Large Deviation Properties for Pattern Statistics in Primitive Rational Models

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

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

Abstract

We present a large deviation property for the pattern statistics representing the number of occurrences of a symbol in words of given length generated at random according to a rational stochastic model. The result is obtained assuming that in the model the overall weighted transition matrix is primitive. In particular we obtain a rate function depending on the main eigenvalue and eigenvectors of that matrix. Under rather mild conditions, we show that the range of validity of our large deviation estimate can be extended to the interval (0,1), which represents in our context the largest possible open interval of validity of the property.

Keywords

large deviations, limit distributions, regular languages, rational formal series, pattern statistics

1. Introduction

Large deviation properties represent a classical subject in probability theory. They yield bounds of exponential decay on the probability that a sequence of random variables differs from the mean values 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, as those deriving for instance from the Central Limit Theorem, that concern asymptotically smaller differences.

In analytic combinatorics large deviation estimates are considered in the study of various relevant structures [15]. In particular they occur in pattern statistics [22] and in the analysis of depth and height of certain classes of trees [7, 14]. In pattern statistics they have been studied with the goal of evaluating the probability of rare events, where a given pattern is over- or under-represented in a random text generated according to a suitable stochastic model [12, 5].

In the present work we prove some properties of this type for sequences 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

ICTCS 2023: International Conference on Theoretical Computer Science, September 13–15, 2023, Palermo, 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)

given length is proportional to the total weight of the accepting transitions labelled by w. This model is quite general, it includes as special cases the traditional Bernoullian and Markovian sources, widely used in the literature to study the number of occurrences of patterns in a random text [24, 25, 16, 5]. We recall that the research concerning pattern statistics has a broad range of motivations and applications [22]. Moreover, it is known that the rational stochastic model allows to generate random words of length n in an arbitrary regular language under uniform distribution: this occurs when the finite automaton defining the model is unambiguous and all transitions have the same weight.

In order to fix our notation, consider a (nondeterministic) 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 analysis of these sequences of random variables is of interest in several contexts. First of all they can represent the number of occurrences of patterns in a random word of length n, generated by a Markovian source, when the set of patterns is given by a regular language [2, 24, 25]. Moreover, they are related to the evaluation of the coefficients of rational formal series (a traditional problem well-studied in the literature [28, 26, 23]) and to the analysis of several problems and properties of regular language. This fact clearly holds for 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]. It also holds for the analysis of additive functions defined on regular languages [20] and for the descriptional complexity of languages and computational models [6]. Further, using the local limit properties of the sequences $\{Y_n\}$, for a wide class of rational series it can be proved that the maximum coefficient of the monomials of degree n 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 value, variance, limit distribution 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 extended 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 in Section 5 that if the transition matrix of $\mathcal{A}$ is primitive, then Y_n satisfies a large deviation property with a rate function depending on the main eigenvalue and the associated eigenvectors. The corresponding proof is rather standard, it relies on traditional tools of analytic combinatorics and the result is implicitly included in the previous literature [21, 11, 15]. However, here our result is significant since it puts in evidence the role played by the main eigenvalue and eigenvectors of the matrix of weights in the definition of both the rate function and the interval of validity of the property. Moreover, in Section 6, we assume a mild condition on the transition matrix of the automaton and, under such hypothesis, we show that the range of validity of the large deviation property can be extended to the entire interval $(0, 1)$, which represents in our context the largest possible open interval where the property may hold.

2. A quick overview on large deviations

Large deviation estimates usually refer to a sequence of random variables (r.v.'s), say $\{X_n\}$, having increasing mean values; it 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 $cE(X_n)$, $c > 0$. Typical situations occur when $E(X_n) \sim \beta n$ for a constant $\beta > 0$, and since this occurs in all our cases, here we start with the following fomal definition [11, 15].

Definition 1. Let $\{X_n\}$ be a sequence of random variables such that $E(X_n) \sim \beta n$ for a constant $\beta > 0$, and let (x_0, x_1) be an 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* relative to 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 $\beta \leq x < x_1$. The first relation concerns the *left tail* $\Pr(X_n \leq xn)$, while the second one refers to the *right tail* $\Pr(X_n \geq xn)$ of the distribution of X_n. It is convenient to keep the two limits separated in the definition since the proofs of a large deviation property often considers one tail at a time.

