



Large deviation properties for pattern statistics in primitive rational models [☆]

Massimiliano Goldwurm ^{*}, Marco Vignati

Dipartimento di Matematica “Federigo Enriques”, Università degli Studi di Milano, Via Saldini 50, Milano, 20133, Italy

ARTICLE INFO

Keywords:

Automata and formal languages
Large deviations
Limit distributions
Pattern statistics
Rational series
Regular languages

ABSTRACT

We present a large deviation property for pattern statistics representing the number of occurrences of a symbol in words of given length generated at random according to a rational stochastic model. This result is proved assuming that the transition matrix of the model is primitive. We show how the rate function of the large deviation property depends on the main eigenvalues and eigenvectors of the transition matrices associated with the different symbols of the alphabet. We also yield general conditions to guarantee that the range of validity of the large deviation estimate coincides with the whole interval $(0, 1)$, which represents in our context the largest possible open interval where the property may hold. The case of smaller intervals of validity is finally examined by means of examples.

1. Introduction

The study of pattern statistics on words is mainly based on the analysis of sequences of integer random variables $\{X_n\}$, where each X_n represents the number of occurrences of one or more strings (called *patterns*), in a word of length n (*text*) generated at random according to a given stochastic model. Traditional goals are the asymptotic expressions of the moments of $\{X_n\}$, its limit distribution, possibly the corresponding local limit properties. The results usually depend on the fixed set of patterns (a symbol or a string or, in the most general case, a language) and on the probabilistic model that generates the text, typically a Bernoullian or a Markovian source or, more generally, a probabilistic algorithm that on input n outputs a random word of length n under a suitable probability distribution [25,24,2,16,22,18].

In this context it is also of interest to evaluate the probability of rare events, like when one or more patterns are over- or under-represented in the random text. To this end a useful tool is represented by the notion of *large deviations*.

We recall that the large deviation properties represent a classical subject of probability theory. They yield bounds of exponential decay on the probability that a sequence of random variables differs from its mean value for an amount of the order of growth of the mean itself [10,11]. Such deviations from the average value are considered “large” with respect to other evaluations, like those deriving for instance from the Central Limit Theorem or from the Law of Large Numbers, which concern the probability of having asymptotically smaller differences from the mean value.

In analytic combinatorics large deviation estimates are considered in the study of various fundamental structures [15]. They occur in the analysis of depth and height of certain classes of trees [7,14], and have been considered also in the estimates of pattern statistics

[☆] This article belongs to Section A: Algorithms, automata, complexity and games, Edited by Paul Spirakis.

^{*} Corresponding author.

E-mail address: massimiliano.goldwurm@unimi.it (M. Goldwurm).

<https://doi.org/10.1016/j.tcs.2024.115055>

Received 9 February 2024; Received in revised form 20 December 2024; Accepted 30 December 2024

Available online 7 January 2025

0304-3975/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

[22]. In that context in particular, they have been studied with the goal of evaluating the probability of a very large or very small number of patterns in a random text [12,5].

In the present work we prove a property of this type for a sequence of pattern statistics representing the number of occurrences of a symbol in a word of length n , belonging to a regular language, generated at random according to a *rational stochastic model*. This model was introduced in [2] and can be formally defined by a nondeterministic finite state automaton with real positive weights on transitions. In this setting, the probability of generating a word w of given length is proportional to the total weight of the accepting computations labelled by w . This model is quite general, it includes as special cases the traditional Bernoullian and Markovian sources, widely used in this context. Moreover, the rational stochastic models can represent the generation of random words of length n in any arbitrary regular language under uniform distribution: this occurs when the finite automaton used to generate the random text is unambiguous and all transitions have the same weight.

In order to describe the known results, consider a weighted finite state automaton \mathcal{A} over the binary alphabet $\{a, b\}$ and, for every $n \in \mathbb{N}$, let Y_n be the number of occurrences of the symbol a in a word of length n generated at random according to the rational model defined by \mathcal{A} . The properties of the sequence $\{Y_n\}$ are of interest in several contexts. First of all they are related to the number of occurrences of patterns in a random word of length n when the set of patterns is given by a regular language. More precisely, given an arbitrary alphabet Σ , for any regular language $R \subset \Sigma^*$ and any Markovian model \mathcal{M} generating random words in Σ^* (of any order), one can construct a weighted finite automaton \mathcal{A} over the alphabet $\{a, b\}$ such that the corresponding statistics Y_n equals the number of (positions) of occurrences of words from R in a word of length n generated by \mathcal{M} [2,24,25]. This means that the analysis of symbol statistics in rational models is a quite general issue, as they can represent a large variety of pattern statistics under several probabilistic assumptions.

A further reason of interest for the rational models is the relationship between the probabilistic evolution of the sequence $\{Y_n\}$ and properties of regular languages and rational formal series. Clearly the asymptotic behaviour of $\{Y_n\}$ is related to the natural problem of estimating the number of words of given length in a regular language having k occurrences of a given symbol [4,13]. This behaviour is also related to the analysis of additive functions defined on regular languages [20], and to the descriptive complexity of languages and computational models [6]. It has also been studied to evaluate the coefficients of rational formal series, which is a traditional problem well-studied in the literature [29,26,23]. In particular, the asymptotic behaviour of the maximum coefficients associated with monomials of degree n in rational bivariate series can be estimated by using the local limit properties of the sequences $\{Y_n\}$ defined by certain automata [2,3]. It is known for instance that for a wide class of these series such a maximum coefficient has an asymptotic growth of the order $\Theta(n^{k/2}/\lambda^n)$ for some $\lambda > 0$ and some integer $k \geq -1$ [8,3].

The asymptotic behaviour of $\{Y_n\}$, i.e. mean values, variance, limit distributions both in the global and in the local sense [17,15], has been studied in the literature under several hypotheses on the corresponding automaton \mathcal{A} . It is known that if \mathcal{A} has a primitive transition matrix then Y_n has a Gaussian limit distribution [2,24] and, under a suitable aperiodicity condition, it also satisfies a local limit theorem [2], which can be generalized to all primitive cases by using a suitable notion of periodicity [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 [9], while the local limit properties in this case are recently studied in [18]. When the automaton \mathcal{A} has several strongly connected components a general analysis of the (global) limit distribution of Y_n can be found in [19].

Here we continue this line of research proving that if the transition matrix of the automaton \mathcal{A} is primitive, then $\{Y_n\}$ satisfies a large deviation property with a rate function that strictly depends on the main eigenvalue and eigenvectors of the same matrix. The corresponding proof is rather standard as it relies on traditional tools of analytic combinatorics, as for instance the so-called “quasi power” theorem, and actually the existence of a large deviation property for $\{Y_n\}$ is consequence of more general results already appeared in the previous literature [21,15]. However, for sake of completeness we here present a direct proof of this result, and show how its interval of validity and its rate function are related to the properties of the transition matrix of the stochastic model (i.e. the automaton \mathcal{A}). The role played by the eigenvalues and eigenvectors of the transition matrices of \mathcal{A} are of particular interest in this context.

We also present some further results concerning the interval of validity of the large deviations property. Assuming rather mild conditions on the eigenvalues of the transition matrices of the automaton, we prove that the range of validity of this estimate coincides with the whole interval $(0, 1)$. In our context this interval represents the largest possible open interval where the property of large deviations may hold. We also show, by means of suitable examples, that those “mild” conditions cannot be omitted in our statement in order to guarantee $(0, 1)$ as interval of validity of the property. In fact, we exhibit examples of primitive rational models where such an interval is strictly included in $(0, 1)$.

