Saddle Point Method in the Analysis of Pattern Statistics for Regular Languages

M. Goldwurm[(1)], J. Lin[(2)], M. Vignati[(1)]

(1) Dipartimento di Matematica, Università degli Studi di Milano, Italy
(2) Department of Mathematics, Khalifa University,
Abu Dhabi - United Arab Emirates

Abstract. In a recent work we have determined the local limit distribution of pattern statistics representing the number of occurrences of a symbol in words of length n in a regular language generated at random according to a suitable stochastic model. Such a model is defined by a finite automaton with weights in $\mathbb{R}_+$, consisting of two primitive components, having some transition from the first to the second component. In the present work we extend those results to the case when there is no communication among the components, and hence the associated formal series is the sum of two rational series recognized by finite state automata with primitive transition matrix. We obtain local limit laws of Gaussian type when there is a dominant component or when, in equipotent case, the main terms of mean value and variance are equal. On the contrary, if these terms are not the same then the local limit distribution is a convex combination of Gaussian laws. All convergence rates of our limits are of the order $O(n^{-1/2})$. This completes the analysis of local limit laws of symbol statistics under a bicomponent stochastic model[1].

Keywords: rational formal series, pattern statistics, limit distributions, local limit laws, Saddle Point Method.

1 Introduction

The Saddle Point Method is a classical tool used to determine the asymptotic expression of integrals, defined over closed curves in the complex plane, depending on an additional parameter. This method is of particular interest in Analytic Combinatorics where, generally, the parameter varies in $\mathbb{N}$ and represents the size of a family of combinatorial structures; in that context several enumeration problems can be reduced to evaluating the coefficients of generating functions, which coincide (by the Cauchy formula) with the integral of the corresponding function along a circle centred at the origin (see for instance [8, Chapter VII]).

A traditional application of this technique concerns the so-called local limit theorems of Gaussian type for sequences of discrete random variables. A well-known example is the classical De Moivre - Laplace Theorem, stating that the

probability functions of sequences of Binomial r.v.'s approximate a Gaussian density function (see for instance [9]). A general framework for this type of theorems considers a sequence $\{X_n\}$ of r.v.'s, each of which takes value in $\{0, 1, \ldots, n\}$. The main goal is to prove that the family of probabilities $\{p_n(k)\}_{k\in\{0,1,\ldots,n\}}$, where $p_n(k) = \Pr(X_n = k)$, suitably standardized converges to the Normal density of mean 0 and variance 1, uniformly with respect to $k \in \{0, 1, \ldots, n\}$. Several results of this type can be proved by using the Saddle Point Method since the coefficients $p_n(k)$ are related to the characteristic function $\Psi_n(t)$ of X_n by the well-known inversion formula:

$$p_n(k) \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} \Psi_n(t)e^{-itk}dt \tag{1}$$

for every $k \in \{0, 1, \ldots, n\}$ [8,9]. One can see the previous expression as a circuit integral over the complex plain, where the curve is given by the unit circle with centre in 0. It may occur that the main contribution to the integral comes from a small portion of the curve passing near a particular point, called "saddle point" since at that position the modulus of the integrand forms a surface that resembles a mountain pass. In these cases a convenient method to compute the coefficient often consists of evaluating the integral near the saddle point and proving that the contribution coming from the remaining part of the curve is negligible. See [8, Ch. VIII] for an appealing introduction to the method and an interesting discussion on its applications. In our cases the curve is just the unit circle (e^{it} with $-\pi \le t \le \pi$) and the "saddle point" always corresponds to $1 = e^0$.

We recall that a local limit theorem does not follow immediately from a traditional convergence in distribution (that is the type of convergence in the usual central limit theorems), since point probabilities are differences of values of the corresponding distribution functions, and hence they may not be detected by a standard analysis of convergence in law. Usually, in order to prove a local limit theorem from a convergence in distribution, some additional regularity or aperiodicity conditions are necessary; standard counterexamples show that such conditions cannot be avoided. Moreover, another evaluation often occurring in local limit properties concerns the convergence rate, which measures the speed of approximation. Finding a tight convergence rate in central limit theorems is a natural goal of literature [12].

In the present paper we adapt the Saddle Point Method to a problem on formal languages, i.e. the analysis of patterns statistics defined over words of regular languages. More precisely, we consider a sequence of r.v.'s $\{Y_n\}$, where each Y_n is the number of occurrences of a symbol a in a word w of length n generated at random in a *rational stochastic model*. Such a model can be formally defined by a finite state automaton with real positive weights on transitions. In this setting the probability of generating a word w is proportional to the weight the automaton associates with w; thus the language recognized by the automaton is the family of all words having non-null probability to be generated. This model is quite general, it includes as a special cases the traditional Bernoullian and Markovian sources and comprises the random generation of words of length n in any regular language under uniform distribution.

The properties of $\{Y_n\}$ are of particular interest for the analysis of regular patterns occurring in words generated by Markovian models [3,13,14] and for the asymptotic estimate of the coefficients of rational series in commutative variables [3,4]. It is also related to various research topics of Computer Science such as the descriptional complexity of languages and computational models [5], and the analysis of additive functions defined on regular languages [11]. Clearly, the asymptotic behaviour of $\{Y_n\}$ depends on properties of the finite automaton $\mathcal{A}$ defining the stochastic model. It is known that if $\mathcal{A}$ has a primitive transition matrix then Y_n has a Gaussian limit distribution [3,13] and, under a suitable aperiodicity condition, it also satisfies a (Gaussian) local limit theorem [3]. The limit distribution of Y_n in the global sense is known also when the transition matrix of $\mathcal{A}$ consists of two primitive components [7] and an analysis of local limit laws in the bicomponent models is presented in [10] in the case when there is some transition from the first to the second component.