A classical example of large deviation property concerns the sequence of binomial random variables $\{X_{n,p}\}_n$ of parameters n and p, where $p \in (0, 1)$ is fixed. 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

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

Such a property implies the following result for a larger deviation

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

which can also be obtained by applying the Law of Large Numbers. The following proposition states a large deviation property for $\{X_{n,p}\}_n$ that improves the last relation, showing that the convergence to 0 is exponential with respect to n and the range of validity coincides with the overall interval $(0, 1)$. Its proof is a consequence of Cramer's Theorem (see e.g. [11, 10]), a classical result we discuss later. However, here we prefer to briefly outline a simpler proof to give the flavour of this property and to compare it with our subsequent results.

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

$$B(x) = x \log \frac{x}{p} + (1 - x) \log \frac{1 - x}{1 - p}\ , \qquad \textit{for every } x \in (0, 1).$$

Proof. We first prove the limit on the left tail. To this end, let $0 < x \leq p$ and let $M_n(x) = \max\{\Pr(X_{n,p} = i) : i \in \mathbb{N}, 0 \leq i \leq xn\}$. Then we have

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

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

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

which replaced in (1) proves $\log \Pr(X_{n,p} \leq xn) = -B(x)n + O(\log n)$.

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

$$N_n(x) \leq \Pr(X_{n,p} \geq xn) \leq (n - nx + 1)N_n(x) \qquad \text{for every } x \in [p, 1)$$

where $N_n(x) = \binom{n}{\lfloor xn \rfloor} p^{\lfloor xn \rfloor}(1-p)^{n-\lfloor xn \rfloor}$. As above, replacing this value in the previous inequalities yields $\log \Pr(X_{n,p} \geq xn) = -B(x)n + O(\log n)$. $\square$

The rate function $B(x)$ is strictly convex in the interval $(0, 1)$, takes a unique minimal value at $x = p$, where $B(p) = 0$, while $\lim_{x \to 0^+} B(x) = log \frac{1}{1-p}$ and $\lim_{x \to 1^-} B(x) = log \frac{1}{p}$. Moreover, $B(x)$ grows vertically for x approaching 0 or 1, since $\lim_{x \to 0^+} B'(x) = -\infty$ and $\lim_{x \to 1^-} B'(x) = +\infty$.

Example 1. Let $p = 5/8$. Then, the shape of the rate function $B(x)$ of the large deviation property for the sequence of binomial random viariables $\{X_{n,p}\}$ is described in Figure 1.

$$p = \frac{5}{8}, \qquad B(x):$$

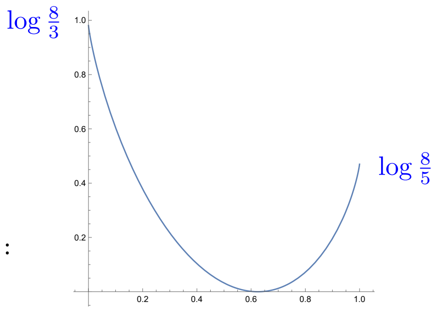

Figure 1: Grafic of rate function $B(x)$ of the large deviation property for $\{X_{n,p}\}$, when $p = \frac{5}{8}$.

We recall that often 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 $+\infty$. A classical situation of this type is established by Cramér's Theorem [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 $R(x) = \sup_{t \in \mathbb{R}}[xt - \log \psi(t)]$, for every $x \in \mathbb{R}$.

3. Symbol statistics for rational models

In order to define our stochastic model consider a *formal series* in the non-commutative variables a, b, that is a function $r : \{a,b\}^* \to \mathbb{R}_+$, where $\mathbb{R}_+ = [0, +\infty)$ and $\{a,b\}^*$ is the free monoid of all words on 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 $\mu : \{a,b\}^* \to \mathbb{R}_+^{m\times m}$ and two (column) arrays $\xi, \eta \in \mathbb{R}_+^m$, such that $(r, w) = \xi'\mu(w)\eta$, for every $w \in \{a,b\}^*$ [1, 27] ([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 the morphism 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. Clearly, such a 4-tuple can be considered as a finite state automaton over the alphabet $\{a,b\}$, with transitions weighted by positive real values. Therefore A (resp. B) represents the matrix of the weights of all transitions labelled by a (resp. b), while ξ (resp. η) is the array of 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 $n \in \mathbb{N}_+$ (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 $\{a,b\}^n$:

$$\sum_{w\in\{a,b\}^n} (r,w) = \xi' \sum_{w\in\{a,b\}^n} \mu(w)\eta = \xi' \left(\prod_{i=1}^{n} \sum_{w_i\in\{a,b\}} \mu(w_i)\right) \eta = \xi' \, (A+B)^n \, \eta \tag{2}$$

