from p to q is given by $\gamma n + \delta_{pq}$.

We conclude this section with an example showing that Proposition 2.21 cannot be extended to the case when M is irreducible but not primitive.

Example 2.22 Consider the a-counting matrix M associated with the state diagram of Figure 2.4. Then M is irreducible with x-period 2, but it is not primitive since also its period equals 2. Consider the path $\ell = q_1 \xrightarrow{b} q_2 \xrightarrow{a} q_1$. We have $|\ell| = 2$ and $|\ell|_a = 1$, hence for any γ, $\gamma|\ell|$ cannot be equal to $|\ell|_a$ modulo 2. Thus, Proposition 2.21 does not hold in this case.



$$M = \begin{pmatrix} 0 & 1 & x & 0 & 0 \\ x & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & 0 \\ 0 & 0 & 0 & 0 & x \\ x & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 2.4: State diagram and matrix of Example 2.22.

2.4.3 Eigenvalues of x-periodic matrices

In this section we consider the semiring $\mathbb{R}_+$ of non–negative real numbers and we study the eigenvalues of primitive matrices $M(x)$ over $\mathbb{R}_+[x]$ when x assumes the complex values z such that $|z| = 1$. The next theorem shows how the eigenvalues of $M(z)$ are related to the x-period of the matrix. To this end we first give two auxiliary lemmata.

Lemma 2.23 *Let M be an irreducible matrix over $\mathcal{S}[x]$ with finite x-period d. Then for every index q there exist an integer n and two exponents h and k appearing in ${M^n}_{qq}$ such that $h - k = d$. If further M is primitive, then the property holds for every n large enough.*

Proof. Consider an arbitrary integer q. By the definition of $d = d(q)$ there exists a finite set of integers s_j such that $d = \sum_j s_j(r_j - k_j)$, where r_j and k_j are exponents appearing in ${M^{n_j}}_{qq}$ for some integer n_j. Observe that, since r_j and k_j can be exchanged, we may assume positive all coefficients s_j. Now set $h = \sum_j s_j r_j$ and $k = \sum_j s_j k_j$. Then $d = h - k$ holds and both h and k are exponents in ${M^n}_{qq}$ where $n = \sum_j s_j n_j$. This proves the statement if M is irreducible.

If further M is primitive, then there exists an integer t such that for each $m \geq t$, ${M^m}_{qq} \neq 0$ and hence ${M^m}_{qq}$ has a non-null coefficient of degree l, for some $l \in \mathbb{N}$. Thus, $h + l$ and $k + l$ are exponents that appear in ${M^{n+m}}_{qq}$ for each integer $m \geq t$ and this completes the proof. □

In the sequel, $|T|$ denotes the matrix with pq-entry given by $|T_{pq}|$.

Lemma 2.24 *Let $M(x)$ be a primitive matrix over $\mathbb{R}_+[x]$ with finite x-period d and set $M = M(1)$. Then, for every integer n large enough and for each $z \in \mathbb{C}$ such that $|z| = 1$, $|\ M(z)^n\ | = M^n$ if and only if z is a d-th root of unity.*

Proof. Given $n \in \mathbb{N}$ and a pair of indices p, q, let ${M(x)^n}_{pq} = \sum_{j=1}^{l} f_j x^{k_j}$. By Proposition 2.19, all the exponents $k_1, k_2, \ldots, k_l$ are congruent modulo d. Then, for every $z \in \mathbb{C}$ we can write

$$
{M(z)^n}_{pq} = z^{k_1}(f_1 + \sum_{j=2}^{l} f_j z^{s_j})
$$

where each $s_j = k_j - k_1$ is multiple of d, for $j = 2, \ldots, l$. As a consequence, if $z = e^{\frac{2k\pi}{d}i}$ for some $k \in \mathbb{Z}$, then ${M(z)^n}_{pq} = z^{k_1} {M^n}_{pq}$ proving the result in one direction.

On the other hand, let n be an integer large enough to satisfy Lemma 2.23 for any index q and consider some diagonal entry ${M(x)^n}_{qq} = \sum_{j=1}^{l} f_j x^{k_j}$. By the previous lemma we may assume $d = k_2 - k_1$ and hence, setting $s_j = k_j - k_1$ we have $d = \text{GCD}\{s_j \mid j = 2, \ldots, l\}$. Now assume $|\ {M(z)^n}_{qq}\ | = \sum_j f_{k_j} = {M^n}_{qq}$ for some $z = e^{i\theta}$ with $0 \leq \theta < 2\pi$. This implies that each θs_j is multiple of 2π and hence for all $j = 2, \ldots, l$ we have

$$
\frac{\theta}{2\pi} = \frac{p_j}{s_j} = \frac{h}{s} \tag{2.2}
$$

where $p_j \in \mathbb{Z}$, $s = \text{LCM}\ \{s_j \mid j = 2, \ldots, l\}$ and $h < s$ is a non-negative integer. Since h is multiple of each s/s_j it is also multiple of $s' = \text{LCM}\ \{s/s_j \mid j = 2, \ldots, l\}$. Now, since GCD $\{s_j \mid j = 2, \ldots, l\} = d$, we have that $s' = s/d$. Thus, being $\theta = 2\pi h/s$ by (2.2), we have that θ is a multiple of $2\pi/d$ and hence $z = e^{i\theta}$ is a d-th root of unity. □

Proposition 2.25 *Let $M(x)$ is a primitive matrix over $\mathbb{R}_+$ with x-period d and denote by γ the constant introduced in Proposition 2.21. If z is a d-th root of unity in $\mathbb{C}$ and ν is an eigenvalue of M, then νz^γ is an eigenvalue of $M(z)$ with the same algebraic multiplicity.*

Proof. The case $d = 1$ is trivial; thus suppose $d > 1$ and assume that z is a $d-$th root of unity. Set $\hat{T} = I\nu z^\gamma - M(z)$ and $T = I\nu - M$. We now verify that $\text{Det}\hat{T} = z^{\gamma m}\ \text{Det}\ T$ holds where m is the size of M. To prove this equality, recall that

$$
\text{Det}\ \hat{T} = \sum_{\rho} (-1)^{\sigma(\rho)} \prod_{q \in Q} \hat{T}_{q\rho(q)}\ .
$$

By Proposition 2.21, since z is a d-th root of 1 in $\mathbb{C}$, we have $\hat{T}_{qq} = (\nu - M_{qq}) z^\gamma = z^\gamma T_{qq}$ for each state q and $\hat{T}_{q_0 q_1} \cdots \hat{T}_{q_{s-1} q_0} = z^{\gamma s}\ T_{q_0 q_1} \cdots T_{q_{s-1} q_0}$ for each simple cycle $(q_0, q_1, \ldots, q_{s-1}, q_0)$ of length $s > 1$. Therefore, for each permutation ρ, we get

$$
\prod_{q \in Q} \hat{T}_{q\rho(q)} = z^{\gamma m} \cdot \prod_{q \in Q} T_{q\rho(q)}
$$

which concludes the proof. □

Theorem 2.26 *Let $M(x)$ be a primitive matrix over $\mathbb{R}_+[x]$ with finite x-period d, set $M = M(1)$ and let λ be the Perron-Frobenius eigenvalue of M. Then, for all $z \in \mathbb{C}$ with $|z| = 1$, the following conditions are equivalent:*
1. $M(z)$ and M have the same set of moduli of eigenvalues;
2. If $\lambda(z)$ is an eigenvalue of maximum modulus of $M(z)$, then $|\lambda(z)| = \lambda$;
3. z is a d-th root of unity in $\mathbb{C}$.

Proof. Clearly condition 1) implies condition 2). To prove that condition 2) implies condition 3) we reason by contradiction, that is we assume that z is not a d-th root of unity. By Lemma 2.24 in this case there exists an integer n such that $| M(z)^n | \neq M^n$. Therefore we can apply Proposition 2.11 and prove that λ^n is greater than the modulus of any eigenvalue of $M(z)^n$. In particular we have $\lambda^n > |\lambda(z)|^n$ which contradicts the hypotheses. Finally condition 3) implies condition 1) as a consequence of Proposition 2.25. □

Example 2.27 Let us consider again the primitive matrix of Figure 1.2. We recall that here $d = 4$; moreover it is easy to see that $\gamma = 3$. Indeed, for each $k = 1, 2$, we have that $|\ell_k| - 3|\ell_k|_a$ is equal to 0 modulo 4. Now consider the characteristic polynomial of the matrix $M(x)$, given by $\chi_x(y) = y^4 - y^2x^2 - yx$ and let ν be a root of χ_1. This implies that $\chi_1(\nu) = \nu^4 - \nu^2 - \nu = 0$ and hence $-i\nu$ is a root of the polynomial χ_i, $-\nu$ is a root of the polynomial χ_{-1} and $i\nu$ is a root of the polynomial χ_{-i}. This is consistent with Theorem 2.26, since 1, i, -1 and $-i$ are the fourth roots of unity. □

2.5 Notations on matrix functions

We conclude this chapter with some general notations on matrix functions. Assume that $A(x)$ is a square matrix whose entries are complex functions in the variable x. The derivative of $A(x)$ with respect of x is the matrix $D_xA(x) = [A'(x)_{ij}]$ of its derivatives. Thus, if $A(x)$ and $B(x)$ are square matrices of the same size, then the following identities can be easily proved:

$$D_x(A(x) \cdot B(x)) = D_xA(x) \cdot B(x) + A(x) \cdot D_xB(x) \ , \tag{2.3}$$

$$D_x(A(x)^n) = \sum_{i=1}^{n} A(x)^{i-1} \cdot D_xA(x) \cdot A(x)^{n-i},$$

$$D_x(A(x)^{-1}) = -A(x)^{-1} \cdot D_xA(x) \cdot A(x)^{-1}. \tag{2.4}$$

Moreover, the traditional big-O notation can be extended to matrix functions: let $A(x)$ be defined in an open domain $E \subseteq \mathbb{C}$, let $g(x)$ be a complex function also defined in E and let x_0 be an accumulation point of E; as x tends to x_0 in E, we write $A(x) = O(g(x))$ to mean that for every pair of indices i, j, $A(x)_{ij} = O(g(x))$, namely there exists a positive constant c such that for every x in E near x_0

$$|A(x)_{ij}| \leq c\,|g(x)| \, .$$

Note that, if the entries of $A(x)$ are analytic at a point $x_0 \in E$, then

$$A(x) = A(x_0) + A'(x_0)(x - x_0) + O\left((x - x_0)^2\right).$$

Further, if some entries of $A(x)$ have a pole of degree 1 at a point x_0 in the boundary of E while the others (if any) are analytic at the same point, then

$$A(x) = \frac{R}{x - x_0} + S + O(x - x_0)$$

for suitable matrices R and S ($R \neq 0$).

Chapter 3

Limit theorems in probability theory

This chapter concerns probability theory, and in particular it focuses on central and local limit theorems for sequences of random variables.

In first section we recall some basic notions and present some typical examples of probability distributions. In Section 3.4 we consider a sequence of Bernoulli trials, that is one of the most classical scheme in probability theory. In particular we focus on DeMoivre–Laplace limit theorems and their extensions to partial sums of more general sequences of random variables. In Section 3.5 we present the Markov processes, which can be seen as generalizations of the Bernoulli scheme. In the last part of the chapter, we take into consideration a general sequence of discrete random variables, without assuming any condition of independence, and we present two criteria to establish central and local limit properties: in Section 3.6 we state the "quasi-power theorem" which provides an useful approach to prove convergence in distributions, while in Section 3.7 we prove a general criterion that guarantees the existence of a local convergence property of a Gaussian type in the sense of Local DeMoivre–Laplace Theorem. In Part II, using such criterion, we show that the same local convergence property holds for certain pattern statistics.

3.1 Probability spaces and random variables

Given a set Ω, a *σ-algebra* $\mathcal{D}$ is a non-empty collections of subsets of Ω such that: the empty set is in $\mathcal{D}$; if D is in $\mathcal{D}$, then its complement D^c is in $\mathcal{D}$; if $\{D_n\}_n$ is a sequence of elements in $\mathcal{D}$, then $\cup_n D_n$ is in $\mathcal{D}$. The pair $(\Omega, \mathcal{D})$ is called a *measurable space* and any set in $\mathcal{D}$ is said to be *measurable*. Moreover, a function $f : \Omega \to \mathbb{R}$ is called *measurable* if, for every real number x, the set $\{a \in \Omega \mid f(a) > x\}$ is measurable.

A *probability measure* on $(\Omega, \mathcal{D})$ is defined as a nonnegative real function $P : \mathcal{D} \to \mathbb{R}$ such that $P(\emptyset) = 0$, $P(\Omega) = 1$ and, if $\{D_n\}_n$ is a finite or countable collection of pairwise disjoint sets in $\mathcal{D}$ and $D = \cup_n D_n$, then $P(D) = \sum_n P(D_n)$. The triple $(\Omega, \mathcal{D}, P)$ is called a *probability space* on the domain Ω, which alone is defined *sample space*. The elements in $\mathcal{D}$ are called *sample points* and the sets in $\mathcal{D}$ are also called *events*.

As an example, consider a countable set Ω and let 2^Ω denote the power set of Ω. For each $a \in \Omega$ let $P(a)$ be a nonnegative real value such that $\sum_{a\in\Omega} P(a) = 1$. Finally, for any subset $D \subseteq \Omega$ set $P(D) = \sum_{a\in D} P(a)$. Then, $(\Omega, 2^\Omega, P)$ turns out to be a probability space, and (Ω, P) is usually called *discrete sample space*.

Given a probability space $(\Omega, \mathcal{D}, P)$, take two arbitrary events D and $E \in \mathcal{D}$. If $P(E) \neq 0$, we define *conditional probability* of D given E the value

$$P(D|E) = \frac{P(D \cap E)}{P(E)} \ .$$

By symmetry, it is clear that $P(D \cap E) = P(D|E) \cdot P(E) = P(E|D) \cdot P(D)$. For a fixed event D, the function $P_D : \mathcal{D} \to \mathbb{R}$ such that $P_D(E) = P(E|D)$ for any $E \in \mathcal{D}$ satisfies the axioms of probability measure. For this reason, the expression *conditional probability* makes sense.

Given a probability space $(\Omega, \mathcal{D}, P)$, a *random variable* (abbreviated "r.v.") is a measurable function $X : \Omega \to \mathbb{R}$. For any real x, the set $\{a \in \Omega \mid X(a) \leq x\}$ is simply denoted by $\{X \leq x\}$ and we also write $P\ \{X \leq x\}$ to indicate the real value $P\ (\{a \in \Omega \mid X(a) \leq x\})$. With obvious meanings, this notation extends to expressions like $\{a \in \Omega \mid X(a) = x\}$, $\{a \in \Omega \mid X(a) < x\}$ and similar ones.

We say that the r.v.'s $X_1, X_2, \ldots X_n$ are *independent* if

$$P\ \{X_1 \leq x_1, X_2 \leq x_2, \ldots X_n \leq x_n\} = P\{X_1 \leq x_1\} \cdot P\{X_2 \leq x_2\} \cdots P\{X_n \leq x_n\}$$

holds for all real values x_i, $i = 1, 2, \ldots, n$.

The *(cumulative) distribution function* F_X of a r.v. X describes the probability that X takes on a value less than or equal to a real number x. More formally, $F_X : \mathbb{R} \to [0, 1]$ is defined by

$$F_X(x) = P\ \{X \leq x\} \ .$$

Note that the distribution function is well-defined for any r.v. X, since the set $\{X \leq x\}$ is always measurable by definition. Moreover, F_X is non-decreasing, right continuous, and we have $lim_{x \to -\infty} F_X(x) = 0$, $lim_{x \to +\infty} F_X(x) = 1$.

A r.v. X is said to be *discrete* if X takes on values in a countable set $\{x_1, x_2, \ldots\}$. In this case, the values x_j are called *mass points* and the function $f_X : \mathbb{R} \to \mathbb{R}_+$ such that

$$f_X(x) = \begin{cases} P\{X = x_j\} & \text{if } x = x_j, j = 1, 2, \ldots \\ 0 & \text{if } x \neq x_j \end{cases}$$

is called *probability function* or *discrete density function* of X. The distribution function of a discrete r.v. is also called *discrete*. We remark that if X is a discrete r.v., then F_X can be derived from f_X and viceversa.

On the contrary, a r.v. X is called *continuous* if there exists a function f_X such that for every real x we have

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\ dx \ .$$

The function $f_X(x)$ is termed the *(probability) density function* of X. If X is continuous, then its distribution is differentiable almost everywhere in $\mathbb{R}$, or $P\{x \in \mathbb{R} | F_X$ is not differentiable in $x\} = 0$. As in the discrete case, F_X can be derived from f_X and viceversa. In particular we have $dF_X(x)/dx = f_X(x)$ almost everywhere. Moreover, $f_X(x) \geq 0$ for every real x and $\int_{-\infty}^{+\infty} f_X(x)\ dx = 1$. In general, if a function f satisfies this two properties, we call it a *density function.* If f_X admits a unique local maximum point x, then f_X is said to be *unimodal.*

3.2 Moments and characteristic function

Let X be a r.v. and consider a function $g : \mathbb{R} \to \mathbb{R}$. We call *expectation* of $g(X)$ the value $\mathbb{E}(g(X))$ defined as follows: if X is discrete with mass points $x_1, x_2, \ldots$, then

$$\mathbb{E}(g(X)) = \sum_j g(x_j) f_X(x_j) \ ;$$

if X is continuous, then

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

provided (in both cases) that the right hand side converges. Clearly, $\mathbb{E}(c) = c$ for each constant c; moreover the expectation is linear, that is $\mathbb{E}(c_1 g_1(X) + c_2 g_2(X)) = c_1 \mathbb{E}(g_1(X)) + c_2 \mathbb{E}(g_2(X))$. One should also notice that the expectation is referred to the density function f_X and it could be defined without reference to the r.v. X.

The *moments* of a r.v. X are the expectation values of the powers of X. More precisely, we define *moment of order r* the value $\mu_r = \mathbb{E}(X^r)$, if it exists. Observe that if X can take on values only in a finite set, then the moments are always well-defined. The moment $\mathbb{E}(X)$ of order 1 is called *mean* of X and it is denoted by μ_X. Thus, if X is discrete we have

$$\mu_X = \mathbb{E}(X) = \sum_j x_j f_X(x_j)$$

while, if X is continuous, then

$$\mu_X = \mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx \ .$$

The *central moments* are defined by $\mu_r = \mathbb{E}((X - \mathbb{E}(X))^r)$. It is easy to see that $\mu_1 = 0$ and

$$\mu_2 = \mathbb{E}((X - \mu_X)^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \ .$$

Such a value μ_2 is called the *variance* of X and is also denoted $\mathbb{V}ar(X)$. The *standard deviation* of X is the square root of the variance and it is usually denoted by σ.

Given a r.v. X with density function f_X, set

$$\Psi_X(z) = \mathbb{E}(e^{zX})$$

for every complex z such that $\mathbb{E}(e^{zX})$ exists. Wherever Ψ_X exists, it is continuously differentiable in a neighbourhood of $t = 0$ and its r-th derivative satisfies the following relation

$$\frac{d\Psi}{dt}(0) = \mathbb{E}(X^r)$$

which holds both in the continuous and in the discrete case. Thus, the moments of a distribution can be obtained from the function Ψ_X which is named *moment generating function* of X just from this fact. In particular, we have

$$\mathbb{E}(X) = \Psi'_X(0) \ , \qquad \mathbb{E}(X^2) = \Psi''_X(0) \ . \tag{3.1}$$

It should be emphasized that, even though the distribution of a r.v. is not uniquely specified by its moments, it is uniquely specified by its moment generating function, Indeed, the following theorem holds.

Theorem 3.1 *If X and Y are two r.v.'s such that $\Psi_X(z)$ and $\Psi_Y(z)$ exist and are equal at $|z| < h$, for some $h > 0$, then the distribution functions F_X and F_Y also coincide.*

The restriction $\Psi_X(i\theta)$, for $\theta \in \mathbb{R}$, is called the *characteristic function* of X. Observe that $\Psi_X(i\theta)$ is well-defined for very real θ and it completely characterizes the distribution of X, thanks to the previous theorem. Moreover, if X takes on values in $\mathbb{N}$, then $\Psi_X(i\theta)$ is periodic in θ of period 2π and assumes the value 1 at $\theta = 0$. We finally remark that the moment generating function and the characteristic function can be defined by means of the Laplace and Fourier Transform, respectively (see [11]).

3.3 Examples of distribution laws

Here we briefly present some typical examples of distribution laws, either discrete or continuous, we will use in the sequel. We just focus on the distributions, thus we do not refer to any specific probability space or r.v., but to the family of all r.v.'s that obey to the given probabilistic law. Some authors use the term *variate* to refer to such a family.

Let us first describe some discrete distributions.

Discrete uniform distribution. The discrete uniform distribution of parameter N is characterized by the following density function

$$f_N(x) = \begin{cases} 1/N & \text{if } x = 1, 2, \cdots N \\ 0 & \text{otherwise} \end{cases}$$

The discrete uniform distribution is also known as the *equally likely outcomes* distribution.

Bernoulli distribution. The Bernoulli distribution is typical of a *Bernoulli trial*, that can be seen as a simple performance of a well-define experiment (as, for instance, the flipping of a coin) having two possible outcomes 0 and 1, in which 1 (*success*) occurs with probability p and 0 (*failure*) occurs with probability q, where $q = 1 - p$. Therefore, it is a discrete distribution having probability function

$$f_p(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

If X has a Bernoulli distribution, then $\mathbb{E}(X) = p$, $\mathbb{V}ar(X) = p(1-p)$ and the moment generating function is given by $\Psi_X(z) = 1 - p + pe^z$.

Binomial distribution. The binomial distribution gives the discrete probability distribution of obtaining exactly m successes out of n independent Bernoulli trials (where the result of each Bernoulli trial is true with probability p and false with probability $1 - p$). The binomial distribution is therefore characterized by the following probability function

$$f_{n,p}(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

Notice that for $n = 1$ we obtain the Bernoulli distribution. If X has a binomial distribution, then $\mathbb{E}(X) = p(1-p)$, $\mathbb{V}ar(X) = np(1-p)$ and the moment generating function is given by $\Psi_X(z) = (1 - p + pe^z)^n$.

Geometric distribution. The geometric distribution (or *Pascal distribution*) of parameter $0 < p \leq 1$ is characterized by the following density function

$$f_p(x) = \begin{cases} pq^x & \text{if } x = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $q = 1 - p$. Then, the distribution function is $F_p(x) = 1 - q^{n+1}$ for $n \leq x < n + 1$. If the r.v. X has geometric distribution, then $\mathbb{E}(X) = q/p$, $\mathbb{V}ar(X) = q/p^2$ and the moment generating function is given by $\Psi_X(z) = p/(1 - qe^z)$.

We now proceed with some examples of continuous distributions.

Uniform distribution. A uniform distribution, sometimes also called *rectangular*, is a distribution that has constant density function over a real interval (a, b). More precisely, the density function and distribution function for a continuous uniform distribution on the interval (a, b) are

$$f_{(a,b)}(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

and

$$F_{(a,b,c)}(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a < x < b \\ 1 & \text{if } x > b \end{cases}$$

If the r.v. X has uniform distribution in the interval (a, b), then $\mathbb{E}(X) = (a+b)/2$, $\mathbb{V}ar(X) = (b-a)^2/12$ and the moment generating function is given by $\Psi_{(a,b)}(z) = \frac{e^{bz} - e^{az}}{(b-a)z}$.

Triangular distribution. The triangular distribution defined over the interval $[a, b]$ with *mode* $c \in [a, b]$ is a continuous distribution with probability density function

$$f_{(a,b,c)}(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{if } a \le x \le c \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{if } c \le x \le b \end{cases}$$

We have

$$F_{(a,b)}(x) = \begin{cases} \frac{(x-a)^2}{(b-a)(c-a)} & \text{if } a \le x \le c \\ \frac{(b-x)^2}{(b-a)(b-c)} & \text{if } c \le x \le b \end{cases}$$

and the mean is $(a + b + c)/3$.

Normal distribution. A normal distribution with parameters μ and $\sigma > 0$ (also called *Gaussian* and denoted by $N(\mu, \sigma)$) is characterized by the following density function

$$f_{N(\mu,\sigma)}(x) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

for $x \in (-\infty, +\infty)$. The values μ and σ^2 turn out to be the mean value and the variance of the distribution, respectively. The corresponding moment generating function is

$$\Psi_{N(\mu,\sigma)}(z) = e^{\,\mu z + \sigma^2 z^2} \; .$$

The so-called *standard normal distribution* is obtained taking $\mu = 0$ and $\sigma = 1$; its density function and moment generating functions are then given by

$$f_{N(0,1)} = \frac{1}{\sqrt{2\pi}}\, e^{x^2/2} \qquad \text{and} \qquad \Psi_{N(0,1)} = e^{z^2/2}.$$

Finally observe that a general normal distribution X can always be converted to a standard normal distribution X by changing variables to $Z = (X - \mu)/\sigma$.

3.4 Bernoulli trials and DeMoivre–Laplace Theorems

Let us now take into consideration one of the most classical scheme in probability theory, that of a sequence of independent *Bernoulli trials*, following the setting-out used in [28, Chapter II]. In the usual setting, one performs a sequence of repeated Bernoulli trials with success probability p, which does not depend on the number of the trial. Such a scheme deserves special interest because of its historical reason and since the laws that rule the behaviour of repeated Bernoulli trials can be generalized to the study of sequences of independent trials with more than two possible outcomes.

Notice that such scheme can be used to generate random word of given length n, over the alphabet whose letter are the possible outcomes. That is, the *word generated by the Bernoulli model* is the sequence of outcomes from the first trial to the n-th one.

The simplest problem concerning a Bernoulli scheme consists in the determination of the probability $P_n(m)$ of having m successes in n trials and $n-m$ failures in the remaining ones. As stated in the previous section, such probability is given by $\binom{n}{m}p^m q^{n-m}$, where $q = 1 - p$ is the failure probability. Thus, we get the binomial probability distribution law.

The computation of $P_n(m)$ for large values of n and m is rather complicated. Thus, asymptotic formulas that would enable one to determine these probabilities to a sufficient degree of accuracy are necessary. A formula of this kind was first discovered by DeMoivre in 1730 for $p = q = 1/2$ and was subsequently generalized by Laplace in the case of arbitrary $p \in (0, 1)$. Intuitively, this formula states that, up to a factor $\Theta(\sqrt{n})$, the probability of having m success in n trials approximates a Gaussian density at m.

Theorem 3.2 (Local DeMoivre–Laplace Theorem) *If the probability of occurrence of some event A in n independent trials is constant and equal to p $(0 < p < 1)$, then the probability $P_m(n)$ that the event A occurs exactly m times in these trials satisfies the relation*

$$\lim_{n\to\infty} \frac{\sqrt{2\pi npq}\ P_n(m)}{e^{-\frac{(m-np)^2}{2npq}}} = 1$$

uniformly with respect to all m such that the values $(m - np)/\sqrt{npq}$ are contained in some finite real interval.

The previous theorem allows one to establish another limit relation of probability theory, known as *central* or *integral limit theorem*, which establishes a weaker property concerning the limit distribution of the sequence of independent trials.

Theorem 3.3 (Integral DeMoivre–Laplace Theorem) *If m is the number of occurrences of an event in n independent trials in each of which the probability of this event is p $(0 < p < 1)$, then the relation*

$$\lim_{n\to 0} P\left\{a \le \frac{m-np}{\sqrt{npq}} < b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}dx \tag{3.2}$$

holds uniformly in (a, b), for $a, b \in \mathbb{R}$.

We just notice that the law of large numbers (stating that X_n/n converges to p) can be proved as a special application of the Integral Theorem of DeMoivre–Laplace stated above. However, this theorem is much more general than the law of large number, since it really provides the limit distribution of $(m - np)/\sqrt{npq}$.

The Integral DeMoivre–Laplace Theorem has served as a basis for a large group of investigations both in the theory of probability and in its numerous applications to natural sciences, engineering and economics. Indeed, notice that equation (3.2) can be also written as

$$\lim_{n\to 0} P\left\{a \le \frac{S_n - \sum_{k=1}^n \mathbb{E}(X_k)}{\sqrt{\mathbb{V}ar S_n}} < b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}dx \tag{3.3}$$

where X_k denotes the r.v. representing the k-th trial and $S_n = \sum_{k=1}^n X_k$ represents the repetition of n independent trials. Thus, the following question naturally arises: is relation (3.3) intrinsically connected to the special choice of the r.v.'s X_k or would it also hold if weaker restrictions were imposed on the distribution functions of this summands? The answer to this question is contained in a number of theorems which go by the generic name of *central limit theorem*. Such theorems usually state that it is merely necessary to impose a very general restriction on the r.v.'s, whose meaning is that the individual terms have a negligible effect on the sum. Analogous generalizations hold for the Local DeMoivre–Laplace Theorem, (they are usually referred to as *local limit theorems*). Clearly, being local limit theorems much stronger, their existence also implies that corresponding central limit theorems hold. For precise results, see [28, Chapter VIII].

3.5 Markov chains

An immediate generalization of the scheme of independent trials is given by the so-called Markov chains. This kind of stochastic process is named after the famous Russian mathematician who first investigated their properties. Here we just recall basic elements of the theory; a complete treatment can be found for instance in [54, 41].

Let us consider a sequence of trials and assume that in the t-th trial, represented by the r.v. X_t, one of the mutually exclusive events $A_1, A_2, \cdots A_n$ may be realized. We say that the sequence of trials forms a *(simple) Markov chain* if, for any $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots$, the conditional probability of occurrence of the event A_i in the $(t+1)$-th trial, given that a known event has occurred in the t-th trial, only depends on which event has occurred in the t-th trial. In other terms, such a conditional probability is not affected by the further information concerning the events that have occurred in the earlier trials.

Notice that, as for the Bernoulli scheme, Markov processes can also be used to generate random words over the alphabet consisting of all possible outcomes.

A different terminology is often used. Indeed, Markov chains serve as theoretical models for describing a *system* which can be in one of the *states* of $\mathcal{A} = A_1, A_2, \cdots A_n$ and which *jumps* at unit intervals from one state to another according to the following probability law: if the system is in state A_i at *time* $t-1$, the next jump takes it to the state A_j with *transition probability* $T_{ij}(t) = P\{X_t = A_j \mid X_{t-1} = A_i\}$. The set of values $T_{ij}(t)$ is prescribed for all i, j, t and determines the probabilistic behaviour of the system, once it is known how it starts off at time 0. The fundamental property of Markov chains states that if $E_0, E_1, \ldots E_{t+1}$ are elements of $\mathcal{A}$ and $P\{X_t = E_t,\ X_{t-1} = E_{t-1}, \ldots, X_0 = E_0\} > 0$, then

$$P\{X_{t+1} = E_{t+1} | X_t = E_t,\ X_{t-1} = E_{t-1}, \ldots, X_0 = E_0\} = P\{X_{t+1} = E_{t+1} | X_t = E_t\} \ .$$

This property express, roughly, that future probabilistic evolution of the process is determined once the *immediate past* is knows.

Let Π_0 denote the vector of initial probabilities $P\{X_0 = A_i\}$, for $i = 1, 2, \ldots n$. Then, the probabilistic structure of a Markov chain is completely determined by the *initial probability distribution* Π_0 and the *transition matrices* $T(t) = (T_{ij}(t))$ at time $t = 1, 2, \ldots$. Indeed, the probability function of X_t is given by the vector Π_t defined as

$$\Pi_t = \Pi_0 \cdot T(1) \cdots T(t) \ . \tag{3.4}$$

Also notice that for any t, $T(t)$ is a square non-negative matrix such that its row sum equal 1. For this reason, $T(t)$ is said to be *stochastic*.