Here we improve those results by presenting some local limit laws for the bicomponent models when there is no communication between the components. As in the previous works [3,10], we have to add suitable aperiodicity conditions to guarantee appropriate local limits. In particular, we prove that the sequence of statistics $\{Y_n\}$ has a local limit law of Gaussian type if the main eigenvalues of the two components are different; in this case the aperiodicity condition only concerns the dominant component, which determines the asymptotic behaviour of the sequence. On the other hand, when the main eigenvalues of the two components coincide (equipotent model) the results depend on the values of four constants: β_1, γ_1 and β_2, γ_2, representing the leading terms of mean value and variance of our statistics associated to the first and second component, respectively. If $\beta_1 \neq \beta_2$ or $\gamma_1 \neq \gamma_2$ then the local limit distribution of $\{Y_n\}$ is a convex combination of two Gaussian laws. On the contrary, if $\beta_1 = \beta_2$ and $\gamma_1 = \gamma_2$, then the local limit density of $\{Y_n\}$ turns out to be Gaussian again. All local limit laws obtained in this work have a convergence rate of the order $O(n^{-1/2})$. These results are similar to the local limit laws obtained in the communicating models [10] when the limit distribution is Gaussian, while they are rather different in the other cases (equipotent models with unequal parameters).

The material we present is organized as follows. In Section 2 we recall the problem and some known results concerning the primitive models. In Section 3 we introduce the non-communicating bicomponent models and prove a local limit law in the dominant case. Then, in Section 4 we study the equipotent (non-communicating bicomponent) models, present the local limit laws in these cases and discuss briefly their meaning. In the last section we compare these results with the previous ones and discuss possible future investigations.

2 Preliminary notions and previous results

Given the binary alphabet $\{a, b\}$, for every word $w \in \{a, b\}^*$ we denote by $|w|$ the length of w and by $|w|_a$ the number of occurrences of a in w. For each $n \in \mathbb{N}$, we also represent by $\{a, b\}^n$ the set $\{w \in \{a, b\}^* : |w| = n\}$. Here a *formal series*

in the non-commutative variables a, b is a function $r : \{a,b\}^* \to \mathbb{R}_+$, where $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$, and for every $w \in \{a,b\}^*$ we denote by (r,w) the value of r at w. Such a series r is called *rational* if for some integer $m > 0$ there is a monoid morphism $\mu : \{a,b\}^* \to \mathbb{R}_+^{m\times m}$ and two arrays $\xi, \eta \in \mathbb{R}_+^m$, such that $(r,w) = \xi'\mu(w)\eta$, for every $w \in \{a,b\}^*$. In this case, as the morphism μ is generated by matrices $A = \mu(a)$ and $B = \mu(b)$, we say that the 4-tuple (ξ, A, B, η) is a *linear representation* of r of size m. Clearly, such a 4-tuple can be considered as a finite state automaton over the alphabet $\{a,b\}$, with transitions (as well as initial and final states) weighted by positive real values. Throughout this work we assume that the set $\{w \in \{a,b\}^n : (r,w) > 0\}$ is not empty for every $n \in \mathbb{N}_+$ (so that $\xi \neq 0 \neq \eta$), and that A and B are not null matrices, i.e. $A \neq [0] \neq B$. Then we can consider the probability measure Pr over the set $\{a,b\}^n$ given by

$$\Pr(w) = \frac{(r,w)}{\sum_{x\in\{a,b\}^n}(r,x)} = \frac{\xi'\mu(w)\eta}{\xi'(A+B)^n\eta} \qquad \forall\, w \in \{a,b\}^n$$

Note that, if r is the characteristic series of a language $L \subseteq \{a,b\}^*$ then Pr is the uniform probability function over the set $L \cap \{a,b\}^n$. Thus we can define the random variable (r.v. for short) $Y_n = |w|_a$, where w is chosen at random in $\{a,b\}^n$ with probability $\Pr(w)$. As $A \neq [0] \neq B$, Y_n is not a degenerate r.v. . It is clear that, for every $k \in \{0,1,\ldots,n\}$,

$$p_n(k) := \Pr(Y_n = k) = \frac{\sum_{|w|=n,|w|_a=k}(r,w)}{\sum_{w\in\{a,b\}^n}(r,w)}$$

Since r is rational also the previous probability can be expressed by using its linear representation. It turns out that

$$p_n(k) = \frac{[x^k]\xi'(Ax+B)^n\eta}{\xi'(A+B)^n\eta} \qquad \forall\, k \in \{0,1,\ldots,n\} \tag{2}$$

For sake of brevity we say that Y_n is *defined* by the linear representation (ξ, A, B, η). The distribution of Y_n can be represented by the map $h_n(z)$ and the characteristic function $\Psi_n(t)$, given respectively by

$$h_n(z) = \xi'(Ae^z + B)^n\eta \qquad \forall\, z \in \mathbb{C} \tag{3}$$

$$\Psi_n(t) = \sum_{k=0}^{n} p_n(k)e^{itk} = \frac{\xi'(Ae^{it}+B)^n\eta}{\xi'(A+B)^n\eta} = \frac{h_n(it)}{h_n(0)} \qquad \forall\, t \in \mathbb{R} \tag{4}$$

In particular mean value and variance of Y_n are determined by

$$\mathrm{E}(Y_n) = \frac{h_n'(0)}{h_n(0)}, \quad \mathrm{Var}(Y_n) = \frac{h_n''(0)}{h_n(0)} - \left(\frac{h_n'(0)}{h_n(0)}\right)^2 \tag{5}$$

Our general goal is to study the limit distribution of $\{Y_n\}$ as n grows to $+\infty$ and in particular its possible local limit law.