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

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

Note that, if r is the characteristic series of a language $L \subseteq \{a,b\}^*$ then Pr is the uniform probability function over the set $L \cap \{a,b\}^n$. 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 $\eta' = (1, 1 \ldots, 1)$.

Then, under the previous hypotheses, we can define the integer random variable (r.v.) $Y_n = |w|_a$, where w is a word chosen at random in $\{a,b\}^n$ with probability $\Pr(w)$, and $|w|_a$ is the number of occurrences of a in w. As $A \neq [0] \neq B$, Y_n is a non-degenerate random variable, taking value in $\{0, 1, \ldots, n\}$. It is clear that the probability function of Y_n is defined by

$$p_n(k) := \Pr(Y_n = k) = \frac{\sum_{w\in\{a,b\}^n, |w|_a = k}(r,w)}{\sum_{w\in\{a,b\}^n}(r,w)} \;, \qquad k \in \{0, 1, \ldots, n\}$$

Since r is rational also the previous probability can be expressed by using its linear representation. The denominator is clearly determined by (2). 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

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

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

As a consequence, if $[x^k]g(x)$ denotes (according to tradition) the coefficient of the monomial of degree k in a polynomial $g(x)$, the probabilities $p_n(k)$'s can be written as

$$p_n(k) = \frac{[x^k]\xi'(Ax+B)^n\eta}{\xi'(A+B)^n\eta} \,, \qquad k \in \{0,1,\ldots,n\} \tag{3}$$

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

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

We have

$$\Psi_n(z) = \sum_{k=0}^{n} p_n(k)e^{zk} = \frac{\xi'(Ae^z+B)^n\eta}{\xi'(A+B)^n\eta} = \frac{h_n(z)}{h_n(0)} \,, \qquad z \in \mathbb{C} \tag{4}$$

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

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

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:

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

Clearly, the r.v.'s $b_i(w)$ ($i = 1, 2, \ldots, n$) 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 $\{0,1,\ldots,n\}$.

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 [29, 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 eigevalue $\lambda > 0$ 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 $A \le T$ (i.e. $a_{ij} \le t_{ij}\ \forall i,j$) and α is an eigenvalue of A then $|\alpha| \le \lambda$. 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

$$\det(Iy - Ae^z - B) = 0 \tag{5}$$

with initial condition $y(0) = \lambda$. Clearly $y(z)$ is eigenvalue of $Ae^z + B$ for every $z \in \mathbb{C}$. Moreover, $y(z)$ is analytic in a neighbourhood of 0 and $y'(0) \neq 0$ since λ is a simple root of the characteristic polynomial of $A + B$.

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) $\mathrm{E}(Y_n) = \beta n + c + O(\varepsilon^n)$, where $|\varepsilon| < 1$, $c \in \mathbb{R}$ and β is a constant satisfying $0 < \beta < 1$ given by $\beta = \frac{y'(0)}{\lambda}$. Moreover $y'(0) = v'Au$, where v' and u are left and right eigenvectors of $A + B$, with respect to λ, such that $v'u = 1$.