If $T(1) = T(2) = \ldots = T(t) = \ldots$, then we say that the Markov chain is *homogeneous*. In this case, we shall refer to the common transition matrix as *the* transition matrix and denote it by T; moreover, recalling the definitions of Section 2.2, the Markov chain is called *irreducible*, *primitive* or *periodic* if P is of this sort. Then, relation (3.4) clearly simplifies to

$$\Pi_t = \Pi_0 \cdot T^t \ ;$$

furthermore, for any $t > h$, we have $\Pi_t = \Pi_h \cdot T^{h-t}$, that is the probability evolution is homogeneous in reference to any initial time h.

Given an homogeneous Markov chain, an initial probability distribution Π_0 is said to be *stationary* if for every $t = 1, 2, \ldots$ one has $\Pi_0 = \Pi_t$. If a Markov chain is irreducible, then it is easy to see that it has a unique stationary distribution given by the solution v of $v_{_T} T = v_{_T}$, $v_{_T} \cdot \mathbf{1} = 1$ (where $\mathbf{1}$ is the vector with 1 in each component). Hence, applying the Perron–Frobenius Theorem, one obtains the following result.

Theorem 3.4 (Ergodic theorem for primitive Markov chains) *Let T be the transition matrix of a primitive homogeneous Markov chain. Then*

$$\lim_{t \to \infty} T^t = \mathbf{1} \cdot v_{_T}$$

elementwise, where v is the unique stationary distribution of the chain; the rate of approach to the limit is geometric.

Let us now focus our attention on sequences of trials with two possible outcomes, in each of which the event E may or may not occur, and assume that the trials are connected in a homogeneous Markov chain. Let α be the probability of the event E occurring at time $(t+1)$ and let β be the probability of the event E occurring at time $(t+1)$ given that E has not occurred at time t; Also assume α and β to be different from 0 and 1 and set $\delta = \alpha - \beta$. Thus, the transition matrix is given by

$$\begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Clearly, such a scheme generalizes the Bernoulli scheme considered in the previous section. Now, let p_t denote the probability of occurrence of E in the t-th trial and set $q_t = 1 - p_t$. It is immediate to see that $p_t = \alpha\, p_{t-1} + \beta\, q_{t-1} = \delta\, p_{t-1} + \beta$. Thus, as t goes to infinity, we have $p_t \to \beta/(1-\delta)$. On the other hand, if ${p_j}^{(i)}$ denotes the probability of the event E in the j-th trial knowing that it occurred in the i-th trial, then we get ${p_j}^{(i)} = \delta\, {p_{j-1}}^{(i)} + \beta$ for all $j > i+1$.

Now consider the probability $P_n(m)$ to find m occurrences of E in n trials. One can prove that

$$P_n(m) = \frac{e^{-z^2/2}}{\sqrt{2\pi npq(1+\delta)/(1-\delta)}} \cdot (1 + \mathrm{o}(1))$$

as $n \to \infty$, which extends the local DeMoivre–Laplace theorem for Bernoulli trials. An integral limit theorem may also be derived. Indeed the relation

$$P\left\{ a \leq \frac{m - np}{\sqrt{2\pi npq(1+\delta)/(1-\delta)}} < b \right\} = \int_a^b e^{-x^2/2} dx$$

holds uniformly as n goes to infinity.

3.6 Quasi-power theorem

Consider a sequence of r.v.'s $\{X_n\}_n$ having distribution functions $\{F_{X_n}\}_n$ and a r.v. X having distribution function F_X. If $\lim_{n\to\infty} F_{X_n}(\tau) = F_X(\tau)$ for every point $\tau \in \mathbb{R}$ of continuity for F_X, then one says that X_n *converges to X in distribution* (or *in law*). Actually, this is an abuse of terminology, since this convergence concept is not referred to the r.v's, but to their distribution functions.

The moment generating functions are an useful tool to prove convergence in law. Indeed, by Theorem 3.1, if $\{\Psi_{X_n}\}_n$ and Ψ_X are defined all over $\mathbb{C}$ and $\Psi_{X_n}(z)$ tends to $\Psi_X(z)$ for every $\theta \in \mathbb{C}$, then X_n converges to X in distribution.

The Integral Theorem of DeMoivre–Laplace establishes the convergence in distribution of the sum of independent Bernoulli trials to a Gaussian r.v. Here we present another convenient approach to prove the convergence in law to a Gaussian r.v. It relies on the so called "quasi-power" theorems introduced in [38] and implicitly used in the previous literature [3] (see also [23]). The main advantage of this theorem, with respect to other classical statements of this kind, is that it does not require any condition of independence concerning the r.v.'s X_n. For instance, the Integral Theorem of DeMoivre–Laplace assume that each X_n is a partial sum of a sequence of independent r.v.'s.

For our purpose here we present a simple variant of such theorem and we also prove an interesting property to be applied to pattern statistics in Part II.

Theorem 3.5 (Quasi-power Theorem) *Let $\{X_n\}$ be a sequence of r.v.'s, where each X_n takes values in $\{0, 1, \ldots, n\}$ and let us assume the following conditions:*

C1 *There exist two functions $u(z)$, $h(z)$, both analytic at $z = 0$ where they take the value $u(0) = h(0) = 1$, and a positive constant c, such that for every $|z| < c$*

$$\Psi_{X_n}(z) = h(z) \cdot u(z)^n \left(1 + O(n^{-1})\right) ; \tag{3.5}$$

C2 *The constant $\sigma = u''(0) - (u'(0))^2$ is strictly positive (variability condition).*

Also set $\mu = u'(0)$. Then $\frac{X_n - \mu n}{\sqrt{\sigma n}}$ converges in distribution to a normal r.v. of mean 0 and variance 1, i.e. for every $x \in \mathbb{R}$

$$\lim_{n \longrightarrow +\infty} P\left\{\frac{X_n - \mu n}{\sqrt{\sigma n}} \leq x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \ .$$

Proof. We argue as in [3]. Let $\Psi_n(z)$ be the moment generating function $\frac{Y_n - \beta n}{\sqrt{\alpha n}}$. Then it suffices to prove that, as n goes to infinity, $\Psi_n(z)$ tends to the moment generating function of the standard normal distribution, i.e. $e^{z^2/2}$. First note that

$$\Psi_n(z) = \mathbb{E}\left(exp\left(\frac{Y_n - \mu n}{\sqrt{\sigma n}} z\right)\right) = exp\left(-z\mu\sqrt{\frac{n}{\sigma}}\right) \cdot \Psi_n\left(\frac{z}{\sqrt{\sigma n}}\right) \ .$$

This implies, by condition C1, that

$$\Psi_n(z) = exp\left(-z\mu\sqrt{\frac{n}{\sigma}}\right) \cdot h\left(\frac{z}{\sqrt{\sigma n}}\right) \cdot exp\left(n \cdot log \ u\left(\frac{z}{\sqrt{\sigma n}}\right)\right) \ . \tag{3.6}$$

Now, as n tends to infinity, $z/\sqrt{\sigma n}$ goes to 0 and in a neighbourhood of $z = 0$ we have

$$\begin{aligned} h(z) &= 1 + \mathrm{O}(z) \\ u(z) &= 1 + u'(0)z + \frac{1}{2}u''(0)z^2 + \mathrm{O}(z^3) = 1 + \mu z + \frac{\sigma + \mu^2}{2} z^2 + \mathrm{O}(z^3) \end{aligned}$$

the last equation implying

$$log(u(z)) = \mu z + \frac{\sigma}{2} z^2 + \mathrm{O}(z^3) \ .$$

Replacing the previous equations into (3.6), we get

$$\Psi_n(z) = \left(1 + \frac{1}{\sqrt{n}}\right) \cdot exp\left(\frac{z^2}{2}\right) + \mathrm{O}\left(\frac{1}{\sqrt{n}}\right)$$

which pointwise converges to $e^{z^2/2}$ as n goes to infinity. □

Proposition 3.6 *Let $\{X_n\}$ be a sequence of r.v.'s such that each X_n takes values in the set $\{0, 1, \ldots, n\}$. Assume that conditions C1 and C2 of Theorem 3.5 hold and let $u(z)$, μ and σ be defined consequently. Then, for every real θ such that $|\theta| \leq n^{-5/12}$, as n grows to infinity, we have*

$$\left| u(i\theta)^n - e^{-(\sigma/2)\theta^2 n + i\mu\theta n} \right| = O\left(n^{-1/2}\right) .$$

Proof. First of all observe that, from our hypotheses, in a neighbourhood of $z = 0$ we have

$$u(z) = 1 + \mu z + \frac{\sigma + \mu^2}{2} z^2 + \mathrm{O}(z^3) . \tag{3.7}$$

This implies that, in a real neighbourhood of $\theta = 0$, the complex function $u(i\theta)$ satisfies the equalities

$$\begin{aligned}
|u(i\theta)| &= \left|1 + i\mu\theta - \frac{\sigma + \mu^2}{2}\theta^2\right| \cdot |1 + \mathrm{O}(\theta^3)| = \\
&= \sqrt{\left(1 - \frac{\sigma + \mu^2}{2}\theta^2\right)^2 + \mu^2\theta^2} \cdot |1 + \mathrm{O}(\theta^3)| = \\
&= \left(1 - \frac{\sigma}{2}\theta^2 + \mathrm{O}(\theta^4)\right) |1 + \mathrm{O}(\theta^3)| , \\
\arg(u(i\theta)) &= \arg\left(1 + i\mu\theta - \frac{\sigma + \mu^2}{2}\theta^2\right) + \arg(1 + \mathrm{O}(\theta^3)) = \\
&= \operatorname{arctg}\left(\frac{\mu\theta}{1 - \frac{\sigma+\mu^2}{2}\theta^2}\right) + \mathrm{O}(\theta^3) = \mu\theta + \mathrm{O}(\theta^3) .
\end{aligned} \tag{3.8}$$

As a consequence, one has

$$u(i\theta)^n = \left(1 - (\sigma/2)\theta^2 + \mathrm{O}(\theta^3)\right)^n \cdot e^{in(\mu\theta + \mathrm{O}(\theta^3))} = e^{-(\sigma/2)n\theta^2 + i\mu n\theta} \cdot e^{n\mathrm{O}(\theta^3)} .$$

Now, for each $|\theta| \leq n^{-5/12}$, we have $|n\theta^3| = \mathrm{O}\ (n^{-1/4})$ and the last expression yields

$$u(i\theta)^n = e^{-(\sigma/2)n\theta^2 + i\mu n\theta} \cdot (1 + \mathrm{O}(n\theta^3)) .$$

Therefore

$$\left| u(i\theta)^n - e^{-(\sigma/2)n\theta^2 + i\mu n\theta} \right| = \mathrm{O}\left(n|\theta^3|e^{-n\theta^2\sigma/2}\right) = \mathrm{O}\left(n^{-1/2}\right) ,$$

the last equality being obtained by deriving with respect to θ. □

3.7 A general criterion for local convergence laws

Convergence in law of a sequence of r.v.'s $\{X_n\}$ does not yield an approximation of the probability that X_n has a specific value. On the other hand, in Section 3.4 we state that, when X_n is the partial sum of a sequence of r.v.'s, approximations to the Gaussian density are provided by local limit theorems.

Let us now present a general criterion that guarantees, for a general sequence of discrete r.v.'s, the existence of a local convergence property of a Gaussian type more general than DeMoivre–Laplace's Theorem mentioned above. In Part II, using such criterion, we show that the same local convergence property holds for certain pattern statistics.

Theorem 3.7 (Local Limit Criterion) *Let $\{X_n\}$ be a sequence of r.v.'s such that, for some integer $d \geq 1$ and every $n \geq d$, X_n takes on values only in the set*

$$\{x \in \mathbb{N} \mid 0 \leq x \leq n, x \equiv \rho(mod\ d)\} \tag{3.9}$$

for some integer $0 \leq \rho < d$. Assume that conditions C1 and C2 of Theorem 3.5 hold and let μ and σ be the positive constants defined in the same theorem. Moreover assume the following property:

C3 *For all $0 < \theta_0 < \pi/d$* $$\lim_{n \to +\infty} \left\{ \sqrt{n} \sup_{|\theta| \in [\theta_0, \pi/d]} |\Psi_{X_n}(i\theta)| \right\} = 0 \ .$$

Then, as n grows to $+\infty$ the following relation holds uniformly for every $j = 0, 1, \ldots, n$.

$$P\{X_n = j\} = \begin{cases} \dfrac{de^{-\frac{(j-\mu n)^2}{2\sigma n}}}{\sqrt{2\pi\sigma n}} \cdot (1 + o(1)) & \text{if } j \equiv \rho \ (mod\ d) \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

Recall that $\Psi_{X_n}(i\theta)$ is the characteristic function of X_n. The condition C3 states that, for every constant $0 < \theta_0 < \pi/d$, as n grows to $+\infty$, the value $\Psi_{X_n}(i\theta)$ is of the order $o(n^{-1/2})$ uniformly with respect to $\theta \in [-\pi/d, -\theta_0] \cup [\theta_0, \pi/d]$. Note that ρ may depend on n even if $\rho = \Theta(1)$.

One can easily show that any sequence $\{X_n\}$ of binomial r.v.'s of parameters n and p, where $0 < p < 1$ (i.e. representing the number of successes over n independent trials of probability p), satisfy the hypotheses of the theorem with $d = 1$. In this case, equation (3.10) coincides with the property stated in DeMoivre–Laplace Local Limit Theorem. Thus our general criterion includes the same theorem as a special case.

Relations like (3.10) already appeared in the literature. In particular in [28, Section 43], (3.10) is proved when X_n is the sum of n independent *lattice* r.v.'s of period d and equal distribution. Note that our theorem is quite general since it does not require any condition of independence of the X_n's.

We also note that, for $d = 1$ a similar criterion for local limit laws has been proposed in [23, Theorem 9.10] where, however, a different condition is assumed, i.e. one requires that the probability generating function $p_n(u)$ of X_n has a certain expansion, for $u \in \mathbb{C}$ belonging to an annulus $1 - \varepsilon \leq |u| \leq 1 + \varepsilon$ $(\varepsilon > 0)$, that corresponds to assuming an equation of the form (3.5) for $z \in \mathbb{C}$ such that $|\Re\mathrm{e}(z)| \leq \delta$ (for some $\delta > 0$).

Before illustrating the proof of the criterion, let us prove the following lemma concerning the characteristic function of the r.v.X_n.

Lemma 3.8 *Under the hypotheses of Theorem 3.7, for every real θ such that $|\theta| \in [0, \pi/d]$ we have*

$$\left| \Psi_{X_n}(i\theta) - e^{-(\sigma/2)\theta^2 n + i\mu\theta n} \right| = \Delta_n(\theta) \ ,$$

where

$$\Delta_n(\theta) = \begin{cases} O(n^{-5/12}) & \text{if } |\theta| \in [0, n^{-5/12}] \\ o(n^{-1/2}) & \text{if } |\theta| \in [n^{-5/12}, \pi/d] \end{cases} \tag{3.11}$$

Proof. For the sake of brevity let Ψ_n stand for Ψ_{X_n}. Let us consider the first interval given in (3.11), i.e. the case $|\theta| \leq n^{-5/12}$. By condition C1 of Theorem 3.5 and Proposition 3.6 we have

$$\Psi_n(i\theta) = h(i\theta) \cdot u(i\theta)^n \ (1 + \mathrm{O}(n^{-1})) = \left(e^{-(\sigma/2)\theta^2 n + i\mu\theta n} + \mathrm{O}(n^{-1/2}) \right) (1 + \mathrm{O}(n^{-5/12}))$$

which proves the relation since $|e^{-(\sigma/2)\theta^2 n + i\mu\theta n}| \leq 1$ for every real θ.

Referring to the second interval, let θ_0 be a constant such that $0 < \theta_0 < c$, where c is defined as in condition C1 of Theorem 3.5 and assume $|\theta| \in [n^{-5/12}, \theta_0]$. Then we have

$$\left|\Psi_n(i\theta) - e^{-(\sigma/2)\theta^2 n + i\mu\theta n}\right| \leq |\Psi_n(i\theta)| + e^{-(\sigma/2)\theta^2 n}.$$

Since $|\theta| \geq n^{-5/12}$, the second term of the right hand side is smaller than or equal to $e^{-(\sigma/2)n^{1/6}} = \mathrm{o}(n^{-1/2})$. Let us show an analogous bound for the first term. To this end, by equations (3.5) and (3.8) we have

$$|\Psi_n(i\theta)| \;=\; |h(i\theta) \cdot u(i\theta)^n \;(1 + \mathrm{O}(n^{-1}))| \;\leq\; h\left|1 - \frac{\sigma}{2}\theta^2 + \mathrm{O}(\theta^3)\right|^n |1 + \mathrm{O}(n^{-1})| \tag{3.12}$$

where $h = \sup_{|\theta| \leq \theta_0} |h(i\theta)|$. By the arbitrariness of θ_0 , for some constant $\varepsilon > 0$ and every $|\theta| \leq \theta_0$ we have

$$\left|1 - \frac{\sigma}{2}\theta^2 + \mathrm{O}(\theta^3)\right| \leq \left|1 - \frac{\sigma}{2}\theta^2\right| + \varepsilon\theta^3 \ .$$

By the same reason we may assume $\theta_0 \leq \min\{\sqrt{2/\sigma},\ \sigma/(4\varepsilon)\}$, which proves

$$\left|1 - \frac{\sigma}{2}\theta^2 + \mathrm{O}(\theta^3)\right| \leq 1 - \frac{\sigma}{2}\theta^2 + \varepsilon\theta_0\theta^2 \leq 1 - \frac{\sigma}{4}\theta^2 \ ,$$

for every $|\theta| \leq \theta_0$. Replacing this value in (3.12) we get

$$|\Psi_n(i\theta)| = \mathrm{O}\left(|1 - (\sigma/4)\theta^2|^n\right) = \mathrm{O}\left(e^{-(\sigma/4)\theta^2 n}\right) \ ,$$

which is again bounded by $\mathrm{o}(n^{-1/2})$ because of the range of θ. This proves relation (3.11) for $|\theta| \in [n^{5/12}, \theta_0]$.

Finally assume $\theta_0 \leq |\theta| \leq \pi/d$. Again, we have

$$\left|\Psi_n(i\theta) - e^{-(\sigma/2)\theta^2 n + i\mu\theta n}\right| \leq \sup_{\theta_0 \leq |\theta| \leq \pi/d} |\Psi_n(i\theta)| + e^{-(\sigma/2)\theta_0^2 n} \ .$$

The first term is bounded by $\mathrm{o}(n^{-1/2})$ by condition C3, while the second one is $\mathrm{O}\ (\tau^n)$ for some $\tau \in (0,1)$ and this completes the proof. □

The proof of Theorem 3.7 is based on the application of the discrete Fourier transform. Here we recall its definition, for more details see for instance [11]. For any positive integer n, the n-th discrete Fourier transform is the transformation $D_n : \mathbb{C}^n \to \mathbb{C}^n$ such that, for every $u = (u_0, \cdots, u_{n-1}) \in \mathbb{C}^n$, $D_n(u) = (v_0, \cdots, v_{n-1})$ where

$$\left(D_n(u)\right)_s = v_s = \sum_{k=0}^{n-1} e^{i\frac{2\pi sk}{n}} u_k$$

for each $s = 0, 1, \ldots n-1$. It is well-known that D_n admits an inverse transformation D_n^{-1} defined by setting for each $k = 0, 1, \ldots n-1$

$$\left(D_n^{-1}(v)\right)_k = \frac{1}{n}\sum_{s=0}^{n-1} e^{-i\frac{2\pi ks}{n}} v_s.$$

Now we are able to present the proof of the criterion for the local convergence.

Proof of Theorem 3.7. First, we apply the Discrete Fourier Transform (see for instance [11]) to the array of probabilities of X_n. Since each X_n assumes values only in (3.9), set $N = \min\{h \in \mathbb{N} \mid n < \rho + hd\}$ and define $p^{(n)}$ as the array $(p_0^{(n)}, p_1^{(n)}, \ldots, p_{N-1}^{(n)})$, where

$$p_h^{(n)} = \mathrm{P}\{X_n = \rho + hd\} \qquad (h = 0, 1, \ldots, N-1).$$

Let $f^{(n)} \in \mathbb{C}^N$ be its Discrete Fourier Transform, i.e. the array of values $f_s^{(n)}$ such that

$$f_s^{(n)} = \sum_{h=0}^{N-1} p_h^{(n)} e^{i\frac{2\pi s}{N}h} = \Psi_{X_n}\left(i\frac{2\pi s}{Nd}\right) e^{-i\frac{2\pi s \rho}{Nd}}$$

where $s = -\lceil N/2 \rceil + 1, -\lceil N/2 \rceil + 2, \ldots, \lfloor N/2 \rfloor$. By antitransforming, each $p_h^{(n)}$ can be obtained by

$$p_h^{(n)} = \frac{1}{N} \sum_{s=-\lceil N/2 \rceil+1}^{\lfloor N/2 \rfloor} f_s^{(n)} e^{-i\frac{2\pi s}{N}h} = \frac{1}{N} \sum_{s=-\lceil N/2 \rceil+1}^{\lfloor N/2 \rfloor} \Psi_{X_n}\left(i\frac{2\pi s}{Nd}\right) e^{-i\frac{2\pi s}{Nd}(\rho + hd)}. \tag{3.13}$$

Now, the previous lemma suggests us to define the function $F_n(\theta) = e^{-(\sigma/2)\theta^2 n + i\mu\theta n}$ for every $-\pi/d < \theta \leq \pi/d$ and to approximate $p_h^{(n)}$ with the following values

$$\hat{p}_h^{(n)} = \frac{1}{N} \sum_{s=-\lceil N/2 \rceil+1}^{\lfloor N/2 \rfloor} F_n\left(\frac{2\pi s}{Nd}\right) e^{-i\frac{2\pi s}{Nd}(\rho + hd)}. \tag{3.14}$$

Clearly, the error associated with the above approximation satisfies the inequality

$$\left|p_h^{(n)} - \hat{p}_h^{(n)}\right| \leq \frac{1}{N} \sum_{s=-\lceil N/2 \rceil+1}^{\lfloor N/2 \rfloor} \left|\Psi_{X_n}\left(i\frac{2\pi s}{Nd}\right) - F_n\left(\frac{2\pi s}{Nd}\right)\right| \leq \frac{1}{N} \sum_{s=-\lceil N/2 \rceil+1}^{\lfloor N/2 \rfloor} \Delta_n\left(\frac{2\pi s}{Nd}\right),$$

which can be computed by splitting the range of s in two intervals as in (3.11). Thus, we get

$$\left|p_h^{(n)} - \hat{p}_h^{(n)}\right| \leq \frac{2}{N}\left\{\lceil Nd/(2\pi n^{5/12}) \rceil \mathrm{O}(n^{-5/12}) + \lceil N/2 \rceil \mathrm{o}(n^{-1/2})\right\} = \mathrm{o}(n^{-1/2}). \tag{3.15}$$

As n grows to $+\infty$ the right hand side of (3.14) tends to the integral of $F_n(x)\, e^{-ix(\rho+hd)}\, d/(2\pi)$ over the interval $x \in (-\pi/d, \pi/d)$. Thus, by standard mathematical tools, one can prove that as n grows to $+\infty$ the relation

$$\hat{p}_h^{(n)} = \frac{d}{\sqrt{2\pi\sigma n}}\, e^{-\dfrac{(\rho + hd - \mu n)^2}{2\sigma n}} + \mathrm{o}(n^{-1/2})$$

holds uniformly for every $h = 0, 1, \ldots, N-1$. Hence, the result is a straightforward consequence of the previous equation, together with relation (3.15). □

Part II

Pattern statistics in rational models

Chapter 4

Rational stochastic models: the primitive case

In this chapter we start off the discussion concerning pattern statistics in rational models. First, we introduce the *frequency problem*: we present different variants of such problem and briefly illustrate some known results, generally concerning classical models as the Bernoulli or the Markov ones. In Section 4.2 we present a new stochastic model, defined by means of a rational formal series in two noncommuting variables, and we introduce the *rational symbol frequency* (RSF) problem. Intuitively it concerns the study of the sequence of r.v.'s $\{Y_n\}_n$ representing the number of occurrences of a symbol a in words of length n chosen at random in $\{a, b\}^*$, according to the probability distribution given by the rational model. In order to compare this problem with those previously dealt with in the literature, in Section 4.3 we show how our model can be viewed as a proper extension of the Markovian model as far as counting the occurrences of a regular set in a random text is concerned.

In Section 4.4, we assume that the transition matrix associated with the series defining the rational model is primitive; we obtain asymptotic estimates for the mean value and the variance of the statistics in exam, showing that they have linear behaviour as n goes to infinity. We also present central and local limit theorems for Y_n (actually, we prove more general results, since these theorems hold for an arbitrary power of any primitive series, too). Indeed, we prove that Y_n converges in distribution to a normal density function; intuitively this means that the occurrence of the letter a in a given position of a "random" word of length n is rather independent of the other occurrences and of the position itself. Thus, the behaviour of Y_n is similar to the sum of n independent Bernoulli r.v.'s of equal parameter. Finally, we establish a local limit theorem for Y_n which turns out to be related to the notion of symbol periodicity introduced in Section 2.4.

As a consequence of our analysis, in Section 4.5 we obtain an asymptotic estimation of the growth of the coefficients for a nontrivial subclass of rational formal series in commuting variables. This problem was actually among the original motivations of this thesis and can be seen as a generalization of classical questions concerning the ambiguity in formal language (see Section 1.8).

4.1 The frequency problem: known results

Probability on pattern occurrences in a random sequence of letters (generally called text) has been widely studied and has applications in many areas of bio-informatics, code theory and data compression, pattern matching, design and analysis of algorithms, games. Here we focus on the frequency of occurrences of a repeated pattern in a random sequence of letters. If we assume to know the probabilistic model (and its parameters) that generates the text, the central question is: *how many occurrences of a given pattern shall we expect in such a random sequence?*

This problem, we refer to as the *frequency problem*, can be studied under different assumptions concerning the source that generates the text, or the pattern to search for through the text. While real sources are often complex objects, pattern statistics only deals with quite idealized sources, described by means of probabilistic models. The simplest model is the Bernoulli one, that represents a memoryless source; if in particular all letters are assigned the same probability, the model is said to be symmetric. Another classical model is defined by Markov processes. Finally, dynamical sources describe non-Markovian processes, characterized by unbounded dependency on past history [58].

As for stochastic models, the choice of pattern can also lead to different settings. String matching is the basic pattern matching problem. Here, one counts the occurrences of a given string as a factor in the text. If the string reduces to a single letter, we shall use the expressions *symbol occurrences* or *symbol frequency* problem. One can also search for a finite set of strings and count the occurrences of all of them. This is useful when one searches for a given pattern but a few mismatches are allowed, too. The approximation string matching is then expressed as matching against the set of words that contains all the valid approximations of the initial strings. Moreover one may be interested in occurrences of the pattern as a subsequence of the text; in this case the letters no longer need to be consecutive. Also, the gaps between successive symbols may be bounded or not. The hidden pattern problem concerns the case where some gaps are bounded while some other are not. A generalization of all these problems is attained when the pattern is defined by a general regular expression, thus including infinite sets of words.

When a pattern is searched for through a text, various constraints can be imposed on the counting of overlapping occurrences; occurrences are considered valid if they satisfy these constraints. In the overlapping model, any occurrences is valid and two overlapping patterns both contribute to the count. On the contrary, in renewal models two overlapping sequences cannot be considered valid simultaneously: one only counts the first occurrence and another occurrence is valid if it does not overlap on the left with any other valid occurrence. Many other constraints can be chosen that define other variants. For instance, one may count overlapping occurrences of different patterns, or one may set a minimal distance between valid occurrences.

Several authors contributed to the study of the frequency problem, generally considering the Bernoulli or the Markov models to generate the random text. Feller already suggested a solution in his book when the pattern is single string [21]. However, the most important recent contributions belong to Guibas and Odlyzko who in a series of seminal papers [32, 33, 34] laid the foundations for the analysis of the symmetric Bernoulli case. In particular, in [34] the authors computed the moment generating function for the number of strings of fixed length that do not contain any one of a given set of patterns. Certainly, this suffices to estimate the probability of at least one pattern occurrences in a random string generated by the symmetric Bernoulli model. Furthermore, in a passing remark the authors presented some basic properties useful for the study of several pattern occurrences in a random text for the symmetric Bernoulli model. In [26] Fudos, Pitoura and Szpankowski extended some of those results to the asymmetric model, computing the probability of a fixed number of occurrences of a string into a random text.

The results have then been extended to the Markovian model, first by Li [44], who considered the problem with no pattern occurrences, and, more recently, by Régnier and Szpankovski. In [50], using a method that treats uniformly both the Bernoulli and the Markov models, they established that the number of occurrences of a string is asymptotically normal, under a primitivity hypothesis of the stochastic model. They also estimated the probability of a fixed number of occurrences of the pattern for three different ranges of such number, also obtaining large deviations results. Moreover, their results allow a symbolic computation of all moments. In another work [49], Régnier also studied the same problem assuming various constraints concerning the overlapping hypothesis.

A further improvement is due to Nicodème, Salvy e Flajolet, that in [45] extended all the previous results considering a text generated by a Markov source and counting the occurrences of a pattern defined by an unrestricted regular expression. In the same paper, the authors also considered some computational aspects concerning the occurrences of the pattern, giving algorithms for computing the parameters of the limit distribution. All these results hold under a primitivity hypothesis on the stochastic matrix defining the Markov process.

Finally, non-Markovian models have been considered by Bourdon and Vallée in [10], where they assumed the text was generated by dynamical sources and they considered generalized pattern, entailing classical and hidden patterns with "don't-care-symbols".

4.2 Stochastic models defined via rational formal series

In this section we present our stochastic model, defined by means of a rational formal series in two noncommuting variables. Given the binary alphabet $\{a, b\}$, let us consider a formal series $r \in \mathbb{R}_+\langle\langle a, b\rangle\rangle$. Let n be a positive integer such that $(r, x) \neq 0$ for some $x \in \{a, b\}^n$, where $\{a, b\}^n$ denotes the set of all words of length n in $\{a, b\}^*$. Consider the probability space of all words in $\{a, b\}^n$ equipped with the probability measure given by

$$\mathrm{P}_n\{\omega\} = \frac{(r, \omega)}{\sum_{x\in\{a,b\}^n}(r, x)} \qquad (\omega \in \{a, b\}^n). \tag{4.1}$$

We define the random variable $Y_n : \{a, b\}^n \to \{0, 1, \dots, n\}$ such that $Y_n(\omega) = |\omega|_a$ for every $\omega \in \{a, b\}^n$ and we say that Y_n *counts the occurrences of* a *in the stochastic model defined by* r. It is clear that, for every $j = 0, 1, \dots, n$, we have

$$\mathrm{P}_n\{Y_n = j\} = \frac{\sum_{|\omega|=n, |\omega|_a=j}(r, \omega)}{\sum_{x\in\{a,b\}^n}(r, x)} \ . \tag{4.2}$$

Observe that if r is the characteristic series χ_L of a language $L \subseteq \{a, b\}^*$, then P_n is just the uniform distribution over the set of words on length n in L, that is $\mathrm{P}_n\{\omega\} = \#(L \cap \{a, b\}^n)^{-1}$ if $\omega \in L$, while $\mathrm{P}_n\{\omega\} = 0$ otherwise. Hence Y_n represents the number of occurrences of a in a word chosen at random in $L \cap \{a, b\}^n$ under uniform distribution.