We recall that a sequence of r.v.'s $\{X_n\}$ *converges in distribution* (or in law) to a random variable X of distribution function F if $\lim_{n\to+\infty} \Pr(X_n \leq x) = F(x)$, for every $x \in \mathbb{R}$ of continuity for F. The central limit theorems yield classical examples of convergence in distribution to a Gaussian random variable.

Instead, the local limit laws establish the convergence of single probabilities to a density function (see for instance [9,8]). More precisely, consider a sequence of r.v.'s $\{X_n\}$ such that each X_n takes value in $\{0, 1, \ldots, n\}$. We say that $\{X_n\}$ *satisfies a local limit law* of Gaussian type if there are two real sequences $\{a_n\}$, $\{s_n\}$, satisfying $\mathrm{E}(X_n) \sim a_n$, $\mathrm{Var}(X_n) \sim s_n^2$, with $s_n > 0$ for all n, such that for some real $\epsilon_n \to 0$, the relation

$$\left| s_n \Pr\left(X_n = k\right) \; - \; \frac{e^{-\left(\frac{k-a_n}{s_n}\right)^2/2}}{\sqrt{2\pi}} \right| \; \leq \; \epsilon_n \tag{6}$$

holds uniformly for every $k \in \{0, 1, \ldots, n\}$ and every $n \in \mathbb{N}$ large enough. Here, ϵ_n yields the *convergence rate* (or the speed) of the law. The best known example of such a property is the de Moivre-Laplace local limit theorem, which concerns sequences of binomial r.v.'s [9].

Similar definitions can be given for other (non-Gaussian) types of local limit laws. In this case the Gaussian density $e^{-x^2/2}/\sqrt{2\pi}$ appearing in (6) is replaced by some density function $f(x)$; clearly, if $f(x)$ is not continuous at some points, the uniformity of k must be adapted to the specific conditions.

2.1 Primitive models

A relevant case occurs when $M = A + B$ is primitive, i.e. $M^k > 0$ for some $k \in \mathbb{N}$ [16]. In this case it is known that Y_n has a local limit law of Gaussian type with a convergence rate $O(n^{-1/2})$ [10] and here we recall some properties useful in subsequent sections.

First note that by Perron-Frobenius Theorem, a primitive matrix M admits a real eigenvalue $\lambda > 0$ greater than the modulus of any other eigenvalue. Moreover, strictly positive left and right eigenvectors ζ, ν of M w.r.t. λ can be defined so that $\zeta'\nu = 1$. Thus, we can consider the function $u = u(z)$ implicitly defined by the equation

$$\mathrm{Det}(Iu - Ae^z - B) = 0$$

such that $u(0) = \lambda$. It turns out that, in a neighbourhood of $z = 0$, $u(z)$ is analytic, is a simple root of the characteristic polynomial of $Ae^z + B$ and $|u(z)|$ is strictly greater than the modulus of all other eigenvalues of $Ae^z + B$. Moreover, a precise relationship between $u(z)$ and function $h(z)$, defined in (3), is well-known and states that for two positive constants c, ρ and a function $r(z)$ analytic and non-null at $z = 0$, one has

$$h_n(z) = r(z)\; u(z)^n + O(\rho^n) \qquad \forall z \in \mathbb{C} : |z| \leq c \tag{7}$$

where $\rho < |u(z)|$ and in particular $\rho < \lambda$.

Mean value and variance of Y_n can be estimated from relations (7) and (5). In turns out [3] that the constants

$$\alpha = \xi' \nu \zeta' \eta \ , \quad \beta = \frac{u'(0)}{\lambda} \quad \text{and} \quad \gamma = \frac{u''(0)}{\lambda} - \left(\frac{u'(0)}{\lambda}\right)^2 \tag{8}$$

are strictly positive and satisfy the relations

$$\mathrm{E}(Y_n) = \beta n + O(1) \quad \text{and} \quad \mathrm{Var}(Y_n) = \gamma n + O(1)$$

Other properties concern function $y(t) = u(it)/\lambda$, defined for real t in a neighbourhood of 0. In particular, there exists a constant $c > 0$, for which relation (7) holds true, satisfying the following relations [3]:

$$|y(t)| = 1 - \frac{\gamma}{2} t^2 + O(t^4), \ \arg y(t) = \beta t + O(t^3), \ |y(t)| \leq e^{-\frac{\gamma}{4} t^2} \qquad \forall\, |t| \leq c \tag{9}$$

The behaviour of $y(t)$ can be estimated precisely when t tends to 0. For any q such that $1/3 < q < 1/2$ it can be proved [3] that

$$y(t)^n = e^{-\frac{\gamma}{2} t^2 n + i\beta t n}(1 + O(t^3)n) \qquad \text{for } |t| \leq n^{-q} \tag{10}$$

The previous properties are used in [10] to prove a local limit theorem for $\{Y_n\}$ when M is primitive, with a convergence rate $O(n^{-1/2})$. The result holds under a further assumption, introduced to avoid periodicity phenomena, defined as follows: consider the transition graph of the finite state automaton defined by matrices A and B, i.e. the directed graph G with vertex set $\{1, 2, \ldots, m\}$ such that, for every $i, j \in \{1, 2, \ldots, m\}$, G has an edge from i to j labelled by a letter a (b, respectively) whenever $A_{ij} > 0$ ($B_{ij} > 0$, resp.). Also denote by d the GCD of the differences in the number of occurrences of a in the (labels of) cycles of equal length of G. the pair (A, B) is said to be *aperiodic* if $d = 1$. Such a property is often verified; for instance it holds true whenever $A_{ij} > 0$ and $B_{ij} > 0$ for two (possibly equal) indices i, j. Moreover, it can be proved [4] that (A, B) is aperiodic if and only if, for every real t such that $0 < t < 2\pi$, we have

$$|\mu| < \lambda \quad \text{for every eigenvalue } \mu \text{ of } Ae^{it} + B \tag{11}$$