2) $\mathrm{Var}(Y_n) = \gamma n + O(1)$, where γ is a positive constant defined by $\gamma = \frac{y''(0)}{\lambda} - \left(\frac{y'(0)}{\lambda}\right)^2$;

3) In a neighbourhood of 0, the function $\Psi_n(z)$ satisfies a "quasi power" condition, that is an equation of the form

$$\Psi_n(z) = r(z) \left(\frac{y(z)}{\lambda}\right)^n (1 + O(\varepsilon^n)) \qquad (|\varepsilon| < 1) \tag{6}$$

where $r(z)$ is 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: therefore $y(t)$ 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). Moreover, all the powers of $Ae^t + B$ satisfy a relation of the form

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

where v_t' and u_t are left and right eigenvectors of $Ae^t + B$ relative to $y(t)$, normed so that $v_t'u_t = 1$ [29, Th. 1.2]. A first consequence is that applying relation (6) to all real z, we obtain (for every $t \in \mathbb{R}$)

$$\Psi_n(t) = \mathrm{E}(e^{tY_n}) = \frac{\xi'(Ae^t + B)^n \eta}{\xi'(A + B)^n \eta} = r(t) \left(\frac{y(t)}{\lambda}\right)^n (1 + O(\varepsilon_t^n)) \tag{7}$$

where the function $r(t) = \frac{\xi' u_t v_t' \eta}{\xi' u_0 v_0' \eta}$ is analytic in $\mathbb{R}$, clearly $r(0) = 1$ and $|\varepsilon_t| < 1$.

5. Large deviations for primitive models

In this section we present a large deviation property for the sequence $\{Y_n\}$ in the primitive case. To this end, assume that $A+B$ is 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 also $Ae^t + B$ is primitive, for any $t \in \mathbb{R}$, we can apply the results of the previous section to all sequences of random variables $\{Y_n(t)\}$. Reasoning as for relation (5), for any $t \in \mathbb{R}$ we can consider the function $y_t(z)$ implicitely defined by the equation

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

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 $y_t'(0) = y'(t)$, $y_t''(0) = y''(t)$.

Applying property **1)** of the previous section to the linear representation (ξ, Ae^t, B, η), for every $t \in \mathbb{R}$ we obtain $\mathrm{E}(Y_n(t)) = \beta(t)n + c_t + O(\varepsilon_t^n)$, where $c_t \in \mathbb{R}$ and $\varepsilon_t \in (0,1)$ are constant and $\beta(t)$ is a real function given by $\beta(t) = \frac{y_t'(0)}{y_t(0)} = \frac{y'(t)}{y(t)}$, for all $t \in \mathbb{R}$. Clearly $\beta(0) = \beta$. Moreover, by the same property, we have $y(t) = v_t'(Ae^t + B)u_t$, $y'(t) = v_t' Ae^t u_t$ and hence

$$\beta(t) = \frac{v_t' Ae^t u_t}{v_t'(Ae^t + B)u_t} \tag{8}$$

which implies

$$0 < \beta(t) < 1 \qquad \forall t \in \mathbb{R} \tag{9}$$

Analogously, applying property **2)** of the previous section to $Y_n(t)$ we get $\mathrm{Var}(Y_n(t)) = \gamma(t)n + O(1)$, for all $t \in \mathbb{R}$, where $\gamma(t)$ is a positive constant given by

$$\gamma(t) = \frac{y_t''(0)}{y_t(0)} - \left(\frac{y_t'(0)}{y_t(0)}\right)^2 = \beta'(t) > 0 \qquad \forall t \in \mathbb{R} \tag{10}$$

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

$$U = \lim_{t \to -\infty} \beta(t), \qquad V = \lim_{t \to +\infty} \beta(t) \tag{11}$$

By relation (9), we have $0 \le U < \beta(0) < V \le 1$, which, together with relation (10), implies the following statement.

Lemma 1. *Let U and V be defined by (11). Then, for every $x \in (U,V)$ there exists a unique $\tau_x \in \mathbb{R}$ such that*

$$\beta(\tau_x) = x \tag{12}$$

Moreover, $\tau_x < 0$ whenever $x < \beta$, $\tau_\beta = 0$ and $\tau_x > 0$ when $x > \beta$.

Now we apply property **3)** of the previous section to the random variable $Y_n(t)$: we get a "quasi power" property for the moment generating function of $Y_n(t)$, that is $\Psi_{Y_n(t)}(z) = r_t(z)\left(\frac{y(t+z)}{y(t)}\right)^n (1 + O(\varepsilon_t^n))$, for some $\varepsilon_t \in (0,1)$, 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 random variables $\left\{ \frac{Y_n(t) - \beta(t)n}{\sqrt{\gamma(t)n}} \right\}_n$ 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

$$\lim_{n \to \infty} \Pr \left(\frac{Y_n(t) - \beta(t)n}{\sqrt{\gamma(t)n}} \leq x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du \;, \qquad \forall t \in \mathbb{R} \tag{13}$$

The previous results allow us to prove a large deviation property for $\{Y_n\}$.

