Frequency of symbol occurrences
in bicomponent stochastic models [*,**,***]

Diego de Falco [a], Massimiliano Goldwurm [a,*], Violetta Lonati [a]

[a]Dipartimento di Scienze dell'Informazione - Università degli Studi di Milano,
Via Comelico 39, 20135 Milano, Italy.

Abstract

We give asymptotic estimates of the frequency of occurrences of a symbol in a random word generated by any bicomponent stochastic model. More precisely, we consider the random variable Y_n representing the number of occurrences of a given symbol in a word of length n generated at random; the stochastic model is defined by a rational formal series r having a linear representation with two primitive components. This model includes the case when r is the product or the sum of two primitive rational formal series. We obtain asymptotic evaluations for the mean value and the variance of Y_n and its limit distribution.

Key words: Automata and formal languages, limit distributions, pattern statistics, rational formal series.

1 Introduction

Estimating the frequency of given patterns in a random text is a classical problem studied in several research areas of computer science and mathematics that has well-known applications in molecular biology [10,15,8,14,17]. Pattern statistics studies this problem in a probabilistic framework: one or more patterns are fixed and a text of length n is randomly generated by a memoryless source (also called *Bernoulli model*) or a Markovian source (the *Markovian model*) where the probability of a symbol in any position only depends on a finite number of previous occurrences [11,15,13,5]. Main goals of the research in this context are the asymptotic expressions of mean value and variance of the number of pattern occurrences in the text and its limit distribution. Several results show a Gaussian limit distribution of these statistics in the sense of the central or local limit theorem [1]. In particular in [13] properties of this kind are obtained for a pattern statistics which represents the number of (positions of) occurrences of words from a regular language in a random string of length n generated in a Bernoulli or a Markovian model.

This approach has been extended in [3,4] to the so-called *rational stochastic model*, where the text is generated at random according to a probability distribution defined by means of a rational formal series in non-commutative variables. In particular cases, this is simply the uniform distribution over the set of words of given length in an arbitrary regular language. We recall that there are well-known linear time algorithms that generate a word at random under such a distribution [6]. The relevance of the rational stochastic model is due to the connection with the classical Markovian random sourses in pattern statistics. This relationship can be stated precisely as follows [3]: the frequency problem of regular patterns in a text generated in the Markovian model (as studied in [13]) is a special case of the frequency problem of a single symbol in a text over a binary alphabet generated in the rational stochastic model; it is also known that the two models are not equivalent.

The symbol frequency problem in the rational model is studied in [3] in the primitive case, i.e. when the matrix associated with the rational formal series (counting the transitions between states) is primitive and hence it has a unique eigenvalue of largest modulus, which is real positive. Under this hypothesis asymptotic expressions for the mean value and the variance of the statistics under investigation are known, together with their limit distributions expressed in the form of both central and local limit theorems [3,4].

In the present paper we study the symbol frequency problem in the *bicomponent rational model*, which is a non-primitive case of the rational model, defined by a formal series that admits a linear representation with two primitive components. In this context there are two special examples of particular

2

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.

We determine 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. The behaviour of this random variable mainly depends on two conditions: whether there exists a communication from the first to the second component and whether one component is dominant, i.e. its main eigenvalue is strictly greater than the main eigenvalue of the other one (if the main eigenvalues are equal we say that the components are equipotent). The analysis of the dominant case splits in two further directions according whether the dominant component is degenerate or not[1]. The equipotent case has several subcases corresponding to the possible differences between the leading terms of the mean values and of the variances of the statistics associated with each component.

Our main results are summarized in a table presented in the last section. It turns out that if one component is dominant and does not degenerate then it determines the main terms of expectation and variance of our statistics, and we get a Gaussian limit distribution. On the contrary, in the dominant degenerate case the limit distribution can assume 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. In the equipotent case, if the leading terms of the mean values (associated with the components) are different, then the overall variance is of a quadratic order showing there is not a concentration phenomenon around the average value of our statistics; in this case the typical situation occurs when there is communication from the first to the second component: here we obtain a uniform limit distribution. On the contrary, when the leading terms of the mean values are equal, we have again a concentration phenomenon with a limit distribution given by a mixture of Gaussian laws, which reduces again to a normal distribution when the local behaviour of our statistics in the two components is asympotically equal.

The main contribution of these results is related to the non-primitive hypothesis. 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.

We think our analysis is significant also from a methodogical point of view:

---

[1] Here, a component is degenerate if all its transitions are labelled by the same symbol.

3

we adapt methods and ideas introduced to deal with the Markovian model to a more general stochastic model, the rational one, which seems to be the natural setting for these techniques.

The material we present is organized as follows. After recalling some preliminaries in Section 2 and the rational stochastic model in Section 3, we revisit the primitive case in Section 4 by using a simple matrix differential calculus. In Section 5 we introduce the bicomponent rational model and then we study the dominant case, i.e. when the main eigenvalue of one component is greater than the main eigenvalue of the other. In Section 7 we consider the equipotent case, when the two main eigenvalues are equal. Finally Section 8 is devoted to the analysis of the sum models while the last section contains the summary and a comparison of the results.

The computations described in our examples are executed by using *Mathematica* [18].

2 Preliminaries

In this section we recall some basic notions and properties concerning nonnegative matrices [16] and probability theory [9].

2.1 Perron–Frobenius theory

The Perron–Frobenius theory is a well-known subject widely studied in the literature (see for instance [16]). To recall its main results we first establish some notation. For every pair of matrices $T = [T_{ij}]$, $S = [S_{ij}]$, the expression $T > S$ means that $T_{ij} > S_{ij}$ for every pair of indices i, j. As usual, we consider any vector v as a column vector and denote by v^T the corresponding row vector. We recall that a nonnegative matrix T is called *primitive* if there exists $m \in \mathbb{N}$ such that $T^m > 0$. The main properties of such matrices are given by the following theorem [16, Sect.1].

Theorem 1 (Perron–Frobenius) *Let T be a primitive nonnegative matrix. There exists an eigenvalue λ of T (called Perron–Frobenius eigenvalue of T) such that: i) λ is real and positive; ii) with λ we can associate strictly positive left and right eigenvectors; iii) $|\nu| < \lambda$ for every eigenvalue $\nu \neq \lambda$; iv) if $0 \leq C \leq T$ and γ is an eigenvalue of C, then $|\gamma| \leq \lambda$, moreover $|\gamma| = \lambda$ implies $C = T$; v) λ is a simple root of the characteristic polynomial of T.*

The following proposition is a first consequence of the theorem above [16, Theorem 1.2].

4

Proposition 2 *If T is a primitive matrix and 1 is its Perron–Frobenius eigenvalue, then*

$$T^n = uv^T + D(n) \cdot \frac{n^s}{h^n}$$

where $s \in \mathbb{N}$, $h > 1$, $D(n)$ is a real matrix such that $|D(n)_{ij}| \leq c$ for all n large enough, every i, j and some constant $c > 0$, while v^T and u are strictly positive left and right eigenvectors of T corresponding to the eigenvalue 1, normed so that $v^T u = 1$.

Moreover, under the same hypotheses, the matrix $D = \sum_{n=0}^{\infty} D(n) n^s / h^n$ is well defined and, by the properties of v and u, satisfies the equality

$$v^T D = Du = 0 \ . \tag{1}$$

2.2 Notations on matrix functions

Assume that $A(x)$ is a square matrix the entries of which are complex functions in the variable x. The derivative of $A(x)$ with respect of x is the matrix $D_x A(x) = [A'(x)_{ij}]$ 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:

$$D_x(A(x) \cdot B(x)) = D_x A(x) \cdot B(x) + A(x) \cdot D_x B(x) \ , \tag{2}$$

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

$$D_x(A(x)^{-1}) = -A(x)^{-1} \cdot D_x A(x) \cdot A(x)^{-1} \ . \tag{3}$$

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 $|A(x)_{ij}| \leq c |g(x)|$, for every x in E near x_0. Thus, if the entries of $A(x)$ are analytic at a point x_0, then $A(x) = A(x_0) + A'(x_0)(x - x_0) + O((x - x_0)^2)$. On the contrary, if some entries of $A(x)$ have a pole of degree 1 at a point x_0, 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 ($R \neq 0$).

5

2.3   Moments and limit distribution of discrete random variables

Let X be a random variable (r.v.) with values in a set $\{x_0, x_1, \ldots, x_k, \ldots\}$ of real numbers and set $p_k = \Pr\{X = x_k\}$, for every $k \in \mathbb{N}$. We denote by F_X its distribution function, i.e. $F_X(\tau) = \Pr\{X \leq \tau\}$ for every $\tau \in \mathbb{R}$. If the set of indices $\{k \mid p_k \neq 0\}$ is finite we can consider the moment generating function of X, given by

$$\Psi_X(z) = \mathbb{E}(e^{zX}) = \sum_{k \in \mathbb{N}} p_k e^{zx_k} \ ,$$

which in our case is well-defined for every $z \in \mathbb{C}$. This function can be used to compute the first two moments of X,

$$\mathbb{E}(X) = \Psi'_X(0) \ , \qquad \mathbb{E}(X^2) = \Psi''_X(0) \ . \tag{4}$$

and yields the characteristic function of X, given by

$$\Phi_X(t) = \mathbb{E}(e^{itX}) = \Psi_X(it) \ .$$

The function $\Phi_X(t)$ (well-defined for every $t \in \mathbb{R}$) completely characterizes the distribution function F_X and represents the classical tool to prove convergence in distribution. We recall that, given a sequence of random variables $\{X_n\}_n$ and a random variable X, X_n converges to X *in distribution* (or *in law*) if $\lim_{n \to \infty} F_{X_n}(\tau) = F_X(\tau)$ for every point $\tau \in \mathbb{R}$ of continuity for F_X. It is well-known that X_n converges to X in distribution if and only if $\Phi_{X_n}(t)$ tends to $\Phi_X(t)$ for every $t \in \mathbb{R}$. Several forms of the central limit theorem are classically proved in this way [9,7].

A convenient approach to prove the convergence in law to a Gaussian random variable relies on the so called "quasi-power" theorems introduced in [12] (see also [7]) and implicitly used in the previous literature [1]. For our purpose it is convenient to recall such a theorem in a simple form (for the proof see [7, Theorem 9.6] or [1, Theorem 1]).