This work is organized as follows: in Section 2 we recall definition and general issues on the large deviations property, together with the relevant example of binomial random variables, which can be seen as a sort of benchmark for our results; in Section 3 we define the rational models and recall basic properties of the associated sequence of symbol statistics $\{Y_n\}$, while in Section 4 we consider the primitive rational models, recalling the known asymptotic results on $\{Y_n\}$ in this case; in Section 5 we prove the large deviations property for $\{Y_n\}$ in the primitive rational models; in Section 6 we present conditions that guarantee the interval $(0, 1)$ as range of validity for such a property, and study examples of primitive rational models where this interval is strictly included in $(0, 1)$. Section 7 closes the present contribution with some comments and hints for further investigations.

2. A quick overview on large deviations

A large deviation estimate usually refers to a sequence of random variables, say $\{X_n\}$, having increasing mean values; it consists 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 every $c > 0$. Typical situations occur when $E(X_n) \sim \beta n$ for a constant $\beta > 0$, and since this occurs in all our examples, here we start with the following formal definition [11,15].

Definition 1. Let $\{X_n\}$ be a sequence of random variables 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:

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \Pr(X_n \geq xn) = -F(x), \quad \text{for } \beta \leq x < x_1. \tag{2}$$

This property is equivalent to require that

$$\Pr(X_n \leq xn) = e^{-F(x)n+o(n)}, \quad \text{for } x_0 < x \leq \beta$$

$$\Pr(X_n \geq xn) = e^{-F(x)n+o(n)}, \quad \text{for } \beta \leq x < x_1.$$

Relation (1) concerns the *left tail* of the distribution of X_n , while (2) refers to the *right tail*. It is convenient to keep the two limits separated since the proofs of large deviation properties may consider one tail at a time. We note that the previous definition differs from the more general Large Deviation Principle considered in the literature [10, Sec.1.2]. We prefer to adopt the above notion because it is simpler, fully represents the phenomenon of large deviations, and it is adequate for the applications we consider. It also corresponds to the traditional results on large deviations concerning pattern statistics [25,12,5,22].

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

A classical example of large deviation bound concerns the sequence of binomial random variables $\{X_{n,p}\}_n$ of parameters n and p , for any fixed $p \in (0, 1)$. In this case, $E(X_{n,p}) = pn$ and by the Central Limit Theorem, we know that $\frac{X_{n,p} - np}{\sqrt{np(1-p)}}$ converges in distribution to a standard Gaussian random variable $\mathcal{N}(0, 1)$. This yields a limit probability concerning “normal” deviations (i.e. of the order \sqrt{n}) from the mean, that is

$$\lim_{n \rightarrow \infty} \Pr\left(|X_{n,p} - np| \geq \varepsilon \sqrt{n}\right) = \Pr\left(|\mathcal{N}(0, 1)| \geq \frac{\varepsilon}{\sqrt{p(1-p)}}\right), \quad \forall \varepsilon > 0.$$

Such a property implies the following result for a larger deviation

$$\Pr\left(|X_{n,p} - np| \geq \varepsilon n\right) = o(1), \quad \forall \varepsilon > 0$$

which can also be obtained by applying the Law of Large Numbers. However, the previous relation can be noticeably improved by the following statement, which shows a large deviation property for $\{X_{n,p}\}_n$, i.e. a convergence to 0 exponential with respect to n . This result is well-known in the theory of Large Deviations [11,10] as consequence of classical Cramér’s Theorem (we recall below), and here we briefly outline a simple direct proof only to compare it with our subsequent outcomes and to give the flavour of the property.

Proposition 1. Any sequence of binomial random variables $\{X_{n,p}\}_n$, where $0 < p < 1$, satisfies a large deviation property in the interval $(0, 1)$ with rate function $B_p(x)$ given by

$$B_p(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}, \quad \text{for every } x \in (0, 1).$$

Proof. First consider the left tail and let $x \in \mathbb{R}$ such that $0 < x \leq p$. We have to prove that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \Pr(X_{n,p} \leq xn) = -B_p(x).$$

To this end, let $M_n(x) = \max\{\Pr(X_{n,p} = i) : i \in \mathbb{N}, 0 \leq i \leq xn\}$. Then we have

$$M_n(x) \leq \Pr(X_{n,p} \leq xn) \leq (xn + 1)M_n(x). \tag{3}$$

Recall that the probability $\Pr(X_{n,p} = i)$ is increasing for integers i such that $0 \leq i \leq pn$; hence $M_n(x) = \binom{n}{\lfloor xn \rfloor} p^{\lfloor xn \rfloor} (1-p)^{n-\lfloor xn \rfloor}$ for every $x \in (0, p]$. Thus, a direct application of Stirling’s formula leads to

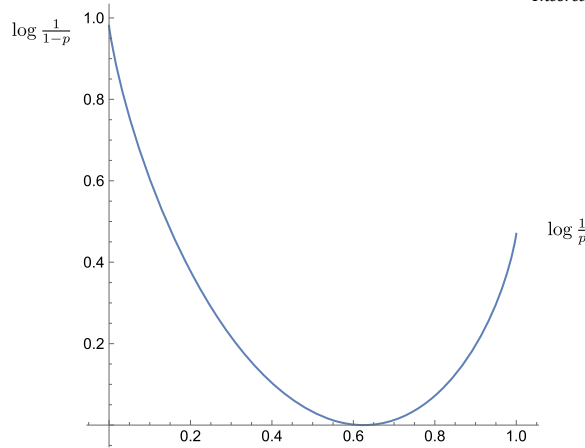


Fig. 1. Graphic of the rate function $B_p(x)$, when $p = \frac{5}{8}$.

$$M_n(x) = \exp \left\{ n \left[x \log \frac{p}{x} + (1-x) \log \frac{1-p}{1-x} \right] + O(\log n) \right\}$$

which replaced in (3) proves that

$$\log \Pr(X_{n,p} \leq xn) = -B_p(x)n + O(\log n). \tag{4}$$

A similar reasoning holds for the right tail. In this case we have

$$N_n(x) \leq \Pr(X_{n,p} \geq xn) \leq (n - nx + 1)N_n(x), \quad \forall x \in [p, 1)$$

where $N_n(x) = \binom{n}{[xn]} p^{[xn]} (1-p)^{n-[xn]}$. As above, replacing this value in the previous inequalities yields

$$\log \Pr(X_{n,p} \geq xn) = -B_p(x)n + O(\log n). \tag{5}$$

Relations (4) and (5) prove the required property. \square

The rate function $B_p(x)$ is strictly convex in the interval $(0, 1)$, it takes a unique minimal value at $x = p$, where $B_p(p) = 0$, while $\lim_{x \rightarrow 0^+} B_p(x) = \log \frac{1}{1-p}$ and $\lim_{x \rightarrow 1^-} B_p(x) = \log \frac{1}{p}$. Moreover, $B_p(x)$ grows vertically for x approaching 0 or 1, that is $\lim_{x \rightarrow 0^+} B'_p(x) = -\infty$ and $\lim_{x \rightarrow 1^-} B'_p(x) = +\infty$, and hence B_p is “essentially smooth” in $(0, 1)$, in the sense of the definition given above, for all $p \in (0, 1)$.

