Some Properties of the Rate Functions of Large Deviations for Symbol Statistics

Massimiliano Goldwurm[1,*,†], Marco Vignati[1,†]

[1]Dipartimento di Matematica "Federigo Enriques", Università degli Studi di Milano, via Cesare Saldini 50, 20133 Milano

Abstract

We present some results concerning the rate functions of large deviations for symbol statistics defined in primitive rational models. In our context these functions are always defined over an open interval $(U, V) \subseteq (0, 1)$. We first prove certain properties of symmetry for such rate functions that allow us to simplify subsequent investigation. Then we show that the limits of these functions at the endpoints U and V are always finite and we prove that, under rather mild conditions, the same U and V are rational numbers of the form i/j such that $i, j \in \{0, 1, \ldots, d\}$, where d is the number of states of the generalized automaton defining the rational model. Under the same hypotheses we yield a precise value for the corresponding limits.

Keywords

automata and formal languages, large deviations, pattern statistics, rational series, regular languages

1. Introduction

In this work we study the properties of large deviations for symbol statistics in rational models. Such statistics represent the number of occurrences of a letter in a word generated at random among all strings of length n, over a binary alphabet, with a probability proportional to the values of a given rational formal power series in non-commutative variables [2, 9, 3, 17]. These probabilistic models can be completely defined by a generalized automaton equipped with positive weights [27]. The study of symbol statistics in rational models is related to classical research topics in the areas of formal languages, descriptional complexity, random generation and analysis of combinatorial structures [13, 4, 6, 20, 8, 15]. Moreover, they are naturally related to the analysis of pattern statistics. Symbol statistics in rational models can represent a large variety of pattern statistics on words in several probabilistic contexts, this occurs for instance when the set of patterns is given by a regular language and the random text is generated by a Markovian source [2, 23, 18]. Typical results on pattern statistics concern the asymptotic evaluation of the moments, the limit distributions and also the properties of large deviations [24, 16, 15, 22]. In particular large deviations estimates are studied in order to evaluate the probability of rare events, when one or more patterns are over- or under-represented in a random text [5, 12, 22].

ICTCS'24: Italian Conference on Theoretical Computer Science, September 11–13, 2024, Torino, Italy

*Corresponding author.

†These authors contributed equally.

massimiliano.goldwurm@unimi.it (M. Goldwurm); marco.vignati@unimi.it (M. Vignati)

http://users.mat.unimi.it/users/goldwurm/ (M. Goldwurm); http://users.mat.unimi.it/users/vignati/ (M. Vignati)

0000-0003-0903-4699 (M. Goldwurm); 0000-0002-4762-0915 (M. Vignati)

In the present contribution we continue the investigation started in [19], where we proved a property of large deviations for any symbol statistics defined in a rational model assuming that the overall matrix of the weights of transitions is primitive. Here, we first prove as an auxiliary tool some properties of symmetry of the rate functions occurring in these large deviations estimates. It is known that these functions are well defined in an open interval (U, V) included in $(0, 1)$. Then, we show that the limits of the rate functions at the endpoints U and V are always finite. Moreover, under a rather mild condition, we prove that such extremes are rational numbers of the form i/j such that $i \leq j$ and $i, j \in \{0, 1, \ldots, d\}$, where d is the size of the model, i.e. the number of states of the generalized automaton defining the model. Under the same hypotheses we also determine the above limits, depending on the coefficients of the characteristic polynomial of the matrix of weights. We conjecture that these results can be extended to all rational models having a primitive matrix of weights.

2. Large deviation properties

The large deviations are typical properties of a sequence of random variables, say $\{X_n\}$, that have increasing mean values. They consist of a bound exponentially decreasing to 0 over the probability that X_n deviates from $E(X_n)$ by an amount greater or equal to $cE(X_n)$, for any $c > 0$. The main situations occur when $E(X_n) \sim \beta n$ for a constant $\beta > 0$, and since this occurs in the contexts considered in this work, here we refer to the following formal definition, inspired by [11, 15], which is rather restrictive with respect to the usual general setting [10].

Definition 1. *Consider a sequence of random variables $\{X_n\}$ such that $E(X_n) = \beta n + o(n)$ for a constant $\beta > 0$, and let (x_0, x_1) be a real interval including β. Assume $F(x)$ is a function defined over (x_0, x_1) taking values in $\mathbb{R}$, such that $F(x) > 0$ for $x \neq \beta$. We say that $\{X_n\}$ satisfies a* large deviation property *in the interval (x_0, x_1) with rate function $F(x)$ if the following limits hold:*

$$\lim_{n\to\infty} \frac{1}{n} \log Pr(X_n \leq xn) = -F(x) \qquad \text{for } x_0 < x \leq \beta$$

$$\lim_{n\to\infty} \frac{1}{n} \log Pr(X_n \geq xn) = -F(x) \qquad \text{for } \beta \leq x < x_1$$

This property is equivalent to require that $\Pr(X_n \leq xn) = e^{-F(x)n+o(n)}$, for $x_0 < x \leq \beta$, and $\Pr(X_n \geq xn) = e^{-F(x)n+o(n)}$, for $\beta \leq x < x_1$. Such a property holds for several quantities of interest in the analysis of various combinatorial structures [14, 7, 15, 21].

The rate functions we encounter in this work are convex over their interval of definition and enjoy the following property, which often occurs in the study of large deviations [10]. For any open interval $(a, b) \subseteq \mathbb{R}$, we say that a convex function $F : (a, b) \longrightarrow \mathbb{R}$ is *essentially smooth* if F is differentiable in (a, b), $\lim_{x\to a^+} F'(x) = -\infty$ and $\lim_{x\to b^-} F'(x) = +\infty$.

3. Symbol statistics in rational models

In order to define the rational stochastic models, consider a *formal series* in the non-commutative variables a, b, that is a function $r : \{a, b\}^* \to \mathbb{R}_+$, where $\mathbb{R}_+ = [0, +\infty)$ and $\{a, b\}^*$ is the free monoid of all words in the alphabet $\{a, b\}$. As usual, we denote by (r, w) the value of r at a word $w \in \{a, b\}^*$. Such a series r is said to be *rational* if for some integer $d > 0$ there exists a monoid morphism $\mu : \{a, b\}^* \to \mathbb{R}_+^{d\times d}$ and two (column) arrays $\xi, \eta \in \mathbb{R}_+^d$, such that $(r, w) = \xi'\mu(w)\eta$, for every $w \in \{a, b\}^*$ [1, 25] [1]. Note that in this case, if $w = w_1 w_2 \cdots w_n$ with $w_i \in \{a, b\}$ for every $i = 1, 2, \ldots, n$, then $\mu(w) = \mu(w_1)\mu(w_2)\cdots\mu(w_n)$. Thus, as μ is generated by the 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 d. Such a 4-tuple can be considered as a *generalized automaton* over the alphabet $\{a, b\}$ as defined in [27], where $\{1, 2\ldots, d\}$ is the set of states and the weights of the transitions are positive real, as well as the weights of the initial and final states. In particular, the matrix A (resp. B) represents the weights of the transitions labelled by a (resp. b), while ξ (resp. η) represents the weights of the initial (resp. final) states.