Theorem 3 *Let $\{X_n\}$ be a sequence of random variables, where each X_n takes values in $\{0, 1, \ldots, n\}$ and assume there exist two functions $r(z)$, $u(z)$, both analytic at $z = 0$, where they take the value $r(0) = u(0) = 1$, and two positive constants c, ρ, such that for every $|z| < c$*

$$\Psi_{X_n}(z) = r(z) \cdot u(z)^n + O(\rho^n) \qquad and \qquad \rho < |u(z)|.$$

Also set

$$\mu = u'(0) \qquad and \qquad \sigma = u''(0) - (u'(0))^2$$

and assume $\sigma > 0$ (variability condition). Then $\frac{X_n - \mu n}{\sqrt{\sigma n}}$ converges in distribution to a normal random variable of mean 0 and variance 1, i.e. for every $x \in \mathbb{R}$

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

Finally, we recall that a sequence of random variable $\{X_n\}$ converges *in probability* to a random variable X if, for every $\varepsilon > 0$, $\Pr\{|X_n - X| > \varepsilon\}$ tends to 0 as n goes to $+\infty$. It is well-known that convergence in probability implies convergence in law.

3 The rational stochastic model

The stochastic model we consider in this work is defined by using the notion of linear representation of a rational formal series [2]. Let $\mathbb{R}_+$ be the semiring of non-negative real numbers. We recall that a *formal series* over Σ with coefficients in $\mathbb{R}_+$ is a function $r : \Sigma^* \longrightarrow \mathbb{R}_+$. Usually, the value of r at ω is denoted by (r, ω) and we write $r = \sum_{\omega \in \Sigma^*} (r, \omega) \cdot \omega$. Moreover, r is called *rational* if it admids a *linear representation*, that is a triple (ξ, μ, η) where, for some integer $m > 0$, ξ and η are (column) vectors in $\mathbb{R}_+^m$ and $\mu : \Sigma^* \longrightarrow \mathbb{R}_+^{m \times m}$ is a monoid morphism, such that $(r, \omega) = \xi^T \mu(\omega) \, \eta$ holds for each $\omega \in \Sigma^*$. We say that m is the *size* of the representation. Observe that considering such a triple (ξ, μ, η) is equivalent to defining a (weighted) non–deterministic automaton, where the state set is given by $\{1, 2, \ldots, m\}$ and the transitions, the initial and the final states are assigned weights in $\mathbb{R}_+$ by μ, ξ and η respectively. Note that (ξ, μ, η) represents a deterministic finite automaton when ξ and η are the characteristic arrays of the initial state and the final states, respectively, and for every $\sigma \in \Sigma$ and every $i = 1, 2, \ldots, m$ there exists an index j such that $\mu(\sigma)_{ij} = 1$, while $\mu(\sigma)_{ij'} = 0$ for any $j' \neq j$: in this case r is the characteristic series of the languages recognized by the automaton.

From now on we assume $\Sigma = \{a, b\}$ and set $A = \mu(a)$, $B = \mu(b)$ and $M = A + B$. Thus, for every positive integer n such that $\xi^T M^n \eta \neq 0$, we can define a probability space as follows. Let us define a *computation path* of length n as a string ℓ of the form

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

where $q_j \in \{1, 2, \ldots, m\}$ and $x_i \in \{a, b\}$ for every $j = 0, 1, \ldots, n$ and every $i = 1, 2, \ldots, n$. We denote by Ω_n the set of all computation paths of length n and, for each $\ell \in \Omega_n$, we define the probability of ℓ as

$$\Pr\{\ell\} = \frac{\xi_{q_0} \mu(x_1)_{q_0 q_1} \mu(x_2)_{q_1 q_2} \cdots \mu(x_n)_{q_{n-1} q_n} \eta_{q_n}}{\xi^T M^n \eta} \ .$$

Denoting by $\mathcal{P}(\Omega_n)$ the family of all subsets of Ω_n, it is clear that $\langle \Omega_n, \mathcal{P}(\Omega_n), \Pr \rangle$ is a probability space.

Now, let us consider the random variable $Y_n : \Omega_n \to \{0, 1, \ldots, n\}$ such that $Y_n(\ell)$ is the number of a occurring in ℓ, for each $\ell \in \Omega_n$. It is clear that, for

7

every integer $0 \le k \le n$, we have

$$\Pr\{Y_n = k\} = \frac{\varphi_k^{(n)}}{\sum_{j=0}^{n} \varphi_j^{(n)}}, \qquad \text{where} \qquad \varphi_k^{(n)} = \sum_{|w|=n,|w|_a=k} \xi^T \mu(w)\eta \ .$$

Note that when (ξ, μ, η) represents a deterministic finite automaton, Y_n is the number of occurrences of a in a word randomly chosen under uniform distribution in the set af all strings of length n in the language recognized by the automaton. This observation may suggest that Y_n could be defined over a sample space simpler than Ω_n (a natural candidate would be Σ^n as in [3]). However, the sample space Ω_n is really necessary in our context, as it will be clear in Sections 5 and 8, since we will have to distinguish different paths having the same labelling word.

We remark that classical probabilistic models as the Bernoulli or the Markov processes, frequently used to study the number of occurrences of regular patterns in random words [11,15,13], are special cases of rational stochastic models. The relationship between Markovian processes and rational stochastic models can be formally stated as follows (for the proof see [3, Sec. 2.1]). Given a regular language R over a finite alphabet and a Markovian process Π generating words at random over the same alphabet, let $O_n(R, \Pi)$ denote the number of (positions of) occurrences of elements of R in a word of length n generated by Π. It turns out that for every such R and Π there exists a linear representation (ξ, μ, η) over the alphabet $\{a, b\}$ such that, for every $n \in \mathbb{N}$, the corresponding random variable Y_n has the same probability function as $O_n(R, \Pi)$, i.e. $\Pr\{Y_n = k\} = \Pr\{O_n(R, \Pi) = k\}$ for any $k = 0, 1, \ldots, n$. The opposite inclusion is not true: there are rational stochastic models which cannot be simulated by any Markovian process. This is due to the fact that the generating function of the bivariate sequence $\{\Pr\{O_n(R, \Pi) = k\}\}_{n,k}$ is a rational analytic function for any R, Π, while there exist linear representations (ξ, μ, η) such that the generating function of the corresponding sequence $\{\Pr\{Y_n = k\}\}_{n,k}$ is not algebraic.

To study the asymptotic behaviour of Y_n, one should consider the moment generating function of the random variable Y_n which is defined as

$$\Psi_{Y_n}(z) = \frac{h_n(z)}{h_n(0)} \qquad \text{where} \qquad h_n(z) = \sum_{k=0}^{n} \varphi_k^{(n)} e^{zk} = \xi^T (Ae^z + B)^n \eta \qquad (5)$$

and observe that by (4) we have

$$\mathbb{E}(Y_n) = \frac{h_n'(0)}{h_n(0)} \quad \text{and} \quad \mathbb{V}ar(Y_n) = \frac{h_n''(0)}{h_n(0)} - \left(\frac{h_n'(0)}{h_n(0)}\right)^2 . \qquad (6)$$

In order to study the asymptotic behaviour of $h_n(0)$, $h_n'(0)$ and $h_n''(0)$, it is useful to introduce the bivariate matrix function $H(z, w)$, well-defined in a

8

neighbourhood of $(0,0)$, given by

$$H(z,w) = \sum_{n=0}^{+\infty}(Ae^z + B)^n w^n = [I - w(Ae^z + B)]^{-1}. \tag{7}$$

Denote by H_z and H_{zz} its partial derivatives $\frac{\partial H}{\partial z}$ and $\frac{\partial^2 H}{\partial z^2}$, respectively, and observe that

$$\sum_{n=0}^{+\infty} h_n(z)w^n = \xi^T \, H(z,w) \, \eta. \tag{8}$$

Finally, the characteristic function of the random variable Y_n is given by

$$\Phi_{Y_n}(t) = \mathbb{E}(e^{itY_n}) = \frac{h_n(it)}{h_n(0)} \ .$$

4   The primitive case

The asymptotic behaviour of Y_n is studied in [3] in the case when (ξ, μ, η) is a *primitive* linear representation, i.e. when the matrix $\mu(a) + \mu(b)$ is primitive. In this section, we present some steps of those proofs by using a more general approach. The discussion will be useful in subsequent sections.

As above, let $A = \mu(a)$ and $B = \mu(b)$. Since the matrix $M = A+B$ is primitive we can consider the Perron–Frobenius eigenvalue λ of M and, by Proposition 2, we have

$$M^n = \lambda^n \left(uv^T + C(n) \right) \tag{9}$$

where $C(n)$ is a real matrix such that $C(n) = O(\varepsilon^n)$ for some $0 \le \varepsilon < 1$ and v^T and u are strictly positive left and right eigenvectors of M corresponding to the eigenvalue λ, normed so that $v^T u = 1$. Moreover, we know that the matrix

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

is well-defined and, by (1), $v^T C = Cu = 0$.

Since $A + B$ is primitive, by Perron–Frobenius Theorem, the function $H(0,w)$ defined in (7) has a unique singularity of smallest modulus at $w = 1/\lambda$ which is a simple pole. Thus, by (3), also $H_z(0,w)$ and $H_{zz}(0,w)$ have a unique singularity of smallest modulus at $w = 1/\lambda$. The following lemma gives a more precise analysis.

Lemma 4 *In a neighbourhood of $w = 1/\lambda$, the matrices $H(0,w)$, $H_z(0,w)$ and $H_{zz}(0,w)$ admit a Laurent expansion of the form*

9

$$H(0, w) = \frac{uv^T}{1 - \lambda w} + C + O(1 - \lambda w) \tag{10}$$

$$H_z(0, w) = \frac{\lambda w}{(1 - \lambda w)^2} \cdot \beta uv^T + \frac{D}{1 - \lambda w} + O(1) \tag{11}$$

$$H_{zz}(0, w) = \frac{2\lambda^2 w^2}{(1 - \lambda w)^3} \cdot \beta^2 uv^T + \tag{12}$$

$$+ \frac{\lambda w}{(1 - \lambda w)^2} \cdot (\beta uv^T + 2\beta D + 2uv^T \frac{ACA}{\lambda^2} uv^T) + O\left(\frac{1}{1 - \lambda w}\right)$$

where the matrix D and the constant β are defined by

$$D = \frac{CA}{\lambda} uv^T + uv^T \frac{AC}{\lambda}, \qquad \beta = \frac{v^T Au}{\lambda}. \tag{13}$$

Proof. First observe that relations (7) and (9) imply the following equalities:

$$H(0, w) = \sum_{n=0}^{+\infty} M^n w^n = \sum_{n=0}^{+\infty} (uv^T + C(n)) \lambda^n w^n$$

$$= \sum_{n=0}^{+\infty} uv^T \lambda^n w^n + \sum_{n=0}^{+\infty} C(n) \lambda^n w^n \tag{14}$$

Since each entry of $\sum_n C(n) x^n$ converges uniformly for x near 1 to a rational function, we have

$$\sum_{n=0}^{\infty} C(n) x^n = C + O(1 - x)$$

and hence the second series in (14) equals $C + O(1 - \lambda w)$, which proves (10).