We also observe that, in this case the relations (1) and (2) also hold for $x = 0$ and $x = 1$, respectively, when one extends $B_p(x)$ with continuity at these points. Therefore, the large deviation property here holds in the entire closed set $[0, 1]$. We will see in the next sections that this extension is not always possible for all the large deviation properties considered in this work.

Example 1. If $p = 5/8$ then $\{X_{n,p}\}$ satisfies a large deviation property in the interval $(0, 1)$, with rate function B_p described in Fig. 1. \diamond

We conclude this section recalling that usually the interval of a large deviation property can be extended to the entire set \mathbb{R} once we allow the rate function $F(x)$ to assume value ∞ . A classical situation of this type is established by Cramér’s Theorem (see for instance [11,10]), stating that if $\{X_n\}$ is a sequence of independent and identically distributed random variables, with bounded moment generating function (i.e. $\psi(t) = E(e^{tX_1}) < \infty$ for any $t \in \mathbb{R}$), then the sequence of partial sums $\{S_n\}$, where $S_n = \sum_{i=1}^n X_i$, satisfies a large deviation property all over \mathbb{R} with rate function

$$R(x) = \sup_{t \in \mathbb{R}} [xt - \log \psi(t)], \quad \forall x \in \mathbb{R}.$$

3. Symbol statistics in rational models

In order to define our stochastic model consider a *formal series* in the non-commutative variables a, b , that is a function $r : \{a, b\}^* \rightarrow \mathbb{R}_+$, 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 $m > 0$ there exists a monoid morphism $\mu : \{a, b\}^* \rightarrow \mathbb{R}_+^{m \times m}$ and two (column) arrays $\xi, \eta \in \mathbb{R}_+^m$, such that $(r, w) = \xi^t \mu(w) \eta$, for every $w \in \{a, b\}^*$ [1,28] ⁽¹⁾. Note that in this case, if $w = w_1 w_2 \cdots w_n$ with $w_i \in \{a, b\}$ for every $i = 1, 2, \dots, n$, then $\mu(w) = \mu(w_1) \mu(w_2) \cdots \mu(w_n)$. Thus, as the morphism μ is

¹ As is customary, we denote by v^t the transpose of an array $v \in \mathbb{R}^m$, i.e. a row array.

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 . Such a 4-tuple can be considered as a *generalized automaton* over the alphabet $\{a, b\}$ as defined in [31], where $\{1, 2, \dots, d\}$ is the set of states, and the weights of the transitions, as well as of the initial and final states, are positive real numbers. Here, 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.

Throughout this work, denoting by $\{a, b\}^n$ the family of all words of length n in $\{a, b\}^*$, we assume that the set $\{w \in \{a, b\}^n : (r, w) > 0\}$ is non-empty for every $n \in \mathbb{N}$ large enough (so that $\xi \neq 0 \neq \eta$), and that A and B are non-zero matrices (i.e., each of them has at least one positive entry). Moreover, for every $n \in \mathbb{N}$, we can easily compute the sum of all values of r associated with words in $\{a, b\}^n$:

$$\sum_{w \in \{a, b\}^n} (r, w) = \xi' \sum_{w \in \{a, b\}^n} \mu(w) \eta = \xi' \left(\prod_{i=1}^n \sum_{w_i \in \{a, b\}} \mu(w_i) \right) \eta = \xi' (A + B)^n \eta. \quad (6)$$

Thus, we can consider the probability measure Pr over the set $\{a, b\}^n$ given by

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

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 $L \cap \{a, b\}^n$. 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 $\eta' = (1, 1, \dots, 1)$.

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 $\text{Pr}(w)$ and $|w|_a$ denotes the number of occurrences of a in w . It is clear that the probability function of Y_n is defined by

$$p_n(k) := \text{Pr}(Y_n = k) = \frac{\sum_{w \in \{a, b\}^n, |w|_a = k} (r, w)}{\sum_{w \in \{a, b\}^n} (r, w)}, \quad \forall k \in \{0, 1, \dots, n\}.$$

Since r is rational, also the previous probability can be expressed by using its linear representation. The denominator is clearly determined by (6). As far as the numerator is concerned, setting $\delta_a(a) = 1$ and $\delta_a(b) = 0$, for a variable x and for every $n \in \mathbb{N}$ one has

$$(Ax + B)^n = \prod_{i=1}^n \sum_{w_i \in \{a, b\}} \mu(w_i) x^{\delta_a(w_i)} = \sum_{k=0}^n \left(\sum_{w \in \{a, b\}^n, |w|_a = k} \mu(w) \right) x^k.$$

As a consequence, denoting by $[x^k]g(x)$ the coefficient of the monomial of degree k in a polynomial $g(x)$, the probability $p_n(k)$ can be written as

$$p_n(k) = \frac{[x^k] \xi' (Ax + B)^n \eta}{\xi' (A + B)^n \eta}, \quad \forall k \in \{0, 1, \dots, n\}. \quad (7)$$

For the sake of brevity, we say that Y_n is *defined* by the linear representation (ξ, A, B, η) . The moment generating function $\Psi_n(z)$ of Y_n can be defined by means of the map $h_n(z)$ given by

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

Then, we have

$$\Psi_n(z) = \sum_{k=0}^n p_n(k) e^{zk} = \frac{\xi' (Ae^z + B)^n \eta}{\xi' (A + B)^n \eta} = \frac{h_n(z)}{h_n(0)}, \quad \forall z \in \mathbb{C} \quad (8)$$

and hence the mean value and the variance of Y_n can be computed by the relations

$$\mathbb{E}(Y_n) = \frac{h'_n(0)}{h_n(0)}, \quad \text{Var}(Y_n) = \frac{h''_n(0)}{h_n(0)} - \left(\frac{h'_n(0)}{h_n(0)} \right)^2.$$

Observe that, in principle, Y_n is the sum of n Bernoullian r.v.'s, which however are neither independent nor identically distributed (and hence traditional Cramer's Theorem cannot be applied in this case). More precisely, Y_n can be seen as a sum of the following form:

$$Y_n = \sum_{i=1}^n b_i(w), \quad \text{where } b_i(w) = \begin{cases} 1 & \text{if } w_i = a \\ 0 & \text{if } w_i = b \end{cases}, \quad w = w_1 \dots w_n, \quad w_i \in \{a, b\}.$$

Clearly, the r.v.'s $b_i(w)$ ($i = 1, 2, \dots, n$) are not independent since each of them strictly depends on the state reached at the i -th step and hence on all the previous transitions; also, they cannot have the same distribution as the weights of the transitions from the various states may be quite different. Therefore, in the general case, Y_n is a random variable very different from a Binomial r.v. $X_{n,p}$, for any $p \in (0, 1)$, even if they both have the same range of values $\{0, 1, \dots, n\}$.

4. Primitive models

In this section we summarize the main properties of Y_n when the matrix $A + B$ is primitive. Recall that a matrix $M \in \mathbb{R}_+^{m \times m}$ is *primitive* if there exists a positive integer n such that $M^n > 0$ (i.e. all entries of M^n are strictly positive). The main properties of these matrices are established by the following well-known theorem (see for instance [30, Sec 1.1]).