A useful tool to study the distribution of the pattern statistics Y_n is given by the generating functions associated with formal series. Given $r \in \mathbb{R}_+\langle\langle a, b\rangle\rangle$, for every $n, j \in \mathbb{N}$ let $r_{n,j}$ be the coefficient of $a^j b^{n-j}$ in the commutative image of r, i.e.

$$r_{n,j} = (\mathcal{F}(r), a^j b^{n-j}) = \sum_{|x|=n, |x|_a=j} (r, x) \ . \tag{4.3}$$

Then, we define the function $r_n(z)$ and the corresponding generating function $\mathbf{r}(z, w)$ by

$$r_n(z) = \sum_{j=0}^{n} r_{n,j}\, e^{jz} \qquad \text{and} \qquad \mathbf{r}(z, w) = \sum_{n=0}^{+\infty} r_n(z)\, w^n = \sum_{n=0}^{+\infty}\sum_{j=0}^{n} r_{n,j}\, e^{jz}\, w^n \tag{4.4}$$

where z, w are complex variables. Thus, from the definition of $r_{n,j}$ and from equation (4.2) we have

$$\mathrm{P}_n\{Y_n = j\} = \frac{r_{n,j}}{r_n(0)} \tag{4.5}$$

and the moment generating function of Y_n is given by

$$\Psi(z) = \sum_{j=1}^{n} \mathrm{P}_n\{Y_n = j\}\ e^{jz} = \frac{r_n(z)}{r_n(0)} \ . \tag{4.6}$$

Moreover, by equation (3.1) the mean and the variance of Y_n can be obtained by evaluating r_n and its derivatives at $z = 0$:

$$\mathbb{E}(Y_n) = \frac{r'_n(0)}{r_n(0)} \ , \qquad \mathbb{V}ar(Y_n^2) = \frac{r''_n(0)}{r_n(0)} - \left(\frac{r'_n(0)}{r_n(0)}\right)^2 \ . \tag{4.7}$$

We also remark that the relation between a series r and its generating function $\mathbf{r}(z, w)$ can be expressed in terms of a semiring morphism. As usual, we denote by $\Sigma^{\otimes}$ the free commutative monoid over the alphabet Σ. Then, consider the monoid morphism

$$\mathcal{H} : \{a, b\}^* \longrightarrow \{e^z, w\}^{\otimes}$$

defined by setting $\mathcal{H}(a) = e^z w$ and $\mathcal{H}(b) = w$. Such a map extends to a semiring morphism from $\mathbb{R}_+\langle\langle a, b\rangle\rangle$ to $\mathbb{R}_+[[e^z, w]]$ so that

$$\mathcal{H}(r) = \sum_{x \in \{a,b\}^*} e^{z|x|_a} w^{|x|} = \mathbf{r}(z, w) \tag{4.8}$$

for every $r \in \mathbb{R}_+\langle\langle a, b\rangle\rangle$. This property translates arithmetic relations among formal series into analogous relations among the corresponding generating functions.

From now on, we assume that the series r is rational. Then the probability spaces given by (4.1) define a stochastic model we call *rational.* The *rational symbol frequency* problem (RSF problem) concerns the study of the distribution properties of the sequence of r.v.'s Y_n counting the occurrences of a in the rational stochastic model defined by r.

Let (ξ, μ, η) be a linear representation for r; set $A = \mu(a)$, $B = \mu(b)$, $M = A + B$ and assume $A \neq 0 \neq B$ to avoid trivial cases. Then, it is easy to verify that the following relations hold:

$$r_n(z) \quad = \quad \xi_T (Ae^z + B)^n \eta \tag{4.9}$$

$$\mathbf{r}(z, w) \quad = \quad \xi_T R(z, w)\, \eta \ , \tag{4.10}$$

where, for z near 0 and w near λ^{-1}, $R(z, w)$ is a matrix satisfying

$$R(z, w) = \sum_{n=0}^{+\infty} (Ae^z + B)^n w^n = [I - w(Ae^z + B)]^{-1} = \frac{\mathrm{Adj}(I - w(Ae^z + B))}{\mathrm{Det}\,(I - w(Ae^z + B))} \ . \tag{4.11}$$

Example 4.1 Consider the following representation:

$$\xi' = (10) \ , \quad \mu(a) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \ , \quad \mu(b) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ , \quad \eta = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

corresponding to the automaton represented in Fig. 4.1. The rational series r thus defined satisfies

$$(r, w) = \begin{cases} 1 & \text{if } |w|_a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

leading to the probability distribution

$$\mathrm{P}_n\{Y_n = k\} = \begin{cases} \frac{1}{n+1} & \text{if } k = 0 \\ \frac{n}{n+1} & \text{if } k = 1 \\ 0 & \text{otherwise.} \end{cases}$$

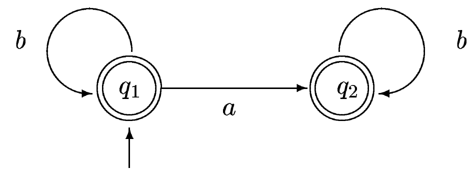


Figure 4.1: State diagram associated with the rational model defined in Example 4.1.

4.3 Rational models and Markovian models

In order to compare the present problem with those previously dealt with in the literature, we show how our model can be viewed as a proper extension of the Markovian model as far as counting the occurrences of a regular set in a random text is concerned.

Thus, let us fix an alphabet Σ and consider an homogeneous Markov chain defined by the initial probability distribution π and the transition matrix $T = (T_{\sigma\tau})_{(\sigma,\tau)\in\Sigma\times\Sigma}$. The pair (π, P) induces a probability measure Π_n over Σ^n

$$\Pi_n(\sigma_1 \dots \sigma_n) = \pi_{\sigma_1} T_{\sigma_1,\sigma_2} \dots T_{\sigma_{n-1},\sigma_n} \ .$$

Now we are given a regular set of patterns $R \subseteq \Sigma^*$ and we are asked to count the number $O_n(\sigma_1 \dots \sigma_n)$ of occurrences of R in a random text $\sigma_1 \dots \sigma_n$ generated by the above Markov process, where by occurrence is meant a position k in the text where a match with an element of R ends. Observe that the values of $O_n(\sigma_1 \dots \sigma_n)$ range from 0 to n. We say that O_n *counts the occurrences of R in the stochastic model defined by* (π, T) and we refer to the study of the distribution

$$\Pi_n\{O_n = k\}$$

associated with some triple (π, T, R) as to the *Markovian pattern frequency problem* (MPF problem). The relationship between the MPF and RSF problems is illustrated by the following theorem.

Theorem 4.2 *Given a finite alphabet Σ, let (π, T) be an homogeneous Markov chain and R a regular expression over Σ. Moreover, let O_n count the number of occurrences of R in the stochastic model defined by (π, T). Then, there exists a series $r \in \mathbb{R}_+\langle\langle a, b\rangle\rangle$ such that the r.v.'s Y_n counting the occurrences of a in the rational model defined by r satisfy the following relation*

$$P_n\{Y_n = j\} = \Pi_n\{O_n = k\}$$

for each $n > 0$ and $j = 0, 1, \dots, n$.

Proof. We first construct a (fully defined) finite deterministic automaton recognizing $\Sigma^* R$ whose set of states is Q, the initial state is p and set of final states is F. As usual, we denote by $\delta(q, \sigma)$ the transition defined by the letter σ in state $q \in Q$. Define the linear representation $\mu : \{a, b\}^* \longrightarrow \mathbb{R}_+^{Q' \times Q'}$ where $Q' = \{p\} \cup \{(q, \sigma) \mid q \in Q, \sigma \in \Sigma\}$ and all entries of the matrices $\mu(a)$ and $\mu(b)$ are zero except the entries

$$\mu(x)_{p,(q',\sigma)} = \pi_\sigma \quad \text{and} \quad \mu(x)_{(q,\sigma),(q',\tau)} = p_{\sigma,\tau}$$

such that $\delta(p, \sigma) = q'$ and $\delta(q, \tau) = q'$ respectively, and (in both cases)

$$x = \begin{cases} a \text{ if } q' \in F \\ b \text{ otherwise.} \end{cases}$$

Denoting by e_p and $\mathbf{1}$ the characteristic vectors of $\{p\}$ and Q' respectively, the triple we are looking for is $(e_p, \mu, \mathbf{1})$. Indeed, let $f : \Sigma^+ \longrightarrow \{a,b\}^+$ be the function such that, for every $\sigma_1\sigma_2\cdots\sigma_n \in \Sigma^+$ (with $\sigma_i \in \Sigma$), $f(\sigma_1\sigma_2\cdots\sigma_n) = x_1x_2\cdots x_n$ where $x_i = a$ if $\delta(p,\sigma_1\sigma_2\cdots\sigma_i) \in F$ and $x_i = b$ otherwise. Then, for every $x \in \{a,b\}^+$, $e_p\mu(x)\mathbf{1} = \sum_{w\in f^{-1}(x)} \Pi_n(w)$. As a consequence, for every n and k, we obtain

$$\mathrm{P}_n\{Y_n = j\} = \frac{\sum_{x\in\{a,b\}^n,\ |x|_a=j} e_p\mu(x)\mathbf{1}}{\sum_{x\in\{a,b\}^n} e_p\mu(x)\mathbf{1}} = \frac{\sum_{w\in\Sigma^n, |w|_R=j} \Pi_n(w)}{\sum_{w\in\Sigma^n} \Pi_n(w)} = \Pi_n\{O_n = j\}.$$

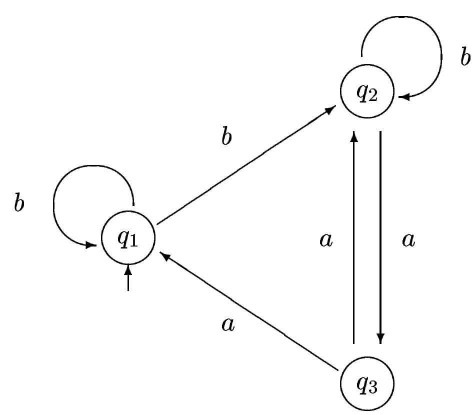
□

Intuitively, the construction carried out in the previous proof consists of four steps:

- building the (fully defined) deterministic automaton recognizing the language Σ^*R;
- re-labelling all the transitions entering in a final states by a and all other transitions by b;
- making final all states;
- using T and π to assign weights to transitions.

In the case of a Bernoulli model a simplified construction is described by the following example.

Example 4.3 Let $\Sigma = \{0,1\}$, define $R = \{10, 100, 101\}$, $\pi = (1/2, 1/2)$ and let $T = (T_{\sigma\tau})$ be given by $T_{\sigma\tau} = 1/2$ for every $\sigma, \tau \in \Sigma$. The construction of Theorem 4.2 provides an automaton which can be reduced by collapsing any pair of equivalent states (i.e. those with equal outgoing transitions) into a unique state. Thus, we get the (weighted) non-deterministic finite automaton represented in Fig. 4.2, where all states are final and all transitions have weight 1/2. The corresponding linear representation (ξ, μ, η) defines a RSF problem equivalent to the original MPF problem. □



$$\xi = (1,0,0)_T \qquad \eta = (1,1,1)_T$$

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

$$B = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Figure 4.2: State diagram and linear representation of Example 4.3.

Remark 4.4 *The converse of Theorem 4.2 does not hold in general.*

Indeed, for an arbitrary $\{O_n\}_n$ defined by a triple (π, P, R), the construction of Theorem 4.2 yields a rational series in non-commutative variables a, b. Applying the morphism $\mathcal{H}$ to such a series, we get

$$\mathbf{r}(z,w) = \sum_{n=0}^{\infty}\sum_{j=0}^{n} \Pi_n\{O_n = j\}\ e^{zj}w^n$$

and considering it as a real valued function in two variables, it can be easily verified that it is rational in the variables e^z, w. On the contrary, there exist rational series that do not satisfy this property, as the following examples show.

Example 4.5 Consider the series r of Example 4.1. It leads to the function

$$\mathbf{r}(z,w) = \sum_{n=0}^{\infty}\sum_{j=0}^{n} \mathrm{P}_n\{Y_n = j\}e^{zj}w^n = (e^z - 1)\frac{\log(1-w)}{w} + \frac{e^z}{1-w}$$

which clearly is not rational in w. □

Example 4.6 Consider the series r defined by the following linear representation

$$\eta = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \mu(a) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \mu(b) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \eta = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

whose corresponding automaton is represented in Fig. 4.3. Observe that the support of r is the regular language $L = \{b, ab\}^*$. Now, let F_n be the number of words in L having length n. Then

$$\mathbf{r}(z,w) = \sum_{n=0}^{\infty}\sum_{j=0}^{n} \mathrm{P}_n\{Y_n = j\}\ e^{zj}w^n = \sum_{n=0}^{\infty}\sum_{j=0}^{n} \frac{s_{n,j}}{F_n}\ e^{zj}w^n\ .$$

It is easy to see that $F_1 = F_2 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for every $n \geq 2$. Hence F_n is the n-th Fibonacci number and we can write

$$F_n = \frac{1}{\sqrt{5}}\left(\phi^n - \overline{\phi}^{\,n}\right),$$

where $\phi = (1+\sqrt{5})/2$ is the golden ratio and $\overline{\phi} = (1-\sqrt{5})/2$. In particular this implies

$$\mathbf{r}(0,w) = \sqrt{5}\ \sum_{n=0}^{\infty}\sum_{j=0}^{n} \frac{s_{n,j}}{\phi^n - \overline{\phi}^{\,n}}\ w^n$$

which is not rational in w. Then, neither $\mathbf{r}(z,w)$ can be rational in the variables e^z and w. □

4.4 Primitive models

In this section we study the RSF problem for a nontrivial subclass of rational models, defined by linear representations with primitive counting matrices. We show that, in these cases, the asymptotic behaviours of the mean and the variance of the r.v. Y_n are strictly linear. Moreover we prove that, when suitably normalized, Y_n converges in law to a standard Gaussian distribution. Finally, we establish a local limit theorem for Y_n which turns out to be related to the notion of

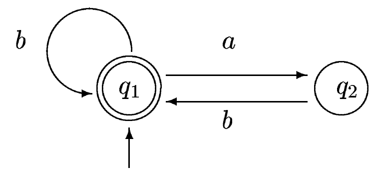


Figure 4.3: State diagram associated with the rational model defined in Example 4.6.

symbol periodicity introduced in Section 2.4. We remark that these results constitute the starting point to deal with more general models in the following chapters.

Our analysis is based on the study of the moment generating function $\Psi_n(z)$ of Y_n; in particular we show that $\Psi_n(z)$ satisfies the assumptions of the criteria presented in Theorems 3.5 and 3.7. Recalling the discussion of Section 4.2, it is clear that the complex functions $r_n(z)$ and their derivatives play a key role.

Still using the notation introduces so far, let r be a series in $\mathbb{R}_+\langle\langle a,b\rangle\rangle$ ad assume that r admits a linear representation (ξ,μ,η) such that $A+B$ is primitive. Then, we say that the rational model defined by r is a *primitive model.* With an abuse of terminology, we also say that (ξ,μ,η) and r are *primitive.* Under these assumptions, the matrix $M = A+B$ verifies the Perron–Frobenius Theorem and all other properties we presented in Section 2.3. Thus, we know that M admits a unique eigenvalue λ of maximum modulus and with λ we can associate strictly positive left and right eigenvectors, say v and u. In particular we can choose v and u normed so that $v_{_T} u = 1$. Moreover by Proposition 2.10 we can write

$$M^n = \lambda^n (u v_{_T} + C(n)) \tag{4.12}$$

where $C(n)$ is a real matrix such that $C(n) = \mathrm{O}\,(\varepsilon^n)$, for some $0<\varepsilon<1$. Note that the matrix

$$C = \sum_{n=0}^{\infty} C(n)$$

is well-defined and satisfies $v_{_T} C = Cu = 0$. Also remark that the value λ is intrinsically linked to the series r and it is invariant with respect to the chosen linear representation of r. Indeed, recalling the definition (4.4) of $r_n(z)$ and by equation (4.9) we have

$$\sum_{|x|=n} (r,x) = r_n(0) = \xi_{_T} M^n \eta = \lambda^n (\xi_{_T} u)(v_{_T}\eta) + O(\rho^n) \tag{4.13}$$

for some $0<\rho<\lambda$.

Now, let us consider the generating function $R(z,w)$ of the sequence $\{r_n(z)\}_n$. We recall that this function is well-defined in a neighbourhood of $(0,\lambda)$, where it satisfies relation (4.11). The singularities analysis of $R(z,w)$ in a neighbourhood of $z=0$ allows us to prove the following proposition.

Proposition 4.7 *Let r be a primitive series and λ be the Perron–Frobenius eigenvalue associated with r. Then there exist an analytic function $y(z)$ at $z=0$, satisfying $y(0)=\lambda$ and a matrix function $F(z)$, having analytic and non-null entries at $z=0$, such that for every z near 0 we have*

$$R(z,w) = \frac{F(z)}{1-y(z)w} + O\left(\frac{1}{1-\rho w}\right), \qquad \textit{as } n\to\infty$$

where $0 < \rho < |y(z)|$.

Proof. By equation (4.11), it is clear that the singularities of $R(z,w)$ are the inverse of the eigenvalues of $Ae^z + B$. The Perron–Frobenius Theorem guarantees that λ is a simple root of the characteristic polynomial of M and this means that it is a simple root of $\mathrm{Det}(yI - (A+B))$. Thus the equation

$$\mathrm{Det}\,(yI - Ae^z - B) = 0 \tag{4.14}$$

defines an implicit function $y = y(z)$ analytic in a neighbourhood of $z = 0$ such that $y(0) = \lambda$. Note that if $A \neq 0$, then $y'(0) \neq 0$. Moreover, by a continuity property, there exists $\rho > 0$ such that, for every z near 0, all roots μ of (4.14) different from 1 (i.e. all other eigenvalues of $Ae^z + B$) satisfy the relation $|\mu| < \rho < |y(z)|$. Hence the singularities of $R(z,w)$, except λ^{-1}, are all greater than ρ^{-1}. By decomposing in partial fractions, we can express the matrix $R(z,w)$ in the form

$$R(z,w) = \frac{F(z)}{1 - y(z)w} + E(z,w) \tag{4.15}$$

where, for every $|z| \leq c$, $E(z,w)$ has singularities μ^{-1} of modulus greater than ρ^{-1}. By l'Hôpital's rule, $F(z)$ is given by

$$F(z) = -\frac{y(z) \cdot \mathrm{Adj}\left(I - y(z)^{-1}(Ae^z + B)\right)}{\frac{\partial}{\partial w}\mathrm{Det}\,(I - w(Ae^z + B))\,|_{w=y(z)^{-1}}}$$

and note that by point 5. of Perron–Frobenius Theorem, $\mathrm{Adj}(\lambda I - A - B) > 0$. Therefore, by continuity, all entries of $F(z)$ are different from 0 for every z near 0 and every w near λ^{-1}. This concludes the proof. □

It should be noted that if $A \neq 0$, then $y'(0) \neq 0$. The previous proposition yields a quasi-power condition for the sequence $\{r_n(z)\}_n$ in a neighbourhood of $z = 0$, as n goes to infinity.

Corollary 4.8 *Assume the hypotheses of Proposition 4.7 and let $y(z)$ and $F(z)$ be the functions defined in the same proposition. Then, for every z near 0 we have*

$$r_n(z) = \xi_T F(z)\eta \cdot y(z)^n + O(\rho^n), \qquad \text{as } n \to \infty$$

where $0 < \rho < |y(z)|$.

Proof. It suffices to observe that $\xi_T R(z,w)\eta$ is the generating function of $\{r_n(z)\}_n$. □

4.4.1 Analysis of mean value and variance in the primitive case

In this section we prove that if $A \neq 0 \neq B$, then the mean and the variance of Y_n have strictly linear behaviour with respect to n.

Theorem 4.9 *Let r be a primitive series, (ξ, μ, η) one of its linear representations and λ its Perron–Frobenius eigenvalue. Also let $y(z)$ and $F(z)$ be the functions defined in Proposition 4.7. If Y_n counts the occurrences of a in the rational model defined by r, then the mean value and the variance of Y_n satisfy the relations*

$$\mathbb{E}(Y_n) = \beta n + \frac{\delta}{\alpha} + O(\varepsilon^n), \qquad \mathbb{V}ar(Y_n) = \gamma n + O(1), \tag{4.16}$$

where $|\varepsilon| < 1$ *and*

$$\beta = \frac{y'(0)}{\lambda}, \qquad \gamma = \frac{y''(0)}{\lambda} - \beta^2, \qquad \alpha = \xi_T F(0)\,\eta, \qquad \delta = \xi_T F'(0)\,\eta. \tag{4.17}$$

Proof. By equation (4.15), we can write

$$r_n(z) = \xi_{T} F(z)\eta \cdot y(z)^n + E_n(z) \tag{4.18}$$

where $E_n(z)$ is the n-th coefficient of $E(z,y)$ In particular, since $y(0) = \lambda$, we have

$$r_n(0) = \alpha\lambda^n + O(\rho^n)$$

for some $0 < \rho < \lambda$. Note that the derivatives of $E(z,y)$ at $z = 0$ have the same singularities of $E(0,y)$ and hence the growth order of $E'_n(0)$ and $E''_n(0)$ is also $O(\rho^n)$. Thus, differentiating both sides of (4.18) ad evaluating at $z = 0$ we obtain

$$\begin{aligned} r'_n(0) &= (\beta\alpha \cdot n + \delta)\,\lambda^n + O(\rho^n) \\ r''_n(0) &= \left(\beta^2\alpha \cdot (n^2 - n) + 2\beta\delta \cdot n + (\gamma + \beta^2)\alpha \cdot n\right)\lambda^n + \ O(\lambda^n)\ . \end{aligned}$$

Then, the statement is proved by recalling formulas (4.7). □

Corollary 4.10 *Under the hypotheses of Theorem 4.9, in a neighbourhood of $z = 0$ we have*

$$y(z) = \lambda\left(1 + \beta z + \frac{\beta^2 + \gamma}{2} z^2 + O(z^3)\right)\ .$$

Remark 4.11 *The constants α, β, γ and δ can be explicitly determined from M and its eigenvectors. Indeed, the following equalities hold:*

$$\alpha = (\xi_{T} u)(v_{T}\eta), \qquad \beta = \frac{v_{T} A u}{\lambda}, \qquad \gamma = \beta - \beta^2 + 2\frac{v_{T} ACAu}{\lambda^2},$$

$$\delta = \xi_{T} D\eta, \qquad \textit{where} \qquad D = \frac{CA}{\lambda} u v_{T} + u v_{T} \frac{AC}{\lambda}.$$

Proof. The first equality follows from equation (4.13). The others can be proved as follows. Observe that (4.12) implies

$$R(0,w) = \sum_{n=0}^{+\infty} M^n w^n = \sum_{n=0}^{+\infty} u v_{T} \lambda^n w^n + \sum_{n=0}^{+\infty} C(n)\lambda^n w^n\ . \tag{4.19}$$

Since each entry of $\sum_n C(n)x^n$ converges uniformly for x near 1 to a rational function, we have $\sum_n C(n)x^n = C + O(1-x)$ and hence the second series in (4.19) equals $C + O(1 - \lambda w)$, which proves

$$R(0,w) = \frac{u v_{T}}{1 - \lambda w}\ +\ C\ +\ \mathrm{O}(1 - \lambda w)\ . \tag{4.20}$$

Now let R_z and R_{zz} denote the partial derivatives $\partial R/\partial z$ and $\partial^2 R/\partial z^2$, respectively. Recalling the derivative formulas (2.3) and (2.4) we get

$$\begin{aligned} R_z(0,w) &= R(0,w)\ Aw\ R(0,w) \\ R_{zz}(0,w) &= R_z(0,w) \cdot [I + 2Aw\ R(0,w)] \end{aligned}$$

and note that by equation (4.4) the sequences $\{r'_n(0)\}$ and $\{r''_n(0)\}$ have generating function given by $\xi_{T} R_z(o,w)\ \eta$ and $\xi_{T} R_{zz}(0,w)\ \eta$. Then, replacing (4.20) in the previous expressions, one can obtain expansions for $r'_n(0)$ and $r''_n(0)$. The result follows by comparing such expansions with those appearing in the proof of Theorem 4.9. □

For the sake of brevity, from now on we say that β and γ are the *mean constant* and the *variance constant* of Y_n. Observe that these constants do not depend on the initial and final states of the weighted automaton associated with r. In particular, if r is the characteristic series of a language L, then $\mathbb{E}(Y_n) = \frac{n}{2} + O(1)$ and hence $\beta = \frac{1}{2}$.

In general, note that $B = 0$ implies $\beta = 1$ and $\gamma = \delta = 0$, while $A = 0$ implies $\beta = \gamma = \delta = 0$; in these cases we say that the linear representation (ξ, μ, η) and the series r are *degenerate*. On the contrary, we show that if $A \neq 0 \neq B$, then the constants of the main terms of mean value and variance of Y_n are non-null. To this aim, let us first prove a technical lemma, where we use the following notation: given a non-null polynomial $p(x) = \sum_k p_k x^k$ having coefficients in $\mathbb{R}_+$, we use $V(p)$ to denote the variance of any r.v. X_p such that $\mathrm{P}\{X_p = k\} = \frac{p_k}{p(1)}$.

Lemma 4.12 *For any pair of non-null polynomials p, q with nonnegative real coefficients, we have*

$$V(pq) = V(p) + V(q) \ , \qquad V(p+q) \geq \frac{p(1)}{p(1)+q(1)} V(p) + \frac{q(1)}{p(1)+q(1)} V(q) \ .$$

In particular, we have $V(p+q) \geq \min\{V(p), V(q)\}$.

Proof. By the definition, we get

$$V(p) = \frac{p''(1) + p'(1)}{p(1)} - \left(\frac{p'(1)}{p(1)}\right)^2 \ . \tag{4.21}$$

Hence, the first equation is immediately proven. Further, observe that

$$(p(1)+q(1))V(p+q) = p''(1) + q''(1) + p'(1) + q'(1) - \frac{(p'(1)+q'(1))^2}{p(1)+q(1)}.$$

Thus, the second relation follows again from (4.21) by recalling that $\frac{(a+b)^2}{c+d} \leq \frac{a^2}{c} + \frac{b^2}{d}$, for every four-tuple of positive values a, b, c, d. □

Theorem 4.13 *Let $r \in \mathbb{R}_+\langle\langle a, b\rangle\rangle$ be a non-degenerate primitive series. If Y_n counts the occurrences of a in the rational model defined by r, then both $\mathbb{E}(Y_n)$ and $\mathbb{V}ar(Y_n)$ have strictly linear behaviour with respect to n.*

Proof. Let β and γ be defined as in Theorem 4.9. Since A is non-null and both v and u are strictly positive (point 2. of Perron–Frobenius Theorem), it is clear that $\beta > 0$. Proving that γ is strictly positive is equivalent to proving that $\mathbb{V}ar(Y_n) \geq cn$ for some $c > 0$ and for infinitely many n. Since $A+B$ is primitive and both A and B are non-null, there exists an integer t such that all the entries of the matrix $C = (Ax+B)^t$ are polynomials with at least two non-null coefficients. This implies that the value

$$c = \min\{V(C_{ij}) \mid i, j = 1, 2, \ldots, m\}$$

is strictly positive. Then, by the previous lemma, for every $n \in \mathbb{N}$ and every pair of indices i, j we have

$$V(C^{n+1}_{ij}) \geq \min\{V(C^n_{ik}) + V(C_{kj}) \mid k = 1, 2, \ldots, m\} .$$

As a consequence, $V(C^{n+1}_{ij}) \geq c + \min\{V(C^n_{ik}) \mid k = 1, 2, \ldots, m\}$ proving that $V(C^n_{ij}) \geq nc$. Since Y_n is a r.v. associated with $\xi'(Ax+B)^n\eta$, we get

$$\mathbb{V}ar(Y_{tn}) \geq \min\{V(C^n_{ij}) \mid i, j = 1, 2, \ldots, m\} \geq nc$$

for every $n \in \mathbb{N}$. Together with (4.16) this proves $\mathbb{V}ar(Y_n) = \Theta(n)$ and hence $\gamma > 0$. □

4.4.2 Limit theorems in the primitive models

In this section we show that, in primitive models, the sequence of normalized r.v.'s $\{\frac{Y_n-\beta n}{\sqrt{\gamma n}}\}_n$ converges in law to a standard Gaussian distribution. Moreover, we establish a local limit theorem for Y_n which turns out to be related to the notion of symbol periodicity introduced in Section 2.4. Actually, we prove a more general result, that is we show that the previous properties hold for an arbitrary power of any primitive series.

As in the previous section, we say that a series $r \in \mathbb{R}_+^{Rat}\langle\langle a,b\rangle\rangle$, is *primitive* if it admits a linear representation (ξ,μ,η) such that $\mu(a)+\mu(b)$ is a primitive matrix; moreover, we say that r is *non-degenerate* if $\mu(a)\neq 0$ and $\mu(b)\neq 0$.

Theorem 4.14 *For any positive integer k and any primitive nondegenerate $r \in \mathbb{R}_+^{Rat}\langle\langle a,b\rangle\rangle$, let s be defined by $s=r^k$ and let Y_n count the occurrences of a in the model defined by s. Then the following properties hold.*

T1 *There exist two constants α and β, satisfying $\alpha>0$ and $0<\beta<1$, such that $\frac{Y_n-\beta n}{\sqrt{\alpha n}}$ converges in distribution to a normal r.v. of mean value 0 and variance 1.*

T2 *If (ξ,μ,η) is a primitive linear representation for r and d is the x-period of $\mu(a)x+\mu(b)$, then there exist d functions $C_i:\mathbb{N}\longrightarrow\mathbb{R}_+$, $i=0,1,\ldots,d-1$, such that $\sum_i C_i(n)=1$ for every $n\in\mathbb{N}$ and further, as n grows to $+\infty$, the relation*

$$P_n\{Y_n=j\} = \frac{d\,C_{\langle j\rangle_d}(n)}{\sqrt{2\pi\alpha n}}\,e^{-\frac{(j-\beta n)^2}{2\alpha n}}\cdot(1+o(1)) \tag{4.22}$$

holds uniformly for every $j=0,1,\ldots,n$ (here $\langle j\rangle_d = j-\lfloor j/d\rfloor d$).

Observe that the primitivity hypothesis cannot be omitted to obtain a Gaussian limit distribution; to see this fact it is sufficient to consider the language a^*b^*. Also notice that in case $k=1$, statement T2 establishes a local limit theorem for any primitive model.

Before proving the previous theorem, let us illustrate its meaning by some examples.

Example 4.15 Consider the RSF problem defined by the weighted automaton of Example 4.3. As shown there, it is equivalent to a MPF problem where the pattern is $R=\{ba,baa,bab\}$ and the stochastic model is given by a Bernoulli process of parameter 1/2. Clearly, the counting matrix $M=A+B$ associated with the automaton is primitive; moreover, one can prove that the x-period of $M(x)=Ax+B$ is $d=2$.

The characteristic polynomial of M is $y^2(y-1)$; hence its Perron–Frobenius eigenvalue is 1. Moreover, $u=(1/2,1/2,1/2)_T$ and $v=(1/2,1,1/2)_T$ are right and left eigenvalues associated with λ, normed so that $v_T u=1$. Thus, recalling (4.13) we can easily get $r_n(0)=1+O(\epsilon)$ for some $\epsilon<1$. Also, the generating function of $\{r_n(z)\}_n$ can be computing using (4.11), so obtaining

$$\mathbf{r}(z,w)=\xi_T R(z,w)\,\eta = \frac{e^z w^2 - e^{2z}w^2+4}{w^2-e^{2z}w^2-4w+4}\ .$$

Hence, considering the expansion of $\mathbf{r}(x,w)$ with respect to e^z and w, we can directly compute its coefficients $r_{n,j}$ for $2\le j\le n$. As a consequence, by relation (4.5) we get

$$\mathrm{P}_n\{Y_n=j\}=\frac{1}{2^n}\cdot\begin{cases}\binom{n+1}{j+1}-\binom{n-1}{j-1} & \text{if } j \text{ is even}\\[2ex] \binom{n-1}{j} & \text{if } j \text{ is odd and } j<n-1\ .\end{cases}$$

This result is consistent with statement T2 of Theorem 4.14, as also shown in Figure 4.4. □

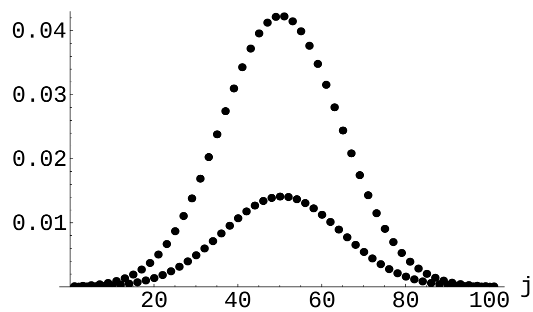


Figure 4.4: Plot of the probability function $\mathrm{P}_n\{Y_n = j\}$ obtained in Example 4.15, for $n = 800$ and $350 \leq j \leq 450$. The limit behaviour is given by the superimposition of two alternating Gaussian densities.