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*

$$G(x) = -\log \left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}} \right) \;, \qquad \forall x \in (U, V)$$

where U and V are defined in relation (11) and τ_x is given by equation (12).

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:

$$\lim_{n \to +\infty} \frac{1}{n} \log \Pr(Y_n \geq xn) = \log \left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}} \right) \tag{14}$$

By Markov inequality, for every $t > 0$, we have

$$\Pr(Y_n \geq xn) = \Pr(e^{tY_n} \geq e^{txn}) \leq \frac{\mathrm{E}(e^{tY_n})}{e^{txn}}$$

and hence, by relation (7) we get

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

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

This bound can be further 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

$$\varphi_x(t) = \log \left(\frac{y(t)}{\lambda e^{xt}} \right) \qquad \forall t \in \mathbb{R} \tag{15}$$

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

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

Also observe that $\varphi_x(t)$ is convex since $\varphi''_x(t) = \beta'(t) > 0$ by relation (10).

An analogous lower bound for $\Pr(Y_n \geq xn)$ can be proved by considering the random variable $Y_n(\tau_x)$. Since $[z^k]\xi'(Ae^{\tau_x}z + B)^n\eta = e^{\tau_x k}[z^k]\xi'(Az + B)^n\eta$, by relations (3) and (4) we have

$$\Pr(Y_n = k) = \frac{\Pr\left(\; Y_n(\tau_x) = k \;\right)\; \Psi_n(\tau_x)}{e^{\tau_x k}} \qquad \forall\, k = 0, 1, \ldots, n \tag{16}$$

Also note that $\mathrm{E}(Y_n(\tau_x)) = \beta(\tau_x)n + O(1) = xn + O(1)$ and by (13) we know that $\{Y_n(\tau_x)\}_n$ has a Gaussian limit distribution. This means that, for every $\varepsilon > 0$, $\Pr\left(\ Y_n(\tau_x) > (x+\varepsilon)n \ \right) = o(1)$ and then

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

Then, from this relation and the identities (16) and (7) we get

$$\begin{aligned} \Pr(Y_n \geq xn) &\geq \Pr\left(\ xn \leq Y_n \leq (x+\varepsilon)n \ \right) \\ &\geq \frac{\Pr\left(\ xn \leq Y_n(\tau_x) \leq (x+\varepsilon)n \ \right) \ \Psi_n(\tau_x)}{e^{\tau_x(x+\varepsilon)n}} = \\ &= \left(\frac{1}{2} + o(1)\right) r(\tau_x) \left(\frac{y(\tau_x)}{\lambda e^{\tau_x(x+\varepsilon)}}\right)^n (1 + O(\varepsilon_{\tau_x}^n)) \end{aligned} \tag{17}$$

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 (14).

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

$$\lim_{n\to+\infty} \frac{1}{n} \log \Pr(Y_n \leq xn) = \log\left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}}\right) \tag{18}$$

where $\tau_x \leq 0$ is defined by equation (12).

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 $\phi_x(t)$ given by (15) is well defined also in this case. For every $x \in (U, \beta]$ and every $t < 0$, by Markov inequality and relation (7) we get

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

By Lemma 1 the minimum of $\varphi_x(t) = \log\left(\frac{y(t)}{\lambda e^{xt}}\right)$ 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 (18).

The corresponding lower bound is obtained as in the analysis of the right tail, leading to relation (17), with obvious changes. $\square$

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

A natural question arising at this point is whether the interval (U, V) can be extended to $(0, 1)$ as in the case of the sequences of binomial random variables considered in section 2.

Let us assume that $A + B$ is a primitive matrix. Since A and B are non-null matrices with entries in $\mathbb{R}_+$, they admit a real non-negative eigenvalue that is 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 $\lambda_B < \lambda$, where λ is the Perron-Frobenius eigenvalue of $A + B$.