To avoid trivial situations, denoting by $\{a, b\}^n$ the family of all words of length n in $\{a, b\}^*$, we assume that the set $\{w \in \{a, b\}^n : (r, w) > 0\}$ contains at least two elements, for every $n \in \mathbb{N}_+$. As a consequence, since $\sum_{w\in\{a,b\}^n}(r, w) = \xi'\,(A + B)^n\,\eta$, we have $\xi \neq \underline{0} \neq \eta$ and both A and B are non-null matrices (i.e., each of them has at least one positive entry). Moreover, we can consider the probability measure Pr over the set $\{a, b\}^n$ given by

$$\Pr(x) = \frac{(r, x)}{\sum_{w\in\{a,b\}^n}(r, w)} = \frac{\xi'\mu(x)\eta}{\xi'\,(A + B)^n\,\eta}\,, \qquad \forall\, x \in \{a, b\}^n \tag{1}$$

Note that, if r is the characteristic series of a (regular) language $L \subseteq \{a, b\}^*$ then Pr is the uniform probability function over the set $L\cap\{a, b\}^n$. Also observe that the traditional Markovian models (to generate a word at random in $\{a, b\}^*$) occur when $A + B$ is a stochastic matrix, ξ is a stochastic array and $\eta' = (1, 1\ldots, 1)$ (for a comparison between the rational models and different types of Markovian models, one may refer to [18], while for their relations with stochastic automata one may consider [27]).

Then, under the previous hypotheses, we can define the integer random variable (r.v.) $Y_n = |w|_a$, where w is chosen at random in $\{a, b\}^n$ with probability $\Pr(w)$ and $|w|_a$ denotes the number of occurrences of a in w. By our assumptions, Y_n is a non-degenerate random variable. One can easily prove [2] that the probability function of Y_n is given by

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

where $[x^k]g(x)$ denotes as usual the coefficient of the monomial of degree k in an arbitrary polynomial $g(x) \in \mathbb{R}[x]$. For sake of brevity we say that Y_n is *defined* by the linear representation (ξ, A, B, η). As shown in [2, 9, 3, 17], its traditional properties, as mean value, variance, limit distributions and local limit properties, can be studied by using its moment generating function $\Psi_n(z) = \sum_{k=0}^n p_n(k)e^{zk} = \frac{\xi'(Ae^z+B)^n\eta}{\xi'(A+B)^n\eta}$ $(z \in \mathbb{C})$.

4. Large deviations for primitive models

In this section we summarize the large deviation properties of Y_n when the matrix $A + B$ is primitive [19]. Here and in the subsequent sections we denote by m_{ij}, a_{ij} and b_{ij} the entries of

[1] As usual, we denote by v' the transpose of an array $v \in \mathbb{R}^d$, i.e. a row array.

indices i, j of the matrices M, A and B, respectively. Recall that a square matrix $M \in \mathbb{R}_+^{d\times d}$ is *primitive* if there exists a positive integer n such that $M^n > 0$, i.e. all entries of M^n are strictly positive. It is well-known that M is primitive if and only it is irreducible and aperiodic [26]. Its main properties are established by the well-known Perron-Frobenius Theorem (see for instance [26, Sec 1.1]) stating that every primitive matrix $M \in \mathbb{R}_+^{d\times d}$ admits a real eigevalue $\lambda > 0$ (called the Perron-Frobenius eigenvalue of M) such that $|\mu| < \lambda$ for any eigenvalue μ of M different from λ, λ is a simple root of the characteristic equation of M and the following property holds:

(i) If a matrix $A =\in \mathbb{R}_+^{d\times d}$ satisfies $A \leq M$ (i.e. $a_{ij} \leq m_{ij}$, $\forall i, j$) and α is an eigenvalue of A then $|\alpha| \leq \lambda$. Moreover, $|\alpha| = \lambda$ implies $A = M$.

Now (still assuming $A \neq 0 \neq B$) let $A + B$ be primitive and let λ be its Perron-Frobenius eigenvalue. In this case it is known that the sequence $\{Y_n\}$ has a Gaussian limit distribution [2] it satisfies a large deviation property [19] and also has several local limit properties [3, 17]. These results are maily obtained by studying the function $y = y(t)$, where $t \in \mathbb{R}$, implicitly defined by the equation

$$\det(Iy - Ae^t - B) = 0 \tag{2}$$

with initial condition $y(0) = \lambda$. Clearly $y(t)$ is analytic and well defined all over $\mathbb{R}$. Since also $Ae^t + B$ is primitive for every $t \in \mathbb{R}$, $y(t)$ is its Perron-Frobenius eigenvalue; thus, $y(t)$ is a positive real function, strictly increasing all over $\mathbb{R}$ (by statement **(i)** above), which implies $y'(t) > 0$ for any $t \in \mathbb{R}$.

Moreover, we can define the constant $\beta = \frac{y'(0)}{\lambda}$ and the function $\beta(t) = \frac{y'(t)}{y(t)}$, for $t \in \mathbb{R}$, which enjoy the following properties [2, 19]:

(a) $0 < \beta < 1$ and $\mathrm{E}(Y_n) = \beta n + c + O(\varepsilon^n)$, for some $|\varepsilon| < 1$ and some constant $c \in \mathbb{R}$;

(b) For every $t \in \mathbb{R}$ we have $0 < \beta(t) < 1$, $\beta'(t) > 0$, and hence the following limits exist and are finite:

$$U = \lim_{t\to-\infty} \beta(t), \qquad V = \lim_{t\to+\infty} \beta(t) \tag{3}$$

Also, these limits imply $0 \leq U < \beta(0) < V \leq 1$.