Theorem 1. (Perron-Frobenius) *If a matrix $T = [t_{ij}] \in \mathbb{R}_+^{m \times m}$ is primitive then it admits a real eigenvalue $\lambda > 0$ such that:*

- (i) $|\mu| < \lambda$ for any eigenvalue μ of T different from λ ;
- (ii) λ can be associated with strictly positive left and right eigenvectors;
- (iii) λ is a simple root of the characteristic equation of T , and hence the associated eigenvectors are unique up to constant multiples;
- (iv) if a matrix $A = [a_{ij}] \in \mathbb{R}_+^{m \times m}$ satisfies $A \leq T$ (i.e. $a_{ij} \leq t_{ij}, \forall i, j$) and α is an eigenvalue of A then $|\alpha| \leq \lambda$. Moreover, $|\alpha| = \lambda$ implies $A = T$.

Usually λ is called the Perron-Frobenius eigenvalue of T .

Then, assume $A + B$ is 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]. Its properties (in particular mean value and variance) can be studied through the function $y = y(z)$ implicitly defined by the equation

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

with initial condition $y(0) = \lambda$. Clearly, $y(z)$ is eigenvalue of $Ae^z + B$ for every $z \in \mathbb{C}$, and is analytic in a neighbourhood of 0. Moreover, in the analysis of the asymptotic properties of $\{Y_n\}$, the following results have been obtained in the literature [2,3] and are useful in our context:

1. $y'(0) = v' Au$, where v' and u are left and right eigenvectors of $A + B$, with respect to λ , such that $v'u = 1$. By properties (ii) and (iii) of Theorem 1, this implies $y'(0) > 0$.
2. $E(Y_n) = \beta n + c + O(\epsilon^n)$, where $|\epsilon| < 1$, $c \in \mathbb{R}$ and β is a constant satisfying $0 < \beta < 1$ given by

$$\beta = \frac{y'(0)}{\lambda}.$$

3. $\text{Var}(Y_n) = \gamma n + O(1)$, where γ is a positive constant defined by

$$\gamma = \frac{y''(0)}{\lambda} - \left(\frac{y'(0)}{\lambda}\right)^2.$$

4. In a neighbourhood of 0, $\Psi_n(z)$ satisfies a “quasi power” condition, that is an equality of the form

$$\Psi_n(z) = r(z) \left(\frac{y(z)}{\lambda}\right)^n (1 + O(\epsilon^n)) \quad (|\epsilon| < 1) \tag{9}$$

where $r(z)$ is also analytic in $z = 0$, and $r(0) = 1$. A consequence of this result is that $\frac{Y_n - \beta n}{\sqrt{\gamma n}}$ converges in distribution to a Gaussian random variable of mean 0 and variance 1.

Some further properties of the moment generating function $\Psi_n(z)$ can be obtained in the case of real z . First observe that for every $t \in \mathbb{R}$ also the matrix $Ae^t + B$ is primitive: then $y(t)$ is its Perron-Frobenius eigenvalue. By the properties of primitive matrices we know that $y(t)$ is a positive real function, analytic and strictly increasing for all $t \in \mathbb{R}$ (statement (iv) in Theorem 1). Therefore, $y(t)$ can be written in the form

$$y(t) = v_t'(Ae^t + B)u_t \tag{10}$$

where v_t' and u_t are left and right eigenvectors of $Ae^t + B$ relative to $y(t)$, normed so that $v_t'u_t = 1$. Moreover, by [30, Th. 1.2], all the powers of $Ae^t + B$ satisfy a relation of the form

$$(Ae^t + B)^n = y(t)^n \cdot u_t v_t' (1 + O(\epsilon_t^n)) \quad (|\epsilon_t| < 1, \forall t \in \mathbb{R})$$

As a consequence, for every $t \in \mathbb{R}$, we obtain

$$\Psi_n(t) = E(e^{tY_n}) = \frac{\xi^t(Ae^t + B)^n \eta}{\xi^t(A + B)^n \eta} = r(t) \left(\frac{y(t)}{\lambda}\right)^n (1 + O(\epsilon_t^n)) \tag{11}$$

where the function $r(t) = \frac{\xi^t u_t v_t' \eta}{\xi^t u_0 v_0' \eta}$ is analytic in \mathbb{R} , clearly $r(0) = 1$ and $|\epsilon_t| < 1$. Note that this relation extends equation (9) to all real z .

5. Large deviations for primitive models

Now assume again $A + B$ primitive and consider the random variable $Y_n(t)$ defined by the linear representation (ξ, Ae^t, B, η) , for any $t \in \mathbb{R}$. Since $Ae^t + B$ is primitive for any $t \in \mathbb{R}$, we can apply the results of the previous section to all sequences of r.v.'s $\{Y_n(t)\}$. To this end, for any $t \in \mathbb{R}$, let us consider the function $y_t(z)$ implicitly defined by the equation

$$\det(Iy_t - Ae^{t+z} - B) = 0, \quad \forall z \in \mathbb{C}$$

with initial condition $y_t(0) = y(t)$. Clearly $y_t(z) = y(t+z)$ and hence $y_t(z)$ is analytic in a neighbourhood of 0 (for any $t \in \mathbb{R}$), it admits derivatives of any order around 0, and

$$y'_t(0) = y'(t), \quad y''_t(0) = y''(t).$$

Applying property 2 of the previous section to the linear representation (ξ, Ae^t, B, η) , for every $t \in \mathbb{R}$, we obtain

$$E(Y_n(t)) = \beta(t)n + c_t + O(\varepsilon_t^n)$$

where $c_t \in \mathbb{R}$ and $\varepsilon_t \in (0, 1)$ are constant, and $\beta(t)$ is a real function given by

$$\beta(t) = \frac{y'_t(0)}{y_t(0)} = \frac{y'(t)}{y(t)}, \quad \forall t \in \mathbb{R}. \tag{12}$$

Clearly $\beta(0) = \beta$. Moreover, by property 1 of the previous section, we have $y'(t) = v'_t Ae^t u_t$, and using (10), from relation (12) we obtain

$$\beta(t) = \frac{v'_t Ae^t u_t}{v'_t (Ae^t + B) u_t} \tag{13}$$

which implies

$$0 < \beta(t) < 1, \quad \forall t \in \mathbb{R}. \tag{14}$$

Analogously, applying property 3 to $Y_n(t)$, we get

$$\text{Var}(Y_n(t)) = \gamma(t)n + O(1), \quad \forall t \in \mathbb{R}$$

where $\gamma(t)$ is a positive constant given by

$$\gamma(t) = \frac{y''_t(0)}{y_t(0)} - \left(\frac{y'_t(0)}{y_t(0)} \right)^2 = \beta'(t) > 0, \quad \forall t \in \mathbb{R}. \tag{15}$$

Therefore $\beta(t)$ is strictly increasing all over \mathbb{R} and the following limits exist and are finite:

$$U = \lim_{t \rightarrow -\infty} \beta(t), \quad V = \lim_{t \rightarrow +\infty} \beta(t). \tag{16}$$

By relation (14), we have

$$0 \leq U < \beta(0) < V \leq 1 \tag{17}$$

which, together with relation (15), implies the following statement.

