Large Deviation Principle for a Markov Chain with a Countable State Space

Arnaud de la Fortelle, Guy Fayolle

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Large Deviation Principle for a Markov Chain with a Countable State Space

Arnaud de La Fortelle et Guy Fayolle

N° 3503

Septembre 1998

THÈME 1

Rapport de recherche

ISSN 0249-6399

Large Deviation Principle for a Markov Chain with a Countable State Space

Arnaud de La Fortelle* et Guy Fayolle†

Thème 1 — Réseaux et systèmes
Projet MEVAL

Rapport de recherche n° 3503 — Septembre 1998 — 39 pages

Abstract: Let E be a denumerable state space, X be an homogeneous Markov chain on E with kernel P. Then the chain X verifies a *weak* Sanov's theorem, i.e. a weak large deviation principle holds for the law of the pair empirical measure. In our opinion this is an improvement with respect to the existing literature since LDP in the Markov case requires in general, either E to be finite, or strong uniformity conditions, which important classes of chains do not verify, e.g. bounded jump networks. Moreover this LDP holds for *any* discrete state space Markov chain including non-ergodic chains.

Key-words: Large deviation, Markov chain, empirical measure, entropy, information, combinatorics, cycle

(Résumé : tsvp)

The work of the first author is a part of a PhD thesis supported by the École Nationale des Ponts et Chaussées and by the French Ministère de l'Équipement.

* E-mail address: Arnaud.De_La_Fortelle@inria.fr
† E-mail address: Guy.Fayolle@inria.fr

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : 01 39 63 55 11 - International : +33 1 39 63 55 11
Télécopie : (33) 01 39 63 53 30 - International : +33 1 39 63 53 30

Principe de grandes déviations pour une chaîne de Markov à espace d'états dénombrable

Résumé : Soit E un espace d'état dénombrable et X une chaîne de Markov homogène sur E de matrice de transitions P. La chaîne X vérifie un principe de Sanov *faible*, c'est-à-dire que la loi des mesures empiriques sur les paires satisfait un principe faible de grandes déviations (PGD). Nous pensons que c'est une amélioration, dans la mesure où les PGD, dans le cas markovien, exigent dans l'état actuel de l'art que P satisfasse une condition d'uniformité très forte, qui n'est hélas pas vérifiée pour d'importantes classes de chaînes, comme par exemple les réseaux à sauts bornés. De plus, ce PGD est valide pour *toute* chaîne de Markov sur un espace d'état discret, y compris les chaînes non-ergodiques.

Mots-clé : Grandes déviations, chaîne de Markov, mesure empirique, entropie, information, combinatoire, cycle

1 Introduction and main ideas

Large deviations theory is a useful tool for analyzing stochastic processes on the long run. Situations where it applies are diverse, so that general results have been specialized in many ways to fit each case. Indeed, the study of stochastic networks has particular features, e.g. the states are countable, the processes are Markov or semi-Markov, the jumps are bounded, etc. LDP around fluid limits describes the behavior of systems when time *and space* are scaled [23], but if one wishes to get more detailed information, for example the probability for a network to visit some state x with a frequency greater than f, we need a different point of view. This paper proposes a proof of a LDP for any Markov chain on a countable space, which provides a general setting encountered in many network models.

General theory ([9, 8]) requires, in the case of Markov chains on Polish spaces, a strong uniformity condition, i.e. there exist integers $0 < l \leq N$ and $M \geq 1$ such that, for *all* states σ, τ,

$$\pi^l(\sigma, .) \leq \frac{M}{N} \sum_{m=1}^{N} \pi^m(\tau, .), \tag{1.1}$$

where $\pi^m(\tau, .)$ denotes the m–step transition probability for a given initial state τ. But this is in fact a very restrictive condition: M/M/1 file, Jackson networks and, more generally, *bounded downward jumps networks do not satisfy it.* Refinements have been obtained [11, 7, 24, 19] but they still do not suffice for our purpose.