(c) For every $x \in (U, V)$ there exists a unique $\tau_x \in \mathbb{R}$ such that $\beta(\tau_x) = x$. Moreover, $\tau_x < 0$ whenever $x < \beta$, $\tau_\beta = 0$ and $\tau_x > 0$ when $x > \beta$.

By using the properties above the following result can be proved [19].

Theorem 1. *Let $\{Y_n\}$ be defined by a linear representation (ξ, A, B, η) where $A+B$ is primitive. Then $\{Y_n\}$ satisfies a large deviation property in the interval (U, V) with rate function*

$$G(x) = -\log\left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}}\right) \tag{4}$$

where τ_x is defined by sentence **(c)** *above. Moreover, G is analytic in the whole interval (U, V), where it is also convex and essentially smooth with a unique minimal value $G(\beta) = 0$.*

A natural question arising from the previous result is whether the interval (U, V) of definition of $G(x)$ coincides with interval $(0, 1)$, as it occurs in the classical case of binomial random variables [10, 11]. One may also ask what the behaviour of $G(x)$ is when x tends to the extremes of the interval. In [19] it is proved that $U = 0$ and $V = 1$ when the main eigenvalues of the matrices A and B are nonnull. A simple check of the proof allows us to strengthen the result, making independent the two conditions that imply, respectively, $U = 0$ and $V = 1$. More precisely, since A and B are non-zero matrices with entries in $\mathbb{R}_+$, both of them admit a real non-negative eigenvalue, we denote by λ_A and λ_B, respectively, that are greater or equal to the modulus of any other eigenvalue. Clearly, it may happen $\lambda_A = 0$ or $\lambda_A = |\mu|$ for some eigenvalue μ of A different from λ_A, and the same may occur for λ_B. However, again by statement (i), we have $\lambda_A < \lambda$ and $\lambda_B < \lambda$. Then, from [19] one can prove the following property.

Theorem 2. *Under the hypotheses of Theorem 1 the following properties hold:*
1) If $\lambda_B > 0$ then $U = 0$ and $\lim_{x\to 0^+} G(x) = \log(\lambda/\lambda_B)$ (whatever λ_A is);
2) If $\lambda_A > 0$ then $V = 1$ and $\lim_{x\to 1^-} G(x) = \log(\lambda/\lambda_A)$ (whatever λ_B is).

This result leaves open the problem of finding the values of U and V, respectively when $\lambda_B = 0$ and $\lambda_A = 0$. In these cases, also the limits $\lim_{x\to U^+} G(x)$ and $\lim_{x\to V^-} G(x)$ are not established. We only know, by Theorem 1, that $0 \le U < V \le 1$ and that $|G'(x)| \to +\infty$ as x approaches U or V.

5. Properties of symmetry

In this section we present a natural property of symmetry between rate functions in our context. Let $A, B \in \mathbb{R}_+^{d\times d}$ be two non-null matrices (i.e. $A \ne [0] \ne B$). Clearly, for every pair of values $y \in \mathbb{C}$ and $t \in \mathbb{R}$, we have $\det(Iy - Ae^t - B) = 0$ if and only if $\det(Iye^{-t} - Be^{-t} - A) = 0$. As a consequence, if $y = y(t)$ and $v = v(u)$ are functions defined, respectively, by the equations $\det(Iy - Ae^t - B) = 0$ and $\det(Iv - Be^u - A) = 0$, with the same initial condition $y(0) = \lambda = v(0)$, then $v(u) = e^u y(-u)$ for every $u \in \mathbb{R}$.

Theorem 3. *Given two non-null matrices $A, B \in \mathbb{R}_+^{d\times d}$, assume that $A + B$ is primitive. Let $G(x)$ and $F(x)$ be the rate functions of the symbol statistics defined, respectively, by (ξ, A, B, η) and (ξ, B, A, η), and suppose $G(x)$ defined over the interval (U, V), for some $0 \le U < V \le 1$. Then the following properties hold:*

1. *$F(x)$ is defined over the interval (U', V') such that $U' = 1 - V$ and $V' = 1 - U$.*
2. *For every $x \in (U', V')$, $F(x) = G(1 - x)$.*
3. *$\lim_{x\to U'^+} F(x) = \lim_{x\to V^-} G(x)$ and $\lim_{x\to V'^-} F(x) = \lim_{x\to U^+} G(x)$.*

Proof. Clearly the third property is consequence of the other two. To prove the first one, let $y(t)$ and $v(u)$ be defined as above. According with the notation of Section 4, set $\beta(t) = \frac{y'(t)}{y(t)}$ and $\overline{\beta}(u) = \frac{v'(u)}{v(u)}$. As stated above, we know that $v(u) = e^u y(-u)$ for every $u \in \mathbb{R}$, and hence $v'(u) = e^u[y(-u) - y'(-u)]$, which implies $\overline{\beta}(u) = 1 - \beta(-u)$. Thus, by relations (3) the last equality shows that $U' = \lim_{u\to-\infty} \overline{\beta}(u) = \lim_{u\to-\infty} 1-\beta(-u) = 1-V$ and, similarly, $V' = \lim_{u\to+\infty} \overline{\beta}(u) = \lim_{u\to+\infty} 1-\beta(-u) = 1-U$. These equalities prove Property 1) since, by Theorem 1, $F(x)$ is defined over the interval (U', V').

To prove Property 2, let $x \in (U', V')$ and set $z = 1-x$. By sentence **(c)** of Section 4, there exists a unique value $\tau_z \in \mathbb{R}$ such that $\beta(\tau_z) = z$. Analogously, there is a unique $u_x \in \mathbb{R}$ such that $\overline{\beta}(u_x) = x$ and, since $\overline{\beta}(u) = 1-\beta(-u)$, we get $u_x = -\tau_z$. Thus, by relation (4), we have

$$\begin{aligned} F(x) &= -\log v(u_x) + \log\lambda + xu_x = -\log e^{u_x} y(-u_x) + \log\lambda + (1-z)u_x = \\ &= -\log y(\tau_z) + \log\lambda + z\tau_z \;=\; G(z) \qquad \square \end{aligned}$$

6. On the limits of $G(x)$ at the endpoints U and V

In this section we study the behaviour of $G(x)$ for x tending to V or U under the assumptions of Theorem 1, without restrictions on the matrices A and B, so including the case $\lambda_A = 0$ and $\lambda_B = 0$. Here we show that under these hypotheses such limits are always finite. In the next section, under more restrictive assumptions, we yield a precise expression of their values.