Now observe that from (2) and (3) we get

$$H_z(0, w) = H(0, w) \, Aw \, H(0, w),$$
$$H_{zz}(0, w) = H_z(0, w) \cdot [I + 2Aw \, H(0, w)].$$

Replacing (10) in the previous expressions, one can easily find equations (11) and (12). $\qquad\square$

Theorem 5 *If M is primitive then the mean value and the variance of Y_n satisfy the relations*

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

where $|\varepsilon| < 1$ and β is defined in (13), while α, γ and δ are given by

$$\alpha = (\xi^T u)(v^T \eta), \qquad \gamma = \beta - \beta^2 + 2\frac{v^T ACAu}{\lambda^2}, \qquad \delta = \xi^T D\eta.$$

Proof. By equation (8), from the previous lemma it is easy to prove that

$$
\begin{aligned}
h_n(0) &= \lambda^n \cdot \alpha + O(\rho^n) \\
h_n'(0) &= n\lambda^n \cdot \alpha\beta + \lambda^n\delta + O(\rho^n) \\
h_n''(0) &= n^2\lambda^n \cdot \alpha\beta^2 + n\lambda^n \cdot \left(\alpha\beta - \alpha\beta^2 + 2\beta\delta + 2\alpha\frac{v^T ACAu}{\lambda^2}\right) + O(\lambda^n)
\end{aligned}
\tag{15}
$$

where $|\rho| < \lambda$ gives the contribution of smaller eigenvalues of M. Then, the result follows from (6). $\qquad\square$

Notice that $B = 0$ implies $\beta = 1$ and $\gamma = \delta = 0$, while $A = 0$ implies $\beta = \gamma = \delta = 0$; on the contrary, if $A \neq 0 \neq B$ then clearly $0 < \beta < 1$ and one can prove that also $0 < \gamma$ [3].

In [3] it is proved that Y_n converges in law to a Gaussian random variable, when M is primitive and $A \neq 0 \neq B$. The proof is based on Theorem 3. To see its main steps, consider the generating function of $h_n(z)$, given by

$$
\xi^T H(z,w)\eta = \frac{\xi^T \mathrm{Adj}\,(I - w(Ae^z + B))\,\eta}{\det\,(I - w(Ae^z + B))} \ .
$$

Since $A + B$ is primitive, its Perron-Frobenius eigenvalue λ is a simple root of $\det(yI - A - B)$. Thus the equation

$$
\det\,(yI - Ae^z - B) = 0
$$

defines an implicit function $y = y(z)$ analytic in a neighbourhood of $z = 0$ such that $y(0) = \lambda$ and $y'(0) \neq 0$.

A further property of primitive matrices (see for instance [16, page 7]) states that $\mathrm{Adj}(\lambda I - A - B) > 0$ and hence, by continuity, all entries of $H(z,w)$ are different from 0 for every z near 0 and every w near λ^{-1}. These properties allow us to prove the following proposition [3].

Proposition 6 *For every z near 0, as n tends to infinity we have*

$$
h_n(z) = \xi^T R(z)\eta \cdot y(z)^n + O(\rho^n),
$$

where $\rho < |y(z)|$ and $R(z)$ is a matrix function given by

$$
R(z) = -\frac{y(z) \cdot Adj\left(I - y(z)^{-1}(Ae^z + B)\right)}{\frac{\partial}{\partial w}\det\left(I - w(Ae^z + B)\right)|_{w = y(z)^{-1}}} \ .
$$

Note that any entry of $R(z)$ is analytic and non-null at $z = 0$. Moreover, from the previous result one can also express the moments of Y_n as function of $y(z)$, obtaining

$$
\beta = \frac{y'(0)}{\lambda} \ , \quad \gamma = \frac{y''(0)}{\lambda} - \left(\frac{y'(0)}{\lambda}\right)^2 \ .
\tag{16}
$$

Since in our case $\gamma > 0$, we can apply Theorem 3 which implies the following

Theorem 7 *If M is primitive and $A \neq 0 \neq B$, then $\frac{Y_n - \beta n}{\sqrt{\gamma n}}$ converges in distribution to a normal random variable of mean 0 and variance 1.*

We conclude this section observing that $\xi^T H(z,w)\eta$ is the generating function of $\{h_n(z)\}$ and hence, by Proposition 6, for every z near 0 we have

$$H(z,w) = \frac{R(z)}{1 - y(z)w} + O(1), \qquad \text{as } w \to y(z)^{-1} . \qquad (17)$$

5 The bicomponent model

Here we consider a linear representation (ξ, μ, η) where the matrix $\mu(a) + \mu(b)$ consists of two primitive components. More formally, we consider a triple (ξ, μ, η) such 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:

$$\xi^T = (\xi_1^T, \xi_2^T), \qquad \mu(x) = \begin{pmatrix} \mu_1(x) & \mu_0(x) \\ 0 & \mu_2(x) \end{pmatrix}, \qquad \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \qquad (18)$$

where $\mu_0(x) \in \mathbb{R}_+^{s \times t}$ for every $x \in \{a, b\}$. In the sequel, we say that (ξ, μ, η) is a *bicomponent* linear representation.

For sake of brevity we use the notations $A_j = \mu_j(a)$, $B_j = \mu_j(b)$ and $M_j = A_j + B_j$ for $j = 0, 1, 2$. Hence, we have

$$A = \mu(a) = \begin{pmatrix} A_1 & A_0 \\ 0 & A_2 \end{pmatrix}, \quad B = \mu(b) = \begin{pmatrix} B_1 & B_0 \\ 0 & B_2 \end{pmatrix}, \quad M = A + B = \begin{pmatrix} M_1 & M_0 \\ 0 & M_2 \end{pmatrix}.$$

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.

To avoid trivial cases, throughout this work we assume $\xi_1 \neq 0 \neq \eta_2$ together with the following *significance* hypothesis:

$$(A_1 \neq 0 \text{ or } A_2 \neq 0) \text{ and } (B_1 \neq 0 \text{ or } B_2 \neq 0) . \qquad (19)$$

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).

In our automaton, a computation path $\ell = q_0 x_1 q_1 x_2 q_2 \cdots q_{n-1} x_n q_n$ 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 $0 \leq s < n$ such that the indices $q_0, q_1, \ldots, q_s$ are in the first component while $q_{s+1}, \ldots, q_n$ 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).

Using the notation introduced in the previous section, from now on we refer the values $h_n(z)$ and $H(z, w)$ to the triple (ξ, μ, η). We also 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 $j = 1, 2$, the values λ_j, C_j, D_j, $h_n^{(j)}(z)$, $H^{(j)}(z, w)$, u_j, v_j, α_j, β_j, γ_j, δ_j, $y_j(z)$ and $R_j(z)$ are well-defined and associated with the linear representation (ξ_j, μ_j, η_j).

Now consider the matrix $H(z, w)$. To express its value as a function of $H^{(1)}(z, w)$ and $H^{(2)}(z, w)$, we use the following identities, which can be proved by induction. For any matrices P, Q, S of suitable sizes, we have

$$\begin{pmatrix} P & Q \\ 0 & S \end{pmatrix}^n = \begin{pmatrix} P^n & \sum_{i=0}^{n-1} P^i Q S^{n-1-i} \\ 0 & S^n \end{pmatrix} ;$$

moreover, also in the case of matrices, for any pair of sequences $\{p_n\}, \{s_n\}$ and any fixed q, we have

$$\sum_{n=0}^{\infty} \left(\sum_{i=0}^{n-1} p_i q s_{n-1-i} \right) w^n = \left(\sum_{n=0}^{\infty} p_n w^n \right) qw \left(\sum_{n=0}^{\infty} s_n w^n \right) .$$

Then, a simple decomposition of $H(z, w)$ follows from the previous equations:

$$H(z, w) = \sum_{n=0}^{+\infty} (Ae^z + B)^n w^n = \begin{bmatrix} H^{(1)}(z, w) & G(z, w) \\ 0 & H^{(2)}(z, w) \end{bmatrix}$$

where

$$G(z, w) = H^{(1)}(z, w) \, (A_0 e^z + B_0) w \, H^{(2)}(z, w) . \tag{20}$$

Thus the function $h_n(z)$ defined in (5) now satisfies the equality

$$\sum_{n=0}^{\infty} h_n(z) w^n = \xi^T H(z, w) \eta = \xi_1^T H^{(1)}(z, w) \eta_1 + \xi_1^T G(z, w) \eta_2 + \xi_2^T H^{(2)}(z, w) \eta_2$$

and setting $\sum_n g_n(z)w^n = \xi_1^T G(z,w)\eta_2$ we obtain

$$h_n(z) = h_n^{(1)}(z) + g_n(z) + h_n^{(2)}(z) . \tag{21}$$

The bicomponent model includes two special cases which 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 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 (18) and such that $M_0 = 0$. As a consequence, the computation paths of type 2 cannot occur and hence

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

Example 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

$$\xi^T = (\pi_1^T, 0), \qquad \mu(x) = \begin{pmatrix} \nu_1(x) \tau_1 \pi_2^T \nu_2(x) \\ 0 \qquad \nu_2(x) \end{pmatrix}, \qquad \eta = \begin{pmatrix} \tau_1 \pi_2^T \tau_2 \\ \tau_2 \end{pmatrix} . \tag{22}$$

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

$$h_n(z) = \sum_{i=0}^{n} \xi_1^T (A_1 e^z + B_1)^i \tau_1 \pi_2^T (A_2 e^z + B_2)^{n-i} \eta_2 .$$

Now let us go back to the general case: we need an asymptotic evaluation of h_n and H. To this end, since M_1 and M_2 are primitive, we can first apply equations (15) to $h_n^{(1)}(0)$ and $h_n^{(2)}(0)$ obtaining asymptotic evaluations for them and their derivatives. As far as $g_n(0)$ and its derivatives are concerned, we have to compute the derivatives of $G(z, w)$ with respect to z, using equations (2) and (3):

$$G_z(z,w) = H_z^{(1)}(z,w) \cdot (A_0 e^z + B_0)w \cdot H^{(2)}(z,w) + H^{(1)}(z,w) \cdot A_0 e^z w \cdot H^{(2)}(z,w) +$$
$$+ H^{(1)}(z,w) \cdot (A_0 e^z + B_0)w \cdot H_z^{(2)}(z,w), \tag{23}$$

$$G_{zz}(z,w) = H_{zz}^{(1)}(z,w) \cdot (A_0 e^z + B_0)w \cdot H^{(2)}(z,w) + 2H_z^{(1)}(z,w) \cdot A_0 e^z w \cdot H^{(2)}(z,w) +$$
$$+ 2H_z^{(1)}(z,w) \cdot (A_0 e^z + B_0)w \cdot H_z^{(2)}(z,w) + H^{(1)}(z,w) \cdot A_0 e^z w \cdot H^{(2)}(z,w) +$$
$$+ 2H^{(1)}(z,w) \cdot A_0 e^z w \cdot H_z^{(2)}(z,w) + H^{(1)}(z,w) \cdot (A_0 e^z + B_0)w \cdot H_{zz}^{(2)}(z,w). \tag{24}$$

We shall see that the properties of Y_n depend on whether the Perron-Frobenius eigenvalues λ_1, λ_2 of M_1 and M_2 are distinct or equal. In the first case the rational representation associated with the largest one determines the main characteristics of Y_n. We say that (ξ_i, μ_i, η_i) is the *dominant* component if $\lambda_1 \neq \lambda_2$ and $\lambda_i = \max\{\lambda_1, \lambda_2\}$; we study this case in the next section. On the contrary, if $\lambda_1 = \lambda_2$ we say that the components are *equipotent* and they both give a contribution to the asymptotic behaviour of Y_n. This case is considered in Section 7.