3 The "sum model"

In this section we study the behaviour of $\{Y_n\}_{n \in \mathbb{N}}$ defined by a linear representation (ξ, A, B, η) of size m consisting of two non-communicating irreducible components. Formally, there are two linear representations, $(\xi_1, A_1, B_1, \eta_1)$ and $(\xi_2, A_2, B_2, \eta_2)$, of size m_1 and m_2 respectively, where $m = m_1 + m_2$, such that:

1. $\xi' = (\xi_1', \xi_2')$, $\quad A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$, $\quad B = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}$, $\quad \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}$
2. $\xi_1 \neq 0 \neq \eta_1$, $\xi_2 \neq 0 \neq \eta_2$, $A_1 \neq 0 \neq B_1$ and $A_2 \neq 0 \neq B_2$;

3. $M_1 = A_1 + B_1$ and $M_2 = A_2 + B_2$ are irreducible matrices and we denote by λ_1 and λ_2 the corresponding Perron-Frobenius eigenvalues.

Note that condition 3 is weaker than a primitivity assumption for M_1 and M_2. Clearly a formal series r with a linear representation of this kind is given by the sum of two rational formal series r_1, r_2, both having an irreducible linear representation, i.e. $(r, w) = (r_1, w) + (r_2, w)$ for every $w \in \{a, b\}^*$.

Assuming these hypotheses, we say that $\{Y_n\}_n$ is defined in a non-communicating bicomponent model or, for sake of brevity, in a *sum model*. As in the corresponding "communicating" case [10], its limit distribution first depends on whether $\lambda_1 \neq \lambda_2$ or $\lambda_1 = \lambda_2$. If $\lambda_1 \neq \lambda_2$ there is a dominant component, corresponding to the greater between λ_1 and λ_2, that determines the asymptotic behaviour of $\{Y_n\}$. If $\lambda_1 = \lambda_2$ the two components are equipotent and they both contribute to the limit behaviour of $\{Y_n\}$.

In both cases, for $j = 1, 2$, we can define $h_n^{(j)}(z)$, $u_j(z)$, $y_j(t)$, α_j, β_j, and γ_j, respectively, as the values $h_n(z)$, $u(z)$, $y(t)$, α, β, γ referred to component j. Note that the hypotheses above guarantee $0 < \beta_j < 1$ and $0 < \gamma_j$, for both $j = 1, 2$). We also define $H(x, y)$ as the matrix-valued function given by

$$H(x,y) = \sum_{n=0}^{+\infty} (Ax + B)^n y^n = \begin{bmatrix} H^{(1)}(x,y) & 0 \\ 0 & H^{(2)}(x,y) \end{bmatrix}, \quad \text{where}$$

$$H^{(1)}(x,y) = \frac{\mathrm{Adj}\,(I - (A_1 x + B_1)y)}{\mathrm{Det}\,(I - (A_1 x + B_1)y)},\ H^{(2)}(x,y) = \frac{\mathrm{Adj}\,(I - (A_2 x + B_2)y)}{\mathrm{Det}\,(I - (A_2 x + B_2)y)} \quad (12)$$

Thus, the generating function of $\{h_n(z)\}_n$ satisfies the identities

$$\sum_{n=0}^{\infty} h_n(z) y^n = \xi' H(e^z, y)\eta = \xi_1' H^{(1)}(e^z, y)\eta_1 + \xi_2' H^{(2)}(e^z, y)\eta_2 \quad (13)$$

and, for every $z \in \mathbb{C}$ and every $t \in \mathbb{R}$, we have

$$h_n(z) = h_n^{(1)}(z) + h_n^{(2)}(z) \qquad \Psi_n(it) = \frac{h_n^{(1)}(it) + h_n^{(2)}(it)}{h_n(0)} \quad (14)$$

3.1 Dominant sum models

First, let us consider the dominant case. Assume $\lambda_1 > \lambda_2$ and let M_1 be aperiodic (and hence primitive). For sake of brevity, we say that $\{Y_n\}$ is defined in a *dominant sum model* with $\lambda_1 > \lambda_2$. In this case we have $0 < \beta_1 < 1$, $0 < \gamma_1$, and it is known that $\frac{Y_n - \beta_1 n}{\sqrt{\gamma_1 n}}$ converges in distribution to a normal r.v. of mean value 0 and variance 1 [7]. Here we show that a Gaussian local limit law holds true at the cost of assuming to be aperiodic only the pair (A_1, B_1).

Theorem 1. *Let $\{Y_n\}$ be defined in a dominant sum model with $\lambda_1 > \lambda_2$ and assume (A_1, B_1) aperiodic. Then, as n tends to $+\infty$, the relation*

$$\left| \sqrt{n} Pr(Y_n = k) - \frac{e^{-\frac{(k-\beta_1 n)^2}{2\gamma_1 n}}}{\sqrt{2\pi\gamma_1}} \right| = O\left(n^{-1/2}\right)$$

holds true uniformly for every $k \in \{0, 1, \ldots, n\}$.