Lemma 1. *Let U and V be defined by relations (16). Then, 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 $x > \beta$.*

At last, let us apply property 4 of the previous section to the r.v. $Y_n(t)$: we get a ‘‘quasi power’’ property for the moment generating function of $Y_n(t)$, that is

$$\Psi_{Y_n(t)}(z) = r_t(z) \left(\frac{y(t+z)}{y(t)} \right)^n (1 + O(\varepsilon_t^n)) \quad (|z| < 1)$$

where $r_t(z)$ is also analytic in $z = 0$ and $r_t(0) = 1$. As a consequence, for every $t \in \mathbb{R}$ the sequence of r.v.'s $\left\{ \frac{Y_n(t) - \beta(t)n}{\sqrt{\gamma(t)n}} \right\}_n$ converges in distribution to a Gaussian random variable of mean 0 and variance 1, i.e. for every constant $x \in \mathbb{R}$ we have

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{Y_n(t) - \beta(t)n}{\sqrt{\gamma(t)n}} \leq x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du, \quad \forall t \in \mathbb{R}. \tag{18}$$

The previous results allow us to prove a large deviation property for our symbol statistics.

Theorem 2. 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

$$G(x) = -\log \left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}} \right) \tag{19}$$

where τ_x is defined in Lemma 1.

Proof. We first study the right tail of $\{Y_n\}$. We have to prove that for every $x \in [\beta, V)$ the following relation holds:

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log \Pr(Y_n \geq xn) = \log \left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}} \right) \tag{20}$$

where $\tau_x \geq 0$ is defined in Lemma 1.

By Markov inequality and relation (11), for every $t > 0$, we have

$$\Pr(Y_n \geq xn) = \Pr(e^{tY_n} \geq e^{txn}) \leq \frac{\mathbb{E}(e^{tY_n})}{e^{txn}} = r(t) \left(\frac{y(t)}{\lambda e^{tx}} \right)^n (1 + O(\varepsilon_t^n))$$

which implies

$$\frac{1}{n} \log \Pr(Y_n \geq xn) \leq \log \left(\frac{y(t)}{\lambda e^{tx}} \right) + O(1/n).$$

This bound can be refined by taking the minimum with respect to $t > 0$ of the first term in the right hand side. To this end let us define the function

$$\varphi_x(t) = \log \left(\frac{y(t)}{\lambda e^{tx}} \right), \quad \forall t \in \mathbb{R}. \tag{21}$$

Note that $\varphi_x(0) = 0$ and $\varphi'_x(t) = \beta(t) - x$. Hence, by Lemma 1, since $x \geq \beta$, $\varphi_x(t)$ takes a unique minimum value at $t = \tau_x \geq 0$. This proves

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log \Pr(Y_n \geq xn) \leq \log \left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}} \right).$$

Also observe that $\varphi_x(t)$ is a convex function since $\varphi''_x(t) = \beta'(t) > 0$ by relation (15).

An analogous lower bound for $\Pr(Y_n \geq xn)$ can be proved by considering the random variable $Y_n(\tau_x)$. Since $[z^k] \xi'(Az + B)^n \eta = e^{\tau_x k} [z^k] \xi'(Az + B)^n \eta$, by relations (7) and (8) we have

$$\Pr(Y_n = k) = \frac{\Pr(Y_n(\tau_x) = k) \Psi_n(\tau_x)}{e^{\tau_x k}}, \quad \forall k = 0, 1, \dots, n. \tag{22}$$

Also note that $\mathbb{E}(Y_n(\tau_x)) = \beta(\tau_x)n + O(1) = xn + O(1)$ and by (18) we know that $\{Y_n(\tau_x)\}_n$ has a Gaussian limit distribution. This means that, for every $\varepsilon > 0$, $\Pr(Y_n(\tau_x) > (x + \varepsilon)n) = o(1)$ and then

$$\Pr(xn \leq Y_n(\tau_x) \leq (x + \varepsilon)n) = \frac{1}{2} - o(1).$$

Thus, from this relation and identities (22) and (11), we get

$$\begin{aligned} \Pr(Y_n \geq xn) &\geq \Pr(xn \leq Y_n \leq (x + \varepsilon)n) \geq \frac{\Pr(xn \leq Y_n(\tau_x) \leq (x + \varepsilon)n) \Psi_n(\tau_x)}{e^{\tau_x(x+\varepsilon)n}} = \\ &= \left(\frac{1}{2} - o(1) \right) r(\tau_x) \left(\frac{y(\tau_x)}{\lambda e^{\tau_x(x+\varepsilon)}} \right)^n (1 + O(\varepsilon_{\tau_x}^n)). \end{aligned} \tag{23}$$

Thus, by the arbitrariness of ε , we have

$$\frac{1}{n} \log \Pr(Y_n \geq xn) \geq \log \left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}} \right) + O(1/n)$$

which yields the required lower bound and concludes the proof of relation (20).

Consider now the left tail. We have to prove that, for every $x \in (U, \beta]$,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log \Pr(Y_n \leq xn) = \log \left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}} \right) \tag{24}$$

where $\tau_x \leq 0$ is defined in Lemma 1.

The reasoning is similar to the previous case. The main difference is that here $U < x \leq \beta$ and one has to use negative values of t . Note that the function $\varphi_x(t)$ given by (21) is well defined also in this case. For every $x \in (U, \beta]$ and every $t < 0$, by Markov inequality and relation (11), we get

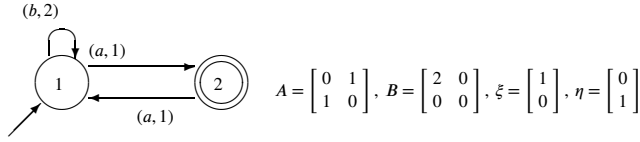


Fig. 2. Primitive rational model with periodic matrix A .

$$\Pr(Y_n \leq xn) = \Pr(e^{tY_n} \geq e^{txn}) \leq \frac{E(e^{tY_n})}{e^{txn}} = r(t) \left(\frac{y(t)}{\lambda e^{tx}} \right)^n (1 + O(\epsilon_t^n)).$$

By Lemma 1, the minimum of $\varphi_x(t) = \log \left(\frac{y(t)}{\lambda e^{tx}} \right)$ is taken at $t = \tau_x \leq 0$, and this proves that

$$\frac{1}{n} \log \Pr(Y_n \leq xn) \leq \log \left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}} \right) + O(1/n)$$

which yields an upper bound to the limit in (24).

The corresponding lower bound is obtained as in the analysis of the right tail, leading to relation (23), with obvious changes. \square

We observe that the previous result could also be obtained by applying the Gärtner-Ellis Theorem [10, Sec. 2.3] and using the Large Deviation Principle mentioned in Section 2. However, in the present one-dimensional context we think that a more direct approach is preferable, since it simplifies the reading of the proof, and brings to light the properties of the parameters of the rational model in the various steps of the deduction.

Now we determine some general properties of the rate function G introduced in (19).