6 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). We also assume $M_0 \neq 0$ since the case $M_0 = 0$, corresponding to Example 1, is treated in Section 8. We first determine asymptotic expressions for mean value and variance of Y_n and then we study its limit distribution.

6.1 Analysis of moments in the dominant case

To study the first two moments of Y_n we develop a singularity analysis for the functions $H(0, w)$, $H_z(0, w)$ and $H_{zz}(0, w)$, which yields asymptotic expressions for $h_n(0)$, $h'_n(0)$ and $h''_n(0)$. In the following analysis a key role is played by the matrix Q defined by

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

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

$$H_w^{(2)}(0, \lambda_1^{-1}) = \lambda_1^2 \cdot Q M_2 Q \qquad \text{and} \qquad H_z^{(2)}(0, \lambda_1^{-1}) = \lambda_1 \cdot Q A_2 Q \ .$$

First of all we can apply Lemma 4 to $H^{(1)}(0, w)$ and $H^{(2)}(0, w)$ and their partial derivatives. Moreover we need asymptotic expression for G and its derivatives. Since $\lambda_1 > \lambda_2$, by using (20) and applying (10) to $H^{(1)}(0, w)$, as w tends to λ_1^{-1}, we get

$$G(0, w) = \left(\frac{u_1 v_1^T}{1 - \lambda_1 w} + C_1 \right) M_0 \left(\frac{1}{\lambda_1} - \frac{1 - \lambda_1 w}{\lambda_1} \right) \left(H^{(2)}(0, \lambda_1^{-1}) + H_w^{(2)}(0, \lambda_1^{-1})(w - \lambda_1^{-1}) \right) +$$

$$+ \ O(1 - \lambda_1 w) \ = \ \frac{1}{1 - \lambda_1 w} \cdot u_1 v_1^T M_0 Q + O(1) \ . \tag{25}$$

In a similar way one can prove that in a neighbourhood of $w = 1/\lambda_1$, the matrices $G_z(0, w)$ and $G_{zz}(0, w)$ admit a Laurent expansion of the form

15

$$G_z(0,w) = \frac{\lambda_1 w}{(1-\lambda_1 w)^2} \cdot \beta_1 u_1 v_1^T M_0 \, Q \; +$$

$$+ \frac{1}{1-\lambda_1 w} \cdot \left[D_1 M_0 Q + u_1 v_1^T (A_0 - \beta_1 M_0) Q + u_1 v_1^T M_0 Q (A_2 - \beta_1 M_2) Q \right] \; + \; \mathrm{O}(1) \quad (26)$$

$$G_{zz}(0,w) = \frac{2\lambda_1^2 w^2}{(1-\lambda_1 w)^3} \cdot \beta_1^2 u_1 v_1^T M_0 Q \; +$$

$$+ \frac{\lambda_1 w}{(1-\lambda_1 w)^2} \cdot \left\{ 2\beta_1 \cdot [u_1 v_1^T (A_0 + M_0 Q A_2) + D_1 M_0] Q - 2\beta_1^2 u_1 v_1^T M_0 (I + Q M_2) Q \right\} +$$

$$+ \frac{\lambda_1 w}{(1-\lambda_1 w)^2} \cdot u_1 \left(\beta_1 + 2 v_1^T \frac{A_1 C_1 A_1}{\lambda_1^2} u_1 \right) v_1^T M_0 \, Q + \mathrm{O}\left(\frac{1}{1-\lambda_1 w} \right) \quad (27)$$

Proposition 8 *If $\lambda_1 > \lambda_2$ then the mean value and variance of Y_n satisfy the following relations:*

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

Proof. By applying elementary identities, the previous expansions yield asymptotic expressions for $g_n(0)$ and its derivatives, which by (21) lead to the following relations:

$$h_n(0) = \lambda_1^n \cdot (\xi_1^T u_1) \cdot v_1^T (\eta_1 + M_0 Q \eta_2) \; + \; O(\rho^n) \, ,$$

$$h_n'(0) = n \lambda_1^n \cdot \beta_1 \, (\xi_1^T u_1) \cdot v_1^T (\eta_1 + M_0 \, Q \, \eta_2) \; +$$

$$+ \lambda_1^n \cdot (\xi_1^T u_1) \cdot v_1^T (A_0 + M_0 \, Q A_2) \, Q \, \eta_2 \; + \; \lambda_1^n \cdot \xi_1^T D_1 (\eta_1 + M_0 \, Q \, \eta_2) \; +$$

$$- \lambda_1^n \cdot \beta_1 (\xi_1^T u_1) \cdot v_1^T M_0 (I + Q M_2) Q \eta_2 \; + \; O(\rho^n) \, ,$$

$$h_n''(0) = n^2 \lambda_1^n \cdot \beta_1^2 \, (\xi_1^T u_1) \cdot v_1^T (\eta_1 + M_0 \, Q \eta_2) \; +$$

$$+ n \lambda_1^n \cdot 2\beta_1 \, \left[\, (\xi_1^T u_1) \cdot v_1^T (A_0 + M_0 \, Q A_2) \, Q \, \eta_2 + \xi_1^T D_1 \cdot (\eta_1 + M_0 \, Q \eta_2) \, \right] \; +$$

$$- n \lambda_1^n \cdot \left[2\beta_1 (\xi_1^T u_1) \cdot v_1^T M_0 (I + Q M_2) Q \eta_2 \right] \; +$$

$$+ n \lambda_1^n \cdot \left(\beta_1 - \beta_1^2 + 2 \, v_1^T \frac{A_1 C_1 A_1}{\lambda_1^2} u_1 \right) \cdot (\xi_1^T u_1) \cdot v_1^T (\eta_1 + M_0 \, Q \eta_2) \; + \; O(\lambda_1^n) \, ,$$

where $|\rho| < \lambda_1$. Then, the result follows from (6). $\qquad\square$

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

$$\mathbb{E}(Y_n) \; = \; n \; - \; E \; + \; O(\varepsilon^n) \, , \quad \text{where } \; E = \frac{v_1^T (B_0 + M_0 Q B_2) Q \eta_2}{v_1^T (\eta_1 + M_0 Q \eta_2)} \; \text{and} \; |\varepsilon| < 1. \tag{28}$$

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

$$\mathbb{E}(Y_n) \; = \; E' \; + \; O(\varepsilon^n) \, , \quad \text{where } \; E' = \frac{v_1^T (A_0 + M_0 Q A_2) Q \eta_2}{v_1^T (\eta_1 + M_0 Q \eta_2)} \quad (|\varepsilon| < 1). \tag{29}$$

Note that both E and E' are stricly positive since $Q > 0$.

Now the problem is to determine conditions that guarantee $\gamma_1 \neq 0$.

6.2 Variability conditions in the dominant case

To answer the previous questions we first recall that, by Theorem 3 in [3], $A_1 \neq 0 \neq B_1$ implies $\gamma_1 \neq 0$. Thus, by Proposition (8), we know that if $\lambda_1 > \lambda_2$ and $A_1 \neq 0 \neq B_1$ then $\mathbb{V}ar(Y_n) = \gamma_1 n + O(1)$ 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. Given a non-null polynomial $p(x) = \sum_k p_k x^k$, where $p_k \geq 0$ for each k, consider the random variable X_p such that $\Pr\{X_p = k\} = \frac{p_k}{p(1)}$. Let $V(p)$ be the variance of X_p and set $V(0) = 0$. Then

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

Moreover, in [3, Theorem 3] it is proved that for any pair of non-null polynomials p, q with positive coefficients, we have

$$V(pq) = V(p) + V(q) , \qquad V(p+q) \geq \frac{p(1)}{p(1) + q(1)} V(p) + \frac{q(1)}{p(1) + q(1)} V(q) . \quad (30)$$

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

A similar approach holds for matrices. Consider 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

$$\overline{V}(M(x) \cdot N(x)) \geq \left[\sum_k \frac{M(1)_{ik} N(1)_{kj}}{M(1)N(1)_{ij}} \{V(M(x)_{ik}) + V(N(x)_{kj})\}\right] . \quad (31)$$

Finally, from Theorem 3 in [3] one can also deduce that, for every primitive matrix $M = A + B$, if $A \neq 0 \neq B$ then

$$\overline{V}(Ax + B)^n_{ij} = \Theta(n) \quad (32)$$

for any pair of indeces i, j.[2]

---

[2] In this work we use the symbol Θ to represent the order of growth of sequences: given two sequences $\{a_n\} \subseteq \mathbb{C}$ and $\{b_n\} \subseteq \mathbb{R}_+$, the relation $a_n = \Theta(b_n)$ means that $c_1 b_n \leq |a_n| \leq c_2 b_n$, for two positive constants c_1 and c_2 and all n large enough.

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

Proposition 9 *If $M_0 \neq 0$, $\lambda_1 > \lambda_2$ 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 8 can be refined as

$$\mathbb{V}ar(Y_n) = \gamma_1 n + c + O(\varepsilon^n) \tag{33}$$

where c is a constant and $|\varepsilon| < 1$. In order to prove it note that the sequences $h_n(0), h_n'(0), h_n''(0)$ have a 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 of the form $c_1 \lambda_1^n + O(\rho^n)$, $b_2 n \lambda^n + c_2 \lambda_1^n + O(\rho^n)$, $a_3 n^2 \lambda^n + b_3 n \lambda^n + c_3 \lambda_1^n + O(\rho^n)$, respectively, for some constants a_i, b_i, c_i and $|\rho| < 1$; thus, equation (33) follows by replacing these expressions in (6) and taking into account Proposition 8.