To present the result in detail, consider the family of real coefficients $\{c_{hj} : h \in \{1, 2\ldots, d\}, j \in \{0, 1, \ldots, h\}\}$ defined by the equality

$$\det(Iy - Ae^t - B) = y^d - \sum_{h=1}^{d}\sum_{j=0}^{h} c_{hj} e^{tj} y^{d-h} \tag{5}$$

Observe that some of them yield the characteristic polynomials of both A and B:

$$\det(Iy - A) = y^d - \sum_{h=1}^{d} c_{hh} y^{d-h} \quad , \qquad \det(Iy - B) = y^d - \sum_{h=1}^{d} c_{h0} y^{d-h} \tag{6}$$

Since some of these coefficients may be null, we define the following sets:

$$\begin{aligned} \mathcal{L} &:= \{(h,j) \,:\, h \in \{1, 2\ldots, d\}, j \in \{0, 1, \ldots, h\}, c_{hj} \neq 0\} \,; \\ \mathcal{Q} &= \mathcal{Q}(V) \;:=\; \{q \in \mathbb{R} \,:\, \exists (h,j) \in \mathcal{L} \text{ such that } q = j - Vh\} \,; \\ \mathcal{L}_q &:= \{(h,j) \in \mathcal{L} \,:\, j - Vh = q\} \,, \qquad \text{for every } q \in \mathcal{Q} \,. \end{aligned}$$

Clearly $\mathcal{L}$ and $\mathcal{Q}$ are not empty, and the family $\{\mathcal{L}_q : q \in \mathcal{Q}\}$ is a partition of $\mathcal{L}$. Moreover, we can also define $\overline{q} := \max\{q \in \mathcal{Q}\}$ and $\underline{q} := \min\{q \in \mathcal{Q}\}$. Thus, up to a division by y^d, equation (2) can be written in the form

$$1 = \sum_{(h,j)\in\mathcal{L}} c_{hj} e^{tj} y^{-h} \tag{7}$$

Now, let us study the properties of the previous equation when t tends to $+\infty$ in order to determine V and the limit $\lim_{x\to V^-} G(x)$. Consider the function $\beta(t) = y'(t)/y(t)$ introduced in Section 4, where $t \in \mathbb{R}$. From the second relation of (3) we deduce that, for $t \to +\infty$, $\frac{d}{dt}\log y(t) = \beta(t) = V - \varepsilon(t)$, where $\varepsilon(t) \to 0^+$ monotonically. This implies

$$\log y(t) - \log\lambda = \int_0^t \frac{d}{dx}\log y(x)dx = Vt - \int_0^t \varepsilon(z)dz = Vt - \eta(t)$$

where $\eta(t) = \int_0^t \varepsilon(z)dz$. Note that $\eta(0) = 0$ and $\eta(t) > 0$ for any $t > 0$; more than that, $\eta(t)$ is strictly increasing in $\mathbb{R}_+$ and one easily sees that $\eta(t) = o(t)$ for $t \to +\infty$. As a consequence, either $\lim_{t\to+\infty} \eta(t) = +\infty$ or $\lim_{t\to+\infty} \eta(t) = c^-$ for some $c > 0$. Moreover, from the last equalities we get

$$y(t) = \lambda e^{Vt-\eta(t)} \tag{8}$$

Replacing this expression in equation (7) and using the sets defined above we get

$$1 \equiv \sum_{q\in\mathcal{Q}} e^{tq} \sum_{(h,j)\in\mathcal{L}_q} c_{hj}\lambda^{-h}e^{h\eta(t)} \qquad\qquad (\forall\, t \in \mathbb{R})$$

Now it is convenient to define the polynomial $p_q(x) := \sum_{(h,j)\in\mathcal{L}_q} c_{hj}\lambda^{-h}x^h$, for every $q \in \mathcal{Q}$. Replacing it in the previous equation, we have

$$1 \equiv e^{t\overline{q}}p_{\overline{q}}(e^{\eta(t)}) + \sum_{q\in\mathcal{Q}\setminus\{\overline{q}\}} e^{tq}p_q(e^{\eta(t)})$$

and hence, for $t \to +\infty$, we obtain

$$1 = e^{t\overline{q}}\left[p_{\overline{q}}(e^{\eta(t)}) + \mathrm{O}(e^{-t\delta})\right] \tag{9}$$

for some $\delta > 0$. This proves $\overline{q} \geq 0$, otherwise since $\eta(t) = \mathrm{o}(t)$ the right hand side of the equality above would go to 0 for $t \to +\infty$. As a consequence, whatever the value of $\overline{q} \geq 0$ is, $\eta(t)$ cannot tend to $+\infty$ for $t \to +\infty$, otherwise the right hand side of the previous equality would not tend to 1. This implies $\lim_{t\to+\infty} \eta(t) = c^-$ for some $c > 0$, and hence $\varepsilon(t)$ is integrable at $+\infty$, that is $\int_0^{+\infty} \varepsilon(t)dt \in \mathbb{R}$, which yields the relation

$$\varepsilon(t) = \mathrm{o}(t^{-1}) \qquad \text{for } t \to +\infty \tag{10}$$

Moreover, setting the constant $C \in (0, 1)$ by

$$C := e^{-c} = \lim_{t\to+\infty} e^{-\eta(t)} \tag{11}$$

we can write the identity (8) in the form

$$y(t) = \lambda e^{Vt}(C + \delta(t)) \tag{12}$$

for some positive function $\delta(t)$ such that $\lim_{t\to+\infty} \delta(t) = 0^+$. Thus, relation (9) becomes

$$1 = e^{t\overline{q}}\left[p_{\overline{q}}(C^{-1}) + \mathrm{o}(1)\right] \tag{13}$$

Note that this does not necessarily imply $\overline{q} = 0$ because C^{-1} could be a root of $p_{\overline{q}}$ and, in that case, the equality above would prove $\overline{q} > 0$.

However, these equalities allow us to evaluate the limit of $G(x)$ for $x \to V^-$. Recall that $\beta(t) = V - \varepsilon(t)$ and, by property **c** of Section 4, $x = \beta(\tau_x)$ and $\lim_{x\to V^-} \tau_x = +\infty$. Thus, from equality (4), by relations (12) and (10), letting $x \to V^-$ we get

$$\begin{aligned} G(x) &= -\log y(\tau_x) + \log\lambda + x\tau_x = -\log(\lambda e^{V\tau_x}(C + \delta(\tau_x))) + \log\lambda + \beta(\tau_x)\tau_x = \\ &= -\log C - \log\left(1 + C^{-1}\delta(\tau_x)\right) - \varepsilon(\tau_x)\tau_x = -\log C + \mathrm{o}(1) \end{aligned} \tag{14}$$

As a consequence $\lim_{x\to V^-} G(x)$ is always finite. By applying Theorem 3 it is easy to verify an analogous property for U. Therefore we have proved the following result, which extends to the cases $\lambda_A = 0$ and $\lambda_B = 0$ the properties presented in Theorem 2.