Corollary 1. *Under the same hypotheses of Theorem 2, the rate function G is real analytic in the whole interval (U, V) . Moreover, G is strictly convex and essentially smooth in (U, V) , with a unique minimal value $G(\beta) = 0$.*

Proof. Note that function $\tau = \tau_x$, implicitly defined by $\beta(\tau) = x$ (for any $x \in (U, V)$), is real analytic in the whole interval (U, V) . Thus, it is easy to see that

$$G'(x) = -\beta(\tau_x)\tau'_x + \tau_x + x\tau'_x = \tau_x \tag{25}$$

implying $G(x)$ decreasing in (U, β) and increasing in (β, V) , with a unique minimal value $G(\beta) = 0$. Moreover, $G''(x) = \tau'_x = 1/\beta'(\tau_x)$, which is always positive because of relation (15), and hence G is strictly convex in (U, V) . At last, equality (25) also proves that

$$\lim_{x \rightarrow U^+} G'(x) = -\infty \text{ and } \lim_{x \rightarrow V^-} G'(x) = +\infty$$

showing that G is essentially smooth in (U, V) . \square

In spite of the previous corollary, a problem that remains to investigate is the behaviour of $G(x)$ as x tends to the extremes of the interval (U, V) , at least in the standard case when $U = 0$ and $V = 1$. A precise evaluation of these limits is obtained in the next section under a further condition on matrices A and B . Here, we only observe that even if function G can be extended to the values $U = 0$ and $V = 1$, however the limits (1) and (2) of Definition 1 may not exist for $x = x_0 = 0$ and $x = x_1 = 1$, respectively. The reason is that, even if $A + B$ is irreducible and aperiodic, matrices A and B may have periodicity greater than 1 and this implies a fluctuation of the probabilities $\Pr(Y_n \leq 0) = \Pr(Y_n = 0)$ and $\Pr(Y_n \geq n) = \Pr(Y_n = n)$, for $n \rightarrow +\infty$, which prevents these quantities from having a precise limit as n grows indefinitely. The following example shows a simple case of such a situation. This is the reason why in the present work large deviations are introduced as limit properties holding over open intervals rather than closed ones as usual in the literature (compare for instance our Definition 1 with those given in [11,15]).

Example 2. The linear representation defined by the finite state automaton of Fig. 2 is primitive, but matrix A has period 2 (a classical “matrice clignotante”), and hence $\xi^t A^n \eta = 0$ when n is even, while $\xi^t A^n \eta = 1$ when n is odd. Therefore $\Pr(Y_n \geq n) = \frac{\xi^t A^n \eta}{\xi^t (A+B)^n \eta} = 0$ for all even n , while it is of the order $\Theta(\lambda^{-n})$ when n is odd, λ being the largest eigenvalue of $A + B$. As a consequence, the value $\frac{1}{n} \log \Pr(Y_n \geq n)$ has no limit for n tending to $+\infty$. By symmetry, exchanging matrices A and B , a similar situation can be arranged for $\Pr(Y_n \leq 0)$. \diamond

6. Large deviations in the interval (0,1)

A natural question arising at this point is whether the interval (U, V) coincides with $(0, 1)$, as in the case of binomial r.v.’s considered in section 2, and, in this case, what the limits of $G(x)$ are as x tends to the extremes of $(0, 1)$.

It is clear that, if A and B are strictly positive matrices, then $U = 0$ and $V = 1$. However, as we show below, this may occur also in less restrictive cases.

Note that A and B are non-zero matrices with entries in \mathbb{R}_+ , and hence they both have a real non-negative eigenvalue greater or equal to the modulus of any other eigenvalue of the respective matrix. We denote such eigenvalues by λ_A and λ_B , respectively. Clearly, as A is not primitive in general, it may occur $\lambda_A = 0$ or $\lambda_A = |\mu|$ for some eigenvalue μ of A such that $\mu \neq \lambda_A$, and the same may happen for λ_B . However, by statement (iv) of Theorem 1, it is clear that $\lambda_A < \lambda$ and $\lambda_B < \lambda$. Moreover, $\lambda_A > 0$ if and only if A has an eigenvalue different from 0, and the same holds for B .

Theorem 3. Assuming $\{Y_n\}$ defined by a linear representation (ξ, A, B, η) with primitive matrix $A + B$, the following properties hold:

1) If $\lambda_B > 0$ then $U = 0$ and

$$\lim_{x \rightarrow 0^+} G(x) = \log(\lambda/\lambda_B). \tag{26}$$

2) If $\lambda_A > 0$ then $V = 1$ and

$$\lim_{x \rightarrow 1^-} G(x) = \log(\lambda/\lambda_A). \tag{27}$$

Proof. In order to prove point 1), let v_B and u_B be left and right eigenvectors of B with respect to λ_B , normed so that $v_B' u_B = 1$. Since $\lambda_B > 0$, v_B and u_B cannot be null. Moreover, for $t \rightarrow -\infty$, the matrix $Ae^t + B$ tends to B and hence the eigenvalue $y(t)$ converges to λ_B , while the matrix $u_t v_t'$ tends to $u_B v_B'$, implying $v_t' B u_t \rightarrow v_B' B u_B = \lambda_B$ and similarly $v_t' A u_t \rightarrow v_B' A u_B \geq 0$. As a consequence, for $t \rightarrow -\infty$, we have

$$y'(t) = v_t' A e^t u_t = (v_B' A u_B + o(1))e^t = O(e^t) = o(1).$$

Therefore, by equality (13) the last relation implies (for $t \rightarrow -\infty$):

$$\beta(t) = \frac{v_t' A e^t u_t}{v_t' (Ae^t + B) u_t} = \frac{O(e^t)}{\lambda_B + o(1)} = o(1) \tag{28}$$

and hence $U = 0$. Note that relation (28) holds whatever $\lambda_A \geq 0$ is.

Now, if $x \rightarrow 0^+$ then $\tau_x \rightarrow -\infty$ and, reasoning as above, we have $y(\tau_x) = v_{\tau_x}' (Ae^{\tau_x} + B) u_{\tau_x} = \lambda_B + o(1)$. Moreover, applying relation (28), we get $x\tau_x = \beta(\tau_x)\tau_x = O(e^{\tau_x})\tau_x = o(1)$; as a consequence

$$\lim_{x \rightarrow 0^+} G(x) = \lim_{x \rightarrow 0^+} -\log(y(\tau_x)) + \log(\lambda) + x\tau_x = \log(\lambda/\lambda_B)$$

and the proof of point 1) is complete.

In order to prove point 2), assume $t \rightarrow +\infty$ and let $L(t) = A + Be^{-t}$. Exchanging A and B in the previous argument and recalling that $L(t)$ and $Ae^t + B$ have the same eigenvectors, we obtain

$$v_t' A u_t = \lambda_A + o(1), \quad v_t' B e^{-t} u_t = O(e^{-t}) = o(1).$$

As a consequence, for $t \rightarrow +\infty$, we have

$$\beta(t) = \frac{v_t' A e^t u_t}{v_t' (Ae^t + B) u_t} = \frac{v_t' A u_t}{v_t' (A + B e^{-t}) u_t} = \frac{\lambda_A + o(1)}{\lambda_A + o(1)} = 1 + o(1)$$

which implies $V = 1$. Also in this case the property holds whatever $\lambda_B \geq 0$ is.

Now, if $x \rightarrow 1^-$ then $\tau_x \rightarrow +\infty$ and, arguing as above, we obtain $y(\tau_x) = (\lambda_A + o(1))e^{\tau_x}$ and $\beta(\tau_x) = 1 + O(e^{-\tau_x})$. This implies $x\tau_x = \beta(\tau_x)\tau_x = \tau_x + o(1)$, and hence we get

$$\lim_{x \rightarrow 1^-} G(x) = \lim_{x \rightarrow 1^-} -\log(y(\tau_x)) + \log(\lambda) + x\tau_x = \log(\lambda/\lambda_A)$$

which proves point 2). \square

Thus, for any primitive matrix $A + B$, the shape of the rate function $G(x)$ is entirely determined whenever both $\lambda_A > 0$ and $\lambda_B > 0$ hold.