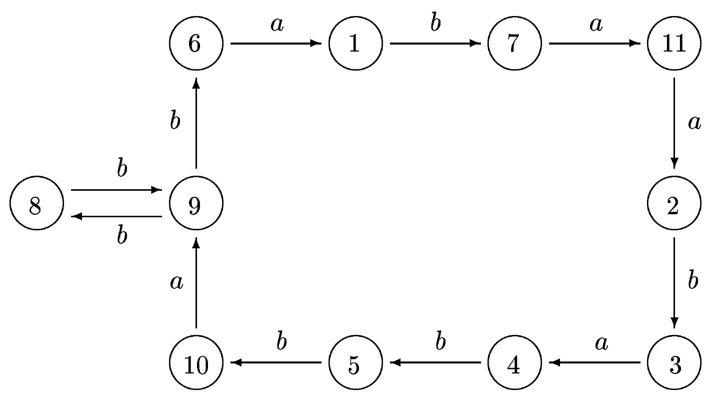


Figure 4.5: State diagram of the automaton defined in Example 4.16.

Example 4.16 Consider the automaton represented in Fig. 4.5. Clearly, it is primitive and its a-counting matrix has x-period 5. Figures from 4.6 to 4.11 represent the probability function and the cumulative distribution of the r.v. Y_n obtained choosing different vectors ξ and η. All plots are drawn for $n = 340$ and j between 60 and 150. It is easy to observe the convergence in distribution to a Gaussian r.v. (statement T1 of Theorem 4.14) and the pointwise superimposition of translated Gaussian behaviours (statement T2 of Theorem 4.14). □

Proof of Theorem 4.14

We split the proof in two separate parts and we use the criteria presented in Theorem 3.5 and in Theorem 3.7.

Proof of T1 Since $s = r^k$, by applying the morphism $\mathcal{H}$ defined in (4.8) we get

$$\mathbf{s}(z, w) = \mathbf{r}(z, w)^k \ .$$

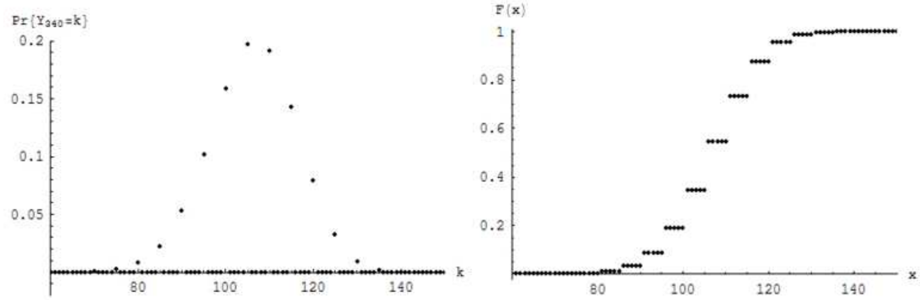


Figure 4.6: Plot of the probability function and the cumulative distribution for the r.v. Y_{340} of Example 4.16 with $\xi = \eta = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)_T$.

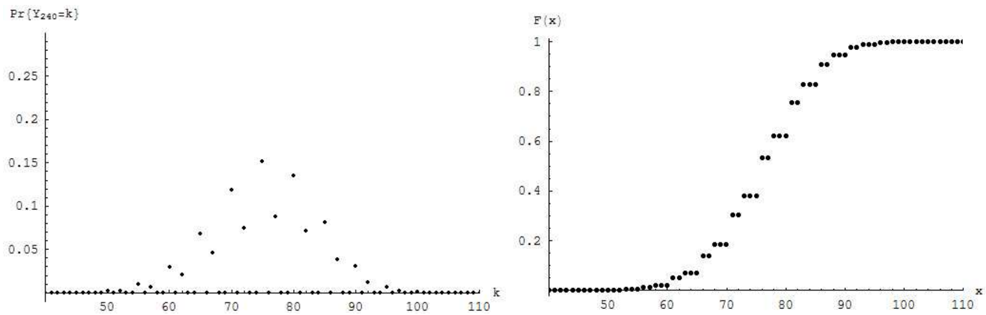


Figure 4.7: Plot of the probability function and the cumulative distribution for the r.v. Y_{340} of Example 4.16 with $\xi = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)_T$ and $\eta = (1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)_T$.

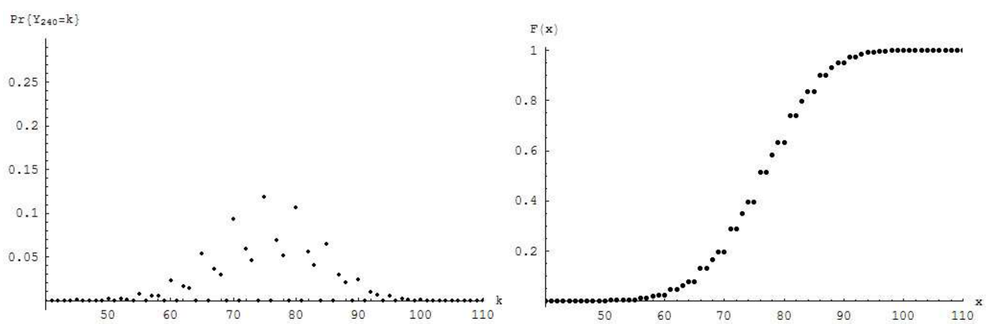


Figure 4.8: Plot of the probability function and the cumulative distribution for the r.v. Y_{340} of Example 4.16 with $\xi = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)_T$ and $\eta = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0)_T$.

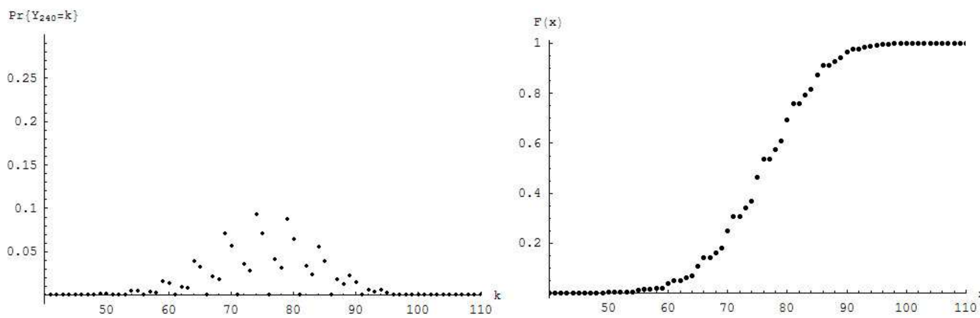


Figure 4.9: Plot of the probability function and the cumulative distribution for the r.v. Y_{340} of Example 4.16 with $\xi = (1,0,0,0,0,0,0,0,0,0,0)_T$ and $\eta = (1,1,1,1,1,1,1,1,1,0,0)_T$.

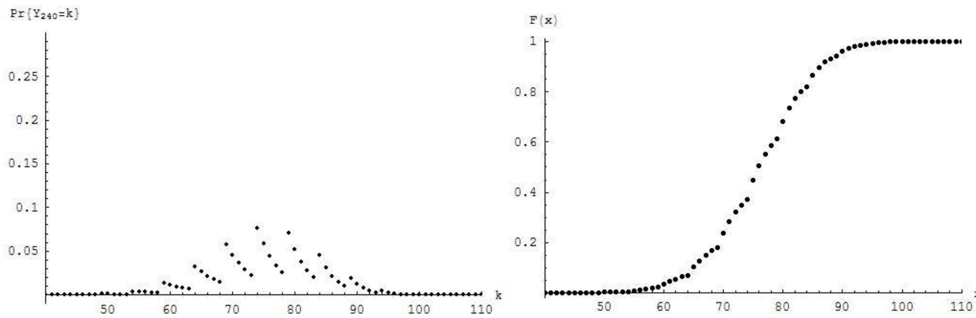


Figure 4.10: Plot of the probability function and the cumulative distribution for the r.v. Y_{340} of Example 4.16 with $\xi = (1,0,0,0,0,0,0,0,0,0,0)_T$ and $\eta = (1,1,1,1,1,1,1,1,1,1,1)_T$.

By Proposition 4.7, we know that near the point $(0, \lambda^{-1})$ the function $\mathbf{r}(z,w) = \xi_T R(z,w)\eta$ admits a Laurent expansion

$$\mathbf{r}(z,w) = \frac{f(z)}{1 - y(z)w} + \mathrm{O}(1)$$

where $f(z)$ and $y(z)$ are complex functions, which are non-null and analytic at $z = 0$; moreover, $y(0) = \lambda$. As a consequence, in a neighbourhood of $(0, \lambda^{-1})$ we have

$$\mathbf{s}(z,w) = \left(\frac{f(z)}{1 - y(z)w}\right)^k + \mathrm{O}\left(\frac{1}{1 - y(z)w}\right)^{k-1}$$

and hence the associated sequence is

$$s_n(z) = f(z)^k \binom{n+k-1}{k-1} y(z)^n + \mathrm{O}\left(n^{k-2} y(z)^n\right).$$

Now, since Y_n counts the occurrences of a in the model defined by the series s, by relation (4.6) its moment generating function is given by $\Psi_{Y_n}(z) = s_n(z)/s_n(0)$ and hence, in a neighbourhood of $z = 0$, it has an expansion

$$\Psi_{Y_n}(z) = \frac{s_n(z)}{s_n(0)} = \left(\frac{f(z)}{f(0)}\right)^k \cdot \left(\frac{y(z)}{\lambda}\right)^n \cdot \left(1 + \mathrm{O}(n^{-1})\right) .$$

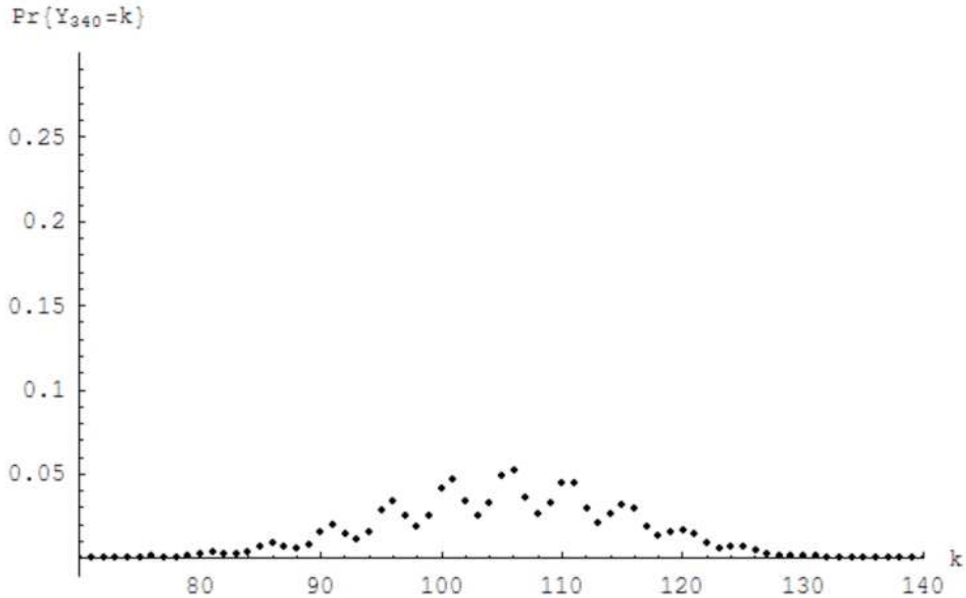


Figure 4.11: Plot of the probability function and the cumulative distribution for the r.v. Y_{340} of Example 4.16 when all states are made both initial and final.

Finally, recall that by Theorem 4.13 $\beta = y'(0)/\lambda$ and $\gamma = y''(0)/\lambda - \beta^2$ are strictly positive. Hence $\{Y_n\}$ satisfies both conditions of Theorem 3.5, with $h(z) = (f(z)/f(0))^k$, $u(z) = y(z)/\lambda$, $\mu = \beta$ and $\sigma = \alpha$. This proves the result. □

Proof of T2 For every $p, q \in \{1, 2, \ldots, m\}$, let $r^{(pq)}$ be the series defined by the linear representation $(\xi_p e_p,\ \mu,\ \eta_q e_q)$, where e_i is the characteristic vector of entry i. Then

$$r = \sum_{p,q=1}^{m} r^{(pq)}$$

Thus, since $s = r^k$, we have

$$s = \sum r^{(p_1 q_1)} \cdot r^{(p_2 q_2)} \cdots r^{(p_k q_k)} \tag{4.23}$$

where the sum is over all sequences $\ell = p_1 q_1 p_2 q_2 \cdots p_k q_k \in \{1, 2, \ldots, m\}^{2k}$. For the sake of brevity, for every such ℓ, let $r^{(\ell)}$ be the series

$$r^{(\ell)} = r^{(p_1 q_1)} \cdot r^{(p_2 q_2)} \cdots r^{(p_k q_k)} \ .$$

By the primitivity hypothesis this series is identically null if and only if $\xi_{p_j} = 0$ or $\eta_{q_j} = 0$ for some $j \in \{1, 2, \ldots, k\}$. For this reason set $Supp = \{\ell \in \{1, 2, \ldots, m\}^{2k} \mid r^{(\ell)} \neq 0\}$. Then, for every $\ell \in Supp$, applying the morphism $\mathcal{R}$ to the previous equation and recalling (4.10), we get

$$\mathbf{r}^{(\ell)}(0, w) = \prod_{i=1}^{k} \mathbf{r}^{(p_i q_i)}(0, w) = \prod_{i=1}^{k} \xi_{p_i} (I - wM)^{-1}_{p_i q_i} \eta_{q_i}$$

with the obvious meaning for the notations $\mathbf{r}^{(\ell)}$ and $\mathbf{r}^{(p_i q_i)}$. This implies, by the primitivity of M, that $\mathbf{r}^{(\ell)}(0,w)$ has a unique pole of smallest modulus at λ^{-1}, which has degree k. As a consequence, the sequence associated with $\mathbf{r}^{(\ell)}(0,w)$ satisfies the following relation

$$r_n^{(\ell)}(0) = c_\ell\, n^{k-1}\lambda^n + \mathrm{O}(n^{k-2}\lambda^n) \tag{4.24}$$

for some $c_\ell > 0$. Moreover, from (4.23) we have $s = \sum_{\ell \in Supp} r^{(\ell)}$ and hence

$$s_{n,j} = \sum_{\ell \in Supp} r_{n,j}^{(\ell)} \qquad \text{and} \qquad s_n(0) = \sum_{\ell \in Supp} r_n^{(\ell)}(0) = c\, n^{k-1}\lambda^n + \mathrm{O}(n^{k-2}\lambda^n) \tag{4.25}$$

for some $c > 0$. Then recalling equation (4.5), for every $j \in \{0, 1, \ldots, n\}$ we have

$$\mathrm{P}_n\{Y_n = j\} = \frac{s_{n,j}}{s_n(0)} = \sum_{\ell \in Supp} \frac{r_{n,j}^{(\ell)}}{s_n(0)} = \sum_{\ell \in Supp} \frac{r_n^{(\ell)}(0)}{s_n(0)}\, \mathrm{P}_n\{Y_n^{(\ell)} = j\}$$

where $Y_n^{(\ell)}$ counts the occurrences of a in the model defined by $r^{(\ell)}$. Finally, from equation (4.25), we get

$$\mathrm{P}_n\{Y_n = j\} = \sum_{\ell \in Supp} C_\ell\, \mathrm{P}_n\{Y_n^{(\ell)} = j\} + \mathrm{O}(n^{-1}) \tag{4.26}$$

where C_ℓ is a positive constant for every $\ell \in Supp$ and $\sum_{\ell \in Supp} C_\ell = 1$.

Thus, to determine the local behaviour of $\{Y_n\}$, we first study $\{Y_n^{(\ell)}\}$. Indeed, by the previous relation, it is sufficient to prove that the equation

$$\mathrm{P}_n\{Y_n^{(\ell)} = j\} = \begin{cases} \dfrac{d\, e^{-\frac{(j-\beta n)^2}{2\alpha n}}}{\sqrt{2\pi\alpha n}} \cdot (1 + o(1)) & \text{if } j \equiv \rho_\ell \pmod d \\ 0 & \text{otherwise} \end{cases}$$

holds uniformly for every $j = 0, 1, \ldots, n$, where α and β are defined as in T1, while ρ_ℓ is a (possibly depending on n) integer such that $0 \le \rho_\ell < d$ (in particular $C_i(n) = \sum_{\rho_\ell = i} C_\ell$ for each i). To this aim, we simply have to show that, for every $n \in \mathbb{N}$, $Y_n^{(\ell)}$ satisfies the hypotheses of Theorem 3.7.

First, we prove that $Y_n^{(\ell)}$ takes values in a set like (3.9), where d is the x-period of $\mu(a)x + \mu(b)$. We provide an integer ρ_ℓ, such that, if $j \not\equiv \rho_\ell \pmod d$, then $r_{n,j}^{(\ell)} = 0$: by equation (4.5), this implies that Prob $\{Y_n^{(\ell)} = j\}$ vanishes, too. By the definition of $r^{(\ell)}$, it is clear that for any $j = 0, 1, \ldots, n$, the values $r_{n,j}^{(\ell)}$ are given by the convolutions

$$r_{n,j}^{(\ell)} = \sum_{\substack{n_1 + n_2 + \cdots + n_k = n \\ j_1 + j_2 + \cdots + j_k = j}} \prod r_{n_i, j_i}^{(p_i q_i)} \ .$$

Now, consider any $r_{n_i,j_i}^{(p_i q_i)}$. By Proposition 2.21, we know that for each pair p_i, q_i there exist an integer δ_i, $0 \le \delta_i < d$ such that

$$r_{n_i,j_i}^{(p_i q_i)} \neq 0 \quad \text{implies} \quad j_i \equiv \gamma n_i + \delta_i \pmod d$$

where $0 \le \gamma < d$ does not depend on p_i and q_i. Thus, choosing ρ_ℓ so that $0 \le \rho_\ell < d$ and $\rho_\ell \equiv \gamma n + \sum_{i=1}^k \delta_i \pmod d$, we have that $r_{n,j}^{(\ell)} \neq 0$ implies $j \equiv \rho_\ell \pmod d$.

As far as condition C1 and C2 are concerned, we can argue (with obvious changes) as in the proof of T1 and observe that the two constants α and β are the same for all series $r^{(\ell)}$ with $\ell \in Supp$, since they depend on the matrices A and B (not on the initial and final arrays).

Finally, to prove condition C3 let us consider the generating function of $\{r_n^{(\ell)}(z)\}$:

$$\mathbf{r}^{(\ell)}(z,w) = \prod_{j=1}^{k} \xi_{p_j}(I - w(Ae^z + B))^{-1}{}_{p_j q_j}\ \eta_{q_j} \ .$$

For every $\theta \in \mathbb{R}$ we have

$$\mathbf{r}^{(\ell)}(i\theta, w) \ = \ \frac{\prod_{j=1}^{k} \ \xi_{p_j}\mathrm{Adj}(I - w(Ae^{i\theta} + B))_{p_j q_j}\ \eta_{q_j}}{[\mathrm{Det}(I - w(Ae^{i\theta} + B))]^k}$$

showing that the singularities of the function are inverses of eigenvalues of $Ae^{i\theta} + B$. As a consequence, by Theorem 2.26, for every $\theta \neq 2k\pi/d$, all singularities of $\mathbf{r}^{(\ell)}(i\theta, w)$ are in modulus greater than λ^{-1}. Hence, by Cauchy's integral formula, for any arbitrary $\theta_0 \in (0, \pi/d)$ we can choose $0 < \tau < \lambda$ such that the associated sequence $\{r_n^{(\ell)}(i\theta)\}$ is bounded by $\mathrm{O}(\tau^n)$ for every $|\theta| \in [\theta_0, \pi/d]$. By (4.24) this implies

$$\Psi_{Y_n^{(\ell)}}(i\theta) \ = \ \frac{r_n^{(\ell)}(i\theta)}{r_n^{(\ell)}(0)} \ = \ \frac{\mathrm{O}(\tau^n)}{\Theta(n^{k-1}\lambda^n)} = \mathrm{O}(\epsilon^n)$$

for some $0 < \epsilon < 1$, which proves condition C3. □

4.5 Estimate of the maximum coefficients of a rational series

The local limit property proved in the last section can be used to study the order of growth of the maximum coefficients of rational formal series in commutative variables. This problem was actually among the original motivations of this thesis and can be seen as an generalization of classical questions concerning the ambiguity in formal language (see Section 1.8).

Formally, given a series $r \in \mathbb{R}_+[[a,b]]$, we define its *maximum function* $g_r : \mathbb{N} \longrightarrow \mathbb{R}_+$ as

$$g_r(n) = \max\{|(r,x)| \ : \ x \in \{a,b\}^{\otimes}, |x| = n\} \qquad (\text{for every } n \in \mathbb{N}).$$

Here we estimate the order of magnitude of $g_r(n)$ for formal series in commuting variables that are powers of primitive rational formal series.

Corollary 4.17 *For any $k \in \mathbb{N}$, $k \neq 0$ and any primitive series $r \in \mathbb{R}_+^{Rat}\langle\langle a,b\rangle\rangle$, let $s = r^k$ and consider its commutative image $S = \varphi(s) \in \mathbb{R}_+^{Rat}[[a,b]]$. Then the maximum function of S satisfies the relation*

$$g_S(n) = \begin{cases} \Theta(n^{k-(3/2)}\lambda^n) & \text{if } r \text{ is not degenerate} \\ \Theta(n^{k-1}\lambda^n) & \text{otherwise} \end{cases}$$

where $\lambda > 0$.

Proof. Let (ξ, μ, η) be a primitive linear representation of r and let λ be the Perron-Frobenius eigenvalue of $\mu(a) + \mu(b)$. To determine $g_S(n)$ we have to compute the maximum of the values $s_{n,j} = (S, a^j b^{n-j})$ for $j = 0, 1, \ldots, n$.

First consider the case when r is not degenerate. Then, let Y_n count the occurrences of a in the model defined by $s = r^k$ and recall that $\mathrm{P}_n(Y_n = j) = s_{n,j}/s_n(0)$. Now, by (4.25) we have

$s_n(0) = \Theta(n^{k-1}\lambda^n)$ and by Theorem 4.14, the set of probabilities $\{\mathrm{P}_n(Y_n = j) \mid j = 0, 1, \ldots, n\}$ has the maximum at some integer $j \in [\beta n - d, \beta n + d]$, where it takes on a value of the order $\Theta(n^{-1/2})$. This proves the first equation.

On the other hand, if r is degenerate, then either $\mu(a) = 0$ or $\mu(b) = 0$. In the first case, all $r_{n,j}$ vanish except $r_{n,0}$ which is of the order $\Theta(\lambda^n)$. Hence for every n, the value $\max_j\{s_{n,j}\} = s_n(0)$ is given by the k-th convolution of $r_{n,0}$, which is of the order $\Theta(n^{k-1}\lambda^n)$. The case $\mu(b) = 0$ is similar. □

If in particular if $k = 1$ and r is the characteristic series of a language $L \subseteq \{a, b\}^*$, then we get

$$\max_{0 \le k \le n} \{ \ \#\{x \in L \cap \{a,b\}^n : |x|_a = k\} \ \} = \Theta\left(\frac{\lambda^n}{\sqrt{n}}\right) .$$

Observe that in the context of trace theory, the previous relation yields the growth of the degree of ambiguity of the trace language generated by L over the commutative monoid with generators $\{a, b\}$.

Example 4.18 Consider the rational function $(1 - a - b)^{-k}$. Its Taylor expansion near the origin yields the series

$$S = \sum_{n=0}^{+\infty} \binom{n+k-1}{k-1} \sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j} .$$

By direct computation, one can verify that

$$g_S(n) = \binom{n+k-1}{k-1} \binom{n}{\lfloor n/2 \rfloor} = \Theta(n^{k-3/2} 2^n) .$$

In fact, S can be obtained as the commutative image of the series $s = r^k$, where $r = \chi_{\{a,b\}^*} \in \mathbb{R}_+\langle\langle a, b\rangle\rangle$. □

Even though the statement of Theorem 4.14 cannot be extended to all rational models, we believe that the property given in Corollary 4.17 well represents the asymptotic behaviour of maximum coefficients of all rational formal series in two commutative variables. We actually think that a similar result holds for all rational formal series in commutative variables. More precisely, let us introduce the symbol $\widehat{\Theta}$ with the following meaning: for any pair of sequences $\{f_n\}, \{g_n\} \subseteq \mathbb{R}_+$, we have $g_n = \widehat{\Theta}(f_n)$ if $g_n = \mathrm{O}(f_n)$ and $g_{n_j} = \Theta(f_{n_j})$ for some monotone strictly increasing sequence $\{n_j\} \subseteq \mathbb{N}$. Then we conjecture that the asymptotic behaviour of the maximum function of every rational formal series $t \in \mathbb{R}_+[[\sigma_1, \cdots, \sigma_\ell]]$, is of the form

$$g_t(n) = \widehat{\Theta}\left(n^{k/2}\lambda^n\right)$$

for some integer $k \geq -\ell + 1$ and some $\lambda \in \mathbb{R}_+$.

Chapter 5

Bicomponent models

In this chapter we improve the analysis of the RSF problem, dropping the primitivity hypothesis assumed so far. More precisely, here we consider *bicomponent rational models*, defined by rational series corresponding to weighted automata with two primitive components. Two special examples are of particular interest: they occur when the formal series defining the model is, respectively, the sum or the product of two primitive formal series. We will call them the *sum* and the *product model*, respectively, and they will represent the leading examples of our discussion.

Our main results concern the asymptotic evaluation of mean value and variance and the limit distribution of the number of symbol occurrences in a word randomly generated according to such a bicomponent rational model. Note that in Chapter 4 we have shown that primitive models are characterized by Gaussian limit behaviour. Moreover, to our knowledge, the pattern frequency problem in the Markovian model is usually studied in the literature under primitive hypothesis and Gaussian limit distributions are generally obtained. On the contrary, here we get in many cases limit distributions quite different from the Gaussian one. Many different situations occur, mainly depending on two conditions: whether there exists a communication from the first to the second component or not (in this case we get a sum model); whether one component is *dominant*, i.e. its Perron–Frobenius eigenvalue is strictly greater than the Perron–Frobenius eigenvalue of the other one (if they are equal we say that the components are *equipotent*).

The chapter is organized as follows. After the formal definition of the problem, in Section 5.2 we analyze the dominant case, which splits in two further directions according whether the dominant component is *degenerate* (that is, all its transitions are labelled by the same symbol) or not. The equipotent case is studied in Section 5.3; here several subcases arise corresponding to the possible differences between the leading terms of the mean values and of the variances of the statistics associated with each component. Both sections 5.2 and 5.3 contain a final part that focuses on the differences occurring in the sum model. All results are discussed in the last section and summarized in Table 5.1.

5.1 Statement of the problem

Let us formally define the model we study in this chapter, named the *bicomponent model*. Consider a rational series r in the non-commuting variables a, b with coefficients in $\mathbb{R}_+$ and let (ξ, μ, η) be one of its linear representations. We assume that there exist two primitive linear representations (ξ_1, μ_1, η_1) and (ξ_2, μ_2, η_2), of size s and t respectively, satisfying the following relations:

$$\xi_T = (\xi_{1T}, \xi_{2T}), \qquad \mu(x) = \begin{pmatrix} \mu_1(x) & \mu_{12}(x) \\ 0 & \mu_2(x) \end{pmatrix}, \qquad \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \tag{5.1}$$

where $\mu_{12}(x) \in \mathbb{R}_+^{s\times t}$ for every $x \in \{a,b\}$.

Intuitively, this linear representation corresponds to a weighted non-deterministic finite state automaton (which may have more than one initial state) such that its state diagram consists of two disjoint strongly connected subgraphs, possibly equipped with some further arrows from the first component to the second one. Thus, a computation path

$$\ell = q_0 \stackrel{x_1}{\to} q_1 \stackrel{x_2}{\to} q_2 \cdots q_{n-1} \stackrel{x_n}{\to} q_n \tag{5.2}$$

can be of three different kinds:

1. All q_j's are in the first component (in which case we say that ℓ is *contained* in the first component);
2. There is an index $0 \le s < n$ such that the indices $q_0, q_1, \dots, q_s$ are in the first component while $q_{s+1}, \dots, q_n$ are in the second one. In this case x_{s+1} is the label of the transition from the first to the second component;
3. All q_j's are in the second component (in which case we say that ℓ is *contained* in the second component).

Clearly, paths of different kinds may be labelled by the same word. In the sequel we need to distinguish different situations, so we refine the probabilistic spaces introduced in Section 4.2. More precisely, let n be a positive integer such that $\xi_T\mu(\omega)\ \eta \neq 0$ for some $\omega \in \{a,b\}^n$ and denote by Ω_n the set of all computation paths of length n. Then, for each $\ell \in \Omega_n$ in the form (5.2), we define the probability of ℓ as

$$\mathrm{P}_n\{\ell\} = \frac{\xi_{q_0}\mu(x_1)_{q_0q_1}\mu(x_2)_{q_1q_2}\cdots\mu(x_n)_{q_{n-1}q_n}\eta_{q_n}}{\xi^T(\mu(a)+\mu(b))^n\eta} .$$

Denoting by 2^{Ω_n} the family of all subsets of Ω_n, it is clear that $\langle\Omega_n, 2^{\Omega_n}, \mathrm{P}_n\rangle$ is a probability space. Moreover, we consider the r.v. $Y_n : \Omega_n \to \{0,1,\dots,n\}$ such that $Y_n(\ell)$ is the number of a occurring in the label of ℓ, for each $\ell \in \Omega_n$. Then, for every integer $0 \le k \le n$, we have

$$\mathrm{P}_n\{Y_n = j\} = \frac{r_{n,j}}{\sum_{k=0}^n r_{n,k}},$$

where the values $r_{n,k}$ are defined as in (4.3). Thus, the probability distribution of the r.v. Y_n is the same as the one studied in Chapter 4, even though the sample space in more subtle.