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

$$\mathbb{V}ar(Y_n) = V\left(\xi_1^T A_1^n \eta_1 x^n + \xi_1^T P_n(x)\eta_2 + \xi_2^T (A_2 x + B_2)^n \eta_2\right)$$

where

$$P_n(x) = \sum_{i=0}^{n-1} A_1^i x^i (A_0 x + B_0)(A_2 x + B_2)^{n-1-i} \; ;$$

hence, by equation (30),

$$\mathbb{V}ar(Y_n) \geq \frac{\xi_2^T M_2^n \eta_2}{\xi^T M^n \eta}(\gamma_2 n + O(1)) \; + \; \frac{\xi_1^T \sum_{i=0}^{n-1} A_1^i M_0 M_2^{n-1-i} \eta_2}{\xi^T M^n \eta} \; V(\xi_1^T P_n(x)\eta_2) \; . \tag{34}$$

Now, applying equations (30) and (31), we get

$$V(\xi_1^T P_n(x)\eta_2) \geq \min_{(j,k)\in I} \left\{ \sum_{i=0}^{n-1} \frac{(A_1^i M_0 M_2^{n-1-i})_{jk}}{\left(\sum_{s=0}^{n-1} A_1^s M_0 M_2^{n-1-s}\right)_{jk}} \; \left(\overline{V}(A_2 x + B_2)^{n-1-i}\right)_{jk} \right\}$$

where $I = \{(j,k) : \xi_{1_j} P_n(x)_{jk} \eta_{2_k} \neq 0\}$. Replacing this value in (34), by relation (32) 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

$$\Pr\{Y_n = n\} = \frac{\xi_1^T M_1^n \eta_1 + \xi_1^T M_1^{n-1} A_0 \eta_2}{\xi^T M^n \eta} = \Theta(1) \ .$$

Moreover, equation (28) implies $\mathbb{E}(Y_n) = n - E + O(\varepsilon^n)$, where $E > 0$, and hence

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

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 (29) we can argue as in the case $B_1 = 0$ with $A_2 = 0$. $\qquad\square$

6.3   Limit distribution in the dominant case

Now we study the limit distribution of $\{Y_n\}$ in the case $\lambda_1 > \lambda_2$ still assuming $M_0 \neq 0$. If the dominant component does not degenerate we obtain a Gaussian limit distribution as in the primitive case [3]. On the contrary, if the dominant component degenerates we obtain a limit distribution that may assume a large variety of forms, mainly depending on the dominated component. In both cases the proof is based on the analysis of the characteristic function of Y_n, that is $h_n(it)/h_n(0)$.

Recalling that $h_n(z) = h_n^{(1)}(z) + g_n(z) + h_n^{(2)}(z)$, we can apply Proposition 6 to $h_n^{(i)}(z)$ for $i = 1, 2$, and we need an analogous result for $g_n(z)$. First consider the generating function of $\{g_n(z)\}$ that is

$$\xi_1^T G(z,w)\eta_2 \ = \ \sum g_n(z) w^n = \xi_1^T H^{(1)}(z,w) \, (A_0 e^z + B_0) w \, H^{(2)}(z,w)\eta_2 \ .$$

By applying equation (17) to $H^{(1)}$, since $\lambda_1 > \lambda_2$, for every z near 0, we get

$$\xi_1^T G(z,w)\eta_2 = \frac{\xi_1^T R_1(z) \, (A_0 e^z + B_0) \, y_1(z)^{-1} \, H^{(2)}(z, y_1(z)^{-1})\eta_2}{1 - y_1(z)w} + O(1)$$

as w tends to $y_1(z)^{-1}$. The contribution of both $h_n^{(1)}$ and g_n yields a quasi-power condition for Y_n.

Proposition 10 *If $M_0 \neq 0$ and $\lambda_1 > \lambda_2$, then for every z near 0, as n tends to infinity we have*

$$h_n(z) = s(z) \, y_1(z)^n + O(\rho^n),$$

where $\rho < |y_1(z)|$ and $s(z)$ is a rational function given by

$$s(z) = \xi_1^T R_1(z) \Big\{ \eta_1 + (A_0 e^z + B_0) \, y_1(z)^{-1} \, H^{(2)}(z, y_1(z)^{-1})\eta_2 \Big\} \ .$$

Observe that the function $s(z)$ is analytic and non-null at $z = 0$.

Then, if $A_1 \neq 0 \neq B_1$ then $\beta_1 > 0$, $\gamma_1 > 0$ and by the previous proposition we can apply Theorem 3 which yields the following

Theorem 11 *If $M_0 \neq 0$, $\lambda_1 > \lambda_2$ 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 random variable of mean 0 and variance 1.*

On the other hand, if either $A_1 = 0$ or $B_1 = 0$ then $\gamma_1 = 0$ and Theorem 3 cannot be applied. Thus, we study two cases separately, dealing directly with the characteristic function of $\{Y_n\}$. First, let $B_1 = 0$ and set $Z_n = n - Y_n$. We have

$$h_n^{(1)}(z) = \xi_1^T (M_1 e^z)^n \eta_1 = (\lambda_1 e^z)^n \xi_1^T (u_1 v_1^T + C_1(n))\eta_1 \ ,$$

$$g_n(z) = \sum_{j=0}^{n-1} (\lambda_1 e^z)^j \xi_1^T (u_1 v_1^T + C_1(n))^j (A_0 e^z + B_0)(A_2 e^z + B_2)^{n-1-j} \eta_2 \ ,$$

$$h_n^{(2)}(z) = \xi_2^T (A_2 e^z + B_2)^n \eta_2 \ .$$

Hence the characteristic function of Z_n can be computed by replacing the previous values in $\mathbb{E}(e^{zZ_n}) = e^{zn} h_n(z)/h_n(0)$. A simple computation shows that, as n goes to $+\infty$, for every $t \in \mathbb{R}$ we have

$$\mathbb{E}(e^{itZ_n}) \ = \ \frac{v_1^T \eta_1 + v_1^T (A_0 + B_0 e^{it})(\lambda_1 I - A_2 - B_2 e^{it})^{-1} \eta_2}{v_1^T (\eta_1 + M_0 Q \eta_2)} \ + \ \mathrm{o}(1) \ .$$

Note that by (19) 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 12 *Let $M_0 \neq 0$ and $\lambda_1 > \lambda_2$. If $B_1 = 0$ then $n - Y_n$ converges in distribution to a random variable W of characteristic function*

$$\Phi_W(t) = \frac{v_1^T \eta_1 + v_1^T (A_0 + B_0 e^{it})(\lambda_1 I - A_2 - B_2 e^{it})^{-1} \eta_2}{v_1^T (\eta_1 + M_0 Q \eta_2)} \ .$$

If $A_1 = 0$, then Y_n converges in distribution to a random variable Z of characteristic function

$$\Phi_Z(t) = \frac{v_1^T \eta_1 + v_1^T (A_0 e^{it} + B_0)(\lambda_1 I - A_2 e^{it} - B_2)^{-1} \eta_2}{v_1^T (\eta_1 + M_0 Q \eta_2)} \ . \tag{35}$$

Now, let us discuss the form of the random variables 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

$$p_r = \frac{M_0 (\lambda_1 - \lambda_2)^{-1} \eta_2}{\eta_1 + M_0 (\lambda_1 - \lambda_2)^{-1} \eta_2} \qquad \text{and} \qquad p_s = B_0/M_0 \ ,$$

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 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 random variables 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 (35); in the product model it can be expressed in the form

$$\Phi_Z(t) = \frac{\pi_2^T(\lambda_1 I - A_2 e^{it} - B_2)^{-1}\tau_2}{\pi_2^T(\lambda_1 I - M_2)^{-1}\tau_2} = \sum_{j=0}^{\infty} \frac{\pi_2^T (M_2/\lambda_2)^j \tau_2 \cdot (\lambda_2/\lambda_1)^j}{\sum_{i=0}^{\infty} \pi_2^T (M_2/\lambda_2)^i \tau_2 \cdot (\lambda_2/\lambda_1)^i} \, \Phi_{Y_j^{(2)}}(t)$$

where π_2 and τ_2 are defined as in Example 2. This characteristic function actually describes the random variable $Y_N^{(2)}$, where N is the random variable with probability law

$$\Pr\{N = j\} = \frac{\pi_2^T (M_2/\lambda_2)^j \tau_2 \cdot (\lambda_2/\lambda_1)^j}{\sum_{i=0}^{\infty} \pi_2^T (M_2/\lambda_2)^i \tau_2 \cdot (\lambda_2/\lambda_1)^i} \; . \tag{36}$$

If $B_2 = 0$ then by (35) 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 with eigenvalues 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 1 we plot the probability law of N defined in (36) 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.
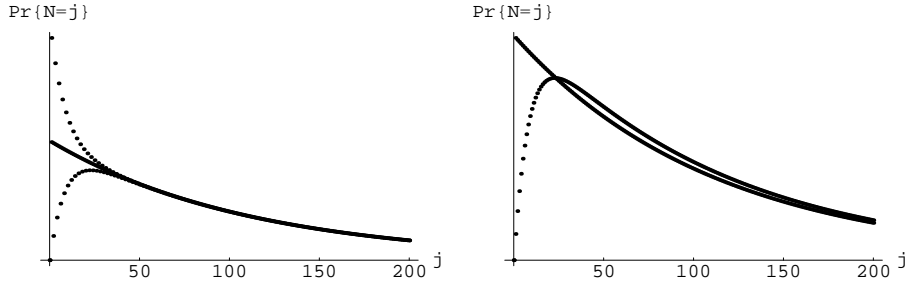
Fig. 1. Probability law of the random variable N defined in (36), for $j = 0, 1, \ldots, 200$. In the first picture we compare the case $\mu = 0.00001$ and $\mu = -0.89$. In the second one we compare the case $\mu = 0.00001$ and $\mu = +0.89$.

7   Equipotent components

Now, we study the behaviour of Y_n in the case $\lambda_1 = \lambda_2$, still assuming $M_0 \neq 0$. Under these hypotheses two main subcases arise. They are determined by the asymptotic mean values associated with each component, namely the constants β_1 and β_2. If they are different the variance of Y_n is of the order $\Theta(n^2)$ and Y_n itself converges in distribution to a uniform random variable. On the contrary, when $\beta_1 = \beta_2$ the order of growth of the variance reduces to $\Theta(n)$ and hence the asymptotic behaviour of Y_n is again concentrated around its expected value. As before we first study the asymptotic behaviour of the moments of Y_n and then we determine the limit distributions.