Proof. First, to simplify the notation set $\beta = \beta_1$ and $\gamma = \gamma_1$ (only in this proof). Then, the main idea is to study the characteristic function $\Psi_n(t)$ for $t \in [-\pi, \pi]$ by splitting this interval into three sets:

$$[-n^{-q}, n^{-q}] \ , \qquad \{t \in \mathbb{R} : n^{-q} < |t| \leq c\} \ , \qquad \{t \in \mathbb{R} : c < |t| \leq \pi\} \ , \tag{15}$$

where $c \in (0, \pi)$ is a constant satisfying relations (9) for both $y_1(t)$ and $y_2(t)$, and q is an arbitrary value such that $\frac{1}{3} < q < \frac{1}{2}$. The behaviour of $\Psi_n(t)$ in these sets is characterized by the following properties:

a. for some $\varepsilon \in (0, 1)$ we have

$$|\Psi_n(t)| = O(\varepsilon^n) \qquad \forall\ t \in \mathbb{R}\ :\ c < |t| \leq \pi \tag{16}$$

b.

$$|\Psi_n(t)| = O\left(e^{-\frac{\gamma}{4} n^{1-2q}}\right) \qquad \forall\ t \in \mathbb{R}\ :\ n^{-q} < |t| \leq c \tag{17}$$

c.

$$\int_{|t| \leq n^{-q}} \left| \Psi_n(t) - e^{-\frac{\gamma}{2} t^2 n + i\beta t n} \right| dt \ = \ O(n^{-1}) \tag{18}$$

Proof of (16). Note that, by relations (12) and (13), for every $t \in \mathbb{R}$ the singularities of the generating function $\xi' H(e^{it}, y)\eta = \sum_{n=0}^{\infty} h_n(it) y^n$ are the inverses of the eigenvalues of $A_1 e^{it} + B_1$ and $A_2 e^{it} + B_2$. Since $c < |t| \leq \pi$, the first ones are in modulus smaller than λ_1 by the aperiodicity of (A_1, B_1) and property (11), while the second ones are in modulus smaller or equal to λ_2 as a consequence of Perron-Frobenius Theorem for irreducible matrices [16, Ex. 1.9]. Thus, since $\lambda_1 > \lambda_2$, for some positive $\tau < \lambda_1$ we have $|h_n(it)| = O(\tau^n)$ for all real t such that $c < |t| \leq \pi$. By the same reason it is clear that $h_n(0) = \Theta(\lambda_1^n)$, and hence for some $\varepsilon \in (0, 1)$ we have $|\Psi_n(t)| = \left| \frac{h_n(it)}{h_n(0)} \right| = \frac{O(\tau^n)}{\Theta(\lambda_1^n)} = O(\varepsilon^n)$, as required.

Proof of (17). By relation (7), there exists $\rho \in (0, \lambda_1)$ such that, for some $\varepsilon \in (0, 1)$,

$$\Psi_n(t) = \frac{h_n(it)}{h_n(0)} = \frac{r_1(it)\lambda_1^n y_1(t)^n + O(\rho^n)}{r_1(0)\lambda_1^n + O(\rho^n)} = [1 + O(t)] y_1(t)^n + O(\varepsilon^n) \tag{19}$$

for all $t \in \mathbb{R}$ satisfying $|t| \leq c$. Also, by inequality (9), we know that $|y_1(t)|^n \leq e^{-\frac{\gamma}{4} t^2 n}$ whenever $|t| \leq c$. Thus, the result follows by replacing this bound in the previous equation and recalling that $n^{-q} \leq |t| \leq c$.

Proof of (18). From equality (19), applying relation (10) and recalling that $nO(t^3) = o(1)$ for $|t| \leq n^{-q}$, we get

$$\Psi_n(t) = (1 + O(t) + nO(t^3)) e^{-\frac{\gamma}{2} t^2 n + i\beta t n} + O(\varepsilon^n) \qquad \forall\ t \in \mathbb{R}\ :\ |t| \leq n^{-q}$$

Then, by a direct computation we obtain

$$\begin{aligned} \int_{|t| \leq n^{-q}} \left| \Psi_n(t) - e^{-\frac{\gamma}{2} t^2 n + i\beta t n} \right| dt &= \int_{|t| \leq n^{-q}} |O(t) + nO(t^3)|\ e^{-\frac{\gamma}{2} t^2 n} dt + O(\varepsilon^n) \\ &= \ O\left(\left[-\tfrac{e^{-\frac{\gamma}{2} t^2 n}}{\gamma n} \right]_0^{n^{-q}} + n \left[-\tfrac{e^{-\frac{\gamma}{2} t^2 n}}{\gamma n} (t^2 + \tfrac{2}{\gamma n}) \right]_0^{n^{-q}} \right) + O(\varepsilon^n) \ = O\left(n^{-1}\right) \end{aligned}$$

Going back to our main goal, recall that the probability $p_n(k) = \Pr\{Y_n = k\}$ can be obtained from the inversion formula (1) and, to evaluate the integral therein, we can split $[-\pi, \pi]$ into the three sets defined in (15). Then, by equalities (16) and (17), for some $\varepsilon \in (0,1)$ we obtain

$$p_n(k) \;=\; \frac{1}{2\pi}\int_{|t|\le n^{-q}} \Psi_n(t)e^{-itk}dt + O\left(e^{-\frac{\gamma}{4}n^{1-2q}}\right) + O(\varepsilon^n) \tag{20}$$

Moreover, by relation (18), defining the variable $v(= v_{k,n}) = \frac{k-\beta n}{\sqrt{\gamma n}}$, we have

$$\begin{aligned}\int_{|t|\le n^{-q}} \Psi_n(t)e^{-itk}dt &= \int_{|t|\le n^{-q}} e^{-\frac{\gamma}{2}t^2 n + i\beta t n}e^{-itk}dt + O(n^{-1}) \\ &= \int_{|t|\le n^{-q}} e^{-\frac{\gamma}{2}t^2 n - itv\sqrt{\gamma n}}dt + O(n^{-1}) \end{aligned} \tag{21}$$