Corollary 2. Assume $A + B$ primitive, $\lambda_A > 0$ and $\lambda_B > 0$. Then $\{Y_n\}$ satisfies a large deviation property in the interval $(0, 1)$ with rate function $G(x)$ defined in (19). Moreover, G is real analytic, strictly convex and essentially smooth in $(0, 1)$, with finite limits at the extremes of the interval given by $\lim_{x \rightarrow 0^+} G(x) = \log(\lambda/\lambda_B)$ and $\lim_{x \rightarrow 1^-} G(x) = \log(\lambda/\lambda_A)$.

Example 3. Consider the linear representation defined by the weighted finite automaton in the left hand side of Fig. 3. The same figure includes the main parameters of the model in this case. Moreover, here we have $y(t) = 2^{-1} \left(1 + 3e^t + \sqrt{1 + 54e^t + 9e^{2t}} \right)$. By using Mathematica one can plot the function $G(x) = \log 6 + xt - \log y(t)$ under condition $y'(t) = xy(t)$, in the interval $(0, 1)$. Its graphic is shown in the right hand side of the figure. By relations (26) and (27), we know that the limits of $G(x)$ for $x \rightarrow 0^+$ and $x \rightarrow 1^-$ are $\log 6$ and $\log 2$, respectively. \diamond

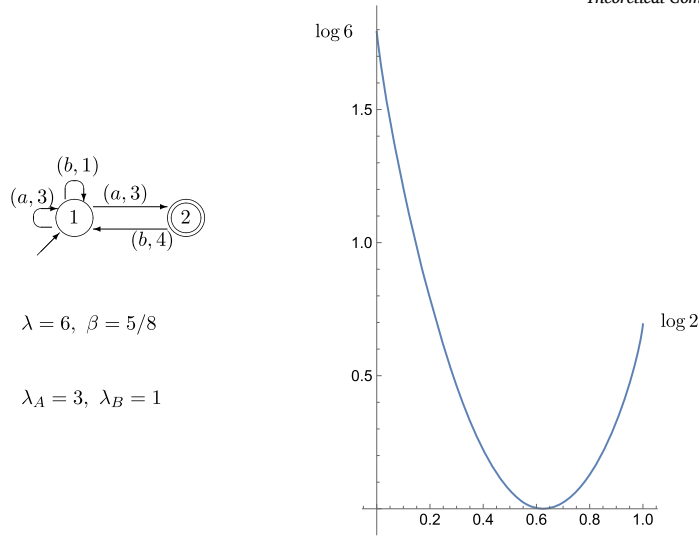


Fig. 3. Rational stochastic model satisfying the hypothesis of Corollary 2. The associated sequence $\{Y_n\}$ satisfies a large deviation property in the interval $(0, 1)$. On the right hand side the graphic of the corresponding rate function G .

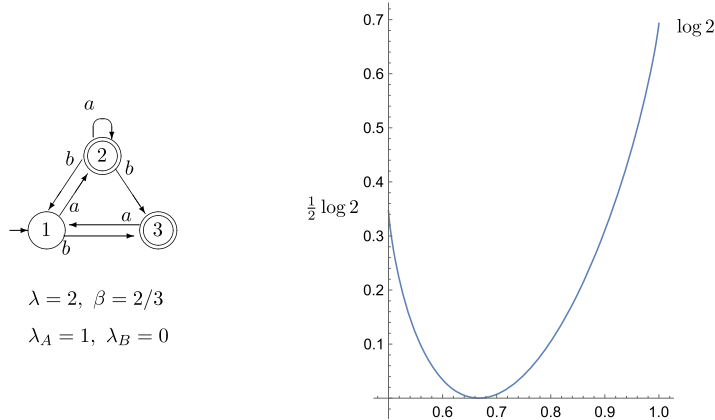


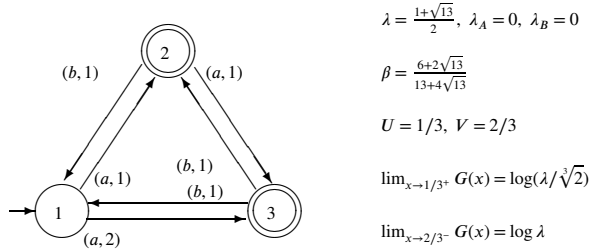
Fig. 4. Finite automaton (with weight 1 for all transitions) defining a primitive rational model where $\lambda_B = 0$ and $U = 1/2$, and graphic of the corresponding rate function G .

The previous example should be compared with Example 1 in order to see differences and similarities between the rate functions $G(x)$ and $B_p(x)$. Note that the probability p corresponds to β , while the limit values $\log \frac{1}{1-p}$ and $\log \frac{1}{p}$ correspond to $\log \frac{\lambda}{\lambda_B}$ and $\log \frac{\lambda}{\lambda_A}$, respectively.

However, Theorem 3 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 $\lim_{x \rightarrow U^+} G(x)$ and $\lim_{x \rightarrow V^-} G(x)$ are not established. We only know, by Corollary 1, that $|G'(x)| \rightarrow +\infty$ as x approaches U or V .

About this problem, we now show examples of primitive rational models where $\lambda_B = 0$ and $0 < U$, or $\lambda_A = 0$ and $V < 1$, or both. Clearly, by reasons of symmetry, exchanging matrices A and B , the first example also yields a case for the second situation. These examples show that, in Theorem 3, the hypothesis $\lambda_B > 0$ cannot be omitted to prove $U = 0$ and, similarly, without the hypothesis $\lambda_A > 0$ the relation $V = 1$ is not true in general. They also show the limit behaviour of $G(x)$ at the extremes of the interval (U, V) in these cases.

Example 4. Consider the linear representation defined by the finite automaton of Fig. 4, where all transitions have weight 1 and the main parameters are shown in the same figure. Here we obtain $y(t) = \frac{e^t}{2} (1 + \sqrt{1 + 8e^{-t}})$, $U = 1/2$, and hence the large deviations property holds in the interval $(\frac{1}{2}, 1)$. The graphic of the rate function $G(x) = \log 2 + xt - \log y(t)$, under condition $y'(t) = xy(t)$, is shown at the right hand side of the figure. In this case it turns out that $\lim_{x \rightarrow 1/2^+} G(x) = \log \sqrt{2}$, while $\lim_{x \rightarrow 1^-} G(x) = \log 2$ according with (27).



$$\lambda = \frac{1+\sqrt{13}}{2}, \lambda_A = 0, \lambda_B = 0$$

$$\beta = \frac{6+2\sqrt{13}}{13+4\sqrt{13}}$$

$$U = 1/3, V = 2/3$$

$$\lim_{x \rightarrow 1/3^+} G(x) = \log(\lambda/\sqrt[3]{2})$$

$$\lim_{x \rightarrow 2/3^-} G(x) = \log \lambda$$

Fig. 5. Finite automaton defining a primitive rational model where $\lambda_A = \lambda_B = 0$, $U = 1/3$, and $V = 2/3$.