Theorem 4. *Under the hypotheses of Theorem 1 both limits of $G(x)$ for $x \to U^+$ and for $x \to V^-$ are finite.*

The previous result is not taken for granted since, from Theorem 2, in case $\lambda_A = 0$ one might expect that the limit of $G(x)$ at V^- is $+\infty$, while we have just proved it is finite. The same holds for the limit at U^+ in case $\lambda_B = 0$.

7. Graphs, matrices and characteristic polynomials

In order to continue the analysis of the extremes U and V of the domain of $G(x)$, we present some properties of the coefficients c_{hj} introduced in (5). We first recall the traditional correspondence between non-negative matrices and weighted graphs.

For every matrix $M \in \mathbb{R}_+^{d\times d}$, let $G(M)$ be the weighted oriented graph with set of nodes $N = \{1, 2, \dots, d\}$, set of edges $E = \{(i, j) \in N \times N \mid m_{ij} > 0\}$ and weight m_{ij} for every $(i, j) \in E$. Further, for any cycle C in $G(M)$, the *weight* of C is defined as the product of the weights of its edges and is denoted by $w(C)$, while the *length* of C is just the number of its edges. A cycle is said to be *simple* if it does not cross any node twice. We denote by $\mathcal{C}_i$ the family of all simple cycles of length i in $G(M)$. Note that $\det(Iy - M) = y^d$ if and only if $G(M)$ has no cycles.

A further key property is that the determinant of M is a sum of products (with sign) of weights of simple cycles in $G(M)$. More precisely, let us denote by $\sharp u$ the cardinality of a set $u \subseteq N$. For every $h \in \{1, 2, \dots, d\}$, define $P_h := \{u \subseteq N : \sharp u = h\}$. Moreover, for every $u \in P_h$, set the submatrix $M_u := [m_{ij}]_{i,j\in u}$ and denote by S_u the family of all permutations of u. Note that any cyclic permutation of a subset $u \subseteq N$ can be interpreted as a simple cycle that crosses every node in u once. In general, any arbitrary permutation σ of a subset $u \subseteq N$ can be represented by a family $U(\sigma)$ of disjoint simple cycles over u, where every vertex in u just belongs to one cycle of $U(\sigma)$. Also observe that the sign of a permutation σ turns out to be $\mathrm{sgn}(\sigma) = (-1)^t$, where t is the number of (simple) cycles of even length in $U(\sigma)$. Thus, for every $u \in P_h$, if $\sigma \in S_u$ is a cyclic permutation then $\mathrm{sgn}(\sigma) = (-1)^{h-1}$. Moreover, for every permutation $\sigma \in S_u$, the value $\prod_{i\in u} m_{i\sigma(i)}$ is different from 0 if and only if all cycles in $U(\sigma)$ are subgraphs of $G(M)$ and, in this case, $\prod_{i\in u} m_{i\sigma(i)} = \prod_{C\in U(\sigma)} w(C)$. Thus, for every $h \in \{1, 2, \dots, d\}$ and $u \in P_h$, we may write

$$\det(M_u) = \sum_{\sigma\in S_u} \mathrm{sgn}(\sigma) \prod_{i\in u} m_{i\sigma(i)} = \sum_{\sigma\in S_u} \mathrm{sgn}(\sigma) \prod_{C\in U(\sigma)} w(C) \tag{15}$$

Then, from the same definition of determinant, one can deduce the following property:

Proposition 1. *Given a matrix $M \in \mathbb{R}_+^{d\times d}$ having a non-null eigenvalue, define the coefficients $a_1, a_2, \dots, a_d$ by the equality*

$$\det(Iy - M) = y^d - \sum_{h=1}^{d} a_h y^{d-h} \tag{16}$$

Then, for every $h \in \{1, 2, \dots, d\}$, we have

$$a_h = (-1)^{h-1} \sum_{u\in P_h} \det(M_u)$$

Moreover, setting $\breve{h} := \min\{h \in \{1, 2, \ldots, d\} : a_h \neq 0\}$*, it turns out that* $\breve{h}$ *is the smallest length of a cycle in* $G(M)$*,* $a_{\breve{h}} > 0$ *and* $a_{\breve{h}} = \sum_{C \in \mathcal{C}_{\breve{h}}} w(C)$*.*

Note in particular that $a_1 = \mathrm{tr}(M)$, i.e. the trace of M, and $a_d = (-1)^{d-1} \det(M)$.

The previous correspondence between matrices and graphs can be extended to pairs of non-negative matrices. Given two arbitrary matrices $A, B \in \mathbb{R}_+^{d \times d}$, let $G(A, B)$ be the labelled oriented graph having set of vertices $N = \{1, 2, \ldots, d\}$ and, for every pair $i, j \in N$, an edge from i to j of weight a_{ij}, labelled by a, whenever $a_{ij} > 0$, and an edge from i to j of weight b_{ij}, labelled by b, whenever $b_{ij} > 0$. Any path (or cycle) in $G(A, B)$ consists of consecutive edges, each of them labelled by either a or b. The weight of a cycle in $G(A, B)$ is the product of the weights of its edges.