Now, assume $\lambda_B > 0$ (which is equivalent to require that B has a nonnull eigenvalue) and let v_B and u_B be left and right eigenvectors of B with respect to λ_B, normed so that $v'_B u_B = 1$. Clearly v_B and u_B cannot be null. Moreover, for $t \to -\infty$, 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 $u_B v'_B$, implying $v'_t B u_t \to v'_B B u_B = \lambda_B$ and similarly $v'_t A u_t \to v'_B A u_B$. As a consequence, for $t \to -\infty$, we have

$$y'(t) = v'_t A e^t u_t = O(e^t) = o(1)$$

Therefore, by equality (8) the last relation implies (for $t \to -\infty$)

$$\beta(t) = \frac{v'_t A e^t u_t}{v'_t (Ae^t + B) u_t} = \frac{O(e^t)}{\lambda_B + o(1)} = o(1) \tag{19}$$

and hence $U = 0$.

Analogously, if $\lambda_A > 0$ we get $V = 1$. In fact, assume $t \to +\infty$ 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 $v'_t A u_t = \lambda_A + o(1)$ and $v'_t B e^{-t} u_t = O(e^{-t}) = o(1)$. As a consequence, for $t \to +\infty$, we have

$$\beta(t) = \frac{v'_t A e^t u_t}{v'_t (Ae^t + B) u_t} = \frac{v'_t A u_t}{v'_t (A + Be^{-t}) u_t} = \frac{\lambda_A + o(1)}{\lambda_A + o(1)} = 1 + o(1) \tag{20}$$

which implies $V = 1$.

The previous argument proves the following result.

Theorem 3. *Let $\{Y_n\}$ be defined by a linear representation (ξ, A, B, η) where $A + B$ is primitive and both A and B have a nonnull eigenvalue. Then $\{Y_n\}$ satisfies a large deviation property in the interval $(0, 1)$ with rate function $G(x) = -\log\left(\frac{y(\tau_x)}{\lambda e^{x\tau_x}}\right)$, for any $x \in (0, 1)$, where τ_x is the unique real value such that $\beta(\tau_x) = x$.*

Under the same hypotheses of the previous theorem we can study the function $G(x)$. Note that the function $\tau = \tau_x$, implicitly defined by $\beta(\tau) - x = 0$, is well defined and analytic for $x \in (0, 1)$. Thus it is easy to see that

$$G'(x) = -\beta(\tau_x)\tau'_x + \tau_x + x\tau'_x = \tau_x$$

and hence $G(x)$ is decreasing in $(0, \beta)$ and increasing in $(\beta, 1)$, with a unique minimal value $G(\beta) = 0$. This also proves that

$$\lim_{x\to 0^+} G'(x) = -\infty \,, \qquad \text{and} \quad \lim_{x\to 1^-} G'(x) = +\infty$$

and hence $G(x)$ grows vertically for $x \to 1^-$ and for $x \to 0^+$. Moreover, $G''(x)$ equals τ'_x, which is always positive since τ_x is strictly increasing in $(0, 1)$, and hence $G(x)$ is a convex function in $(0, 1)$.

Finally, let us determine the behaviour $G(x)$ for $x \to 0^+$ and for $x \to 1^-$. Letting $x \to 0^+$, we have $\tau_x \to -\infty$ and, arguing as for equation (19), we obtain $y(\tau_x) = \lambda_B + o(1)$. Moreover, applying the same equation (19), we get

$$x\tau_x = \beta(\tau_x)\tau_x = O(e^{\tau_x})\tau_x = o(1)$$

As a consequence, we can derive the following limit

$$\lim_{x\to 0^+} G(x) = \lim_{x\to 0^+} -\log(y(\tau_x)) + \log(\lambda) + x\tau_x = \log(\lambda/\lambda_B) \tag{21}$$

Analogously, let $x \to 1^-$. Then $\tau_x \to +\infty$ and, reasoning as for equation (20), we get $y(\tau_x) = e^{\tau_x}(\lambda_A + o(1))$ and $x\tau_x = \tau_x + o(1)$, which implies

$$\lim_{x\to 1^-} G(x) = \lim_{x\to 1^-} -\log(y(\tau_x)) + \log(\lambda) + x\tau_x = \log(\lambda/\lambda_A) \tag{22}$$

Example 2. As an example, consider the linear representation defined by the weighted finite automaton of Figure 2 (left hand side). In this case we have $\lambda = 6$ and $y(t) = 2^{-1}\left(1 + 3e^t + \sqrt{1 + 54e^t + 9e^{2t}}\right)$. By using Mathematica and plotting function $G(x) = \log 6 + xt - \log y(t)$ in the interval $(0, 1)$, with condition $y'(t) = xy(t)$, we obtain the graphic at the right hand side of the same figure. The numerical computation also confirms that the limit of $G(x)$ for $x \to 0^+$ and $x \to 1^-$ are $\log 6$ and $\log 2$, respectively, as proved by relations (21) and (22).