7.1   Analysis of moments in the equipotent case

For sake of brevity let $\lambda = \lambda_1 = \lambda_2$. As in the dominant case, to study the first two moments of Y_n we can apply equations (15) to get asymptotic evaluations for $h_n^{(1)}(0)$, $h_n^{(2)}(0)$ and their derivatives. We need an analogous result concerning the function $g_n(0)$. In this case, since $M_0 \neq 0$, $G(0, w)$ has a pole of degree 2 in λ^{-1} and then it gives the main contribution to $h_n(0)$.

Proposition 13 *Assume $\lambda_1 = \lambda_2 = \lambda$ and let $M_0 \neq 0$. 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. We argue as in the proof of Proposition 8. For this reason we avoid many details and give a simple outline of the proof. First consider the case $\beta_1 \neq \beta_2$. From relations (15) one gets the asymptotic expressions of $h_n^{(1)}(0)$, $h_n^{(2)}(0)$ and corresponding derivatives. In order to evaluate $g_n(0)$, $g_n'(0)$ and $g_n''(0)$, one can proceed as in the dominant case: use equations (20), (23) and

22

(24) and apply Lemma 4 to $H^{(1)}(0,w)$ and $H^{(2)}(0,w)$. It turns out that, in a neighbourhood of $w = 1/\lambda$, the matrices $G(0,w)$, $G_z(0,w)$ and $G_{zz}(0,w)$ admit a Laurent expansion of degree 2, 3 and 4, respectively. This leads to asymptotic expressions for $g_n(0)$, $g'_n(0)$ and $g''_n(0)$, which can be used together with (21) to get the following expressions:

$$h_n(0) = n\lambda^n \cdot \xi_1^T u_1 \ v_1^T \frac{M_0}{\lambda} \ u_2 \ v_2^T \eta_2 + O(\lambda^n) \ ,$$

$$h'_n(0) = n^2\lambda^n \cdot \frac{\beta_1 + \beta_2}{2} \cdot \xi_1^T u_1 \ v_1^T \frac{M_0}{\lambda} \ u_2 \ v_2^T \eta_2 + O(n\lambda^n) \ ,$$

$$h''_n(0) = n^3\lambda^n \cdot \frac{\beta_1^2 + \beta_1\beta_2 + \beta_2^2}{3} \cdot \xi_1^T u_1 \ v_1^T \frac{M_0}{\lambda} \ u_2 \ v_2^T \eta_2 + O(n^2\lambda^n) \ .$$

Point 1 now follows by applying (6).

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 $G(0,w)$, $G_z(0,w)$ and $G_{zz}(0,w)$ and recalling that $\beta C_1 = \beta C_2 = 0$. The overall computation leads to the following relations:

$$\mathbb{E}(Y_n) = n \cdot \beta + \left\{ \frac{v_1^T M_0 D_2 \eta_2}{v_1^T M_0 u_2 v_2^T \eta_2} + \frac{\xi_1^T D_1 M_0 u_2}{\xi_1^T u_1 v_1^T M_0 u_2} + \frac{v_1^T A_0 u_2}{v_1^T M_0 u_2} - \beta \right\} + O(\varepsilon^n) \ ,$$

$$\mathbb{V}ar(Y_n) = n \cdot \left(\beta - \beta^2 + v_2^T \frac{A_2 C_2 A_2}{\lambda^2} u_2 + v_1^T \frac{A_1 C_1 A_1}{\lambda^2} u_1 \right) + O(1) = \frac{\gamma_1 + \gamma_2}{2} \ n + O(1) \ .$$

Finally observe that, since $\beta_1 = \beta_2$ equation (19) implies $A_i \neq 0 \neq B_i$ for each $i = 1,2$ and hence also $\gamma_i \neq 0$, which proves point 2. $\qquad\square$

7.2 Limit distribution in the equipotent case

To study the limit distribution in the equipotent case ($\lambda_1 = \lambda_2 = \lambda$) with the assumption $M_0 \neq 0$, we consider again the characteristic function of Y_n, that is $h_n(it)/h_n(0)$. In this case, we do not obtain a quasi-power Theorem, since the contribution of $g_n(z)$ to the behaviour of $h_n(z)$ has a different form. In fact, consider the generating function

$$\xi_1^T G(z,w)\eta_2 = \sum g_n(z)w^n = \xi_1^T H^{(1)}(z,w) \ (A_0 e^z + B_0)w \ H^{(2)}(z,w)\eta_2 \ .$$

We study its behaviour for z near 0 and w near λ^{-1}. To this end, first define the analytic function

$$s(z) = \xi_1^T R_1(z)(A_0 e^z + B_0)R_2(z)\eta_2 \tag{37}$$

and observe that $s(0) \neq 0$. Then apply equation (17) to $H^{(1)}$ and $H^{(2)}$. Since $\lambda_1 = \lambda_2 = \lambda$, for every z near 0 we get

23

$$\xi_1^T G(z,w)\eta_2 = \frac{s(z)w}{(1-y_1(z)w)(1-y_2(z)w)} + O\left(\frac{1}{1-y_1(z)w}\right) + O\left(\frac{1}{1-y_2(z)w}\right) + O(1) \qquad (38)$$

$$= s(z)\sum_{n=1}^{\infty}\sum_{k=0}^{n-1} y_1(z)^k y_2(z)^{n-1-k}w^n + O\left(\frac{1}{1-y_1(z)w}\right) + O\left(\frac{1}{1-y_2(z)w}\right) + O(1)$$

as w tends to λ^{-1}. Thus, at $z=0$, since $y_1(0) = y_2(0) = \lambda$ by (21) we have

$$h_n(0) = s(0)\cdot n\lambda^{n-1} + O(\lambda^n) . \qquad (39)$$

However, for $z \neq 0$, the asymptotic behaviour of $g_n(z)$ depends on the condition $\beta_1 \neq \beta_2$.

Proposition 14 *If $M_0 \neq 0$, $\lambda_1 = \lambda_2 = \lambda$ and $\beta_1 \neq \beta_2$, then for every z near 0, different from 0, we have*

$$h_n(z) = s(z)\cdot\frac{y_1(z)^n - y_2(z)^n}{\lambda(\beta_1 - \beta_2)z + O(z^2)} + O\left(y_1(z)^n\right) + O\left(y_2(z)^n\right) + O(\rho^n),$$

where $0 \leq \rho < \lambda$.

Proof. Since $\beta_1 \neq \beta_2$, from (38) we get, for any z near 0 different from 0

$$g_n(z) = s(z)\cdot\frac{y_1(z)^n - y_2(z)^n}{y_1(z) - y_2(z)} + O\left(y_1(z)^n\right) + O\left(y_2(z)^n\right) + O(\rho^n) . \qquad (40)$$

Also observe that, by (16), for any $i = 1,2$ and every z near 0 we can write

$$y_i(z) = \lambda + \lambda\beta_i z + O(z^2) . \qquad (41)$$

Hence, the result follows by replacing the previous relations into (40) and recalling that the contribution of $h_n^{(1)}(z)$ and $h_n^{(2)}(z)$ is of the order $O(y_1(z)^n)$ and $O(y_2(z)^n)$, respectively. $\qquad\square$

Theorem 15 *If $M_0 \neq 0$, $\lambda_1 = \lambda_2 = \lambda$ and $\beta_1 \neq \beta_2$, then Y_n/n converges in distribution to a random variable 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 14 and equation (41), for every non-null $t \in \mathbb{R}$, we have

$$h_n\left(\frac{it}{n}\right) = s(0)\cdot n\lambda^{n-1}\frac{\left(1 + \frac{it\beta_1}{n} + O\left(\frac{1}{n^2}\right)\right)^n - \left(1 + \frac{it\beta_2}{n} + O\left(\frac{1}{n^2}\right)\right)^n}{it(\beta_1 - \beta_2) + O\left(\frac{1}{n}\right)} + O(\lambda^n)$$

which, by (39), yields the following expression for the characteristic function of Y_n/n:

$$\mathbb{E}(itY_n/n) = \frac{h_n(it/n)}{h_n(0)} = \frac{e^{it\beta_1} - e^{it\beta_2}}{it(\beta_1 - \beta_2)} + O\left(\frac{1}{n}\right) .$$

Observe that the main term of the right hand side is the characteristic function of a uniform distribution in the required interval. $\qquad\square$

Now, let us consider the case $\beta_1 = \beta_2 = \beta$. Then point 2) of Proposition 13 holds and hence there is a concentration phenomenon around the mean value of Y_n. The limit distribution can be deduced from equation (40), which still holds in our case but assumes different forms according whether $\gamma_1 \neq \gamma_2$ or not. In the following theorem, let γ be defined by $\gamma = \frac{\gamma_1 + \gamma_2}{2}$.

Theorem 16 If $M_0 \neq 0$, $\lambda_1 = \lambda_2$, $\beta_1 = \beta_2$ and $\gamma_1 \neq \gamma_2$ then $\frac{Y_n - \beta n}{\sqrt{\gamma n}}$ converges in distribution to a random variable T of characteristic function

$$\Phi_T(t) = \frac{e^{-\frac{\gamma_2}{2\gamma}t^2} - e^{-\frac{\gamma_1}{2\gamma}t^2}}{(\frac{\gamma_1}{2\gamma} - \frac{\gamma_2}{2\gamma})t^2} \ . \tag{42}$$

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, replacing these values into (40), for each $t \in \mathbb{R}$ different from 0 we get

$$h_n \left(\frac{it}{\sqrt{\gamma n}} \right) = s(0) \cdot n\lambda^{n-1} \cdot e^{i\beta t\sqrt{n/\gamma}} \cdot \frac{e^{-\frac{\gamma_2}{2\gamma}t^2} - e^{-\frac{\gamma_1}{2\gamma}t^2}}{(\frac{\gamma_1}{2\gamma} - \frac{\gamma_2}{2\gamma})t^2} \left(1 + O(n^{-1/2}) \right) \ ,$$

where $s(z)$ is defined as in (37). The required result follows from the previous equation and from relation (39), recalling that $e^{-it\beta\sqrt{\frac{n}{\gamma}}} \cdot h_n \left(\frac{it}{\sqrt{\gamma n}} \right) / h_n(0)$ is the characteristic function of $\frac{Y_n - \beta n}{\sqrt{\gamma n}}$. $\square$

By direct inspection, one can see that the probability density corresponding to the characteristic function (42) is a mixture of Gaussian densities of mean 0, with variances uniformly distributed over the interval with extremes $\frac{\gamma}{\gamma_1}$ and $\frac{\gamma}{\gamma_2}$. Indeed, we have

$$\Phi_T(t) = \frac{\gamma}{\gamma_2 - \gamma_1} \int_{\frac{\gamma_1}{\gamma}}^{\frac{\gamma_2}{\gamma}} e^{-\frac{1}{2}vt^2} dv \ . \tag{43}$$

Finally we deal with the case where also the main terms of the variances are equal.