Now, setting $t\sqrt{\gamma n} = x$, the last integral becomes

$$\begin{aligned}\int_{|t|\le n^{-q}} e^{-\frac{\gamma}{2}t^2 n - itv\sqrt{\gamma n}}dt &= \frac{1}{\sqrt{\gamma n}}\int_{|x|\le n^{\frac{1}{2}-q}\sqrt{\gamma}} e^{-ivx-\frac{x^2}{2}}dx \\ &= \tfrac{1}{\sqrt{\gamma n}}\left[\textstyle\int_{-\infty}^{+\infty} e^{-ivx-\frac{x^2}{2}}dx - \int_{|x|>n^{\frac{1}{2}-q}\sqrt{\gamma}} e^{-ivx-\frac{x^2}{2}}dx\right] \\ &= \tfrac{1}{\sqrt{\gamma n}}\left[\sqrt{2\pi}\; e^{-\frac{v^2}{2}} + O\left(\textstyle\int_{\sqrt{\gamma}n^{\frac{1}{2}-q}}^{+\infty} x e^{-\frac{x^2}{2}}dx\right)\right] \\ &= (\gamma n)^{-1/2}\left(\sqrt{2\pi}\; e^{-\frac{v^2}{2}} + O(e^{-\frac{\gamma}{2}n^{1-2q}})\right) \end{aligned} \tag{22}$$

Thus the result follows by replacing (22) in (21) and (21) in (20). □

4 Equipotent sum models

Now, let us study the local limit properties of our statistics for non-communicating bicomponent models in the equipotent case. More precisely, let $\{Y_n\}$ be defined by a linear representation (ξ, A, B, η) satisfying the above conditions 1, 2, 3. Assume $\lambda_1 = \lambda_2 = \lambda$ and let both matrices M_1, M_2 be aperiodic (and hence primitive). Under these hypotheses we say that $\{Y_n\}$ is defined in an *equipotent sum model*. The limit distribution of $\{Y_n\}$ in this case is studied in [7] and depends on the parameters α_1, α_2, β_1, β_2, γ_1, γ_2. Here we extend those results to local limit properties, with a convergence rate $O(n^{-1/2})$, under the further assumption that both pairs (A_1, B_1) and (A_2, B_2) are aperiodic. To this end we first determine some identities for function $h_n(z)$ in the present case, using the notation introduced in Section 2.

From the properties of primitive matrices [16] it is easy to see that

$$\begin{aligned} h_n(0) = \xi' M^n \eta &= \xi_1'\nu_1\zeta_1'\eta_1 \cdot \lambda^n + \xi_2'\nu_2\zeta_2'\eta_2 \cdot \lambda^n + O(\rho^n) \\ &= (\alpha_1 + \alpha_2)\lambda^n + O(\rho^n) \qquad 0 \le \rho < \lambda \end{aligned}$$

where α_j's, for $j = 1, 2$, are defined in (8). Also note that $\alpha_j = r_j(0)$ for each j. Using these facts function $\Psi_n(t)$ can be evaluated from (14).

Here we obtain a local limit law of a type depending on the parameters α_1, α_2, β_1, β_2, γ_1, γ_2. In particular, if $\beta_1 = \beta_2$ or $\gamma_1 = \gamma_2$ we again obtain a local limit law of Gaussian type, similar to Theorem 1 (independent of α_1, α_2). Otherwise we get a convex combination of two Gaussian distribution.

Theorem 2. *Let $\{Y_n\}$ be defined in an equipotent sum model and assume that both pairs (A_1, B_1), (A_2, B_2) are aperiodic. Then, as n tends to $+\infty$, the relation*

$$\left| \sqrt{n} Pr(Y_n = k) \; - \; \left(\frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{e^{-\frac{(k-\beta_1 n)^2}{2\gamma_1 n}}}{\sqrt{2\pi\gamma_1}} + \frac{\alpha_2}{\alpha_1 + \alpha_2} \frac{e^{-\frac{(k-\beta_2 n)^2}{2\gamma_2 n}}}{\sqrt{2\pi\gamma_2}} \right) \right| \; = \; O\left(n^{-1/2}\right)$$

holds true uniformly for every $k \in \{0, 1, \ldots, n\}$.

Proof. Again the main idea is to study the characteristic function $\Psi_n(t)$ for $t \in [-\pi, \pi]$ by splitting this interval into the three sets given in (15), where $c \in (0, \pi)$ is a constant satisfying relations (9) for both $y_1(t)$ and $y_2(t)$, and q is an arbitrary value such that $\frac{1}{3} < q < \frac{1}{2}$. The behaviour of $\Psi_n(t)$ in these sets is characterized by the following properties:

d. For some $\varepsilon \in (0, 1)$ we have

$$|\Psi_n(t)| = O(\varepsilon^n) \qquad \forall\, t \in \mathbb{R} \;:\; c < |t| \leq \pi \tag{23}$$

e. There exists $a > 0$ such that

$$|\Psi_n(t)| = O\left(e^{-an^{1-2q}}\right) \qquad \forall\, t \in \mathbb{R} \;:\; n^{-q} < |t| \leq c \tag{24}$$

f. $\int_{|t| \leq n^{-q}} \left| \Psi_n(t) - \frac{\alpha_1}{\alpha_1+\alpha_2} e^{-\frac{\gamma_1}{2} t^2 n + i\beta_1 tn} - \frac{\alpha_2}{\alpha_1+\alpha_2} e^{-\frac{\gamma_2}{2} t^2 n + i\beta_2 tn} \right| dt \; = \; O(n^{-1})$ (25)

Proof of (23). The reasoning is similar to proving relation (16). The only difference is that now also the eigenvalues of $A_2 e^{it} + B_2$ are smaller than $\lambda = \lambda_2$, and this actually simplifies the argument.