For the sake of brevity, we use $r^{(i)}$ $(i = 1,2)$ to denote the series such that $(r^{(i)},\omega) = \xi_i\mu_i(\omega)\eta_i$, for all $\omega \in \{a,b\}^*$. Moreover, we use the notations $A_i = \mu_i(a)$, $B_i = \mu_i(b)$, $M_i = A_i + B_i$ for $i = 1,2$ and $A_{12} = \mu_{12}(a)$, $B_{12} = \mu_{12}(b)$, $M_{12} = A_{12} + B_{12}$. Hence, we have

$$A = \mu(a) = \begin{pmatrix} A_1 & A_{12} \\ 0 & A_2 \end{pmatrix}, \quad B = \mu(b) = \begin{pmatrix} B_1 & B_{12} \\ 0 & B_2 \end{pmatrix}, \quad M = A + B = \begin{pmatrix} M_1 & M_{12} \\ 0 & M_2 \end{pmatrix}.$$

We always assume $A \neq 0 \neq B$ and $\xi_1 \neq 0 \neq \eta_2$. Anyway, we take into consideration models such that $A_i = 0$ or $B_i = 0$ for some $i = 1,2$; in this case, consistently with the terminology used so far, we say that the i-th component is *degenerate*. To avoid trivial cases, we also assume the following *significance* hypothesis:

$$(A_1 \neq 0 \text{ or } A_2 \neq 0) \text{ and } (B_1 \neq 0 \text{ or } B_2 \neq 0) . \tag{5.3}$$

Note that if the last condition is not true, then Y_n may assume two values at most (either $\{0,1\}$ or $\{n-1,n\}$). Assuming the significance hypothesis means to forbid the cases when both components only have transitions labelled by the same letter (either a or b). For analogous reasons, when $M_{12} = 0$ we always assume $\xi_2 \neq 0 \neq \eta_1$.

Still using the notations introduced in the previous chapter, from now on the functions $r_n(z)$ and $R(z,w)$ are referred to the triple (ξ, μ, η), while the expressions ${r_n}^{(i)}(z)$ and $R^{(i)}(z,w)$ are referred to the triple (ξ_i, μ_i, η_i) for $i = 1,2$. Thus, the decomposition of the linear representation induces a decompositions of $r_n(z)$ and $R(z,w)$, too. In particular we obtain

$$R(z,w) = \begin{pmatrix} R^{(1)}(z,w) & S(z,w) \\ 0 & R^{(2)}(z,w) \end{pmatrix} \tag{5.4}$$

where the matrix $S(z,w)$ is defined by

$$S(z,w) = R^{(1)}(z,w) \cdot M_{12}(z) \cdot R^{(2)}(z,w)\ . \tag{5.5}$$

In other terms, $R^{(i)}$ gives the contribution of the i-th component, while $S(z,w)$ represents the interconnection between the components.

Now, by the decomposition (5.4) we get

$$\sum_{n=0}^{\infty} r_n(z) w^n = \xi_{\scriptscriptstyle T}\ R(z,w)\ \eta = \xi_{1\scriptscriptstyle T}\ R^{(1)}(z,w)\ \eta_1\ +\ \xi_{1\scriptscriptstyle T}\ S(z,w)\ \eta_2\ +\ \xi_{2\scriptscriptstyle T}\ R^{(2)}(z,w)\ \eta_2$$

and setting $\sum_n s_n(z) w^n = \xi_{1\scriptscriptstyle T} S(z,w) \eta_2$ we obtain

$$r_n(z) = r_n^{(1)}(z) + s_n(z) + r_n^{(2)}(z). \tag{5.6}$$

To estimate the mean and the value of Y_n and to determine its limit distribution, we follow the line developed for the primitive case. First, we need an asymptotic evaluation of $r_n(z)$, $R(z,w)$ and their derivatives. Referring to the terms $r_n^{(i)}(z)$, $i = 1,2$, since M_1 and M_2 are primitive, we can apply all results of Section (4.4) to the series $r^{(1)}$ and $r^{(2)}$. Moreover, we agree to append indices 1 and 2 to the values associated with the linear representations (ξ_1, μ_1, η_1) and (ξ_2, μ_2, η_2), respectively. Thus, for each $i = 1,2$, the Perron-Frobenius eigenvalues λ_i, the eigenvectors u_i, v_i, the functions $y_i(z)$, $f_i(z)$, the matrices C_i, D_i, $F_i(z)$ and the constants α_i, β_i, γ_i, δ_i are well-defined and associated with the linear representation (ξ_i, μ_i, η_i). On the other hand, to obtain asymptotic approximations of $s_n(z)$, we can exploit equations (2.3) and (2.4) to compute the derivatives of $S(z,w)$ with respect to z. So, it is clear that the behaviour of $s_n(z)$ and hence the properties of Y_n depend on the way the components mix together. This combination depends mainly on two conditions.

- (i) Whether there is a communication from the first to the second component (i.e. $M_{12} \neq 0$); if there is no communication, then we get the sum model (see Example 5.1).
- (ii) Whether there exists a *dominant* component (i.e. $\lambda_1 > \lambda_2$ or viceversa) or both components have the same eigenvalues, in which case we say that the components are *equipotent*.

In Section 5.2 we analyze the dominant case, while the equipotent case is studied in Section 5.3. We first assume $M_{12} \neq 0$; anyway both sections contain a final part that focuses on the differences occurring when the matrix M_{12} vanishes.

5.1.1 Sum and product models

Before setting out the analysis, we present two special cases that occur respectively when the formal series r defined by (ξ, μ, η) is the sum or the product of two rational formal series that have primitive linear representation.

Example 5.1 (Sum) Let r be the series defined by

$$(r,\omega) = \xi_{1_T}\mu_1(\omega)\eta_1 + \xi_{2_T}\mu_2(\omega)\eta_2 \qquad \forall\omega \in \{a,b\}^*$$

where (ξ_j, μ_j, η_j) is a primitive linear representation for $j = 1, 2$. Clearly, r admits a bicomponent linear representation (ξ, μ, η) which satisfies (5.1) and such that $M_{12} = 0$. As a consequence, the computation paths of type 2 cannot occur and hence

$$r_n(z) = r_n^{(1)}(z) + r_n^{(2)}(z) \ .$$

Example 5.2 (Product) Consider the formal series

$$(r,\omega) = \sum_{\omega=xy} \pi_{1_T}\ \nu_1(x)\ \tau_1 \cdot \pi_{2_T}\ \nu_2(y)\ \tau_2 \qquad \forall\omega \in \{a,b\}^*$$

where (π_j, ν_j, τ_j) is a primitive linear representation for $j = 1, 2$. Then, r admits a bicomponent linear representation (ξ, μ, η) such that

$$\xi_T = (\pi_{1_T}, 0)\,, \qquad \mu(x) = \begin{pmatrix} \nu_1(x) & \tau_1\ \pi_{2_T}\ \nu_2(x) \\ 0 & \nu_2(x) \end{pmatrix}, \qquad \eta = \begin{pmatrix} \tau_1\ \pi_{2_T}\ \tau_2 \\ \tau_2 \end{pmatrix}. \tag{5.7}$$

In this case, the three terms of $r_n(z)$ can be merged in a unique convolution

$$r_n(z) = \sum_{i=0}^{n} \xi_{1_T}(A_1e^z + B_1)^i\ \tau_1\ \pi_{2_T}\ (A_2e^z + B_2)^{n-i}\eta_2 \ .$$

5.2 Dominant component

In this section we study the behaviour of $\{Y_n\}$ assuming $\lambda_1 > \lambda_2$ (the case $\lambda_1 < \lambda_2$ is symmetric). If the dominant component is not degenerate, then it determines the main terms of expectation and variance of our statistics and we get a Gaussian limit distribution. We also describe the limit distribution in the case of a dominant degenerate component. Apparently, this has a large variety of possible forms depending even on the other (non-main) eigenvalues of the dominated component and including the geometric law in some simple cases.

We first assume $M_{12} \neq 0$ while the case $M_{12} = 0$, corresponding to Example 5.1, is treated in Section 5.2.4.

5.2.1 Analysis of moments

As for the primitive case, to study the first two moments of Y_n we develop a singularity analysis for the functions $R(0,w)$, $R_z(0,w)$ and $R_{zz}(0,w)$, which yields asymptotic expressions for $r_n(0)$, $r_n'(0)$ and $r_n''(0)$. A key role is played by the matrix Q defined by

$$Q = (\lambda_1 I - M_2)^{-1} = \lambda_1^{-1}R^{(2)}(0,\lambda_1^{-1}) \ .$$

Note that Q is well-defined since $\lambda_1 > \lambda_2$. Moreover, we have

$$R_w^{(2)}(0,\lambda_1^{-1}) = \lambda_1^2 \cdot QM_2Q \qquad \text{and} \qquad R_z^{(2)}(0,\lambda_1^{-1}) = \lambda_1 \cdot QA_2Q \ .$$

M_1 and M_2 being primitive, we can apply the results of Section 4.4.1 to $R^{(1)}(0,w)$, $R^{(2)}(0,w)$ and their partial derivatives. Moreover we need asymptotic expression for S and its derivatives. Since $\lambda_1 > \lambda_2$, by using (5.5) and applying (4.20) to $R^{(1)}(0,w)$, as w tends to λ_1^{-1}, we get

$$S(0,w) = \frac{1}{1-\lambda_1 w} \cdot u_1 v_{1_T} M_{12} Q + \mathrm{O}(1) \ .$$

In a similar way one can obtain the Laurent expansions of the matrices $S_z(0,w)$ and $S_{zz}(0,w)$ in a neighbourhood of $w = 1/\lambda_1$. Thus, recalling equation (5.6) and summing up the contributions of all terms, one comes to the following equations:

$$\begin{aligned}
r_n(0) &= r_n^{(1)}(0) + s_n(0) + r_n^{(2)}(0) = \\
&= \lambda_1^n \cdot (\xi_{1_T} u_1) \cdot v_{1_T}(\eta_1 + M_{12} Q \eta_2) \;+\; O(\rho^n)\ , \\
r_n'(0) &= n\lambda_1^n \cdot \beta_1\ (\xi_{1_T} u_1) \cdot v_{1_T}(\eta_1 + M_{12}\ Q\ \eta_2)\ + \\
&+\ \lambda_1^n \cdot (\xi_{1_T} u_1) \cdot v_{1_T}(A_{12} + M_{12}\ QA_2)\ Q\ \eta_2\ +\ \lambda_1^n \cdot \xi_{1_T} D_1(\eta_1 + M_{12}\ Q\ \eta_2)\ + \\
&-\ \lambda_1^n \cdot \beta_1 (\xi_{1_T} u_1) \cdot v_{1_T} M_{12}(I + QM_2) Q\eta_2\ +\ O(\rho^n)\ , \\
r_n''(0) &= n^2 \lambda_1^n \cdot \beta_1^2\ (\xi_{1_T} u_1) \cdot v_{1_T}(\eta_1 + M_{12}\ Q\eta_2)\ + \\
&+\ n\lambda_1^n \cdot 2\beta_1\ [\ (\xi_{1_T} u_1) \cdot v_{1_T}(A_{12} + M_{12}\ QA_2)\ Q\ \eta_2 + \xi_{1_T} D_1 \cdot (\eta_1 + M_{12}\ Q\eta_2)\ ]\ + \\
&-\ n\lambda_1^n \cdot [2\beta_1 (\xi_{1_T} u_1) \cdot v_{1_T} M_{12}(I + QM_2) Q\eta_2]\ + \\
&+\ n\lambda_1^n \cdot \left(\beta_1 - \beta_1^2 + 2\ v_{1_T} \frac{A_1 C_1 A_1}{\lambda_1^2} u_1\right) \cdot (\xi_{1_T} u_1) \cdot v_{1_T}(\eta_1 + M_{12}\ Q\eta_2)\ +\ O(\lambda_1^n)\ ,
\end{aligned}$$

where $|\rho| < \lambda_1$. Thus, the relationship (4.7) between the function $r_n(z)$ and the moments of the r.v. Y_n yields to the following result.

Proposition 5.3 *If $\lambda_1 > \lambda_2$, then the mean value and variance of Y_n satisfy the following relations:*

$$\mathbb{E}(Y_n) = \beta_1 n + O(1)\ , \qquad \mathbb{V}ar(Y_n) = \gamma_1 n + O(1)\ .$$

From this proposition we easily deduce expressions of the mean value for degenerate cases, too. If $B_1 = 0$, then $\beta_1 = 1$, $D_1 = 0$ and, by the significance hypothesis, $B_2 \neq 0$; thus we get

$$\mathbb{E}(Y_n) = n\ -\ E\ +\ O(\varepsilon^n)\ , \qquad \text{where } E = \frac{v_{1_T}(B_{12} + M_{12} Q B_2) Q \eta_2}{v_{1_T}(\eta_1 + M_{12} Q \eta_2)} \text{ and } |\varepsilon| < 1. \tag{5.8}$$

On the contrary, if $A_1 = 0$, then $\beta_1 = 0$, $D_1 = 0$, $A_2 \neq 0$ and we get

$$\mathbb{E}(Y_n)\ =\ E'\ +\ O(\varepsilon^n)\ , \qquad \text{where } E' = \frac{v_{1_T}(A_{12} + M_{12} Q A_2) Q \eta_2}{v_{1_T}(\eta_1 + M_{12} Q \eta_2)} \quad (|\varepsilon| < 1). \tag{5.9}$$

Note that both E and E' are strictly positive since $Q > 0$. Now the problem is to determine conditions that guarantee $\gamma_1 \neq 0$.

5.2.2 Variability conditions

To answer the previous questions we first recall that, by Theorem 4.13, $A_1 \neq 0 \neq B_1$ implies $\gamma_1 \neq 0$. Thus, by Proposition 5.3, we know that if $\lambda_1 > \lambda_2$ and $A_1 \neq 0 \neq B_1$, then $\mathbb{V}ar(Y_n) = \gamma_1 n + \mathrm{O}(1)$ with $\gamma_1 > 0$.

Clearly, if either $A_1 = 0$ or $B_1 = 0$, then $\gamma_1 = 0$ and the question is whether $\mathbb{V}ar(Y_n)$ keeps away from 0. To study the variability condition in this case (the degenerate dominant case), it is convenient to express the variance by means of polynomials and to extend Lemma 4.12 to matrices. We recall that given a non-null polynomial $p(x) = \sum_k p_k x^k$, where $p_k \in \mathbb{R}_+$ for each k, we use $V(p)$ to denote the variance of any r.v. X_p such that $\mathrm{P}\{X_p = k\} = \frac{p_k}{p(1)}$. Analogously, we introduce the following notation: given a matrix $M(x)$ of polynomials in the variable x with non-negative coefficients, we can define its matrix of variances as

$$\overline{V}(M(x)) = [V(M(x)_{ij})]\ .$$

Then, for each finite family of matrices $\{M^{(k)}(x)\}_{k\in I}$ having equal size and non-null polynomial entries, the following relation holds

$$\overline{V}\left(\sum_{k\in I} M^{(k)}(x)\right) \geq \left[\sum_{k\in I} \frac{M^{(k)}(1)_{ij}}{\sum_{s\in I} M^{(s)}(1)_{ij}} V(M^{(k)}(x)_{ij})\right].$$

Moreover, if $M(x)$ and $N(x)$ are matrices of non-null polynomials of suitable sizes, then

$$\overline{V}\left(M(x)\cdot N(x)\right) \geq \left[\sum_{k} \frac{M(1)_{ik}N(1)_{kj}}{M(1)N(1)_{ij}} \left\{V(M(x)_{ik}) + V(N(x)_{kj})\right\}\right]. \tag{5.10}$$

We are able to establish the variability condition in the dominant degenerate case.

Proposition 5.4 *If $M_{12} \neq 0$, $\lambda_1 > \lambda_2$ and either $B_1 = 0$ or $A_1 = 0$, then $\mathbb{V}ar(Y_n) = c + O(\varepsilon^n)$ for some $c > 0$ and $|\varepsilon| < 1$.*

Proof. First observe that the asymptotic expression of the variance given in Proposition 5.3 can be refined as

$$\mathbb{V}ar(Y_n) = \gamma_1 n + c + \mathrm{O}(\varepsilon^n) \tag{5.11}$$

where c is a constant and $|\varepsilon| < 1$. This is due to the fact that, by equation (4.7), the variance depends on the sequences $r_n(0), r_n'(0), r_n''(0)$, which have generating function with a pole of smallest modulus at λ_1^{-1} of degree (at most) 1, 2, 3, respectively: hence their asymptotic expressions are $c_1\lambda_1^n + \mathrm{O}(\rho^n)$, $b_2 n\lambda^n + c_2\lambda_1^n + \mathrm{O}(\rho^n)$, $a_3 n^2\lambda^n + b_3 n\lambda^n + c_3\lambda_1^n + \mathrm{O}(\rho^n)$, respectively, for some constants a_i, b_i, c_i and $|\rho| < 1$; thus, equation (5.11) follows from Proposition 5.3 by replacing the previous expressions in (4.7). In particular the constant c is determined by the values c_i, $i = 1, 2, 3$.

Now, by our hypothesis, since either $B_1 = 0$ or $A_1 = 0$ we have $\gamma_1 = 0$ and we only have to prove $c > 0$. To this end we show that $\mathbb{V}ar(Y_n) \geq \Theta(1)$. Consider the case $B_1 = 0$ and first assume $A_2 \neq 0$. Note that, by the significance hypothesis also $B_2 \neq 0$ holds, and hence $\gamma_2 > 0$. Moreover, we have

$$\mathbb{V}ar(Y_n) = V\left(\xi_{1_T} A_1^n \eta_1 x^n + \xi_{1_T} G_n(x)\eta_2 + \xi_{2_T}(A_2 x + B_2)^n \eta_2\right)$$

where

$$G_n(x) = \sum_{i=0}^{n-1} A_1^i x^i (A_{12}x + B_{12})(A_2 x + B_2)^{n-1-i};$$

hence, by Lemma 4.12 we get

$$\mathbb{V}ar(Y_n) \geq \frac{\xi_{2_T} M_2^n \eta_2}{\xi_T M^n \eta}(\gamma_2 n + O(1)) \ + \ \frac{\xi_{1_T}\sum_{i=0}^{n-1} A_1^i M_{12} M_2^{n-1-i}\eta_2}{\xi_T M^n \eta} V(\xi_{1_T} G_n(x)\eta_2). \tag{5.12}$$

Applying again Lemma 4.12 and equation (5.10), we also obtain

$$V(\xi_{1_T} G_n(x)\eta_2) \geq \min_{(j,k)\in I}\left\{\sum_{i=0}^{n-1} \frac{(A_1^i M_{12} M_2^{n-1-i})_{jk}}{\left(\sum_{s=0}^{n-1} A_1^s M_{12} M_2^{n-1-s}\right)_{jk}} \left(\overline{V}(A_2 x + B_2)^{n-1-i}\right)_{jk}\right\}$$

where $I = \{(j,k) \mid \xi_{1_j} G_n(x)_{jk}\eta_{2_k} \neq 0\}$. Now, note that from Theorem 4.13 one can deduce that, for every primitive matrix $M = A + B$, if $A \neq 0 \neq B$, then $\overline{V}(Ax+B)^n{}_{ij} = \Theta(n)$ for any pair of indices i, j. Thus, replacing the previous value in (5.12), we get

$$\mathbb{V}ar(Y_n) \geq \Theta\left(\frac{\sum_{i=0}^{n-1}\lambda_1^i \lambda_2^{n-i}(n-i)}{\lambda_1^n}\right) = \Theta(1).$$

On the other hand, if $A_2 = 0$ we have

$$\mathrm{P}_n\{Y_n = n\} = \frac{\xi_{1T} M_1^n \eta_1 + \xi_{1T} M_1^{n-1} A_{12}\eta_2}{\xi_T M^n \eta} = \Theta(1) \ .$$

Moreover, equation (5.8) implies $\mathbb{E}(Y_n) = n - E + O(\varepsilon^n)$, where $E > 0$, and hence

$$\mathbb{V}ar(Y_n) = \sum_{k=0}^{n} (E-k)^2 \mathrm{P}_n\{Y_n = n-k\} + O(\varepsilon^n) \geq E^2 \mathrm{P}_n\{Y_n = n\} + O(\varepsilon^n) = \Theta(1)$$

which completes the proof in the case $B_1 = 0$.

Now, let us study the case $A_1 = 0$. If $B_2 \neq 0$, then $\mathbb{V}ar(Y_n^{(2)}) = \Theta(n)$ and the result can be proved as in the case $B_1 = 0$ with $A_2 \neq 0$. If $B_2 = 0$, then by using (5.9) we can argue as in the case $B_1 = 0$ with $A_2 = 0$. □

5.2.3 Limit distribution

Now we study the limit distribution of $\{Y_n\}$ in the case $\lambda_1 > \lambda_2$ still assuming $M_{12} \neq 0$. If the dominant component is not degenerate we obtain a Gaussian limit distribution as in the primitive case. On the contrary, if the dominant component is degenerate we obtain a limit distribution that may assume a large variety of forms, mainly depending on the second component. In both cases the proof is based on the analysis of the characteristic function of Y_n, that is $r_n(it)/r_n(0)$.

Again, recalling that $r_n(z) = r_n^{(1)}(z) + s_n(z) + r_n^{(2)}(z)$, we can apply Corollary 4.8 to $r_n^{(i)}(z)$ for $i = 1, 2$, and we need an analogous result for $s_n(z)$. First consider the generating function of $\{s_n(z)\}$, that is

$$\xi_{1T} S(z,w)\eta_2 = \sum s_n(z) w^n = \xi_{1T} R^{(1)}(z,w) \ (A_{12}e^z + B_{12}) w \ R^{(2)}(z,w)\eta_2 \ . \qquad (5.13)$$

By applying Proposition 4.7 to $R^{(1)}$, since $\lambda_1 > \lambda_2$, for every z near 0, we get

$$\xi_{1T} S(z,w)\eta_2 = \frac{\xi_{1T} F_1(z) \ (A_{12}e^z + B_{12}) \ y_1(z)^{-1} \ R^{(2)}(z, y_1(z)^{-1})\eta_2}{1 - y_1(z) w} + \mathrm{O}(1)$$

as w tends to $y_1(z)^{-1}$. The contribution of both $r_n^{(1)}$ and s_n yields a quasi-power condition for Y_n.

Proposition 5.5 *If $M_{12} \neq 0$ and $\lambda_1 > \lambda_2$, then for every z near 0, as n tends to infinity we have*

$$r_n(z) = f(z) \ y_1(z)^n + O(\rho^n),$$

where $\rho < |y_1(z)|$ and $f(z)$ is a rational function given by

$$f(z) = \xi_{1T} F_1(z) \left\{ \eta_1 + (A_{12}e^z + B_{12}) \ y_1(z)^{-1} \ R^{(2)}(z, y_1(z)^{-1})\eta_2 \right\} \ .$$

Observe that the function $f(z)$ is analytic and non-null at $z = 0$. If $A_1 \neq 0 \neq B_1$, then $\beta_1 > 0$, $\gamma_1 > 0$ and by the previous proposition we can apply the Quasi-power Theorem which yields the following

Theorem 5.6 *If $M_{12} \neq 0$, $\lambda_1 > \lambda_2$ and $A_1 \neq 0 \neq B_1$, then $\frac{Y_n - \beta_1 n}{\sqrt{\gamma_1 n}}$ converges in distribution to a normal r.v. of mean 0 and variance 1.*

On the other hand, if either $A_1 = 0$ or $B_1 = 0$, then $\gamma_1 = 0$ and the Quasi-power Theorem cannot be applied. Thus, we study two cases separately, dealing directly with the characteristic function of $\{Y_n\}$.

Let $B_1 = 0$ and set $Z_n = n - Y_n$. We have $r_n(z) = r_n^{(1)}(z) + s_n(z) + r_n^{(2)}(z)$, where

$$\begin{aligned} r_n^{(1)}(z) &= \xi_{1_T}(M_1 e^z)^n \eta_1 = (\lambda_1 e^z)^n \xi_{1_T}(u_1 v_{1_T} + C_1(n))\eta_1 \ , \\ s_n(z) &= \sum_{j=0}^{n-1} (\lambda_1 e^z)^j \xi_{1_T}(u_1 v_{1_T} + C_1(n))^j (A_{12} e^z + B_{12})(A_2 e^z + B_2)^{n-1-j}\eta_2 \ , \\ r_n^{(2)}(z) &= \xi_{2_T}(A_2 e^z + B_2)^n \eta_2 \ . \end{aligned}$$

Hence the characteristic function of Z_n can be computed by replacing the previous values in $\mathbb{E}(e^{zZ_n}) = e^{zn} r_n(z)/r_n(0)$. A simple computation shows that, as n goes to $+\infty$, for every $t \in \mathbb{R}$ we have

$$\mathbb{E}(e^{itZ_n}) = \frac{v_1^T \eta_1 + v_1^T (A_0 + B_0 e^{it})(\lambda_1 I - A_2 - B_2 e^{it})^{-1}\eta_2}{v_1^T(\eta_1 + M_0 Q \eta_2)} + o(1) \ .$$

Note that by the significance hypothesis (5.3) this function cannot reduce to a constant. The case $A_1 = 0$ can be treated in a similar way. Hence we have proved the following

Theorem 5.7 *Let $M_{12} \neq 0$ and $\lambda_1 > \lambda_2$. If $B_1 = 0$, then $n - Y_n$ converges in distribution to a random variable W of characteristic function*

$$\Phi_W(t) = \frac{v_{1_T}\eta_1 + v_{1_T}(A_{12} + B_{12}e^{it})(\lambda_1 I - A_2 - B_2 e^{it})^{-1}\eta_2}{v_{1_T}(\eta_1 + M_{12} Q \eta_2)} \ .$$

If $A_1 = 0$, then Y_n converges in distribution to a random variable Z of characteristic function

$$\Phi_Z(t) = \frac{v_{1_T}\eta_1 + v_{1_T}(A_{12}e^{it} + B_{12})(\lambda_1 I - A_2 e^{it} - B_2)^{-1}\eta_2}{v_{1_T}(\eta_1 + M_{12} Q \eta_2)} \ . \tag{5.14}$$

Now, let us discuss the form of the r.v.'s W and Z introduced in the previous theorem. The simplest cases occur when the matrices M_1 and M_2 have size 1×1 and hence $M_1 = \lambda_1$, $M_2 = \lambda_2$ and both A_2 and B_2 are constants. In this case $W = R(S + G)$, where R and S are Bernoullian r.v. of parameter p_r and p_s, respectively given by

$$p_r = \frac{M_{12}(\lambda_1 - \lambda_2)^{-1}\eta_2}{\eta_1 + M_{12}(\lambda_1 - \lambda_2)^{-1}\eta_2} \qquad \text{and} \qquad p_s = B_{12}/M_{12} \ ,$$

while G is a geometric r.v. of parameter $B_2/(\lambda_1 - A_2)$. Clearly a similar expression holds for Z.

Moreover, in the product model W and Z further reduce to simple geometric r.v.'s (still in the monodimensional case). More precisely, if (ξ, μ, η) is defined as in Example 5.2 and both M_1 and M_2 have size 1×1, then one can prove that

$$\Phi_Z(t) = \frac{1 - \frac{A_2}{\lambda_1 - B_2}}{1 - \frac{A_2}{\lambda_1 - B_2} e^{it}} \qquad \text{and} \qquad \Phi_W(t) = \frac{1 - \frac{B_2}{\lambda_1 - A_2}}{1 - \frac{B_2}{\lambda_1 - A_2} e^{it}}$$

which are the characteristic functions of geometric r.v.'s of parameter $\frac{A_2}{\lambda_1 - B_2}$ and $\frac{B_2}{\lambda_1 - A_2}$ respectively.

However, the range of possible forms of W and Z is much richer than a simple geometric behaviour. To see this fact consider the function $\Phi_Z(t)$ in (5.14); in the product model it can be expressed in the form

$$\Phi_Z(t) \quad = \quad \frac{\pi_{2_T}(\lambda_1 I - A_2 e^{it} - B_2)^{-1}\tau_2}{\pi_{2_T}(\lambda_1 I - M_2)^{-1}\tau_2} = \sum_{j=0}^{\infty} \frac{\pi_{2_T}(M_2/\lambda_2)^j \tau_2 \cdot (\lambda_2/\lambda_1)^j}{\sum_{i=0}^{\infty} \pi_{2_T}(M_2/\lambda_2)^i \tau_2 \cdot (\lambda_2/\lambda_1)^i} \ \Phi_{Y_j^{(2)}}(t)$$

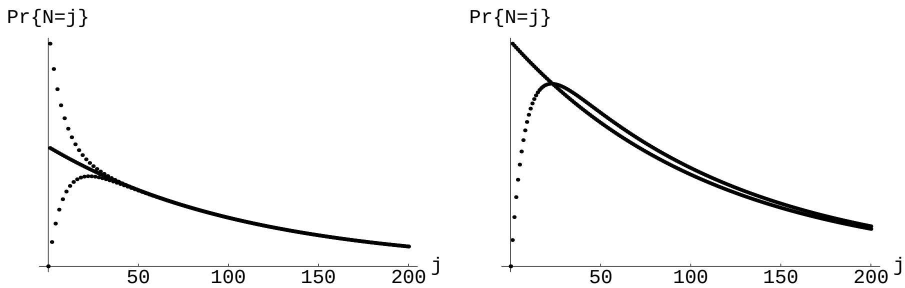


Figure 5.1: Probability law of the r.v. N defined in (5.15), for $j = 0, 1, \ldots, 200$. In the first picture we compare the case $\mu = 0.00001$ and $\mu = -0.89$. In the second one we compare the case $\mu = 0.00001$ and $\mu = +0.89$.

where π_2 and τ_2 are defined as in Example 5.2. This characteristic function actually describes a r.v. $Y_N^{(2)}$, where N is a r.v. with probability law

$$\mathrm{P}_n\{N = j\} = \frac{\pi_{2_T} \left(M_2/\lambda_2\right)^j \tau_2 \cdot (\lambda_2/\lambda_1)^j}{\sum_{i=0}^{\infty} \pi_{2_T} \left(M_2/\lambda_2\right)^i \tau_2 \cdot (\lambda_2/\lambda_1)^i} \,. \tag{5.15}$$

If $B_2 = 0$, then by (5.14) Z reduces to N, and an example of the rich range of its possible forms is shown by considering the case where $(A_1 = 0 = B_2)$ $\lambda_1 = 1.009$, $\lambda_2 = 1$, and the second component is represented by a generic (2×2) - matrix whose eigenvalues are 1 and μ such that $-1 < \mu < 1$. In this case, since the two main eigenvalues have similar values, the behaviour of $\Pr\{N = j\}$ for small j depends on the second component and in particular on its smallest eigenvalue μ. In Figure 5.1 we plot the probability law of N defined in (5.15) for $j = 0, 1, \ldots, 200$ in three cases: $\mu = -0.89$, $\mu = 0.00001$ and $\mu = 0.89$; the first picture compares the curves in the cases $\mu = -0.89$ and $\mu = 0.00001$, while the second picture compares the curves when $\mu = 0.00001$ and $\mu = 0.89$. Note that in the second case, when μ is almost null, we find a distribution similar to a geometric law while, for $\mu = -0.89$ and $\mu = 0.89$, we get a quite different behaviour which approximates the previous one for large values of j.

5.2.4 What changes in the sum model?

Under the hypothesis $\lambda_1 > \lambda_2$ and $A_1 \neq 0 \neq B_1$, the condition $M_{12} \neq 0$ does not play a fundamental role. Indeed, the proofs of Sections 5.2.1 and 5.2.3 can be perfectly adapted to the case $M_{12} = 0$. Thus the following theorem holds.

Theorem 5.8 *In the sum model, if $\lambda_1 > \lambda_2$ and $A_1 \neq 0 \neq B_1$, then*

$$\mathbb{E}(Y_n) = \beta_1 n + \frac{\delta_1}{\alpha_1} + O(\varepsilon^n) \,, \qquad \mathbb{V}ar(Y_n) = \gamma_1 n + O(1) \qquad (|\varepsilon| < 1),$$

where α_1, β_1, γ_1, δ_1 are the constants associated with the first component defined as in Theorem 4.9. Moreover, $\frac{Y_n - \beta_1 n}{\sqrt{\gamma_1 n}}$ converges in distribution to a normal r.v. of mean 0 and variance 1.