Theorem 17 If $M_0 \neq 0$, $\lambda_1 = \lambda_2$, $\beta_1 = \beta_2$ and $\gamma_1 = \gamma_2$ then $\frac{Y_n - \beta n}{\sqrt{\gamma n}}$ converges in distribution to a normal random variable of mean 0 and variance 1.

Proof. In this case, for $z = \Theta(n^{-1/2})$, the convolution in (38) satisfies the relation

$$\sum_{j=0}^{n-1} y_1(z)^j y_2(z)^{n-1-j} = n\lambda^{n-1} \left(1 + \beta z + \frac{\gamma + \beta^2}{2} z^2 \right)^{n-1} (1 + \mathrm{O}(z^3))^{n-1} \ .$$

Replacing this value in the same equation, we get

$$h_n\left(\frac{it}{\sqrt{\gamma n}}\right) = s(0) \cdot n\lambda^{n-1}\exp\left\{i\beta t\sqrt{n/\gamma} - \frac{t^2}{2}\right\}\left(1 + O(n^{-1/2})\right) \ .$$

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 $e^{-t^2/2}$. $\qquad\qquad\qquad\square$

We conclude this section with some examples which illustrate the result obtained in the equipotent case when $\beta_1 = \beta_2$. In Figure 2 we illustrate the form of the limit distributions obtained in Theorems 16 and 17. We represent the density of the random variable having characteristic function (42), 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 17; if γ_2 is much greater than γ_1, then we find a density with a cuspid in the origin corresponding to Theorem 16.
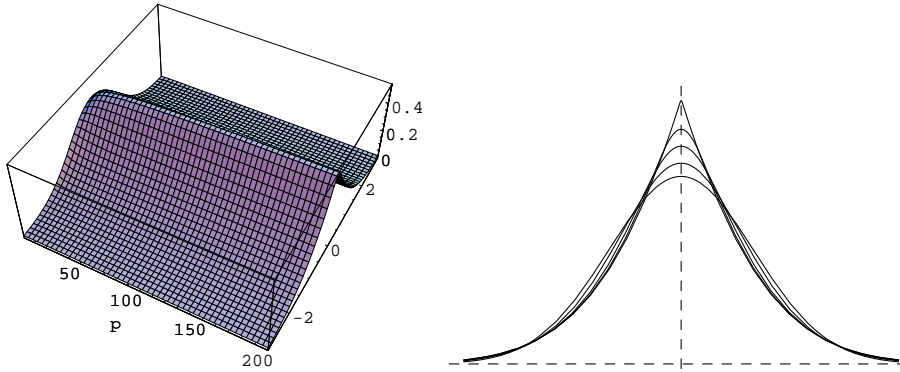

Fig. 2. The first picture represents the density of the random variable having characteristic function (42), according to the parameter $p = \gamma_2/\gamma_1$. The second picture represents some sections obtained for $p = 1.0001, 5, 15, 50, 20000$.

One may also ask whether the hypotheses of Theorem 16 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

$$\xi_1 = \eta_1 = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}, \ A_1 = \mu_1(a) = \begin{pmatrix} 3/20 & 1 \\ 1/16 & 9/40 \end{pmatrix}, \ B_1 = \mu_1(b) = \begin{pmatrix} 3/5 & 0 \\ 0 & 21/40 \end{pmatrix}$$

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 $M_0 \neq 0$.

8 The sum model

In this section we study the behaviour of Y_n assuming $M_0 = 0$. This case corresponds to Example 1, where the stochastic model is defined by the sum of two primitive formal series, having linear representations (ξ_1, μ_1, η_1) and (ξ_2, μ_2, η_2), respectively. Since here $M_0 = 0$, to avoid trivial cases, we also assume $\xi_2 \neq 0 \neq \eta_1$.

The main difference with respect to the general analysis is that here $g_n(z)$ disappears and hence

$$h_n(z) = h_n^{(1)}(z) + h_n^{(2)}(z) . \tag{44}$$

Thus, if $\lambda_1 > \lambda_2$ the leading term is $h_n^{(1)}(z)$ and hence $h_n(z)$ behaves almost as in the case $M_0 \neq 0$. On the other hand, if $\lambda_1 = \lambda_2$ the analysis of the sum model differs significantly from the general case: without the contribution of $g_n(z)$ the function $h_n(z)$ now has a simple pole in the main singularity, rather than a pole of order 2.

8.1 Dominant case in the sum model

Let us assume $\lambda_1 > \lambda_2$. First, reasoning as in the proof of Proposition 8, we easily get the following expressions for mean value and variance:

$$\mathbb{E}(Y_n) = \beta_1 n + \frac{\delta_1}{\alpha_1} + O(\varepsilon^n) , \qquad \mathbb{V}ar(Y_n) = \gamma_1 n + O(1) \qquad (|\varepsilon| < 1)$$

where, according to our notation, α_1, β_1, γ_1, δ_1 are the constants associated with the first component defined as in Theorem 5.

As far as the limit distribution is concerned, observe that $h_n^{(1)}(z)$ satisfies Proposition 6 and hence the following theorem holds.

Theorem 18 *If $M_0 = 0$, $\lambda_1 > \lambda_2$ 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 random variable of mean 0 and variance 1.*

Now consider the degenerate cases $A_1 = 0$ or $B_1 = 0$. 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 $c > 0$

$$\Pr\{|Y_n - n| > c\} \leq \frac{\mathbb{V}ar(Y_n)}{c^2} = O(\varepsilon^n)$$

and hence, $Y_n - n = o(1)$ in probability. A similar result can be obtained in the case $A_1 = 0$.

Theorem 19 *Assume $M_0 = 0$ and $\lambda_1 > \lambda_2$. If $B_1 = 0$ (resp. $A_1 = 0$) then $n - Y_n$ (resp. Y_n) tends to 0 in probability.*

8.2 Equipotent case in the sum model

Here we study the equipotent case $\lambda_1 = \lambda_2 = \lambda$. The first moments of Y_n can be obtained from (6) and (44) by recalling that $h_n^{(1)}(0)$, $h_n^{(2)}(0)$ and their derivatives satisfy (15). Thus, we get the following

Proposition 20 *Assume $M_0 = 0$ and $\lambda_1 = \lambda_2$. 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

$$\mathbb{E}(Y_n) = n \cdot \beta + O(1) , \quad \mathbb{V}ar(Y_n) = n \cdot \frac{\alpha_1\gamma_1 + \alpha_2\gamma_2}{\alpha_1 + \alpha_2} + O(1) .$$

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

$$\Pr\{U_n = x\} = \begin{cases} \frac{\xi_1^T M_1^n \eta_1}{\xi^T M^n \eta} & \text{if } x = 1 , \\ \\ \frac{\xi_2^T M_2^n \eta_2}{\xi^T M^n \eta} & \text{if } x = 0 . \end{cases}$$

We also define $L_n = \beta_1 U_n + \beta_2(1 - U_n)$ and observe that if $\beta_1 = \beta_2$ then $L_n = \beta_1 = \beta_2$. Moreover, it is clear 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 $p = \alpha_1/(\alpha_1 + \alpha_2)$. These random variables occur in the following

Proposition 21 *If $M_0 = 0$ and $\lambda_1 = \lambda_2$ then the random variable $\frac{Y_n}{n} - L_n$ converges to 0 in probability.*

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

$$Y_n - nL_n = U_n \cdot (Y_n^{(1)} - n\beta_1) + (1 - U_n) \cdot (Y_n^{(2)} - n\beta_2) .$$

28

Moreover, by the previous proposition $\mathbb{E}(Y_n - nL_n) = \mathrm{O}(1)$ and hence

$$\mathbb{V}ar(Y_n - nL_n) = \sum_{i=0,1} \mathbb{E}((Y_n - nL_n)^2 \mid U_n = i) \cdot \Pr\{U_n = i\} + \mathrm{O}(1) =$$

$$= \sum_{j=1,2} \mathbb{E}((Y_n^{(j)} - n\beta_j)^2) \cdot \frac{\alpha_j}{\alpha_1 + \alpha_2} + \mathrm{O}(1) = n \cdot \frac{\alpha_1\gamma_1 + \alpha_2\gamma_2}{\alpha_1 + \alpha_2} + \mathrm{O}(1) .$$

Thus, by Chebyshev's inequality, for every $c > 0$ one gets

$$\Pr\left\{ \left| \frac{Y_n}{n} - L_n \right| \geq c \right\} = \mathrm{O}\left(\frac{1}{n}\right) .$$

$\square$

Corollary 22 *If $M_0 = 0$ and $\lambda_1 = \lambda_2$ 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.*

The above results intuitively state that Y_n asymptotically behaves like 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 random variable 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

$$\Upsilon = U \cdot N_1 + (1 - U) \cdot N_2 \tag{45}$$

where we assume U, N_1, N_2 independent of one another. Note that, if $\gamma_1 = \gamma_2$ then Υ has a normal distribution of mean 0 and variance γ_1. The characteristic function of Υ is given by

$$\mathbb{E}(e^{it\Upsilon}) = \frac{\alpha_1}{\alpha_1 + \alpha_2} e^{-\frac{\gamma_1}{2}t^2} + \frac{\alpha_2}{\alpha_1 + \alpha_2} e^{-\frac{\gamma_2}{2}t^2} .$$

It turns out that $\frac{Y_n - nL_n}{\sqrt{n}}$ converges in distribution to Υ.