Proof of (24). By relation (7), for some $\varepsilon \in (0, 1)$ and all $t \in \mathbb{R}$ satisfying $|t| \leq c$, we have

$$\Psi_n(t) = \frac{h_n(it)}{h_n(0)} = \frac{r_1(it)u_1(it)^n + r_2(t)u_2(it)^n}{(r_1(0) + r_2(0))\lambda^n} + O(\varepsilon^n) = \sum_{j=1,2} c_j y_j(t)^n + O(\varepsilon^n) \tag{26}$$

where c_1 and c_2 are positive constants. Also, setting $a = \min\{\gamma_1/4, \gamma_2/4\}$, by inequality (9) recalling $n^{-q} \leq |t| \leq c$ we obtain $|y_j(t)|^n \leq e^{-an^{1-2q}}$, for each $j = 1, 2$. Thus, the result follows by replacing this bound in the previous relation.

Proof of (25). From equality (26), applying relation (10) and recalling that $nO(t^3) = o(1)$ for $|t| \leq n^{-q}$, in the same interval for t we get

$$\Psi_n(t) = \sum_{j=1,2} \frac{r_j(0) + O(t)}{r_1(0) + r_2(0)} (1 + nO(t^3)) e^{-\frac{\gamma_j}{2} t^2 n + i\beta_j tn} + O(\varepsilon^n)$$

Thus, since $r_j(0) = \alpha_j$ for each j, reasoning as in the proof of (18) we obtain

$$\int_{|t|\le n^{-q}} \left| \Psi_n(t) - \sum_{j=1,2} \frac{\alpha_j}{\alpha_1+\alpha_2} e^{-\frac{\gamma_j}{2}t^2 n + i\beta_j t n} \right| dt =$$
$$= \sum_{j=1,2} \int_{|t|\le n^{-q}} |O(t) + nO(t^3)|\, e^{-\frac{\gamma_j}{2}t^2 n} dt + O(\varepsilon^n) = O\left(n^{-1}\right) \qquad (27)$$

Now consider our main goal. Defining $p_n(k) = \Pr\{Y_n = k\}$, from the inversion formula (1), by relations (23), (24) and (25), we obtain

$$p_n(k) = \frac{1}{2\pi} \int_{|t|\le n^{-q}} \Psi_n(t) e^{-itk} dt + O\left(e^{-an^{1-2q}}\right) + O(\varepsilon^n)$$
$$= \frac{1}{2\pi} \sum_{j=1,2} \frac{\alpha_j}{\alpha_1+\alpha_2} \int_{|t|\le n^{-q}} e^{-\frac{\gamma_j}{2}t^2 n + i\beta_j t n - itk} dt + O(n^{-1}) \qquad (28)$$

Moreover, defining the variables $v_j = \frac{k-\beta_j n}{\sqrt{\gamma_j n}}$, for $j = 1, 2$, the last integrals can be evaluated as in (21) and (22), obtaining

$$\int_{|t|\le n^{-q}} e^{-\frac{\gamma_j}{2}t^2 n + i\beta_j t n - itk} dt = \frac{1}{\sqrt{\gamma_j n}} \left(\sqrt{2\pi}\, e^{-\frac{v_j^2}{2}} + O(e^{-\frac{\gamma_j}{2} n^{1-2q}}) \right)$$

which replaced in (28) yields the result. □

We observe at once that if $\beta_1 = \beta_2$ and $\gamma_1 = \gamma_2$ then the limit density given in Theorem 2 reduces to a Gaussian law. Then, we can state the following

Corollary 3. *Let $\{Y_n\}$ be defined in an equipotent sum model with $\beta_1 = \beta_2 = \beta$, $\gamma_1 = \gamma_2 = \gamma$ and assume aperiodic both pairs (A_1, B_1), (A_2, B_2). Then, as n tends to $+\infty$, the relation*

$$\left| \sqrt{n} Pr(Y_n = k) \; - \; \frac{e^{-\frac{(k-\beta n)^2}{2\gamma n}}}{\sqrt{2\pi\gamma}} \right| \; = \; O\left(n^{-1/2}\right)$$

holds true uniformly for every $k \in \{0, 1, \ldots, n\}$.

On the contrary, when $\beta_1 \neq \beta_2$ or $\gamma_1 \neq \gamma_2$ the previous result yields a local limit law toward a convex combination of two Gaussian distributions that differ by their mean value or by their variance. More precisely, in this case the limit distribution obtained by Theorem 2 is that one of a r.v. $\mathcal{L}$ of the form

$$\mathcal{L} = [\beta_1 \mathcal{B}_p + \beta_2(1 - \mathcal{B}_p)]\, n + [\mathcal{B}_p \mathcal{N}_{0,\gamma_1} + (1 - \mathcal{B}_p)\mathcal{N}_{0,\gamma_2}]\, \sqrt{n}$$

where $\mathcal{B}_p$ is a Bernoullian r.v. of parameter $p = \frac{\alpha_1}{\alpha_1+\alpha_2}$, and $\mathcal{N}_{0,\gamma_j}$ is a Gaussian r.v. of mean 0 and variance γ_j, assuming mutually independent all variables $\mathcal{B}_p$, $\mathcal{N}_{0,\gamma_1}$, $\mathcal{N}_{0,\gamma_2}$. In particular it is clear from the analysis of $\Psi_n(t)$ that

$$\frac{Y_n - [\beta_1 \mathcal{B}_p + \beta_2(1 - \mathcal{B}_p)]\, n}{\sqrt{n}} \text{ tends in distribution to } [\mathcal{B}_p \mathcal{N}_{0,\gamma_1} + (1 - \mathcal{B}_p)\mathcal{N}_{0,\gamma_2}]$$

A curious fact in this case is that $\mathcal{L}$ also depends on the weights of initial and final states (ξ, η). This does not occur in any other bicomponent model and seems to state that the model is not ergodic.