By applying Proposition 1 to the matrix $Aw + B$, for an arbitrary variable w, we obtain

$$\det(Iy - Aw - B) = y^d - \sum_{h=1}^{d} (-1)^{h-1} \sum_{u \in P_h} \det(A_u w + B_u)\, y^{d-h} \tag{17}$$

Moreover, for every $u \in P_h$, setting $P_j(u) := \{v \subseteq u : \sharp v = j\}$ for any $j = 0, 1, \ldots, h$, by relation (15) and the previous arguments we have

$$\begin{aligned} \det(A_u w + B_u) &= \sum_{\sigma \in S_u} \mathrm{sgn}(\sigma) \prod_{i \in u} (a_{i\sigma(i)} w + b_{i\sigma(i)}) \\ &= \sum_{\sigma \in S_u} \mathrm{sgn}(\sigma) \sum_{j=0}^{h} \left(\sum_{v \in P_j(u)} \prod_{i \in v} a_{i\sigma(i)} \prod_{i \in u \setminus v} b_{i\sigma(i)} \right) w^j \end{aligned}$$

Note that, for $j = 0$, the sum included in round brackets of the last expression reduces to $\prod_{i \in u} b_{i\sigma(i)}$, while for $j = h$ it becomes $\prod_{i \in u} a_{i\sigma(i)}$. Thus, from equalities (5) and (17), we get

$$c_{hj} = (-1)^{h-1} \sum_{u \in P_h} \sum_{\sigma \in S_u} \mathrm{sgn}(\sigma) \sum_{v \in P_j(u)} \prod_{i \in v} a_{i\sigma(i)} \prod_{i \in u \setminus v} b_{i\sigma(i)} \tag{18}$$

For our purposes it is relevant here to show that $c_{hj} > 0$ for certain pairs of indices $(h, j) \in \mathcal{L}$ that have maximum or minimum ratio j/h. To prove a property of this kind we introduce a (possibly) different partition of the coefficients set $\mathcal{L}$, i.e. let us define

$$\begin{aligned} \mathcal{R} &:= \{r \in \mathbb{Q} \;:\; \exists (h, j) \in \mathcal{L} \text{ such that } r = j/h\}\,, \\ \mathcal{L}_r &:= \{(h, j) \in \mathcal{L} \;:\; j/h = r\}\,, \qquad \text{for every } r \in \mathcal{R}\,, \\ \overline{r} &:= \max\{r \in \mathcal{R}\}\,, \qquad \underline{r} := \min\{r \in \mathcal{R}\}\,. \end{aligned}$$

Observe that, if $A + B$ has a non-null eigenvalue then $\mathcal{L}$ is not empty, $0 \leq r \leq 1$ for every $r \in \mathcal{R}$, the family of sets $\{\mathcal{L}_r : r \in \mathcal{R}\}$ forms a partition of $\mathcal{L}$, and also $\mathcal{L}_{\overline{r}}$ and $\mathcal{L}_{\underline{r}}$ are not empty. As in the previous sections, we denote by λ_A and λ_B the largest real non-negative eigenvalue of A and B, respectively. Then, we can prove the following property.

Proposition 2. *Given two matrices* $A, B \in \mathbb{R}_+^{d \times d}$ *such that* $A + B$ *has a non-null eigenvalue, assume* $\lambda_A = 0$ *and let* $(\bar{h}, \bar{j})$ *be the element in* $\mathcal{L}_{\overline{r}}$ *with smallest value* $\bar{h}$*. Then* $\overline{r} < 1$ *and* $c_{\bar{h}\bar{j}} > 0$*.*

Proof. First observe that, since $\lambda_A = 0$, by relation (6) $c_{hh} = 0$ for all $h \in \{1, \ldots, d\}$ and hence $\overline{r} < 1$. The positive sign of $c_{\bar{h}\bar{j}}$ follows from an analysis of identity (18). Note that, just by (18), the value of each non-null coefficient c_{hj} is given by the weights of permutations of sets of h nodes whose cycles in $G(A, B)$ have j many edges labelled by a and $h - j$ edges labelled by b. Since $\overline{r}$ is the maximum value in $\mathcal{R}$, all (possible) simple cycles in $G(A, B)$ of length $h < \bar{h}$ have a number of occurrences of a striclty smaller than $\bar{j}$: hence they cannot contribute to determine the value $c_{\bar{h}\bar{j}}$. As a consequence, the only permutations σ that contribute to $c_{\bar{h}\bar{j}}$ are cyclic permutations of $\bar{h}$ nodes whose corresponding cycle in $G(A, B)$ has exactly $\bar{j}$ many edges labelled by a (and the others by b). Thus, the sign of these permutations always is $\mathrm{sgn}(\sigma) = (-1)^{\bar{h}-1}$, which implies $c_{\bar{h}\bar{j}} > 0$ by the same identity (18). □

An analogous symmetric reasoning (exchanging A and B, a and b) allows us to state a similar result for $\mathcal{L}_{\underline{r}}$. Note that, also in this case, if $(\breve{h}, \breve{j})$ is the element of $\mathcal{L}_{\underline{r}}$ with smallest value $\breve{h}$ then all permutations that contribute to $c_{\breve{h}\breve{j}}$ are cyclic, and hence reasoning as above $c_{\breve{h}\breve{j}} > 0$.

Proposition 3. *Given two matrices $A, B \in \mathbb{R}_+^{d \times d}$ such that $A + B$ has a non-null eigenvalue, assume $\lambda_B = 0$ and let $(\breve{h}, \breve{j})$ be the element in $\mathcal{L}_{\underline{r}}$ with smallest value $\breve{h}$. Then $\underline{r} > 0$ and $c_{\breve{h}\breve{j}} > 0$.*

8. On the endpoints of the domain of $G(x)$

The results of Section 6 are obtained under the assumptions of Theorem 1, i.e. $A + B$ primitive and both A and B non-null. Now we introduce further hypotheses in our analysis. Our goal is to prove that, under a mild condition, the endpoints of the domain of $G(x)$ are of the form j/h for some integers $j, h \in \{0, 1, \ldots, d\}$ such that $j \leq h$ ($h \neq 0$) and, under the same hypothesis, we determine a precise value for the limits of $G(x)$ at these extremes when $\lambda_A = 0$ or $\lambda_B = 0$. The proofs are based on the properties of characteristic polynomials presented in Section 7, and here we use the notions introduced therein.

Theorem 5. *Under the hypotheses of Theorem 1, let $\lambda_A = 0$ and assume that $\mathcal{L}_{\overline{r}}$ is a singleton defined by $\mathcal{L}_{\overline{r}} = \{(\bar{h}, \bar{j})\}$ for a given $(\bar{h}, \bar{j}) \in \mathcal{L}$. Then $V = \overline{r} < 1$, $c_{\bar{h}\bar{j}} > 0$ and*

$$\lim_{x \to V^-} G(x) = \log \frac{\lambda}{\sqrt[\bar{h}]{c_{\bar{h}\bar{j}}}}$$

Proof. First note that, by Proposition 2, $\overline{r} < 1$ and $c_{\bar{h}\bar{j}} > 0$. Moreover, for every $r \in \mathcal{R}$ different from $\overline{r}$ and every $(h, j) \in \mathcal{L}_r$, one has $j - \overline{r}h < 0$. Now, let us define the function

$$F(t, y) := 1 - c_{\bar{h}\bar{j}} e^{t\bar{j}} y^{-\bar{h}} - \sum_{(h,j) \in \mathcal{L} \setminus \{(\bar{h}, \bar{j})\}} c_{hj} e^{tj} y^{-h}$$