Other evaluations obtained in this case, for $t \rightarrow -\infty$, are the following:

$$y(t) = v_t(Ae^t + B)u_t = \sqrt{2}e^{t/2}(1 + \Theta(e^{t/2}))$$

$$y'(t) = v_t Au_t e^t = \frac{e^{t/2}}{\sqrt{2}}(1 + \Theta(e^t)).$$

This means (still for $t \rightarrow -\infty$) that $v_t B u_t \sim \frac{e^{t/2}}{\sqrt{2}}$, while $v_t A u_t \sim \frac{e^{-t/2}}{\sqrt{2}}$. Therefore, even if $v'_t u_t = 1$ for every $t \in \mathbb{R}$, for $t \rightarrow -\infty$ some entries of the matrix $u_t v'_t$ are of the order $O(e^{t/2}) \rightarrow 0$, while others are of the order $\Omega(e^{-t/2}) \rightarrow +\infty$. \diamond

Example 5. Consider the linear representation defined by the finite automaton in the left hand side of Fig. 5. Again, a direct computation, together with an appropriate use of the Implicit Function Theorem (see for instance [27, Ch. 9]), proves the equalities in the right hand side of the same figure. \diamond

7. Conclusions

In conclusion we can see that, under the hypotheses of Corollary 2, the rate function G is rather similar to the map B_p defined in Section 2, i.e. the rate function of large deviations for the sequence $\{X_{n,p}\}$ of binomial r.v.'s. In this case, the interval of validity of the property is the same. Moreover, the constant $\beta = \frac{v^t A u}{\lambda}$, where $G(x)$ takes its minimal value 0, corresponds to the success probability p of $X_{n,p}$, while the limits of $G(x)$ for $x \rightarrow 0^+$ and $x \rightarrow 1^-$, i.e. the values $\log(\lambda/\lambda_B)$ and $\log(\lambda/\lambda_A)$, correspond to $-\log(1 - p)$ and $-\log p$, respectively.

However, from Examples 4 and 5, we have seen that if $\lambda_B = 0$ or $\lambda_A = 0$, the interval of validity (U, V) may be different from $(0, 1)$. Hence natural problem for further investigation is to determine a general law for the values U and V in these cases, together with the limit of the rate function $G(x)$ for x tending to U or V . Other subjects for further studies are the validity of large deviation properties for non-primitive rational models, in particular for models consisting of two or more components, like those studied in [9,18] and in [19].

CRedit authorship contribution statement

Massimiliano Goldwurm: Writing – original draft. **Marco Vignati:** Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We wish to thank Elena Villa and Claudio Macchi for interesting and useful discussions on several aspects of the theory of large deviations.

This work has been supported by Università degli Studi di Milano, under project PSR 2021.

References

[1] J. Berstel, 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, *Theor. Comput. Sci.* 302 (2003) 431–456.
 [3] A. Bertoni, C. Choffrut, M. Goldwurm, V. Lonati, Local limit properties for pattern statistics and rational models, *Theory Comput. Syst.* 39 (2006) 209–235.

- [4] A. Bertoni, P. Massazza, R. Radicioni, Random generation of words in regular languages with fixed occurrences of symbols, in: Proc. WORDS'03, Fourth Intern. Conf. on Combinatorics of Words, Turku (Finland), 2003, pp. 332–343.
- [5] J. Bourdon, M. Régnier, Large deviation properties for patterns, *J. Discret. Algorithms* 24 (2014) 2–11.
- [6] S. Broda, A. Machiavelo, N. Moreira, R. Reis, A hitchhiker's guide to descriptonal complexity through analytic combinatorics, *Theor. Comput. Sci.* 528 (2014) 85–100.
- [7] N. Broutin, L. Devroye, Large deviations for the weighted height of an extended class of trees, *Algorithmica* 46 (2006) 271–297.
- [8] C. Choffrut, M. Goldwurm, V. Lonati, On the maximum coefficients of rational formal series in commuting variables, in: Proc. 8th DLT Conference, in: Lecture Notes in Computer Science, vol. 3340, Springer, 2004, pp. 114–126.
- [9] D. de Falco, M. Goldwurm, V. Lonati, Frequency of symbol occurrences in bicomponent stochastic models, *Theor. Comput. Sci.* 327 (3) (2004) 269–300.
- [10] A. Dembo, O. Zeitouni, *Large Deviation Techniques 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, M. Régnier, Rare events and conditional events on random strings, *Discrete Math. Theor. Comput. Sci.* 6 (2004) 191–214.
- [13] A. Denise, O. Roques, M. Termier, Random generation of words of context-free languages according to the frequency of letters, in: D. Gardy, A. Mokkaedem (Eds.), *Mathematics and Computer Science*, in: Trends in Mathematics, Birkäuser, 2000, pp. 113–125.
- [14] L. Devroye, Universal limit laws for depths in random trees, *SIAM J. Comput.* 28 (1999) 409–432.
- [15] P. Flajolet, R. Sedgewick, *Analytic Combinatorics*, Cambridge Univ. Press, 2009.
- [16] P. Flajolet, W. Szpankowski, B. Vallée, Hidden word statistics, *J. ACM* 53 (2006) 147–183.
- [17] B.V. Gnedenko, *Theory of Probability*, Gordon and Breach Science Publ., 1997.
- [18] M. Goldwurm, J. Lin, M. Vignati, Local limit laws for symbol statistics in bicomponent rational models, *Theor. Comput. Sci.* 970 (2023) 114051, <https://doi.org/10.1016/j.tcs.2023.114051>.
- [19] M. Goldwurm, V. Lonati, Pattern statistics and Vandermonde matrices, *Theor. Comput. Sci.* 356 (2006) 153–169.
- [20] P. Grabner, M. Rigo, Distribution of additive functions with respect to numeration systems on regular languages, *Theory Comput. Syst.* 40 (2007) 205–223.
- [21] H.-K. Hwang, Large deviations for combinatorial distributions. I: central limit theorems, *Ann. Appl. Probab.* 6 (1) (1996) 297–319.
- [22] P. Jacquet, W. Szpankowski, *Analytic Pattern Matching: From DNA to Twitter*, Cambridge University Press, 2015.
- [23] P. Massazza, R. Radicioni, On computing the coefficients of bivariate holonomic formal series, *Theor. Comput. Sci.* 346 (2005) 418–438.
- [24] P. Nicodeme, B. Salvy, P. Flajolet, Motif statistics, *Theor. Comput. Sci.* 287 (2) (2002) 593–617.
- [25] M. Régnier, W. Szpankowski, On pattern frequency occurrences in a Markovian sequence, *Algorithmica* 22 (4) (1998) 621–649.
- [26] C. Reutenauer, *Propriétés arithmétiques et topologiques de séries rationnelles en variables non commutatives*, Thèse Sc. Maths, Doctorat troisième cycle, Université Paris VI, 1977.
- [27] W. Rudin, *Principles of Mathematical Analysis*, 3rd edition, McGraw-Hill, 1976.
- [28] A. Salomaa, M. Soittola, *Automata-Theoretic Aspects of Formal Power Series*, Springer–Verlag, 1978.
- [29] M.-P. Schützenberger, Finite counting automata, *Inf. Control* 5 (1962) 91–107.
- [30] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer–Verlag, New York Heidelberg Berlin, 1981.
- [31] P. Turakainen, Generalized automata and stochastic languages, *Proc. Am. Math. Soc.* 21 (1969) 303–309.