As an example, consider the rational model defined by the weighted finite automaton of Figure 1, together with $\xi = (1, 0, 1, 0)$ and $\eta = (0, 1, 1, 1)$. Such an automaton recognizes the set $\{w \in \{a, b\}^* \mid w$ has not pattern aa or pattern $bb\}$. Clearly this is a bicomponent model, with both pairs (A_1, B_1) and (A_2, B_2) aperiodic. Moreover $M_1 = M_2$, while $A_1 \neq A_2$. Hence the two components are equipotent and $\beta_1 \neq \beta_2$. This implies a local limit law towards a convex combination of two Gaussian laws. Note that simple changes may modify the limit distribution: for instance, setting to 2 the weight of transition $2 \xrightarrow{b} 1$ makes dominant the first component, implying a Gaussian local limit law (Theorem 1).

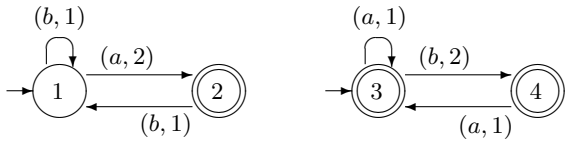

Fig. 1. Weighted finite automaton defining a non-communicating bicomponent model with $\lambda_1 = \lambda_2 = 2$, $\alpha_1 = 2/3$, $\alpha_2 = 4/3$, $\beta_1 = 1/3$, $\beta_2 = 2/3$, $\gamma_1 = \gamma_2 = 2/27$.

5 Conclusions

The analysis of local limit laws of symbol statistics Y_n's presented in this work concerns the rational stochastic models consisting of two primitive components without communications. An analogous study has been presented in [10] for the case when there is some transition from the first to second component.

It is interesting to discuss similarities and differences between the results obtained in the two cases. Note that in both cases a dominant component implies a Gaussian local limit law, with a convergence rate $O(n^{-1/2})$, at the cost of adding an aperiodicity condition only for the main component. The same occurs in the equipotent models when both parameters of mean value and variance for the two components are equal ($\beta_1 = \beta_2$, $\gamma_1 = \gamma_2$). On the contrary, if one of the parameters is different then here we obtain a limit distribution of the form $\mathcal{L}$ given above, while in [10] a uniform r.v. or a continuous mixture of Gaussian laws are obtained according to whether $\beta_1 \neq \beta_2$ or $\beta_1 = \beta_2$, $\gamma_1 \neq \gamma_2$. Another difference is that in [10] no limit distribution depends on the weights of initial and final states. Intuitively those models are "ergodic". On the contrary in the present work $\mathcal{L}$ depends on ξ and η and hence the model is not "ergodic". Studying reasons and further consequences of these differences seems to be a natural goal for future investigations.

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, V. Lonati. On the number of occurrences of a symbol in words of regular languages. *Theoret. Comput. Sci.*, 302:431–456, 2003.
4. A. Bertoni, C. Choffrut, M. Goldwurm, V. Lonati. Local limit properties for pattern statistics and rational models. *Theory Comput. Systems*, 39:209–235, 2006.
5. S. Broda, A. Machiavelo, N. Moreira, and R. Reis. A hitchhiker's guide to descriptional complexity through analytic combinatorics. *Theoret. Comput. Sci.*, 528:85–100, 2014.
6. J.R. Cannon. *The one-dimensional Heat Equation*. Encyclopedia of Mathematics and its Applications, vol. 23, Addison–Wesley Publishing Company, 1984.
7. D. de Falco, M. Goldwurm, V. Lonati. Frequency of symbol occurrences in bicomponent stochastic models. *Theoret. Comput. Sci.*, 327 (3):269–300, 2004.
8. P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge Univ. Press, 2009.
9. B.V. Gnedenko. *Theory of probability*. Gordon and Breach Science Publ., 1997.
10. M. Goldwurm, J. Lin, M. Vignati. Analysis of symbol statistics in bicomponent rational models. Accepted for presentation at DLT 2019, 23rd International Conference on Developments in Language Theory, to appear in LNCS, Springer. Available at http://users.mat.unimi.it/users/goldwurm .
11. P. Grabner, M. Rigo. Distribution of additive functions with respect to numeration systems on regular languages. *Theory Comput. Systems*, 40:205-223, 2007.
12. H.-K. Hwang. On convergence rates in the Central Limit Theorem for combinatorial structures. *Europ. J. Combinatorics*, 19:329–343, 1998.
13. P. Nicodeme, B. Salvy, and P. Flajolet. Motif statistics. *Theoret. Comput. Sci.*, 287(2): 593–617, 2002.
14. M. Régnier and W. Szpankowski. On pattern frequency occurrences in a Markovian sequence. *Algorithmica*, 22 (4):621–649, 1998.
15. A. Salomaa and M. Soittola. *Automata-Theoretic Aspects of Formal Power Series*. Springer–Verlag, 1978.
16. E. Seneta. *Non-negative matrices and Markov chains*. Springer–Verlag, New York Heidelberg Berlin, 1981.