On the contrary, if the dominant component is degenerate, then the sum model differs significantly from the general model. If $B_1 = 0$, then $\beta_1 = 1$ and $\gamma_1 = \delta_1 = 0$, hence we get $\mathbb{E}(Y_n) = n + O(\varepsilon^n)$. On the other hand, if $A_1 = 0$, then $\beta_1 = \gamma_1 = \delta_1 = 0$ and hence we get $\mathbb{E}(Y_n) = O(\varepsilon^n)$. In both cases we have $\gamma_1 = 0$ and a direct computation proves $\mathbb{V}ar(Y_n) = O(\varepsilon^n)$, showing that Y_n almost surely reduces to a single value (n or 0, respectively). In fact, by Chebyshev's inequality, if $B_1 = 0$ we have for every $c > 0$

$$\mathrm{P}_n\{|Y_n - n| > c\} \leq \frac{\mathbb{V}ar(Y_n)}{c^2} = \mathrm{O}(\varepsilon^n) \ .$$

A similar result can be obtained in the case $A_1 = 0$.

Theorem 5.9 *In the sum model, assume* $\lambda_1 > \lambda_2$*. If* $B_1 = 0$*, then* $\lim_{n\to\infty} P_n\{n - Y_n > c\} = 1$*; if* $A_1 = 0$*, then* $\lim_{n\to\infty} P_n\{Y_n > c\} = 1$.

5.3 Equipotent components

Now, we study the behaviour of Y_n in the case $\lambda_1 = \lambda_2$. Then two main subcases arise. The first one occurs when the constants β_1 and β_2 are different. Then the variance is of a quadratic order showing there is not a concentration phenomenon around the average value of our statistics. In this case, except for the sum model, we get a uniform limit distribution between the mean constants associated with the two components.

However, if the mean constants are equal, then the variance reduces to a linear order of growth and we have again a concentration phenomenon. In this case the limit distribution depends on the main terms of the variances associated with the two components: if they are equal we obtain a Gaussian limit distribution again; if they are different we obtain a limit distribution defined by a mixture of Gaussian random variables.

We first assume $M_{12} \neq 0$, while the sum model is considered in Section 5.3.3. In this case, the contributions of both components are isolated and this yields to different results with respect to the case $M_{12} \neq 0$.

As before we begin studying the asymptotic behaviour of the moments of Y_n and then we determine the limit distributions.

5.3.1 Analysis of moments

We argue as in Section 5.2.1; for this reason we avoid many details and give simple outline of the proofs. For the sake of simplicity let $\lambda = \lambda_1 = \lambda_2$.

Proposition 5.10 *Assume* $\lambda_1 = \lambda_2 = \lambda$ *and let* $M_{12} \neq 0$*. Then the following statements hold:*

1. *If* $\beta_1 \neq \beta_2$*, then* $\mathbb{E}(Y_n) = \frac{\beta_1+\beta_2}{2}\, n + O(1)$ *and* $\mathbb{V}ar(Y_n) = \frac{(\beta_1-\beta_2)^2}{12}\, n^2 + O(n)$*;*
2. *If* $\beta_1 = \beta_2 = \beta$*, then* $\mathbb{E}(Y_n) = \beta\, n + O(1)$ *and* $\mathbb{V}ar(Y_n) = \frac{\gamma_1+\gamma_2}{2}\, n + O(1)$*, where* $\gamma_i > 0$ *for each* $i \in \{1, 2\}$.

Proof. First consider the case $\beta_1 \neq \beta_2$. In order to evaluate $s_n(0)$, $s'_n(0)$ and $s''_n(0)$, one can proceed as in the dominant case, considering the bivariate function $S(z, w)$ defined in (5.5) and then applying the results of Section 4.4.1 to $R^{(1)}(0, w)$ and $R^{(2)}(0, w)$. It turns out that, in a neighbourhood of $w = 1/\lambda$, both $R^{(1)}(0, w)$ and $R^{(2)}(0, w)$ admit Laurent expansions of degree 1, while the matrices $S(0, w)$, $S_z(0, w)$ and $S_{zz}(0, w)$ admit a Laurent expansion of degree 2, 3 and 4, respectively. Thus, in equation (5.6) the main contribution to the behaviour of $r_n(0)$ is given by the term $s_n(0)$ and this is true for derivatives, too. This leads to the following expressions, holding as n tends to infinity:

$$
\begin{aligned}
r_n(0) &= n\lambda^n \cdot \xi_{1_T} u_1 \, v_{1_T} \, \frac{M_{12}}{\lambda} \, u_2 \, v_{2_T} \eta_2 + O(\lambda^n) \ , \\
r_n'(0) &= n^2 \lambda^n \cdot \frac{\beta_1 + \beta_2}{2} \cdot \xi_{1_T} u_1 \, v_{1_T} \, \frac{M_{12}}{\lambda} \, u_2 \, v_{2_T} \eta_2 + O(n\lambda^n) \ , \\
r_n''(0) &= n^3 \lambda^n \cdot \frac{\beta_1^2 + \beta_1\beta_2 + \beta_2^2}{3} \cdot \xi_{1_T} u_1 \, v_{1_T} \, \frac{M_{12}}{\lambda} \, u_2 \, v_{2_T} \eta_2 + O(n^2\lambda^n) \ .
\end{aligned}
$$

Point 1. now follows from relation (4.7).

If $\beta_1 = \beta_2 = \beta$, the previous evaluations yield $\mathbb{E}(Y_n) = \beta n + O(1)$ but $\mathbb{V}ar(Y_n) = O(n)$. Then, terms of lower order are now necessary to evaluate the variance. These can be obtained as above by a singularity analysis of $S(0,w)$, $S_z(0,w)$ and $S_{zz}(0,w)$ and observing that $\beta C_1 = \beta C_2 = 0$. The overall computation leads to the following relations:

$$
\begin{aligned}
\mathbb{E}(Y_n) &= n \cdot \beta \ + \left\{ \frac{v_{1_T} M_{12} D_2 \eta_2}{v_{1_T} M_{12} u_2 v_{2_T} \eta_2} + \frac{\xi_{1_T} D_1 M_{12} u_2}{\xi_{1_T} u_1 v_{1_T} M_{12} u_2} + \frac{v_{1_T} A_{12} u_2}{v_{1_T} M_{12} u_2} - \beta \right\} \ + \ O(\varepsilon^n) \\
\mathbb{V}ar(Y_n) &= n \cdot \left(\beta - \beta^2 + v_{2_T} \, \frac{A_2 C_2 A_2}{\lambda^2} \, u_2 + v_{1_T} \, \frac{A_1 C_1 A_1}{\lambda^2} \, u_1 \right) \ + \ O(1) \ = \ \frac{\gamma_1 + \gamma_2}{2} \, n + O(1) \ .
\end{aligned}
$$

for $|\varepsilon| < 1$. Finally observe that, since $\beta_1 = \beta_2$ the significance condition (5.3) implies $A_i \neq 0 \neq B_i$ for each $i = 1, 2$ and hence also $\gamma_i \neq 0$. □

Referring to the variability condition, observe that the previous proposition states that if $\beta_1 \neq \beta_2$, then the variance is of order $\Theta(n^2)$, while if $\beta_1 = \beta_2$ the variance is of order $\Theta(n)$ and moreover it is intuitively an average value of the variances associated with the two components.

5.3.2 Limit distribution

To study the limit distribution in the equipotent case ($\lambda_1 = \lambda_2 = \lambda$) with the assumption $M_{12} \neq 0$, we consider again the characteristic function of Y_n, that is $r_n(it)/r_n(0)$. In this case, we do not obtain a quasi-power condition, since the contribution of $s_n(z)$ to the behaviour of $r_n(z)$ has a different form.

Recalling that $r_n(z) = r_n^{(1)}(z) + s_n(z) + r_n^{(2)}(z)$, we can apply Corollary 4.8 to $r_n^{(i)}(z)$ for $i = 1, 2$. As far as the sequence $\{s_n(z)\}$ is concerned, consider its generating function $\xi_{1_T} S(z,w)\eta_2$ defined in (5.13). Then, let us apply Proposition (4.7) to $R^{(1)}$ and $R^{(2)}$ and define the analytic function

$$
f(z) = \xi_{1_T} F_1(z)(A_{12} e^z + B_{12}) F_2(z) \eta_2 \ . \tag{5.16}
$$

We have $f(0) \neq 0$ and, since $\lambda_1 = \lambda_2 = \lambda$, for every z near 0 we get

$$
\begin{aligned}
\xi_{1_T} S(z,w)\eta_2 &= \frac{f(z)w}{(1 - y_1(z)w)\,(1 - y_2(z)w)} + \mathrm{O}\left(\frac{1}{1 - y_1(z)w}\right) + \mathrm{O}\left(\frac{1}{1 - y_2(z)w}\right) + \mathrm{O}(1) \qquad (5.17) \\
&= f(z) \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} y_1(z)^k y_2(z)^{n-1-k} w^n + \mathrm{O}\left(\frac{1}{1 - y_1(z)w}\right) + \mathrm{O}\left(\frac{1}{1 - y_2(z)w}\right) + \mathrm{O}(1)
\end{aligned}
$$

as w tends to λ^{-1}. Thus, at $z = 0$, since $y_1(0) = y_2(0) = \lambda$ we get

$$
r_n(0) = f(0) \cdot n\lambda^{n-1} + \mathrm{O}(\lambda^n). \tag{5.18}
$$

However, for $z \neq 0$, the asymptotic behaviours of $s_n(z)$ depends on the condition $\beta_1 \neq \beta_2$.

Proposition 5.11 *If $M_{12} \neq 0$, $\lambda_1 = \lambda_2 = \lambda$ and $\beta_1 \neq \beta_2$, then for every z near 0, different from 0, we have*

$$r_n(z) = f(z) \cdot \frac{y_1(z)^n - y_2(z)^n}{\lambda(\beta_1 - \beta_2)z + O(z^2)} + O(y_1(z)^n) + O(y_2(z)^n) + O(\rho^n),$$

where $0 \leq \rho < \lambda$.

Proof. Since $\beta_1 \neq \beta_2$, from (5.17) we get, for any z near 0 different from 0

$$s_n(z) = f(z) \cdot \frac{y_1(z)^n - y_2(z)^n}{y_1(z) - y_2(z)} + \mathrm{O}\left(y_1(z)^n\right) + \mathrm{O}\left(y_2(z)^n\right) + \mathrm{O}(\rho^n) \ . \tag{5.19}$$

Also observe that, by Corollary (4.10), for any $i = 1, 2$ and every z near 0 we can write

$$y_i(z) = \lambda + \lambda\beta_i z + \mathrm{O}(z^2). \tag{5.20}$$

Hence, the result follows by replacing the previous relations into (5.19) and observing that the contribution of $r_n^{(1)}(z)$ and $r_n^{(2)}(z)$ is of the order $O(y_1(z)^n)$ and $O(y_2(z)^n)$, respectively. □

In the following theorem we determine the limit distribution of Y_n/n.

Theorem 5.12 *If $M_{12} \neq 0$, $\lambda_1 = \lambda_2 = \lambda$ and $\beta_1 \neq \beta_2$, then Y_n/n converges in distribution to a r.v. uniformly distributed over the interval $[b_1, b_2]$, where $b_1 = \min\{\beta_1, \beta_2\}$ and $b_2 = \max\{\beta_1, \beta_2\}$.*

Proof. By Proposition 5.11 and equation (5.20), for every $t \in \mathbb{R}, t \neq 0$, as n tends to infinity, we have

$$r_n\left(\frac{it}{n}\right) = f(0) \cdot n\lambda^{n-1} \frac{\left(1 + \frac{it\beta_1}{n} + \mathrm{O}\left(\frac{1}{n^2}\right)\right)^n - \left(1 + \frac{it\beta_2}{n} + \mathrm{O}\left(\frac{1}{n^2}\right)\right)^n}{it(\beta_1 - \beta_2) + \mathrm{O}\left(\frac{1}{n}\right)} + \mathrm{O}(\lambda^n).$$

Thus, the last equation yields the following expression for the characteristic function of Y_n/n:

$$\mathbb{E}(itY_n/n) = \frac{r_n(it/n)}{r_n(0)} = \frac{e^{it\beta_1} - e^{it\beta_2}}{it(\beta_1 - \beta_2)} + \mathrm{O}\left(\frac{1}{n}\right).$$

The theorem is proved observing that the main term of the right hand side is the characteristic function of a uniform distribution in the required interval. □

Now, let us consider the case $\beta_1 = \beta_2 = \beta$. Then point 2. of Proposition 5.10 holds and hence there is a concentration phenomenon around the mean value of Y_n. The limit distribution can be deduced from equation (5.19), which still holds in our case but assumes different forms according whether $\gamma_1 \neq \gamma_2$ or not. In the following, let γ be defined by $\gamma = \frac{\gamma_1 + \gamma_2}{2}$.

Theorem 5.13 *If $M_{12} \neq 0$, $\lambda_1 = \lambda_2$, $\beta_1 = \beta_2$ and $\gamma_1 \neq \gamma_2$, then $\frac{Y_n - \beta n}{\sqrt{\gamma n}}$ converges in distribution to a r.v. T of characteristic function*

$$\Phi_T(t) = \frac{e^{-\frac{\gamma_2}{2\gamma}t^2} - e^{-\frac{\gamma_1}{2\gamma}t^2}}{(\frac{\gamma_1}{2\gamma} - \frac{\gamma_2}{2\gamma})t^2} \ . \tag{5.21}$$

Proof. First observe that in our case, for $i = 1, 2$,

$$y_i(z) = \lambda\left(1 + \beta z + \frac{\gamma_i + \beta^2}{2}z^2 + \mathrm{O}(z^3)\right).$$

Hence, by replacing these values into (5.19), we get for each $t \in \mathbb{R}$ different from 0

$$r_n\left(\frac{it}{\sqrt{\gamma n}}\right) = f(0) \cdot n\lambda^{n-1} \cdot e^{i\beta t\sqrt{n/\gamma}} \cdot \frac{e^{-\frac{\gamma_2}{2\gamma}t^2} - e^{-\frac{\gamma_1}{2\gamma}t^2}}{(\frac{\gamma_1}{2\gamma} - \frac{\gamma_2}{2\gamma})t^2}\left(1 + O(n^{-1/2})\right)$$

where $f(z)$ is defined as in (5.16). The required result follows from the previous equation and from relation (5.18), recalling that $e^{-it\beta\sqrt{\frac{n}{\gamma}}} \cdot r_n\left(\frac{it}{\sqrt{\gamma n}}\right)/r_n(0)$ is the characteristic function of $\frac{Y_n - \beta n}{\sqrt{\gamma n}}$.
□

By direct inspection, one can see that the probability density corresponding to the characteristic function (5.21) is a mixture of Gaussian densities of mean 0 and variances uniformly distributed in the interval $[c_1, c_2]$, where $c_1 = \min\{\gamma_i, \gamma_j\}$ and $c_2 = \max\{\gamma_i, \gamma_j\}$ In other words, $\frac{Y_n - \beta n}{\sqrt{n}}$ converges in law to a random variable with density function

$$\Phi_T(t) = \frac{1}{c_2 - c_1}\int_{c_1}^{c_2} \frac{e^{-x^2/(2v)}}{\sqrt{2\pi v}}dv \ . \tag{5.22}$$

Finally we deal with the case where also the main terms of the variances are equal.

Theorem 5.14 *If $M_{12} \neq 0$, $\lambda_1 = \lambda_2$, $\beta_1 = \beta_2$ and $\gamma_1 = \gamma_2$, then $\frac{Y_n - \beta n}{\sqrt{\gamma n}}$ converges in distribution to a normal r.v. of mean 0 and variance 1.*

Proof. In this case, for $z = \Theta(n^{-1/2})$, the convolution in (5.17) satisfies the equation

$$\sum_{j=0}^{n-1} y_1(z)^j y_2(z)^{n-1-j} = n\lambda^{n-1}\left(1 + \beta z + \frac{\gamma + \beta^2}{2}z^2\right)^{n-1}(1 + \mathrm{O}(z^3))^{n-1}.$$

Replacing this value in the same equation, we get

$$r_n\left(\frac{it}{\sqrt{\gamma n}}\right) = f(0) \cdot n\lambda^{n-1}\exp\left\{i\beta t\sqrt{n/\gamma} - \frac{t^2}{2}\right\}\left(1 + O(n^{-1/2})\right).$$

Hence, reasoning as in the previous proof one can see that the characteristic function of $\frac{Y_n - \beta n}{\sqrt{\gamma n}}$ converges to $e^{-t^2/2}$. □

We conclude with some examples which illustrate the result obtained in the equipotent case when $\beta_1 = \beta_2$.

Example 5.15 In Figure 5.2 we illustrate the form of the limit distributions obtained in Theorems 5.13 and 5.14. We represent the density of the r.v. having characteristic function (5.21), for different values of the ratio $p = \gamma_2/\gamma_1$. When p approaches 1, the curve tends to a Gaussian density according to Theorem 5.14; if γ_2 is much greater than γ_1, then we find a density with a cuspid in the origin corresponding to Theorem 5.13. □

Example 5.16 One may also ask whether the hypotheses of Theorem 5.13 are satisfied for some pairs of primitive linear representations. As an example of such a pair, consider the triple (ξ_1, μ_1, η_1) where

$$\xi_1 = \eta_1 = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}, \qquad A_1 = \mu_1(a) = \begin{pmatrix} 3/20 & 1 \\ 1/16 & 9/40 \end{pmatrix}, \qquad B_1 = \mu_1(b) = \begin{pmatrix} 3/5 & 0 \\ 0 & 21/40 \end{pmatrix}$$

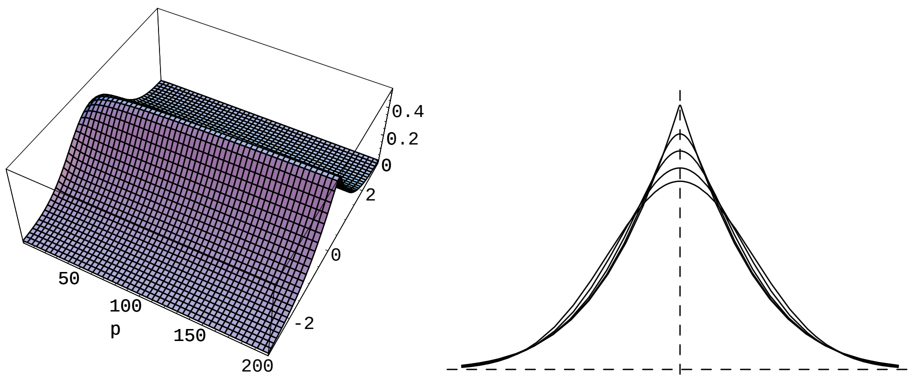


Figure 5.2: The first picture represents the density of the r.v. having characteristic function (5.21), according to the parameter $p = \gamma_2/\gamma_1$. The second picture represents some sections obtained for $p = 1.0001, 5, 15, 50, 20000$.

and the triple (ξ_2, μ_2, η_2) such that $\xi_2 = \xi_1 = \eta_2$,

$$A_2 = \mu_2(a) = \begin{pmatrix} 3/40 & 1 \\ 1/16 & 3/10 \end{pmatrix}, \quad \text{and} \quad B_2 = \mu_2(b) = \begin{pmatrix} 27/40 & 0 \\ 0 & 9/20 \end{pmatrix}.$$

In this case $M_1 = M_2$ and hence $\lambda_1 = \lambda_2$; moreover, by direct computation one can show that $\beta_1 = \beta_2 = 7/16$, while $\gamma_1 = \frac{1611}{6400}$ and $\gamma_2 = \frac{1899}{6400}$. Thus the hypotheses of the theorem are satisfied for any possible non-negative value of $M_{12} \neq 0$. □

5.3.3 What changes in the sum model?

When $\lambda_1 = \lambda_2$, the separation of the components expressed by the condition $M_{12} = 0$ plays a relevant role. Indeed, when $M_{12} \neq 0$, the main contribution to the behaviour of $r_n(0)$ was given by $s_n(0)$, which here vanishes since

$$r_n(z) = r_n^{(1)}(z) + r_n^{(2)}(z).$$

Proposition 5.17 *In the sum model, assume $\lambda_1 = \lambda_2$. If $\beta_1 \neq \beta_2$, then*

$$\mathbb{E}(Y_n) = n \cdot \frac{\alpha_1\beta_1 + \alpha_2\beta_2}{\alpha_1 + \alpha_2} + O(1), \quad \mathbb{V}ar(Y_n) = n^2 \cdot \frac{\alpha_1\alpha_2\,(\beta_1 - \beta_2)^2}{(\alpha_1 + \alpha_2)^2} + O(n).$$

If $\beta_1 = \beta_2 = \beta$, then

$$\mathbb{E}(Y_n) = n \cdot \beta + O(1), \quad \mathbb{V}ar(Y_n) = n \cdot \frac{\alpha_1\gamma_1 + \alpha_2\gamma_2}{\alpha_1 + \alpha_2} + O(1).$$

Proof. First note that $r_n^{(1)}(0)$, $r_n^{(2)}(0)$ and their derivatives satisfy the properties stated in Section 4.4.1. Hence, the result follows from equation (4.7), by a simple computation. □

Now, let us study the limit distribution. Let U_n be the Bernoullian random variable $U_n : \Omega_n \to \{0, 1\}$ such that for each $\ell \in \Omega_n$

$$U_n(\ell) = \begin{cases} 1 & \text{if } \ell \text{ is entirely contained in the first component,} \\ 0 & \text{if } \ell \text{ is entirely contained in the second component.} \end{cases}$$

It is easy to show that

$$\mathrm{P}_n\{U_n = x\} = \begin{cases} \frac{\xi_{1T} M_1^n \eta_1}{\xi_T M^n \eta} & \text{if } x = 1 \; , \\ \\ \frac{\xi_{2T} M_2^n \eta_2}{\xi_T M^n \eta} & \text{if } x = 0 \; . \end{cases}$$

Furthermore, let $L_n = \beta_1 U_n + \beta_2(1 - U_n)$ and observe that if $\beta_1 = \beta_2$, then $L_n = \beta_1 = \beta_2$. Also notice that L_n converges in distribution to a random variable $\beta_1 U + \beta_2(1-U)$, where U is a Bernoullian r.v. of parameter $p = \alpha_1/(\alpha_1 + \alpha_2)$.

Proposition 5.18 *In the sum model, if $\lambda_1 = \lambda_2$, then the distribution of Y_n/n converges to the distribution having probability mass $\frac{\alpha_1}{\alpha_1+\alpha_2}$ at β_1 and probability mass $\frac{\alpha_2}{\alpha_1+\alpha_2}$ at β_2.*

Proof. We first evaluate the variance of $Y_n - nL_n$. Clearly Y_n and L_n are not independent, but we can express their dependence by writing $Y_n = U_n Y_n^{(1)} + (1-U_n)Y_n^{(2)}$ and hence

$$Y_n - nL_n \;=\; U_n \cdot (Y_n^{(1)} - n\beta_1) \;+\; (1-U_n)\cdot(Y_n^{(2)} - n\beta_2) \; .$$

Moreover, by the previous proposition $\mathbb{E}(Y_n - nL_n) = \mathrm{O}(1)$ and so

$$\begin{aligned} \mathbb{V}ar(Y_n - nL_n) &= \mathbb{E}((Y_n - nL_n)^2) + \mathrm{O}(1) = \sum_{i=0,1} \mathbb{E}((Y_n - nL_n)^2 \mid U_n = i)\cdot \mathrm{P}_n\{U_n = i\} + \mathrm{O}(1) \\ &= \sum_{j=1,2} \mathbb{E}((Y_n^{(j)} - n\beta_j)^2)\cdot \frac{\alpha_j}{\alpha_1+\alpha_2} + \mathrm{O}(1) = n\cdot\frac{\alpha_1\gamma_1 + \alpha_2\gamma_2}{\alpha_1+\alpha_2} + \mathrm{O}(1) \; . \end{aligned}$$

The result is a consequence of Chebyshev's inequality: for every $c > 0$ one gets

$$\mathrm{P}_n\left\{\left|\frac{Y_n}{n} - L_n\right| \geq c\right\} = \mathrm{O}\left(\frac{1}{n}\right)$$

□

The above proposition intuitively states that Y_n asymptotically behaves as nL_n, where L_n may only assume two values. Thus, a natural question concerns the limit distribution of $Y_n - nL_n$. To deal with this problem assume $\gamma_1 \neq 0 \neq \gamma_2$ and consider the r.v. Υ constructed by considering a Bernoullian r.v. U of parameter $p = \alpha_1/(\alpha_1+\alpha_2)$, two normal r.v.'s N_1, N_2 of mean 0 and variance γ_1 and γ_2, respectively, and setting

$$\Upsilon = U\cdot N_1 + (1-U)\cdot N_2 \qquad (5.23)$$

where we assume U, N_1, N_2 independent of one another. Note that, if $\gamma_1 = \gamma_2$, then Υ is a normal r.v. of mean 0 and variance γ_1. The characteristic function of Υ is given by

$$\mathbb{E}(e^{it\Upsilon}) = \frac{\alpha_1}{\alpha_1+\alpha_2}e^{-\frac{\gamma_1}{2}t^2} \;+\; \frac{\alpha_2}{\alpha_1+\alpha_2}e^{-\frac{\gamma_2}{2}t^2} \; .$$

Proposition 5.19 *In the sum model, if $\lambda_1 = \lambda_2$ and $\gamma_1 \neq 0 \neq \gamma_2$, then the distribution of $\frac{Y_n - nL_n}{\sqrt{n}}$ converges to the mixture, with weights $\frac{\alpha_1}{\alpha_1+\alpha_2}$ and $\frac{\alpha_2}{\alpha_1+\alpha_2}$, of two normal distributions with mean zero and variance γ_1 and γ_2 respectively. In particular, if $\gamma_1 = \gamma_2 = \gamma$, then $\frac{Y_n - nL_n}{\sqrt{n\gamma}}$ converges in law to a standard normal random variable.*

Proof. Let us define the r.v. $\Upsilon_n = \frac{Y_n - nL_n}{\sqrt{n}}$. Its characteristic function is given by

$$\begin{aligned} \mathbb{E}(e^{it\Upsilon_n}) &= \sum_{i=0,1}\mathbb{E}(e^{it\Upsilon_n} \mid U_n = i)\cdot \mathrm{P}_n\{U_n = i\} = \sum_{j=1,2}\mathbb{E}\left(e^{it\frac{Y_n^{(j)} - n\beta_j}{\sqrt{n}}}\right)\cdot\left(\frac{\alpha_j}{\alpha_1+\alpha_2} \;+\; \mathrm{O}(\varepsilon^n)\right) \\ &= \frac{\alpha_1}{\alpha_1+\alpha_2}e^{-\frac{\gamma_1}{2}t^2} \;+\; \frac{\alpha_2}{\alpha_1+\alpha_2}e^{-\frac{\gamma_2}{2}t^2} \;+\; \mathrm{O}\left(n^{-1/2}\right) \; . \end{aligned}$$

□

The previous results hold even if $\beta_1 = \beta_2 = \beta$; notice that in that case L_n reduces to the constant β and $\gamma_1 \neq 0 \neq \gamma_2$ otherwise either $A = 0$ or $B = 0$. Hence we obtain the following

Corollary 5.20 *In the sum model, assume $\lambda_1 = \lambda_2$ and $\beta_1 = \beta_2 = \beta$. Then the distribution of $\frac{Y_n - n\beta}{\sqrt{n}}$ converges to the mixture, with weights $\frac{\alpha_1}{\alpha_1+\alpha_2}$ and $\frac{\alpha_2}{\alpha_1+\alpha_2}$, of two normal distributions with mean zero and variance γ_1 and γ_2 respectively. In particular, if $\gamma_1 = \gamma_2 = \gamma$, then $\frac{Y_n - n\beta}{\sqrt{n\gamma}}$ converges in law to a standard normal random variable.*

5.4 Summary

The results presented in this chapter are summarized in Table 5.1. To explain them intuitively, recall that in our model each component is primitive and hence, considered separately, it yields a Gaussian limit distribution. Thus, the behaviour of the overall model derives from the relationship between these two components. Depending on how their separate contributions mix together, we get quite different limit distributions. This combination depends on two main conditions: (i) whether there is a communication from the first to the second component (i.e. $M_{12} \neq 0$) and (ii) whether there exists a dominant component (i.e. $\lambda_1 > \lambda_2$ or viceversa). The analysis of the dominant case splits in two further directions according whether the dominant component is degenerate or not. The equipotent case (occurring when $\lambda_1 = \lambda_2$) has several subcases corresponding to the possible differences between the leading terms of the mean values and of the variances associated with each component.

We obtain Gaussian limit distributions only when the dominant component does not degenerate and hence we can neglect the other component, or when the two components essentially have the same asymptotic behaviour (i.e. in the equipotent case with equal leading terms of mean values and variances).

Notice that the existence of a connection between the two components is less relevant when one is dominant. Therefore, condition (i) concerning the matrix M_{12} is meaningful mainly in the equipotent case. Here, if $M_{12} \neq 0$ the main contribution to the bivariate generating function is given by $S(z, w)$, which represents the connection from the first to the second component and is essentially given by the convolution of the two contributions. On the contrary, when $M_{12} = 0$ the function $S(z, w)$ vanishes and the two components contribute separately to the overall behaviour of the system.

As a consequence, when the leading terms of the mean values are different, we get a uniform limit distribution in the case $M_{12} \neq 0$, while, if $M_{12} = 0$, we obtain a limit distribution concentrated in two values that correspond to the separate components. Analogously, when the main terms of the average values are equal but the leading terms of the variances are different, we get a mixture of Gaussian distributions having the same mean value: if $M_{12} \neq 0$ such distributions have variance uniformly distributed over a given interval; on the contrary, if $M_{12} = 0$ they reduce to two Gaussian distributions, with variances corresponding to the separate components.

We observe that the dominance condition (ii) plays a key role to determine the limit distribution in two main cases of the previous classification: the dominant non-degenerate case and the equipotent case with different leading terms of the mean values. To illustrate its role we present the following example.