For a qualitative comparison with the rate function associated with the sequence of binomial r.v.'s $\{X_{n.p}\}$, when $p = 5/8$, such a graphic should be compared with that one of Figure 1.

7. Conclusion

In conclusion we can see that the qualitative behaviour of rate function $G(x)$ is rather similar to that one of $B(x)$ discussed in section 2, that is the rate function of the large deviation property for the sequence $\{X_{n,p}\}$ of binomial random variables. Note that the interval of validity of the property is the same. The constant $\beta = \frac{v'Au}{\lambda}$, where $G(x)$ takes its minimal value 0, corresponds to success probability p of $X_{n,p}$. Moreover, the limits of $G(x)$ for $x \to 0^+$ and $x \to 1^-$, i.e. $\log(\lambda/\lambda_B)$ and $\log(\lambda/\lambda_A)$ respectively, correspond to $-\log(1-p)$ and $-\log p$.

Natural problems for further investigation concern the case when λ_A or λ_B are null; in this case Theorem 3 does not apply immediately but one may guess that a similar property holds. In particular a natural question is whether in this case the large deviation property still holds in the interval $(0, 1)$. 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].

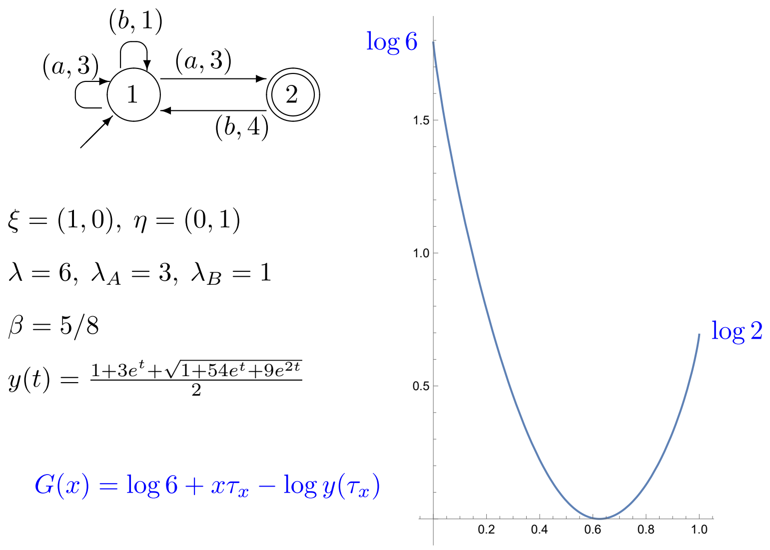

Figure 2: Weighted finite automaton defining a sequence of r.v.'s $\{Y_n\}$ with the corresponding key values λ, λ_A, λ_B, β, $y(t)$, and the graphic of rate function $G(x)$ of the associated large deviation property.

Acknowledgments

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

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 height 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] 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. *Theoret. Comput. Sci.*, 970:114051, 2023, https://doi.org/10.1016/j.tcs.2023.114051.

[19] M. Goldwurm and V. Lonati. Pattern statistics and Vandermonde matrices. *Theoret. Comput. Sci.*, 356:153–169, 2006.

[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. Massazza and R. Radicioni. On computing the coefficients of bivariate holonomic formal series. *Theoret. Comput. Sci.*, 346: 418–438, 2005.

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

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

[26] C. Reutenauer. Propriétés arithmétiques et topologiques de séries rationelles en variables non commutatives. Thèse Sc. Maths, Doctorat troisieme cycle, Université Paris VI, 1977.

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

[28] M.-P. Schützenberger. Finite counting automata. *Inform. and Control*, 5: 91–107, 1962.

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