Proposition 23 *If $M_0 = 0$, $\lambda_1 = \lambda_2$ 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

$$\mathbb{E}(e^{it\Upsilon_n}) = \sum_{i=0,1} \mathbb{E}(e^{it\Upsilon_n} \mid U_n = i) \cdot \Pr\{U_n = i\} = \sum_{j=1,2} \mathbb{E}\left(e^{it\frac{Y_n^{(j)} - n\beta_j}{\sqrt{n}}}\right) \cdot \left(\frac{\alpha_j}{\alpha_1 + \alpha_2} + \mathrm{O}(\varepsilon^n)\right)$$

$$= \frac{\alpha_1}{\alpha_1 + \alpha_2} e^{-\frac{\gamma_1}{2}t^2} + \frac{\alpha_2}{\alpha_1 + \alpha_2} e^{-\frac{\gamma_2}{2}t^2} + \mathrm{O}\left(n^{-1/2}\right) .$$

$\square$

The previous results hold also when $\beta_1 = \beta_2 = \beta$; clearly 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 24 *Assume $M_0 = 0$, $\lambda_1 = \lambda_2$ 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.*

9 Summary and conclusions

Most results presented in this work are summarized in Table 1. To explain them intuitively, first recall that in a primitive rational model the limit distribution of our statistics is Gaussian. Thus, in a model consisting of two primitive components the limit behaviour is determined by the way how their separate contributions mix together. This combination may produce quite different limit distributions depending on two main conditions: (i) whether there is a communication from the first to the second component (i.e. $M_0 \neq 0$) 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. On the contrary, condition (i) concerning M_0 is particularly meaningful in the equipotent case. Here, if $M_0 \neq 0$ the main contribution to the bivariate generating function is given by $G(z, w)$, 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 $M_0 = 0$ the function $G(z, w)$ 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 $M_0 \neq 0$, while, if $M_0 = 0$, we obtain a limit distribution concentrated in two values that correspond to

Conditions		Results		
Dominance	Degeneracy	Mean value	Variance	Limit distribution
$M_0 \neq 0$ — $\lambda_1 > \lambda_2$	$A_1 \neq 0 \neq B_1$	$\beta_1 n + O(1)$ $0 < \beta_1 < 1$	$\gamma_1 n + O(1)$ $0 < \gamma_1$	$\frac{Y_n - \beta_1 n}{\sqrt{\gamma_1 n}} \longrightarrow_d N_{0,1}$
	$B_1 = 0$	$n - E + O(\varepsilon^n)$ $E > 0$	$c + O(\varepsilon^n)$ $c > 0$	$n - Y_n \longrightarrow_d W$ Theorem 12
	$A_1 = 0$	$E' + O(\varepsilon^n)$ $E' > 0$	$c' + O(\varepsilon^n)$ $c' > 0$	$Y_n \longrightarrow_d Z$ Theorem 12
$\lambda_1 = \lambda_2$	$\beta_1 \neq \beta_2$	$\frac{\beta_1 + \beta_2}{2} n + O(1)$	$\frac{(\beta_1 - \beta_2)^2}{12} n^2 + O(n)$	$\frac{Y_n}{n} \longrightarrow_d \mathrm{Unif}(b_1, b_2)$ Theorem 15
	$\beta_1 = \beta_2 = \beta$ $\gamma_1 \neq \gamma_2$	$\beta n + O(1)$	$\gamma n + O(1)$ $\gamma = \frac{\gamma_1 + \gamma_2}{2}$	$\frac{Y_n - \beta n}{\sqrt{\gamma n}} \longrightarrow_d T$ Theorem 16
	$\beta_1 = \beta_2 = \beta$ $\gamma_1 = \gamma_2 = \gamma$	$\beta n + O(1)$	$\gamma n + O(1)$	$\frac{Y_n - \beta n}{\sqrt{\gamma n}} \longrightarrow_d N_{0,1}$
$M_0 = 0$ — $\lambda_1 > \lambda_2$	$A_1 \neq 0 \neq B_1$	$\beta_1 n + O(1)$ $0 < \beta_1 < 1$	$\gamma_1 n + O(1)$ $0 < \gamma_1$	$\frac{Y_n - \beta_1 n}{\sqrt{\gamma_1 n}} \longrightarrow_d N_{0,1}$
	$B_1 = 0$	$n + O(\varepsilon^n)$	$O(\varepsilon^n)$	$n - Y_n \longrightarrow_p 0$ Theorem 19
	$A_1 = 0$	$O(\varepsilon^n)$	$O(\varepsilon^n)$	$Y_n \longrightarrow_p 0$ Theorem 19
$\lambda_1 = \lambda_2$	$\beta_1 \neq \beta_2$	$c_1 n + O(1)$ $c_1 = \frac{\alpha_1 \beta_1 + \alpha_2 \beta_2}{\alpha_1 + \alpha_2}$	$c_2 n^2 + O(n)$ $c_2 = \frac{\alpha_1 \alpha_2 (\beta_1 - \beta_2)^2}{(\alpha_1 + \alpha_2)^2}$	$\frac{Y_n}{n} \longrightarrow_d$ $\beta_1 U + \beta_2 (1 - U)$ Corollary 22
	$\beta_1 = \beta_2 = \beta$ $\gamma_1 \neq \gamma_2$	$\beta n + O(1)$	$c_3 n + O(1)$ $c_3 = \frac{\alpha_1 \gamma_1 + \alpha_2 \gamma_2}{\alpha_1 + \alpha_2}$	$\frac{Y_n - \beta n}{\sqrt{n}} \longrightarrow_d$ $U N_{0,\gamma_1} + (1 - U) N_{0,\gamma_2}$ Corollary 24
	$\beta_1 = \beta_2 = \beta$ $\gamma_1 = \gamma_2 = \gamma$	$\beta n + O(1)$	$\gamma n + O(1)$	$\frac{Y_n - \beta n}{\sqrt{\gamma n}} \longrightarrow_d N_{0,1}$ Corollary 24

Table 1

This picture summarizes most results presented in this paper. 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 r.v. of mean value m and variance s and a Bernoullian r.v. of parameter $p = \alpha_1/(\alpha_1 + \alpha_2)$.

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 $M_0 \neq 0$ such distributions have variance uniformly distributed over a given interval; on the contrary, if $M_0 = 0$ 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. We present the following example to show how the equipotent case can be considered as a sort of equilibrium point between two

(opposite) dominant cases.

Consider the product model of Example 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 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 to build from (22) 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 4 illustrates the probability function of the random variable 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 11. 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 15.
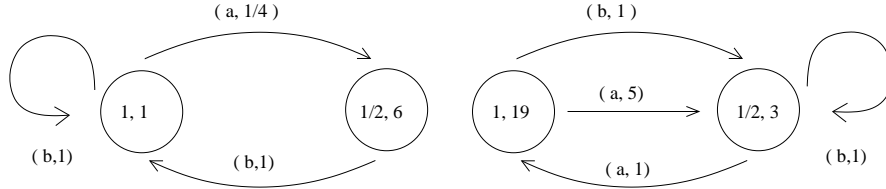

Fig. 3. Two weighted finite automata over the alphabet $\{a, b\}$, defining the primitive linear representations (π_i, ν_i, τ_i), $i = 1, 2$.
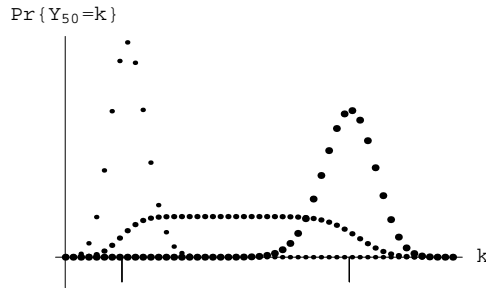

Fig. 4. Probability functions of Y_{50} in the product model where the two factor components are defined by Figure 3 with weighted expanded by a constant factor. The vertical bars have abscissas $50\beta_1$ and $50\beta_2$. The curves correspond to the cases where (λ_1, λ_2) are equal to (2,1), (1,2) and (1,1), respectively.

We conclude observing that some of the previous results clearly extend to rational stachostic models given by more than two primitive components. For instance the result given in Theorem 11 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 15 or Proposition 23 holds (according

whether there exists a communication from the first to the second component). However, in the multicomponent model the number of subcases grows exponentially since more than two equipotent components can dominate the others; then the limit distribution depends on several parameters and properties: the number of dominant components, the geometry of communication among them, the values of the main constants of mean value and variance associated with these components and the occurrence of degenerate cases. For this reason, we think that the general multicomponent model should be first studied by considering a set of typical situations, rather than by an exhaustive analysis of all possible subcases (as done in this work).

References

[1] E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory*, 15:91–111, 1973.

[2] J. Berstel and C. Reutenauer. *Rational series and their languages*, Springer-Verlag, New York - Heidelberg - Berlin, 1988.

[3] A. Bertoni, C. Choffrut, M. Goldwurm, and V. Lonati. On the number of occurrences of a symbol in words of regular languages. *Theoretical Computer Science*, 302(1-3):431–456, 2003.

[4] A. Bertoni, C. Choffrut, M. Goldwurm, and V. Lonati. Local limit distributions in pattern statistics: beyond the Markovian models. *Proceedings 21st STACS*, V. Diekert and M. Habib Eds., L.N.C.S. vol. n. 2996, Springer, 117–128, 2004.

[5] J. Bourdon and B. Vallée. Generalized pattern matching statistics. *Mathematics and computer science II: algorithms, trees, combinatorics and probabilities*. Proc. of Versailles Colloquium, Birkhauser, 249–265, 2002.

[6] A. Denise. Génération aléatoire uniforme de mots de langages rationnels. *Theoretical Computer Science*, 159:43–63, 1996.

[7] P. Flajolet and R. Sedgewick. The average case analysis of algorithms: multivariate asymptotics and limit distributions. *Rapport de recherche* n. 3162, INRIA Rocquencourt, May 1997.

[8] M. S. Gelfand. Prediction of function in DNA sequence analysis. *J. Comput. Biol.*, 2:87–117, 1995.

[9] B.V. Gnedenko. *The theory of probability* (translated by G. Yankovsky). Mir Publishers - Moscow, 1976.

[10] L. J. Guibas and A. M. Odlyzko. Maximal prefix-synchronized codes. *SIAM J. Appl. Math.*, 35:401–418, 1978.

[11] L. J. Guibas and A. M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *Journal of Combinatorial Theory. Series A*, 30(2):183–208, 1981.

[12] H.K. Hwang. *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques*. Ph.D. Dissertation, École polytechnique, Palaiseau, France, 1994.

[13] P. Nicodeme, B. Salvy, and P. Flajolet. Motif statistics. *Theoretical Computer Science*, 287(2): 593–617, 2002.

[14] B. Prum, F. Rudolphe and E. Turckheim. Finding words with unexpected frequencies in deoxyribonucleic acid sequence. *J. Roy. Statist. Soc. Ser. B*, 57: 205–220, 1995.

[15] M. Régnier and W. Szpankowski. On pattern frequency occurrences in a Markovian sequence. *Algorithmica*, 22 (4):621–649, 1998.

[16] E. Seneta. *Non-negative matrices and Markov chains*, Springer–Verlag, New York Heidelberg Berlin, 1981.

[17] M. Waterman. *Introduction to computational biology*, Chapman & Hall, New York, 1995.

[18] S. Wolfram. *The Mathematica book*, Fourth Edition, Wolfram Media - Cambridge University Press, 1999.