Example 5.21 Consider the product model of Example 5.2 and define the "factor" components (π_i, ν_i, τ_i), $i = 1, 2$, by means of the weighted finite automata described in Figure 5.3. The matrices $A_i = \nu_i(a)$ and $B_i = \nu_i(b)$ are defined by the labels associated with transitions in the pictures. The values of the components of the arrays π_i and τ_i are included in the corresponding states. Multiplying the matrices $A_i = \nu_i(a)$ and $B_i = \nu_i(b)$ (for $i = 1, 2$) by suitable factors, it is possible

Conditions			*Results*		
	Dominance	Degeneracy	Mean value	Variance	Limit distribution
$M_{12} \neq 0$	$\lambda_1 > \lambda_2$	$A_1 \neq 0 \neq B_1$	$\beta_1 n + O(1)$ $0 < \beta_1 < 1$	$\gamma_1 n + O(1)$ $0 < \gamma_1$	$\frac{Y_n - \beta_1 n}{\sqrt{\gamma_1 n}} \longrightarrow_d N_{0,1}$
		$B_1 = 0$	$n - E + O(\varepsilon^n)$ $E > 0$	$c + \mathrm{O}(\varepsilon^n)$ $c > 0$	$n - Y_n \longrightarrow_d W$ Theorem 5.7
		$A_1 = 0$	$E' + O(\varepsilon^n)$ $E' > 0$	$c' + \mathrm{O}(\varepsilon^n)$ $c' > 0$	$Y_n \longrightarrow_d Z$ Theorem 5.7
	$\lambda_1 = \lambda_2$	$\beta_1 \neq \beta_2$	$\frac{\beta_1+\beta_2}{2} n + O(1)$	$\frac{(\beta_1-\beta_2)^2}{12} n^2 + O(n)$	$\frac{Y_n}{n} \longrightarrow_d \mathrm{Unif}(b_1, b_2)$ Theorem 5.12
		$\beta_1 = \beta_2 = \beta$ $\gamma_1 \neq \gamma_2$	$\beta n + O(1)$	$\gamma n + O(1)$ $\gamma = \frac{\gamma_1+\gamma_2}{2}$	$\frac{Y_n - \beta n}{\sqrt{\gamma n}} \longrightarrow_d T$ Theorem 5.13
		$\beta_1 = \beta_2 = \beta$ $\gamma_1 = \gamma_2 = \gamma$	$\beta n + O(1)$	$\gamma n + O(1)$	$\frac{Y_n - \beta n}{\sqrt{\gamma n}} \longrightarrow_d N_{0,1}$
$M_{12} = 0$	$\lambda_1 > \lambda_2$	$A_1 \neq 0 \neq B_1$	$\beta_1 n + O(1)$ $0 < \beta_1 < 1$	$\gamma_1 n + O(1)$ $0 < \gamma_1$	$\frac{Y_n - \beta_1 n}{\sqrt{\gamma_1 n}} \longrightarrow_d N_{0,1}$
		$B_1 = 0$	$n + O(\varepsilon^n)$	$O(\varepsilon^n)$	$n - Y_n \longrightarrow_p 0$ Theorem 5.9
		$A_1 = 0$	$O(\varepsilon^n)$	$O(\varepsilon^n)$	$Y_n \longrightarrow_p 0$ Theorem 5.9
	$\lambda_1 = \lambda_2$	$\beta_1 \neq \beta_2$	$c_1 n + O(1)$ $c_1 = \frac{\alpha_1\beta_1+\alpha_2\beta_2}{\alpha_1+\alpha_2}$	$c_2 n^2 + O(n)$ $c_2 = \frac{\alpha_1\alpha_2(\beta_1-\beta_2)^2}{(\alpha_1+\alpha_2)^2}$	$\frac{Y_n}{n} \longrightarrow_d$ $\beta_1 U + \beta_2(1-U)$ Proposition 5.18
		$\beta_1 = \beta_2 = \beta$ $\gamma_1 \neq \gamma_2$	$\beta n + O(1)$	$c_3 n + O(1)$ $c_3 = \frac{\alpha_1\gamma_1+\alpha_2\gamma_2}{\alpha_1+\alpha_2}$	$\frac{Y_n - \beta n}{\sqrt{n}} \longrightarrow_d$ $U N_{0,\gamma_1} + (1-U) N_{0,\gamma_2}$ Corollary 5.20
		$\beta_1 = \beta_2 = \beta$ $\gamma_1 = \gamma_2 = \gamma$	$\beta n + O(1)$	$\gamma n + O(1)$	$\frac{Y_n - \beta n}{\sqrt{\gamma n}} \longrightarrow_d N_{0,1}$ Corollary 5.20

Table 5.1: This table summarizes most results presented in this chapter. To specify the limit distributions in some cases we refer to theorems proved in the previous sections. Moreover, we use $N_{m,s}$ and U to denote, respectively, a normal random variable of mean m and variance s and a Bernoullian r.v. of parameter $p = \alpha_1/(\alpha_1 + \alpha_2)$.

to build from (5.7) a family of primitive linear representations (ξ, μ, η) where we may have $\lambda_1 = \lambda_2$ or $\lambda_1 \neq \lambda_2$. In all cases, it turns out that $\beta_1 = (1 + (1 + \sqrt{2})^2)^{-1} \simeq 0.146$ and $\beta_2 = 11/15 \simeq 0.733$ (and hence $\beta_1 \neq \beta_2$). Figure 5.4 illustrates the probability function of the r.v. Y_{50} in three different cases. If $\lambda_1 = 2$ and $\lambda_2 = 1$ we find a normal density of mean asymptotic to $50\,\beta_1$. If $\lambda_1 = 1$ and $\lambda_2 = 2$ we have a normal density of mean asymptotic to $50\,\beta_2$. Both situations correspond to Theorem 5.6. If $\lambda_1 = \lambda_2 = 1$, we recognize the convergence to the uniform distribution in the interval $[50\,\beta_1, 50\,\beta_2]$ according to Theorem 5.12. □

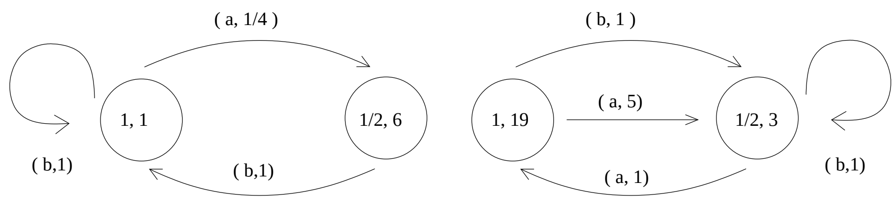


Figure 5.3: Two weighted finite automata over the alphabet $\{a, b\}$, defining the primitive linear representations (π_i, ν_i, τ_i), $i = 1, 2$.

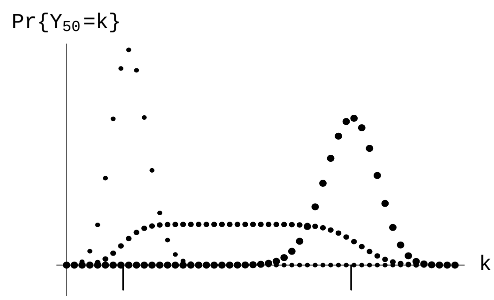


Figure 5.4: Probability functions of Y_{50} in the product model where the two factor components are defined by Figure 5.3 with weighted expanded by a constant factor. The vertical bars have abscissas $50\beta_1$ and $50\beta_2$. The curves correspond to the cases where (λ_1, λ_2) are equal to (2,1), (1,2) and (1,1), respectively.

We conclude observing that some of the previous results clearly extend to rational stochastic models given by more than two primitive components. For instance the result given in Theorem 5.6 also holds in the multicomponent case when only one dominant component exists and this is not degenerate. Analogously, if two (non-degenerate) equipotent components dominate the others, then a result similar to Theorem 5.12 or Proposition 5.19 holds (according whether there exists a communication from the first to the second component). However, it is clear that in the multicomponent model the number of subcases grows exponentially: more than two equipotent components can dominate the others and the limit distribution depends also on the geometry of communication among them; further, with more than one dominant component, several degenerate cases can occur and the limit distribution might depend on the dominated components too. For these reasons, a significant analysis of the general multicomponent model have be to based on an approach quite different from those used so far. This will be the goal of next chapter.

Chapter 6

Multicomponent models

Up to now, the RSF has been studied only in the primitive models or in models consisting in two primitive components. In this chapter we present a general approach to the analysis of arbitrary rational models, explicitly establishing the limit distribution in the most significant cases. Such analysis is based on the decomposition of the linear representation defining the model into strongly connected components. This is a usual approach in the analysis of counting problems on regular languages (see for instance [29] for an application concerning trace languages).

The section is organized as follows. In Section 6.1 we show how an arbitrary model can be decomposed and we introduce the notions of main chain and simple model. The role of main chains is then the object of Section 6.2. Under a special assumption on the main chain, in Section 6.3 we determine the limit distributions of pattern statistics for simple models; they are characterized by an interesting family of unimodal density functions defined by polynomials over adjacent intervals. Finally in Section 6.4 we extend the results to all simple models and also provide a natural method to determine the limit distribution in the general case.

6.1 Decomposition of a rational model

Let (ξ, μ, η) be a linear representation over the alphabet $\{a, b\}$ and let Y_n count the number of occurrences of a in the stochastic model defined by the corresponding rational series. Using the notation of the previous chapters, set $A = \mu(a)$, $B = \mu(b)$, $M = A + B$. Then, consider the incidence graph of M and let $C_1, C_2, \ldots, C_s$ be its strongly connected components. We define C_i *initial* (resp. *final*) if $\xi_p \neq 0$ (resp. $\eta_p \neq 0$) for some $p \in C_i$. As in Section 2.2, the *reduced graph* of (ξ, μ, η) is then the directed acyclic graph G where $C_1, C_2, \ldots, C_s$ are the vertices and any pair (C_i, C_j) is an edge if and only if $i \neq j$ and $M_{pq} \neq 0$ for some $p \in C_i$ and some $q \in C_j$.

Hence, up to a permutation of indices, the matrix M can be represented as a triangular block matrix of the form

$$M = \begin{pmatrix} M_1 & M_{12} & M_{13} & \cdots & M_{1s} \\ 0 & M_2 & M_{23} & \cdots & M_{2s} \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & M_s \end{pmatrix} \tag{6.1}$$

where each M_i corresponds to the strongly connected component C_i and every M_{ij} corresponds to the transitions from vertices of C_i to vertices of C_j in the incident graph of M. Also A, B, ξ and η admit similar decompositions: we define the matrices A_i, A_{ij}, B_i, B_{ij} and the vectors ξ_i, η_i in the corresponding way and we say that the component C_i is *degenerate* if $A_i = 0$ or $B_i = 0$. To each M_i we can apply the Perron–Frobenius Theorem for irreducible matrices, hence we know that each M_i has a nonnegative real eigenvalue λ_i of maximum modulus. We call *main eigenvalue* of M the value $\lambda = \max\{\lambda_i \mid i = 1, 2, \ldots, s\}$ and we say that C_i is a *dominant component* if $\lambda_i = \lambda$. Observe that $\lambda_i = 0$ only if C_i reduces to a loopless single node and hence from now on we assume $\lambda > 0$. If further M_i is primitive, we say that C_i is a *primitive component.*

The block decomposition of M induces a decomposition of the matrix $R(z, w)$ defined in (4.11). More precisely, the blocks under the diagonal are all null, while the upper triangular part is composed by a family of matrices, say $R_{ij}(z, w)$, $1 \leq i \leq j \leq s$. Note that the bivariate generating function $\mathbf{r}(z, w) = \xi_T R(z, w)\, \eta$, which is the main tool of our investigation, is now given by

$$\mathbf{r}(z, w) = \sum_{n=0}^{\infty} \xi^T (Ae^z + B)^n \eta \cdot w^n = \sum_{1 \leq i \leq j \leq s} \xi_i^T R_{ij}(z, w) \eta_j \ . \tag{6.2}$$

Setting $M_{ij}(z) = A_{ij}e^z + B_{ij}$ and reasoning by induction on $j - i$, one can prove that, for each $1 \leq i \leq j \leq s$,

$$R_{ij}(z, w) = \begin{cases} (I - w(A_i e^z + B_i))^{-1} & \text{if } j = i \\ \sum R_{i_1 i_1}(z, w) M_{i_1 i_2}(z) R_{i_2 i_2}(z, w) \cdots M_{i_{\ell-1} i_\ell}(z) R_{i_\ell i_\ell}(z, w) \cdot w^{\ell-1} & \text{if } j \neq i \end{cases} \tag{6.3}$$

where the sum is extended over all sequences of integers $(i_1, i_2, \ldots, i_\ell)$, $\ell \geq 2$ such that $i_1 = i$, $i_t < i_{t+1}$ for each $t = 1, \ldots, \ell - 1$ and $i_\ell = j$.

Equation (6.3) suggests us to introduce the notion of chain of the reduced graph G associated with (ξ, μ, η). A *chain* is a simple path in G, i.e. any sequence of distinct components $\kappa = (C_{i_1}, C_{i_2}, \ldots, C_{i_\ell})$, $\ell \geq 1$, such that $M_{i_j i_{j+1}} \neq 0$ for every $j = 1, 2, \ldots, \ell - 1$. We say that ℓ is the *length* of κ while the *order* of κ is the number of its dominant components. We denote by Γ the family of all chains in G starting with an initial component and ending with a final component. We say that a chain κ is *main chain* if $\kappa \in \Gamma$ and its order is maximal in Γ. We denote by Γ_m the set of all main chains in G.

In Section 6.2 we show how the main chains determine the limit distribution of the sequence $\{Y_n\}$ associated with the linear representation (ξ, μ, η). Intuitively, this is a consequence of two facts. First, the characteristic function of (a normalization of) Y_n depends on the sequences $\{r_n(z)\}$ for z near 0, and hence on the generating function $\mathbf{r}(z, w)$. Second, by (6.2), this function is a sum of products like those in (6.3), each of which is identified by a chain: the products corresponding to the main chains have singularities of smallest modulus with the largest degree, and hence they yield the main asymptotic contribution to the associated sequence $\{r_n(z)\}$.

The relevance of main chains leads to study the simple but representative case when the model has just one main chain, say κ. In this case, the properties of Y_n first depend on the order k of κ, i.e. the number of its dominant components. We first determine the limit distribution of Y_n when all dominant components of κ are primitive, non-degenerate and have distinct mean constants. A similar approach can be developed when the above mean constants are partially or totally coincident.

For this reason we introduce the notion of simple model. Formally, we say that (ξ, μ, η) is a *simple* linear representation, or just a *simple model*, if Γ_m contains only one chain κ and, for every dominant component C_i in κ, M_i primitive and $A_i \neq 0 \neq B_i$ (for instance, if M is primitive, then it defines a simple model with main chain of length 1). Note that, for such matrix M_i, we can define the mean constant β_i and the variance constants γ_i as in (4.17), $0 < \beta_i < 1$ and $\gamma_i > 0$.

If the order of κ is 2 or lower than 2, then the limit distribution derives from the analysis of primitive and bicomponent models given in Chapters 4 and 5, respectively:

- If κ has only one dominant component C_i, then the limit distribution of $\frac{Y_n - \beta_i n}{\sqrt{\gamma_i n}}$ is a Gaussian distribution of mean value 0 and variance 1 (see Theorem 4.14).

- If κ has two dominant components C_i, C_j, then we have the following three subcases:

  1. If $\beta_i \neq \beta_j$, then Y_n/n converges in law to a r.v. uniformly distributed in the interval $[b_1, b_2]$, where $b_1 = \min\{\beta_i, \beta_j\}$ and $b_2 = \max\{\beta_i, \beta_j\}$ (see Theorem 5.12).
  2. If $\beta_i = \beta_j = \beta$ but $\gamma_i \neq \gamma_j$, then the limit distribution of $\frac{Y_n - \beta n}{\sqrt{n}}$ is a mixture of normal distributions of mean value 0 and variance uniformly distributed in the interval $[c_1, c_2]$, where $c_1 = \min\{\gamma_i, \gamma_j\}$ and $c_2 = \max\{\gamma_i, \gamma_j\}$ (see Theorem 5.13). In other words, $\frac{Y_n - \beta n}{\sqrt{n}}$ converges in law to a random variable with density function

  $$f(x) = \frac{1}{c_2 - c_1} \int_{c_1}^{c_2} \frac{e^{-x^2/(2v)}}{\sqrt{2\pi v}} dv \ .$$

  3. If $\beta_i = \beta_j = \beta$ and $\gamma_i = \gamma_j = \gamma$, then the distribution of $\frac{Y_n - \beta n}{\sqrt{\gamma n}}$ again converges to a Gaussian distribution of mean value 0 and variance 1 (see Theorem 5.14).

Notice that in Chapter 5 we considered non-simple models, too. For instance, in the sum model with two equipotent components one has two main chains of length 1 and one obtains different results according to the values of mean and variance constants.

6.2 The role of main chains

Now we study the properties of main chains and in particular we show that the behaviour of the sequence $\{r_n(z)\}$ as z tends to 0 is mainly determined by the contribution of main chains.

To this aim, let us examine the terms in the decomposition (6.2) of the generating function $\xi^T R(z, w)\eta$. First we consider the case $i = j$; by relation (6.3), we have

$$R_{jj}(z, w) = (I - w(A_j e^z + B_j))^{-1} = \frac{\text{Adj}(I - w(A_j e^z + B_j))}{\text{Det}(I - w(A_j e^z + B_j))}$$

and hence, as z tends to 0, the singularities of each entry approach the inverses of eigenvalues of M_j. We can distinguish three cases according to the properties of M_j:

i) M_j is primitive and dominant. Then λ is its (unique) eigenvalue of largest modulus. The equation $\text{Det}(yI - (A_j e^z + B_j)) = 0$ implicitly defines a function $y = y_j(z)$ in a neighbourhood of $z = 0$ such that $y_j(0) = \lambda$. Such a function is analytic at the point $z = 0$ and admits an expansion

$$y_j(z) = \lambda \left(1 + \beta_j z + \frac{\gamma_j + \beta_j^2}{2} z^2 + \mathrm{O}(z^3) \right) \tag{6.4}$$

where β_j and γ_j are the mean and variance constants of M_j. This implies, for z near 0 and some $0 < \rho < \lambda$, that

$$R_{jj}(z, w) = \frac{R_j(z)}{1 - y_j(z)w} + \mathrm{O}\left(\frac{1}{1 - \rho w}\right)$$

where $R_j(z)$ is a matrix of functions analytic and non-null at $z = 0$.

ii) M_j is dominant (but not necessarily primitive). Then we can consider the family E_j of the eigenvalues of M_j of largest modulus. We know that $\lambda \in E_j$; moreover, by the Perron–Frobenius Theorem for irreducible matrices, every $\alpha \in E_j$ is a simple root of the characteristic polynomial of $A_j e^z + B_j$; hence the equation $\text{Det}(yI - (A_j e^z + B_j)) = 0$ implicitly defines a function $y = y_\alpha(z)$ in a neighbourhood of $z = 0$ such that $y_\alpha(0) = \alpha$. Also $y_\alpha(z)$ is analytic at $z = 0$ where it admits an expansion

$$y_\alpha(z) = \alpha\left(1 + m_\alpha z + s_\alpha z^2 + \mathrm{O}(z^3)\right) \tag{6.5}$$

with $m_\alpha \in \mathbb{R}_+$ and $\Re s_\alpha \geq 2m_\alpha^2$ Again this implies, for z near 0 and some $0 < \rho < \lambda$, the equation

$$R_{jj}(z,w) = \sum_{\alpha \in E_j} \frac{R_\alpha(z)}{1 - y_\alpha(z)w} + \mathrm{O}\left(\frac{1}{1-\rho w}\right) \tag{6.6}$$

where $R_\alpha(z)$ are matrices of functions analytic and non-null at $z = 0$.

iii) M_j is not dominant. Then all its eigenvalues are smaller than λ in modulus and, reasoning as above, as z is near to 0 all singularities of $R_{jj}(z,w)$ are in modulus greater than λ^{-1}. This implies, for some $0 < \rho < \lambda$ and all z near 0

$$R_{jj}(z,w) = \mathrm{O}\left(\frac{1}{1-\rho w}\right). \tag{6.7}$$

Now, let us examine the behaviour of $R_{ij}(z,w)$ for $i \neq j$. Recalling (6.3), we consider an arbitrary chain $\kappa = (C_{i_1}, C_{i_2}, \ldots, C_{i_\ell})$ with $\ell \geq 2$ and we denote $R_\kappa(z,w)$ the corresponding matrix given by

$$R_\kappa(z,w) = R_{i_1 i_1}(z,w)M_{i_1 i_2}(z)R_{i_2 i_2}(z,w)\cdots M_{i_{\ell-1} i_\ell}(z)R_{i_\ell i_\ell}(z,w)\cdot w^{\ell-1}. \tag{6.8}$$

We also define the sequence $\{r_n^{(\kappa)}(z)\}$ by

$$\xi_{i_1}^T R_\kappa(z,w)\eta_{i_\ell} = \sum_{n=0}^{\infty} r_n^{(\kappa)}(z) w^n. \tag{6.9}$$

Then, next proposition can be proved by replacing the expansions (6.6) and (6.7) into (6.8).

Proposition 6.1 *The following statements hold for every chain κ and every $c \in \mathbb{C}$, as n tends to $+\infty$:*

1. *If the order of κ is 0, then ${r_n}^{(\kappa)}(c/n) = O(\tau^n)$ for some $0 < \tau < \lambda$.*
2. *If the order of κ is $k \geq 1$, then ${r_n}^{(\kappa)}(c/n) = O(\lambda^n\ n^{k-1})$. Further, if the dominant components of κ are all primitive and nondegenerate, then ${r_n}^{(\kappa)}(c/n) = \Theta(\lambda^n\ n^{k-1})$.*

Proof. Let κ be a chain of order $k \geq 1$. Without loss of generality, we can assume that $\kappa = (C_1, C_2, \ldots, C_\ell)$. Then we have

$$R_\kappa(z,w) = R_{11}(z,w)M_{12}(z)R_{22}(z,w)\cdots M_{\ell-1\ell}(z)R_{\ell\ell}(z,w)\cdot w^{\ell-1} \tag{6.10}$$

and it is clear that the singularities of $\xi_1^T R_\kappa(z,w)\eta_\ell$ are those of the matrices $R_{jj}(z,w)$ for $j = 1, 2, \ldots, \ell$. Set $J = \{j : C_j \text{ is dominant }\}$, note that $\#J = k$ and observe that, as z goes to 0, $R_{jj}(z,w)$ can be expressed in the form (6.6) for each $j \in J$, while if $j \in J^c$ relation (6.7) holds.

First consider the case where κ has order 0. Then $J = \emptyset$ and, as z goes to 0, we have $R_{jj}(z,w) = \mathrm{O}((1-\rho w)^\ell)$ for some $0 < \rho < \lambda$. Therefore there exists a constant $\rho < \tau < \lambda$ such that

$$r_n^{(\kappa)}(c/n) = \mathrm{O}(n^{\ell-1}\rho^n) = \mathrm{O}(\tau^n)$$

holds. This conclude the proof of point 1.

Now suppose that κ has order $k \geq 1$. Replacing the expansions (6.6) and (6.7) in (6.10) one verifies that $\xi_1^T R_\kappa(z,w)\eta_\ell$ can be expressed as sum of terms of the form

$$\frac{H(z) \cdot w^{\ell-1}}{\prod_{j=1}^{\ell}(1-\epsilon_j(z)w)} \tag{6.11}$$

where $H(z)$ is a complex functions, analytic and nonnull at $z=0$, and each $\epsilon_j(z)$ is an eigenvalue of $(A_j e^z + B_j)$. For any z near 0 each of these terms can be split as sum of two fractions obtained by isolating the main singularities. Setting $I = \{j \in J : \lim_{z\to 0} |\epsilon_j(z)| = \lambda\}$, the previous fraction can be written as

$$\frac{H_1(z,w)}{\prod_{j\in I}(1-\epsilon_j(z)w)} + \frac{H_2(z,w)}{\prod_{j\in I^c}(1-\epsilon_j(z)w)} \tag{6.12}$$

where $H_1(z,w)$ is a polynomial in w of degree $\#I - 1$ whose coefficients are functions in z analytic at $z=0$, and $H_2(z,w)$ is defined similarly. We can ignore the second fraction because (for z near 0) its singularities are in modulus strictly greater than λ^{-1} and hence its contribution to $r_n^{(\kappa)}(c/n)$ is of the order $\mathrm{O}(\tau^n)$, for some $0 < \tau < \lambda$. On the contrary, in a neighbourhood of $z=0$, the first fraction can be written in the form

$$\frac{H_1(z,w)}{\prod_{j\in I}(1-y_{\alpha_j}(z)w)} \tag{6.13}$$

where each α_j belongs to the set E_j of eigenvalues of maximum modulus of M_j. The sequence associated with each of these fractions is related to the convolution of the sequences $\{y_{\alpha_j}(z)^n\}$, for $j \in I$. More precisely, the n-th element of such a sequence is a linear combination of the elements of index $n-i$ of that convolution, for $0 \leq i < \#I$. For $z = c/n$, using the expansion in (6.5), the modulus of each of these terms can be bounded by

$$C_1 \cdot \left| \sum_{\sum_{j\in I} n_j = n-i} \prod_{j\in I} (y_{\alpha_j}(c/n))^{n_j} \right| \leq C_1 \sum_{\sum_{j\in I} n_j = n-i} |\alpha_j|^{n-i} \cdot |1 + C_2/n|^{n-i} = \mathrm{O}\left(\lambda^n n^{\#I-1}\right) \tag{6.14}$$

for some positive constants C_1 and C_2. Thus, since $I \subseteq J$, we have $\#I \leq k$ and hence equation (6.14) proves that $r_n^{(\kappa)}(c/n) = \mathrm{O}(\lambda^n\ n^{k-1})$ for every $c \in \mathbb{C}$, as n goes to infinity.

Now, assume that all dominant components of κ are primitive and non-degenerate. Then $E_j = \{\lambda\}$ for all $j \in J$ and in $\xi_1^T R_\kappa(z,w)\eta_\ell$ there is only one term of the form (6.13) such that $I = J$. Call $U(z,w)$ such a term. Note that, by equation (6.14), all the other terms ($I \neq J$ being true) give a contribution of the order $\mathrm{O}(\lambda^n\ n^{k-2})$. Replacing the expansion (6.4) in $U(z,w)$ and reasoning as in (6.14), we get $r_n{}^{(\kappa)}(c/n) = \Theta(\lambda^n\ n^{k-1})$ for every $c \in \mathbb{C}$, as n goes to infinity. □

Since by equation (6.2) we have $r_n(z) = \sum_{\kappa\in\Gamma} r_n^{(\kappa)}(z)$, we obtain the following result, which shows the key role of the main chains.

Theorem 6.2 *If all dominant components of the main chains are primitive and non-degenerate, then for every constant $c \in \mathbb{C}$ we have*

$$r_n(c/n) = \sum_{\kappa\in\Gamma_m} r_n^{(\kappa)}(c/n)\ (1 + O(1/n)) = \Theta(\lambda^n n^{k-1})$$

where k is the order of the main chains.

We conclude this section observing that Theorem 6.2 does not hold if the main chains admit non-primitive dominant components.

6.3 Limit distribution in simple models

In this section we determine the limit distribution of Y_n in the simple models that satisfy the following additional property: the dominant components of the main chain have (pairwise) distinct mean constants. Such analysis will be extended in Section 6.4 to all simple models (with partially or totally coincident mean constants of dominant components) and also to all models whose main chains have primitive, non-degenerate dominant components.

Under our hypotheses, the limit distribution turns out to be related to a special family of distribution functions we call *polynomial* since their density is defined by polynomials over adjacent intervals. Formally, consider an array $b = (b_1, b_2, \ldots, b_k)$ of $k \geq 2$ real numbers such that $0 < b_1 < b_2 < \cdots < b_k < 1$ and let $f_b : \mathbb{R} \longrightarrow \mathbb{R}$ be the function defined by

$$f_b(x) = \begin{cases} 0 & \text{if } x < b_1 \\ (k-1)\sum_{j=r}^{k} c_j (b_j - x)^{k-2} & \text{if } b_{r-1} \leq x < b_r \text{ for some } 1 < r \leq k \\ 0 & \text{if } x \geq b_k \end{cases} \tag{6.15}$$

where $c_j = \prod_{i \neq j}(b_j - b_i)^{-1}$ for every $j = 1, 2, \cdots, k$. Note that if $k = 2$, then f_b is the uniform density function over the interval (b_1, b_2). In Section 6.3.2 below we show that, for every $k \geq 3$, f_b is a density function, i.e. $f_b(x) \geq 0$ for every $x \in \mathbb{R}$, and $\int_{-\infty}^{+\infty} f_b(x)dx = 1$; moreover, we determine its characteristic function and show that f_b is unimodal. In the following we say that a r.v. X is a *polynomial* r.v. of parameters $b_1, b_2, \ldots, b_k$ if $f_b(x)$ is its density function.

Theorem 6.3 *Let Y_n count the number of a in a simple model with main chain κ having order k. Moreover, let $\beta_1, \ldots, \beta_k$ be the mean constants of dominant components in κ in non-decreasing order. If $k \geq 2$ and all β_j's are distinct, then Y_n/n converges in law to a polynomial r.v. of parameters $\beta_1, \ldots, \beta_k$.*

The proof of Theorem 6.3 is presented in the following subsections; it is based on the properties of convolutions of sequences of powers of complex numbers.

6.3.1 Multiple convolutions

Given an array $a = (a_1, a_2, \ldots, a_k)$ of $k \geq 2$ nonnull complex numbers, consider the function

$$G_a(w) = \frac{w^{k-1}}{\prod_{i=1}^{k}(1 - a_i w)} .$$

This is the generating function of the convolution of the sequences $\{a_1^n\}_n, \{a_2^n\}_n, \ldots, \{a_k^n\}_n$ shifted of $k-1$ indices. More precisely, at the point $w = 0$ such a function admits the power series expansion $G_a(w) = \sum_{n=0}^{+\infty} g_a(n)w^n$ such that

$$g_a(n) = \begin{cases} 0 & \text{if } \; 0 \leq n \leq k-2 \\ \sum_{*} a_1^{i_1} a_2^{i_2} \cdots a_k^{i_k} & \text{if } \; n \geq k-1 \end{cases} \tag{6.16}$$

where we mean that the sum (*) is extended over all k-tuples $(i_1, \ldots, i_k) \in \mathbb{N}^k$ such that $i_1 + \cdots + i_k = n - k + 1$. When all a_j's are distinct, the following proposition allows us to express the terms of the sequence $\{g_a(n)\}_{n \geq 0}$ in a useful form.

Proposition 6.4 *Let $a = (a_1, a_2, \dots, a_k)$ be an array of $k \geq 2$ distinct nonnull complex numbers and let the sequence $\{g_a(n)\}_n$ be defined by (6.16). Then, for every $n \in \mathbb{N}$, we have*

$$g_a(n) = \sum_{j=1}^{k} c_j \ a_j^n \tag{6.17}$$

where $c_j = \prod_{i \neq j}(a_j - a_i)^{-1}$ for every $j = 1, 2, \cdots, k$.

Proof. First observe that the generating function $G_a(w)$ of $\{g_a(n)\}_n$ can be decomposed in partial fractions. Hence there exist k complex coefficients $c_1, c_2, \dots, c_k$ such that

$$G_a(w) = \frac{w^{k-1}}{\prod_{i=1}^{k}(1 - a_i w)} = \sum_{i=1}^{k} \frac{c_i}{1 - a_i w} \ .$$

Now, multiplying the last term in the previous equation by $(1 - a_j w)$, we obtain

$$\frac{w^{k-1}}{\prod_{i \neq j}(1 - a_i w)} = c_j + \sum_{i \neq j} \frac{c_i(1 - a_j w)}{1 - a_i w}$$

which leads to $c_j = \prod_{i \neq j}(a_j - a_i)^{-1}$ when evaluated in $w = 1/a_j$. This completes the proof. □

Corollary 6.5 *Let $a = (a_1, a_2, \dots, a_k)$ be an array of $k \geq 2$ distinct nonnull complex numbers and set $c_j = \prod_{i \neq j}(a_j - a_i)^{-1}$ for every $j = 1, 2, \cdots, k$. Then for each $0 \leq s \leq k - 2$ the polynomial $\sum_{j=1}^{k} c_j(a_j - x)^s$ is identically null and in particular $\sum_{j=1}^{k} c_j a_j^s = 0$. Moreover we have $\sum_{j=1}^{k} c_j a_j^{k-1} = 1$.*

Proof. By equations (6.16) and (6.17) we derive $\sum_j c_j a_j^{k-1} = g_a(k-1) = 1$ and $\sum_j c_j a_j^s = g_a(s) = 0$ for every $0 \leq s \leq k - 2$. Thus, by the binomial formula, also the polynomial $\sum_j c_j(a_j - x)^s$ is identically null. □

6.3.2 Polynomial distributions

Let f_b be defined as in equation (6.15), where b is a k-tuple of real numbers $(b_1, b_2, \dots, b_k)$ such that $0 < b_1 < b_2 < \cdots < b_k < 1$ and $k \geq 3$. Here we prove that f_b is a unimodal density function and we determine its characteristic function. These results are consequences of the key property stated in Corollary 6.5, in particular the next proposition is easily proved from equation (6.15).