Under our hypotheses, equation (7) reduces to $F(t, y) = 0$. Thus, for any constant $\alpha > 0$ and for $t \to +\infty$, the value $F(t, \alpha e^{t\overline{r}})$ satisfies the relation

$$F(t, \alpha e^{t\overline{r}}) = 1 - c_{\bar{h}\bar{j}} \alpha^{-\bar{h}} - \sum_{(h,j) \in \mathcal{L} \setminus \{(\bar{h}, \bar{j})\}} c_{hj} \alpha^{h} e^{t(j - \overline{r}h)} = 1 - c_{\bar{h}\bar{j}} \alpha^{-\bar{h}} + \mathrm{o}(1)$$

Then, for t large enough, $F(t, \alpha e^{t\overline{r}})$ is greater or smaller than 0 according to whether α is greater or smaller than $\sqrt[\bar{h}]{c_{\bar{h}\bar{j}}}$. Since $y = y(t)$ is solution of equation $F(t, y) = 0$, this implies the following relation for every $\epsilon > 0$ and every t large enough:

$$\left(\sqrt[\bar{h}]{c_{\bar{h}\bar{j}}} - \epsilon\right) e^{t\overline{r}} \leq y(t) \leq \left(\sqrt[\bar{h}]{c_{\bar{h}\bar{j}}} + \epsilon\right) e^{t\overline{r}}$$

Since $y(t)$ has the form (12), the inequalities above prove $V = \overline{r} = \bar{j}/\bar{h}$. Clearly, $\bar{j} > 0$ because $V > 0$.

At last, let us consider the behaviour of $G(x)$ for x tending to V^-. Since $V = \overline{r}$ we have $\overline{q} = 0$ and equality (13) in our hypotheses becomes $1 = c_{\bar{h}\bar{j}}(\lambda C)^{-\bar{h}} + \mathrm{o}(1)$, which implies $C = \sqrt[\bar{h}]{c_{\bar{h}\bar{j}}}/\lambda$. Replacing this value in (14) one gets

$$G(x) = \log\left(\frac{\lambda}{\sqrt[\bar{h}]{c_{\bar{h}\bar{j}}}}\right) + \mathrm{o}(1) \qquad \square$$

Again, applying Theorem 3 to the previous result, a similar property can be proved for U.

Theorem 6. *Under the hypotheses of Theorem 1, let $\lambda_B = 0$ and assume that $\mathcal{L}_{\underline{r}}$ is a singleton defined by $\mathcal{L}_{\underline{r}} = \{(\breve{h}, \breve{j})\}$ for a given $(\breve{h}, \breve{j}) \in \mathcal{L}$. Then $U = \underline{r} > 0$, $c_{\breve{h}\breve{j}} > 0$ and*

$$\lim_{x \to U^+} G(x) = \log \frac{\lambda}{\sqrt[\breve{h}]{c_{\breve{h}\breve{j}}}}$$

Clearly, under the hypotheses of the last two theorems, both U and V turn out to be of the form j/h for some $(h, j) \in \mathcal{L}$. Moreover, it is easy to verify that if the size d of the linear representation of the rational model belongs to $\{2, 3\}$ then all sets $\mathcal{L}_r$ are singleton, and hence Theorems 6 and 5 apply in case $\lambda_A = 0$ and $\lambda_B = 0$, respectively.

Example 1. Consider the linear representation defined by the finite automaton of Figure 1, where all transitions have weight 1. In this case $\lambda = 2$, $\beta = 2/3$, $\lambda_A = 1$ while $\lambda_B = 0$, $\det(Iy - Ae^t - B) = y^3 - y^2e^t - 2ye^t$ and $y(t)$ can be computed explicitly as $y(t) = \frac{e^t}{2}\left(1 + \sqrt{1 + 8e^{-t}}\right)$. Moreover, the non-null coefficients c_{hj} are $c_{11} = 1$ and $c_{21} = 2$, and hence $\underline{r} = 1/2$, which implies $U = 1/2$ by Theorem 6. By the same theorem we have $\lim_{x\to 1/2^+} G(x) = (1/2)\log 2$. Since $\lambda_A > 0$ the behaviour of $G(x)$ near V is established by Theorem 2: $V = 1$ and $\lim_{x\to 1^-} G(x) = \log 2$. These evaluations are confirmed by analysis of the rate function $G(x) = \log 2 + xt - \log y(t)$, where t satisfies the relation $y'(t) = xy(t)$, the graphic of which is plotted at the right hand side of the figure. $\Diamond$

Here is an example of size $d = 4$ where a set $\mathcal{L}_r$ is not singleton, but $\mathcal{L}_{\overline{r}}$ is. However, in a similar way, it is easy to produce examples with 4 states where also $\mathcal{L}_{\overline{r}}$ is not singleton.

Example 2. Consider the linear representation defined by the automaton in the left hand side of Figure 2. Clearly $A+B$ is primitive and a direct computation shows that its Perron-Frobenius eigenvalue is a constant $\lambda = 3.38\ldots$. Moreover, here we have $\lambda_A = 0$, $\lambda_B = 2$ and

$$\det(Iy - Aw - B) = y^4 - 3y^3 - 5y^2w + 2y^2 + 7yw - 6w^3 + 2w^2$$

showing the values of non-null coefficients c_{hj}'s. Note that here $\mathcal{L}_{1/2} = \{(2, 1), (4, 2)\}$ is not singleton, while $\overline{r} = 3/4$, $\mathcal{L}_{\overline{r}} = \{(4, 3)\}$ and $c_{\bar{h}\bar{j}} = 6$. Then, by Theorems 6, we have $V = 3/4$ and $\lim_{x\to 3/4^-} G(x) = \log(\lambda/\sqrt[4]{6})$. $\Diamond$

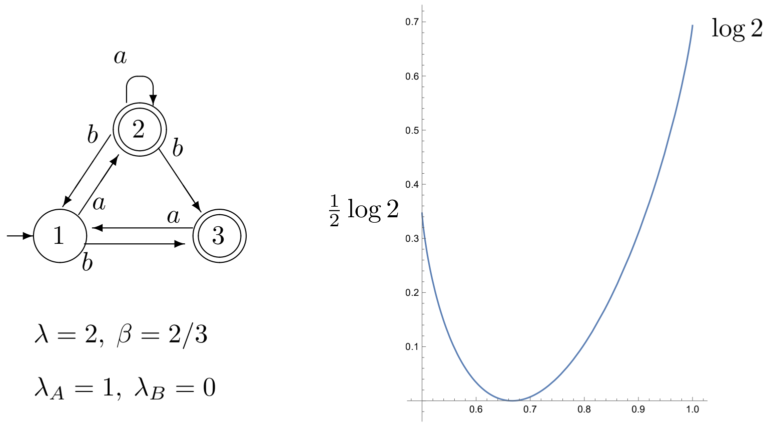

Figure 1: Generalized automaton of size 3, where all transitions have weight 1 and the matrix of transition weights is primitive. The right hand side shows a graphic of the corresponding rate function $G(x)$.

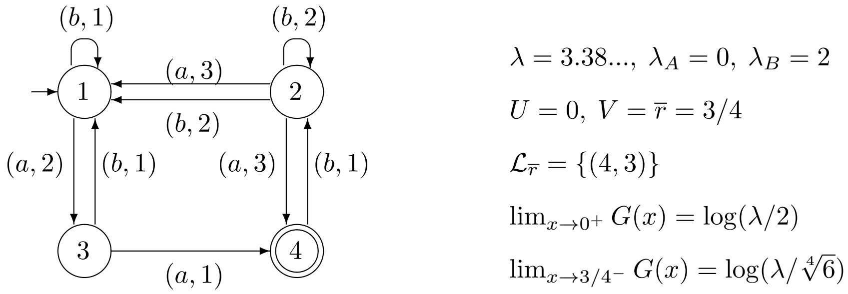

Figure 2: Generalized automaton with primitive transition matrix where $\lambda_A = 0$ and $\mathcal{L}_{\overline{r}}$ is a singleton.

9. Future work

A natural goal for possible subsequent investigations concerns the extension of the results presented in Theorems 5 and 6 to all rational models (ξ, A, B, η) having primitive matrix $A + B$ (with both A and B not null). In fact, we think that analogous results can be obtained also when $\mathcal{L}_{\overline{r}}$ and $\mathcal{L}_{\underline{r}}$ are not singleton. In these cases, the limits of $G(x)$ at V and U should be related to the roots of the polynomials $p_{\overline{r}}(x)$ and $p_{\underline{r}}(x)$ (defined as in Section 6) which can be studied by using the properties of the primitive matrices.

Acknowledgments

This work has been supported by Università degli Studi di Milano, under project PSR 2023.

References

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

[2] 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.

[3] A. Bertoni, C. Choffrut, M. Goldwurm and V. Lonati. Local limit properties for pattern statistics and rational models. *Theory Comput. Systems*, 39:209–235, 2006.

[4] A. Bertoni, P. Massazza and R. Radicioni. Random generation of words in regular languages with fixed occurrences of symbols. Proc. WORDS'03, Fourth Intern. Conf. on Combinatorics of Words, Turku (Finland), 332–343, 2003.

[5] J. Bourdon and M. Régnier. Large deviation properties for patterns. *J. of Discrete Algorithms*, 24:2–11, 2014.

[6] 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.

[7] N. Broutin and L. Devroye. Large deviations for the weighted heigth of an extended class of trees. *Algorithmica*, 46:271-297, 2006.

[8] C. Choffrut, M. Goldwurm and V. Lonati. On the maximum coefficients of rational formal series in commuting variables. *Proc. 8th DLT Conference*, Lecture Notes in Computer Science vol. 3340, pp. 114-126, Springer, 2004.

[9] D. de Falco, M. Goldwurm, V. Lonati. Frequency of symbol occurrences in bicomponent stochastic models. *Theoret. Comput. Sci.*, 327 (3):269–300, 2004.

[10] A. Dembo and O. Zeitouni. *Large Deviation Techinques and Applications.* Corrected reprint of the second (1998) edition. Springer–Verlag, 2010.

[11] F. den Hollander. *Large Deviations.* American Mathematical Society, 2000.

[12] A. Denise and M. Régnier. Rare events and conditional events on random strings. *Discrete Math. Theor. Comput. Sci.*, 6:191–214, 2004.

[13] A. Denise, O. Roques and M. Termier. Random generation of words of context-free languages according to the frequency of letters. In D. Gardy, A. Mokkadem (eds), Mathematics and computer Science, Trends in Mathematics, 113–125, Birkäuser, 2000.

[14] L. Devroye. Universal limit laws for depths in random trees. *SIAM Journal on Computing*, 28:409-432, 1999.

[15] P. Flajolet and R. Sedgewick. *Analytic Combinatorics.* Cambridge Univ. Press, 2009.

[16] P. Flajolet, W. Szpankowski and B. Vallée. Hidden word statistics. *Journal of the ACM*, 53:147–183, 2006.

[17] M. Goldwurm, J. Lin, M. Vignati. Local limit laws for symbol statistics in bicomponent rational models. *Theoret. Comput. Sci.*, 970:114051, 2023, https://doi.org/10.1016/j.tcs.2023.114051.

[18] M. Goldwurm, R. Radicioni. Probabilistic models for pattern statistics. *RAIRO Theoretical Informatics and Applications*, 40:207-225, 2006.

[19] M. Goldwurm, M. Vignati. Large deviation properties for pattern statistics in primitive rational models. *Proc. 24th ICTCS*, CEUR-WS.ORG, vol. 3587, pp. 192-205, 2023 (available at https://ceur-ws.org/Vol-3587/1908.pdf).

[20] P. Grabner and M. Rigo. Distribution of additive functions with respect to numeration systems on regular languages. *Theory Comput. Systems*, 40:205-223, 2007.

[21] H.-K. Hwang. Large deviations for combinatorial distributions. I: Central limit theorems. *The Annal of Applied Probability*, 6(1):297–319, 1996.

[22] P. Jacquet and W. Szpankowski. *Analytic Pattern Matching: From DNA to Twitter.* Cambridge University Press, 2015.

[23] P. Nicodeme, B. Salvy, and P. Flajolet. Motif statistics. *Theoret. Comput. Sci.*, 287(2): 593–617, 2002.

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

[25] A. Salomaa and M. Soittola. *Automata-Theoretic Aspects of Formal Power Series.* Springer–Verlag, 1978.

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

[27] P. Turakainen. Generalized automata and stochastic languages. *Proc. Amer. Math. Soc.*, 21: 303–309, 1969.