Proposition 6.6 *If $k \geq 3$, then f_b is continuously differentiable all over $\mathbb{R}$ up to the order $k - 3$. Moreover the $(k-2)$-th derivative of f_b is well defined in $\mathbb{R} \backslash \{b_1, \dots, b_k\}$ and is constant in each of the intervals (b_i, b_{i+1}), $i = 1, \cdots, k - 1$.*

Lemma 6.7 *Let $f : \mathbb{R} \to \mathbb{R}$ be a function admitting j-th derivative all over $\mathbb{R}$ for some $j \geq 1$. Also assume that, for some reals $a < b$, f has m zeros in (a, b) and $f(x) = 0$ for each $x \leq a$ or $x \geq b$. Then, for every $i = 1, \dots, j$, the i-th derivative of f admits at least $m + i$ zeros in (a, b).*

Proof. We reason by induction on $i = 1, \dots, j$. If $i = 1$, then consider the $m + 1$ intervals determined by the zeros of f in $[a, b]$. For each of them, say (x_1, x_2), Rolle's Theorem guarantees that $f'(x) = 0$ for at least one $x \in (x_1, x_2)$.

Now assume $1 < i < j$ and consider the i-th derivative of f, that is $h = f^{(i)}$. By the properties of f, we have $h(a) = h(b) = 0$ and by the inductive hypothesis h admits $m + i$ zeros in (a, b). Therefore, by applying the previous argument to h, one proves that $h' = f^{(i+1)}$ admits $m + i + 1$ zeros in (a, b). □

Proposition 6.8 *For every $k \geq 3$, the function f_b admits a unique maximum all over $\mathbb{R}$.*

Proof. If $k = 3$ the property follows by a direct inspection of the function, which is linear and nonnull in the intervals (b_1, b_2) and (b_2, b_3). If $k \geq 4$, let us consider the $(k-3)$-th derivative ${f_b}^{(k-3)}(x)$ of $f_b(x)$. It is immediate to see that ${f_b}^{(k-3)}(x)$ is linear with respect to x in each of the $k-1$ intervals (b_i, b_{i+1}), $i = 1, \ldots k-1$. Moreover, by Corollary 6.5, it does not vanish in $(b_1, b_2) \cup (b_{k-1}, b_k)$. Thus, ${f_b}^{(k-3)}$ has at most $k-3$ many zeros in (b_1, b_k).

Now, assume by ab adsurdum that f_b is not unimodal. Then its derivative f_b' vanishes in at least 3 points in the interval (b_1, b_k) and hence f_b' satisfies the hypotheses of Lemma 6.7 with $i = k-4$ and $m = 3$. As a consequence, ${f_b}^{(k-3)}$ admits at least $k-1$ zeros in (b_1, b_k), which contradicts the previous property. $\square$

Fig.6.1 and Fig.6.2 show the plots of the functions f_b having parameters $b = (0.1, 0.3, 0.4, 0.8)$ and $b = (0.008, 0.95, 0.96, 0.97, 0.98, 0.99)$, respectively. In each figure the first picture represents the entire curve, while the others show the details of the function in some subintervals. The vertical bars indicate the values of b_j's. Note that if $k = 4$ the maximum necessarily lays in the intermediate interval (b_2, b_3). On the contrary, if $k > 4$ the maximum can lay in any interval between b_2 and b_{k-1}. For instance in Fig. 6.2, because of the asymmetric position of the points b_j's, it lays in the second interval (b_2, b_3).

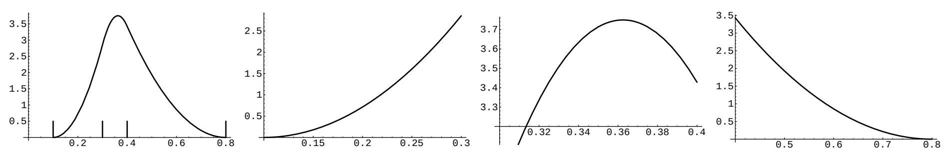


Figure 6.1: Plot of the function $f_b(x)$, where $b_1 = 0.1$, $b_2 = 0.3$, $b_3 = 0.4$, $b_4 = 0.8$. The vertical bars indicate the values of b_j's.

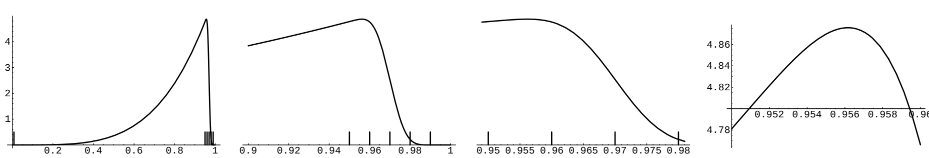


Figure 6.2: Plot of the function $f_b(x)$, where $b_1 = 0.008$, $b_2 = 0.95$, $b_3 = 0.96$, $b_4 = 0.97$, $b_5 = 0.98$, $b_6 = 0.99$. The vertical bars indicate the values of b_j's.

From Proposition 6.8 it is clear that $f_b(x) \geq 0$ for all $x \in \mathbb{R}$. To prove that f_b is a density function we still have to show that its integral over $\mathbb{R}$ equals 1. This is a consequence of Proposition 6.9 below where we determine the characteristic function of f_b, defined by

$$\Phi_b(t) = \frac{(k-1)!}{(it)^{k-1}} \sum_{j=1}^{k} \frac{e^{ib_j t}}{\prod_{i \neq j}(b_j - b_i)} \, . \tag{6.18}$$

Proposition 6.9 *For every $b = (b_1, b_2, \ldots, b_k) \in \mathbb{R}^k$ such that $0 < b_1 < b_2 < \cdots < b_k < 1$ and $k \geq 2$, $f_b(x)$ is a density function and $\Phi_b(t)$ is its characteristic function.*

Proof. By using Corollary 6.5 one can show, by a direct computation, that $\lim_{t \to 0} \Phi_b(t) = 1$. Therefore, it suffices to show that $\int_{-\infty}^{+\infty} f_b(x) e^{itx} dx = \Phi_b(t)$ for every $t \in \mathbb{R}$. We prove this equality by using Corollary 6.5 again. Set $I(t) = \int_{-\infty}^{\infty} f_b(x) e^{itx} dx$ and $c_j = \prod_{i \neq j} (b_j - b_i)^{-1}$ for every $j = 1, \ldots, k$. Observe that

$$I(t) = (k-1) \sum_{r=2}^{k} \sum_{j=r}^{k} c_j \int_{b_{r-1}}^{b_r} (b_j - x)^{k-2} e^{itx} dx \ .$$

Integrating by parts one can verify that for $t \neq 0$ the function $e^{itx}(c-x)^p$ admits the antiderivative

$$\frac{e^{itx}}{it} \sum_{s=0}^{p} \frac{p!\,(c-x)^{p-s}}{(p-s)!\,(it)^s} \ .$$

Hence we can write $I(t) = \sum_{r=2}^{k} \sum_{j=r}^{k} c_j (A_{r,j} - A_{r-1,j})$ where

$$A_{r,j} = e^{itb_r} \sum_{s=0}^{k-2} \frac{(k-1)!\,(b_j - b_r)^{k-2-s}}{(k-2-s)!\,(it)^{s+1}} \quad \text{and in particular} \quad A_{r,r} = \frac{(k-1)!}{(it)^{k-1}} e^{itb_r} \ .$$

Now set $B_r = \sum_{j=r}^{k} c_j A_{r,j}$ and $C_r = \sum_{j=r}^{k} c_j A_{r-1,j}$. For each $2 \leq r \leq k-1$ we have $B_r - C_{r+1} = c_r A_{r,r}$ and moreover $B_k = c_k A_{k,k}$. Finally, by Corollary 6.5 we have $C_2 = \sum_{j=1}^{k} c_j A_{1,j} - c_1 A_{1,1} = -c_1 A_{1,1}$. As a consequence the integral can be computed as follows

$$I(t) = \sum_{r=2}^{k} (B_r - C_r) = \sum_{j=1}^{k} c_j A_{j,j} = \frac{(k-1)!}{(it)^{k-1}} \sum_{j=1}^{k} c_j e^{itb_j} = \Phi_b(t)$$

and this concludes the proof. □

6.3.3 Polynomial limit theorem

We are now able to prove Theorem 6.3. By Proposition 6.9 it is sufficient to show that the characteristic function of Y_n/n converges to $\Phi_\beta(t)$ for every $t \in \mathbb{R}$, where β is the k-tuple of the mean constants β_j in increasing order. Observe that, by equation (4.6), the characteristic function of Y_n/n is given by

$$\Psi_{Y_n/n}(it) = \Psi_{Y_n}(it/n) = \frac{r_n(it/n)}{r_n(0)} \ ;$$

thus, let us first prove the following lemma.

Lemma 6.10 *Consider a simple model with main chain κ of order k and let $\beta_1, \beta_2, \ldots, \beta_k$ be the mean constants of its dominant components. Then, for every $t \in \mathbb{R}$, as n grows to $+\infty$ we have*

$$r_n\left(\frac{it}{n}\right) = \sum_{j=0}^{k-1} S_j\left(\frac{it}{n}\right) \lambda^{n-j} D_j\left(\frac{it}{n}\right) \cdot (1 + O(1/n)) \tag{6.19}$$

where, for each j, $S_j(z)$ is an analytic function at $z = 0$ and $D_j(z)$ is defined by

$$D_j(z) = \sum_{n_1 + \ldots + n_k = n-j} (1+\beta_1 z)^{n_1} (1+\beta_2 z)^{n_2} \cdots (1+\beta_k z)^{n_k} \ .$$

Proof. By Theorem 6.2 we have $r_n(it/n) = r_n^{(\kappa)}(it/n)(1 + \mathrm{O}(1/n))$ and hence we have to show that $r_n^{(\kappa)}(it/n)$ equals the right hand side of (6.19). Without loss of generality we may assume $\kappa = (C_1, C_2, \dots, C_\ell)$ where $C_1, C_2, \dots, C_k$ are its dominant components. Moreover, by our hypotheses, M_j is primitive and non-degenerate for all $j = 1, 2, \cdots, k$. Then, reasoning as in the proof of Proposition 6.1, for any z near 0 the leading term of the generating function $\xi_1 R_\kappa(z, w)\ \eta_\ell$ is given by an expression

$$U(z, w) = \sum_{i=0}^{k-1} \frac{S_i(z) \cdot w^i}{\prod_{j=1}^{k}(1 - y_j(z)w)} \tag{6.20}$$

where the functions $y_j(z)$'s are defined as in (6.4) and the functions $S_i(z)$'s are analytic at $z = 0$. Now, let $u_n(z)$ and $h_n(z)$ be defined by

$$U(z, w) = \sum_{n=0}^{+\infty} u_n(z) w^n \quad \text{and} \quad \prod_{j=1}^{k} (1 - y_j(z)w)^{-1} = \sum_{n=0}^{+\infty} h_n(z) w^n \ ,$$

respectively. Then, by equation (6.20), for every $n \geq k$ we have

$$u_n(z) = \sum_{j=0}^{k-1} S_j(z) h_{n-j}(z) \ .$$

Since $h_n(z) = \sum_{n_1 + \dots n_k = n} y_1(z)^{n_1} y_2(z)^{n_2} \cdots y_k(z)^{n_k}$, replacing (6.4) in the previous equation and taking $z = it/n$ a simple computation proves that $u_n(it/n)$ equals the right hand side of (6.19).

The result now follows by observing that, by Proposition 6.1, the other additive terms of $\xi_1 R_\kappa(z, w)\ \eta_\ell$, for $z = it/n$, yield a sequence of the order $\mathrm{O}(\lambda^n n^{k-2})$. □

Proposition 6.11 *Assume the hypotheses of Theorem 6.3 and let β be the k-tuple of the mean constants β_j in increasing order. Then, for every $t \in \mathbb{R}$, $\Psi_{Y_n/n}(it)$ tends to $\Phi_\beta(t)$ as n grows to $+\infty$.*

Proof. In our hypotheses Lemma 6.10 holds. For $t = 0$ this implies $D_j(0) = n^{k-1}/(k-1)! \cdot (1 + \mathrm{O}(1/n))$ and hence

$$r_n(0) = \frac{n^{k-1}}{(k-1)!} \cdot \sum_{j=0}^{k-1} S_j(0) \lambda^{n-j} \cdot (1 + \mathrm{O}(1/n)) \ .$$

Moreover, for $t \neq 0$, since the β_j's are distinct we have $D_j(it/n) = g_a(n + k - j - 1)$, where a is the array of the complex values $1 + it\beta_\ell/n$, $\ell = 1, \dots, k$. Then, by Proposition 6.4, for every nonnull $t \in \mathbb{R}$ we get

$$D_j(it/n) = \sum_{\ell=1}^{k} \frac{(1 + it\beta_\ell/n)^{n-j+k-1}}{\prod_{i \neq \ell}(it\beta_\ell/n - it\beta_i/n)}$$

and hence from (6.19) we obtain

$$r_n\left(\frac{it}{n}\right) = \sum_{j=0}^{k-1} S_j\left(\frac{it}{n}\right) \frac{\lambda^{n-j} n^{k-1}}{(it)^{k-1}} \sum_{\ell=1}^{k} \frac{(1 + it\beta_\ell/n)^{n-j+k-1}}{\prod_{i \neq \ell}(\beta_\ell - \beta_i)} \cdot (1 + \mathrm{O}(1/n)) \ .$$

Now, as n tends to ∞, it is easy to see that

$$\Psi_{Y_n/n}(it) = \frac{r_n(it/n)}{r_n(0)} \longrightarrow \frac{(k-1)!}{(it)^{k-1}} \sum_{j=1}^{k} \frac{e^{i\beta_j t}}{\prod_{\ell \neq j}(\beta_j - \beta_\ell)} = \Phi_\beta(t) \ .$$

□

6.4 Further developments

The analysis presented in the previous section can be extended to all simple models, also when the mean constants β_j's (associated with the dominant components of the main chain) are partially or totally coincident. The limit distributions of our statistics in this more general case generalize in a natural way the behaviour known for the bicomponent models studied in Section 5. These distributions are defined extending the notion of polynomial density function given in (6.15) by allowing multiplicities in the associated array b. Here we simply state our results avoiding the long proofs; these can be developed along the same line of Section 6.3 and are based on a study of convolutions with multiplicities analogous to the discussion presented in Section 6.3.1.

To state these results precisely we only have to introduce the main characteristic function occurring in this general approach, which is defined as follows. Let $b = (b_1, b_2, \ldots, b_r)$ be an array of $r \geq 2$ distinct real numbers lying in the interval $(0, 1)$ and let $m = (m_1, m_2, \ldots, m_r) \in \mathbb{N}^r$ be a tuple of multiplicities, where $m_j \geq 1$ for each j and $m_1 + \ldots + m_r = k$. Then define the function

$$\Phi_{b,m}(t) = (k-1)! \sum_{j=1}^{r} \sum_{s=1}^{m_j} c_{j,s} \cdot \frac{e^{itb_j}}{(it)^{k-s}(s-1!)} \tag{6.21}$$

where the $c_{j,s}$'s are constants given by

$$c_{j,s} = (-1)^{m_j - s} \sum_{\sum_{\ell \neq j} h_\ell = m_j - s} \prod_{\ell \neq j} \binom{m_\ell + h_\ell - 1}{m_\ell - 1} \cdot \frac{1}{(b_j - b_\ell)^{m_\ell + h_\ell}} . \tag{6.22}$$

One can prove that this is a characteristic function and the corresponding density function can be obtained from (6.15) by a continuity argument. The main difference is that the new density may be non-continuous at the points $x = b_j$ such that $m_j > 1$, $j = 1, \ldots, k$.

Now, let Y_n count the occurrences of a in a simple model having main chain κ of order k. Let $\beta_1, \ldots, \beta_k$ and $\gamma_1, \ldots, \gamma_k$ be, respectively, the mean and variance constants of the dominant components in κ. We also denote by β and γ the arrays of distinct β_j's and γ_j's in increasing order and by u and v the arrays of the corresponding multiplicities. Clearly, if $\beta_1, \ldots, \beta_k$ are pairwise distinct, then Theorem 6.3 holds. Otherwise we have the following cases:

1. If $\beta_1, \ldots, \beta_k$ are partially but not totally coincident (i.e. $\beta_i = \beta_j$ and $\beta_s \neq \beta_t$ for some indices i, j, s, t, $i \neq j$), then Y_n/n converges in distribution to a r.v. of characteristic function $\Phi_{\beta,u}(t)$;

2. If $\beta_j = \beta_1$ for all $j = 2, \ldots, k$ and all γ_j's are pairwise distinct, then $\frac{Y_n - \beta_1 n}{\sqrt{n}}$ converges in distribution to a r.v. of characteristic function $\Phi_\gamma(-t^2/(2i))$;

3. If $\beta_j = \beta_1$ for all $j = 2, \ldots, k$ and $\gamma_1, \ldots, \gamma_k$ are partially but not totally coincident, then $\frac{Y_n - \beta_1 n}{\sqrt{n}}$ converges in distribution to a r.v. of characteristic function $\Phi_{\gamma,v}(-t^2/(2i))$;

4. If $\beta_j = \beta_1$ and $\gamma_j = \gamma_1$ for all $j = 2, \ldots, k$, then $\frac{Y_n - \beta_1 n}{\sqrt{\gamma_1 n}}$ converges in distribution to a normal r.v. of mean 0 and variance 1.

The previous results can be further extended by a standard conditioning argument (already used in chapter 5 for the degenerate cases) to all rational models (ξ, μ, η) such that for every $\kappa \in \Gamma_m$ all dominant components in κ are primitive and non-degenerate. In this case, by equations (6.8) and (6.9), for every $\kappa \in \Gamma_m$ one can easily see that

$$r_n^{(\kappa)}(z) = s_\kappa(z)\lambda^n n^{k-1} + O(\lambda^n n^{k-2})$$

where k is the degree of κ and $s_\kappa(z)$ is a non-null analytic function at $z = 0$. Then, by Theorem 6.2, we have

$$r_n(0) = S\lambda^n n^{k-1} + O(\lambda^n n^{k-2})$$

where $S = \sum_{\kappa\in\Gamma_m} s_\kappa(0)$. We can also associate each $\kappa \in \Gamma_m$ with the probability value p_κ, given by $p_\kappa = s_\kappa(0)/S$. Note that the values $\{p_\kappa\}_{\kappa\in\Gamma_m}$ define a discrete probability measure and they can be explicitly computed from (ξ, μ, η).

Moreover, each $\kappa \in \Gamma_m$ defines a simple rational model in its own right, with an associated sequence of r.v.'s $\{Y_n{}^{(\kappa)}\}$ having its own limit distribution according to Theorem 6.3 and points 1-4 above. In particular, $Y_n^{(\kappa)}/n$ always converges in distribution to a random variable of distribution function $F_\kappa(x)$ defined according to the previous results. Note that if all constants β_j's are here equal, then $F_\kappa(x)$ reduces to the degenerate distribution of mass point β_1. Now it is not difficult to see that the overall statistics Y_n/n converges in distribution to a r.v. of distribution function $F(x)$ defined by $F(x) = \sum_{\kappa\in\Gamma_m} F_\kappa(x)p_\kappa$.

Conclusions

In this thesis we considered the *frequency problem* that studies, from a probabilistic point of view, the number of occurrences of a pattern into a text randomly generated by a stochastic source. In particular, we used rational formal series (or, equivalently, weighted automata) in two non-commuting variables a and b to define a new kind of probabilistic model we called *rational.* We then took in exam the *rational symbol frequency* problem; intuitively this concerns the study of the sequence of random variables $\{Y_n\}_n$ representing the number of occurrences of the symbol a in words of length n chosen at random in $\{a, b\}^*$, according to the probability distribution given by the rational model.

We showed how our model can be viewed as a proper extension of the Markovian model usually considered in the literature. Indeed, we prove that the question of studying the number of occurrences of a regular pattern in a text generated by a Markovian source can always be translated into the rational symbol frequency problem for a suitable rational series over two non-commuting variables, while the converse does not hold in general.

We first assumed that the transition matrix associated with the series defining the model was primitive. We showed that in this case the mean and the variance of Y_n are asymptotically linear and we provided precise expressions for the constants appearing in their asymptotic formulas. We also showed that a central limit theorem holds and we provided a condition that guarantees the existence of a Gaussian local limit theorem; to state this condition, we introduced a notion of symbol periodicity for weighted automata which extends the classical periodicity theory of Perron–Frobenius for non-negative matrices. As an application of the previous analysis, we obtained an asymptotic estimation of the growth of the coefficients for a subclass of rational formal series in two commuting variables.

We then extended the results, dropping the primitive hypothesis usually assumed in the literature. First, we studied bicomponent models, defined by weighted automaton with two strongly connected components, obtaining in many cases limit distributions quite different from the Gaussian one. We also presented a general approach to deal with arbitrary non-primitive models. Again, we started from the decomposition of the weighted automaton defining the model into strongly connected components, in order to detect the elements that mainly determine the limit distribution. In the most relevant cases we established the limit distribution, that is characterized by a unimodal density function defined by polynomials over adjacent intervals.

Among the possible developments of this thesis, let us mention the question of generalizing the analysis and the techniques used here to the *multivariate* case, where different patterns (or simply letters) are considered. In particular, for instance, one could fix an integer k_i for each pattern P_i and study the problem of counting the random texts having exactly k_i occurrences of the pattern P_i, for every index i. Other open questions are related to the *inverse problem*: given some experimental results on the occurrences of a letter into a random text, it would be interesting to derive structural properties of the model that generated the text, like its primitivity, its periodicity or the number of its irreducible components.

Bibliography

[1] D. Barbara, T. Imielinski. Sleepers and Workaholics: Caching Strategies in Mobile Environments. *Proceedings of the 1994 ACM SIGMOD*, ACM Press, 1-15, 1994.

[2] M.-P. Béal. *Codage Symbolique*, Masson, 1993.

[3] E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory*, 15:91–111, 1973.

[4] E. A. Bender, F. Kochman. The distribution of subword counts is usually normal. *European Journal of Combinatorics*, 14:265–275, 1993.

[5] J. Berstel, C. Reutenauer. *Rational series and their languages*, Springer-Verlag, New York - Heidelberg - Berlin, 1988.

[6] A. Bertoni, C. Choffrut, M. Goldwurm, V. Lonati. On the number of occurrences of a symbol in words of regular languages. *Theoretical Computer Science*, 302(1-3):431–456, 2003.

[7] A. Bertoni, C. Choffrut, M. Goldwurm, V. Lonati. Local limit distributions in pattern statistics: beyond the Markovian models. *Proceedings STACS 2004*, *Lecture Notes in Computer Science*, Vol 2996, Springer-Verlag, 117-128, 2004.

[8] A. Bertoni, M. Goldwurm, M. Santini. Random generation and approximate counting of ambiguously described combinatorial structures. *Proceedings STACS 2000*, H. Reichel e S. Tison Eds., *Lecture notes in Computer Science*, Vol 1770, Springer-Verlag, 567-580, 2000.

[9] A. Bertoni, M. Goldwurm, M. Santini. Random generation for finitely ambiguous context-free languages. *R.A.I.R.O. Theoretical Informatics and Applications*, 35:499-512, 2001.

[10] J. Bourdon, B. Vallée. Generalized pattern matching statistics. *Mathematics and computer science II, algorithms, trees, combinatorics and probabilities*, Proc. of Versailles Colloquium, Birkhauser, 249–265, 2002.

[11] R.N. Bracewell. *The Fourier transform and its applications*, McGraw-Hill Book Company, 1986.

[12] P. Cartier, D. Foata. *Problèmes combinatoires de commutation et rearrangements.* Springer-Verlag, 1969.

[13] C. Choffrut, M. Goldwurm, V. Lonati. On the maximum coefficients of rational formal series in commuting variables. To appear in *Proceedings 8th DLT*, *Lecture Notes in Computer Science*, Springer-Verlag, 2004.

[14] D. de Falco, M. Goldwurm, V. Lonati. Frequency of symbol occurrences in simple non-primitive stochastic models. *Proceedings 7th DLT*, Z. Esig and Z. Fülop editors, *Lecture Notes in Computer Science*, vol. 2710, Springer-Verlag, 242–253, 2003.

[15] D. de Falco, M. Goldwurm, V. Lonati. Pattern statistics in bicomponent models. *Proceedings Words 03*, Turku (Finland), TUCS General Publication vol. 27, 344–357, 2003.

[16] D. de Falco, M. Goldwurm, V. Lonati. Frequency of symbol occurrences in bicomponent stochastic models. *Theoretical Computer Science*, 327(3):269–300, 2004.

[17] A. Denise. Génération aléatoire uniforme de mots de langages rationnels. *Theoretical Computer Science*, 159:43–63, 1996.

[18] A. Denise, O. Roques, M. Termier. Random generation of words of context-free languages according to the frequencies of letters. *Trends in Mathematics*, Birkauser, 2000.

[19] V. Diekert. *Combinatorics on traces.* Springer–Verlag, New York Heidelberg Berlin Tokyo, 1990.

[20] V. Diekert, G. Rozenberg (editors). *The book of traces.* World Scientific, 1995.

[21] W. Feller. *An introduction to probability and its applications*, Vol 1. John Wiley and Sons, New York, 1968.

[22] P. Flajolet, R. Sedgewick. The average case analysis of algorithms: saddle point asymptotics. *Rapport de recherche* n. 2376, INRIA Rocquencourt, October 1994.

[23] P. Flajolet, R. Sedgewick. The average case analysis of algorithms: multivariate asymptotics and limit distributions. *Rapport de recherche* n. 3162, INRIA Rocquencourt, May 1997.

[24] P. Flajolet, P. Zimmermann, B. Van Cutsem, A calculus for the random generation of combinatorial structures. *Theoretical Computer Science*, vol. 132(1-2):1–35, 1994.

[25] F.G. Frobenius. Über Matrizen aus positiven Elementen. in *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin.* 1908 (471-6); 1909 (514-8).

[26] I. Fudos, E. Pitoura, W. Szpankowski. On pattern occurrences in a random text. *Information and Processing Letters*, 57:307–312, 1996.

[27] M. S. Gelfand. Prediction of function in DNA sequence analysis. *Journal of Computational Biology*, 2:87–117, 1995.

[28] B.V. Gnedenko. *The theory of probability* (translated by G. Yankovsky). Mir Publishers - Moscow, 1976.

[29] M. Goldwurm. Probabilistic estimation of the number of prefixes of a trace, *Theoretical Computer Science*, 92:249–268, 1992.

[30] M. Goldwurm, V. Lonati. Pattern occurrences in multicomponent models. To appear in *Proceedings STACS 2005*, *Lecture Notes in Computer Science*, Springer-Verlag, 2005.

[31] P. Grabner, M. Rigo. Additive functions with respect to numeration systems on regular languages. *Monatshefte für Mathematik*, 139:205–219, 2003.

[32] L. Guibas, A. Odlyzko. Maximal Prefix-Synchronized Codes. *SIAM Journal on Applied Mathematics*, 35:401-418, 1978.

[33] L. Guibas, A. Odlyzko. Periods in strings. *Journal of Combinatorial Theory*, Ser A, 30: 19-43, 1981.

[34] L. J. Guibas, A. M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *Journal of Combinatorial Theory*, Ser A, 30(2):183–208, 1981.

[35] P. Henrici. *Elements of numerical analysis*, John Wiley, 1964.

[36] J.E. Hopcroft, J.P. Ullman. *Formal languages and their relation to automata.* Addison-Wesley, 1969.

[37] J.E. Hopcroft, J.P. Ullman. *Introduction to automata theory, language and computation.* Addison-Wesley, 1979.

[38] H.K. Hwang. *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques.* Ph.D. Dissertation, École polytechnique, Palaiseau, France, 1994.

[39] M. R. Jerrum, L. G. Valiant, V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43(2-3):169–188, 1986.

[40] P. Jokinen, E. Ukkonen, Two algorithms for approximate string matching in static texts. *Proceedings MFCS 91*, *Lecture Notes in Computer Science* n.520, Springer-Verlag, 240-248, 1991.

[41] J. Kemeny, J. L. Snell. *Finite Markov chains.* D. Van Nostrand Co., Inc., Princeton, Toronto, New-York, 1960.

[42] W. Kuich, G. Baron. Two papers on automata theory. *Technical Report* n. 253, Institute fur Informationsverarbeitung, Technische Universität Graz und Österreich computer gesellschaft, June 1988.

[43] W. Kuich, A. Salomaa. *Semirings, automata, languages.* Springer–Verlag, New York Heidelberg Berlin Tokyo, 1986.

[44] S. Li. A martingale approach to the study of occurrence of sequences patterns in repeated experiments. *The Annals of Probability*, 8:1171–1176, 1980.

[45] P. Nicodeme, B. Salvy, P. Flajolet. Motif statistics. *Theoretical Computer Science*, 287(2):593–617, 2002.

[46] P. Pezvner, M. Borodovsky, A. Mironov. Linguistic of nucleotide sequences: the significance of deviations from mean. Statistical characteristic and prediction of the frequency of occurrence of words. *Journal of Biomol. Struct. Dynamics* 6:1013-1026, 1991.

[47] B. Prum, F. Rudolphe, E. Turckheim. Finding words with unexpected frequencies in deoxyribonucleic acid sequence. *Journal of Royal Statistic Society*, Serie B, 57:205-220, 1995

[48] A. Bertoni, P. Massazza, R. Radicioni. Random generations of words in regular languages with fixed occurrences of symbols. *Proceedings of Words* 4th International Conference on Combinatorics of Words, Turku, Finland, pages 332-343, 2003.

[49] M. Régnier. A unified approach to word statistics. *Proceedings of the second annual international conference on Computational molecular biology*, ACM Press, 207–213, 1998.

[50] M. Régnier, W. Szpankowski. On pattern frequency occurrences in a Markovian sequence. *Algorithmica* 22:621–649, 1998.

[51] C. Reutenauer. *Propriétés arithmétiques et topologiques de séries rationnelles en variables non commutatives*, These Sc. Maths, Doctorat troisieme cycle, Université Paris VI, 1977.

[52] A. Salomaa, M. Soittola. *Automata-theoretic aspects of formal power series*. Springer-Verlag, 1978.

[53] M.P. Schützenberger. Certain elementary family of automata. *Proceedings of the Symposium on Mathematical Theory of Automata*, 139–153, 1962

[54] E. Seneta. *Non-negative matrices and Markov chains*, Springer–Verlag, New York Heidelberg Berlin, 1981.

[55] I. Simon. The nondeterministic complexity of a finite automata. In *Mots*, M. Lothaire editor, Lecture Notes in Mathematics, Hermes, 384–400, Paris 1990.

[56] E. Sutinen, W. Szpankowski. On the collapse of q-gram filtration. *Proc. Fun with Algorithms*, Elba 1998.

[57] E. Ukkonen. Approximate string-matching with q-grams and maximal matchings. *Theoretical Computer Science*, 92:191–211, 1992.

[58] B. Vallée. Dynamical sources in Information Theory: fundamental intervals and word prefixes. *Algorithmica*, 29:262–306, 2001.

[59] M. Waterman. *Introduction to computational biology*, Chapman & Hall, New York, 1995.

[60] A. Weber. On the valuedness of finite transducers. *Acta Informatica*, 27:749–780, 1990.

[61] A. Weber, H. Seidl. On the degree of ambiguity of finite automata. *Theoretical Computer Science*, 88:325–349, 1991.

[62] K. Wich. Exponential ambiguity of context-free grammars. *Proceedings of the 4th DLT*, G. Rozenberg and W. Thomas editors. World Scientific, Singapore, 1999, 125–138.

[63] K. Wich. Sublinear ambiguity. *Proceedings of the 25th MFCS*, M. Nielsen and B. Rovan editors. *Lecture Notes in Computer Science*, vol. n.1893, Springer-Verlag, 690–698, 2000.

[64] K. Wich. Characterization of Context-free languages with polynomially bounded ambiguity. *Proceedings 26th MFCS*, J. Sgall, A. Pultr and P. Kolman editors. *Lecture Notes in Computer Science*, vol. n.2136, Springer-Verlag, 669–680, 2001.

[65] S. Wolfram. *The Mathematica book*, Fourth Edition, Wolfram Media - Cambridge University Press, 1999.