The case of finite state space is well-known and the rate function is expressed as an *entropy* [10]. The link between large deviations and entropy has already been used [4, 20], but earlier work on information theory underlines the meaning of entropy as a *mean information gain* [22, 13], and this interpretation has often proved useful [18, 17, 14, 21]. We recall briefly this meaning in Section 2.1.

Our approach is mainly based on Ellis' work [10, 11], who considers the *pair empirical measure*

$$L_n(\omega) \stackrel{\text{def}}{=} \frac{1}{n}\left(\sum_{i=1}^{n-1} \delta_{X_i(\omega),X_{i+1}(\omega)} + \delta_{X_n(\omega),X_1(\omega)}\right) \in M_s(E^2).$$

where $M_s(E^2)$ is the set of *balanced* measures[1]. A LDP can be achieved, in the finite case, by purely combinational arguments. In the countable case, we are able

[1] The notation used throughout the paper is given in Appendix D.

to bound in this way the probability of drawing a given number n of transitions (Proposition A.7)

$$\begin{aligned}\frac{1}{n}\ln \mathbb{P}\left[L_n = A\right] &\leq -\frac{1}{n}\min_{i,j:a_{ij}>0}\ln p_{ij} + \frac{2\ln n}{n},\\ \frac{1}{n}\ln \mathbb{P}\left[L_n = A\right] &\geq -H(A\|P) - R_q(n),\end{aligned}$$

where R_q is a quantity such that $\lim_{n\to\infty} R_q(n) = 0$, H is the entropy function[2] and $A \in M_s(E^2)$.

Unfortunately, such a bound cannot be used directly to deduce a LDP, as in the finite case. The analysis of the entropy function H (i.e. the *rate function*) in Section 2 shows important differences that change the nature of the LDP. In the infinite case, H is lower semi-continuous and the level sets $\Psi(\alpha) \stackrel{\text{def}}{=} \left\{A \in M_s(E^2) : H(A\|P) \leq \alpha\right\}$ are *not necessarily compact*. This means a *full* LDP cannot be achieved, but we prove nevertheless that a *weak* LDP holds.

Large deviation principle *Let X be an irreducible Markov chain with kernel P. Then the pair empirical measure L_n verifies a weak LDP with rate function $H(.\|P)$, i.e. for all open sets $O \subset M_s(E^2)$ and all compact sets $K \subset M_s(E^2)$,*

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\ln \mathbb{P}\left[L_n \in O\right] \geq -\inf_{A\in O} H(A\|P),\\ \limsup_{n\to\infty}\frac{1}{n}\ln \mathbb{P}\left[L_n \in K\right] \leq -\inf_{A\in K} H(A\|P).\end{aligned}$$

The idea to get the lower bound (Section 3.2) is to construct successive approximations of the measure A, in order to prove $\ln \mathbb{P}\left[L_n = A\right] \simeq -nH(A\|P)$. Then for any open set $O \subset M_s(E^2)$, we have

$$\liminf_{n\to\infty}\frac{1}{n}\ln \mathbb{P}\left[L_n \in O\right] \geq \sup_{A\in O}\liminf_{n\to\infty}\frac{1}{n}\ln \mathbb{P}\left[L_n = A\right] \geq -\inf_{A\in O} H(A\|P).$$

Since H is not continuous, convergence problems might occur. To avoid them, we study the link between balanced measures and cycles in Appendix C, and the continuity of the entropy is analyzed in Section 2. First, for all measures $A \in$

[2] The indeterminate forms $-\infty + \infty$ on the right side of the inequality can be solved.

O verifying $H(A\|P) < \infty$ and having a *finite* support, we construct a sequence $\{A(m), m \geq 0\}$ such that

$$A(m) \to A \quad \text{and} \qquad \liminf_{m\to\infty} \frac{1}{n(m)} \ln \mathbb{P}\left[L_{n(m)} = A(m)\right] \geq -H(A\|P).$$

Secondly, we extend this result by continuity, relaxing the finite support condition.

Since we can not achieve a full LDP, the upper bound is obtained on compact sets K instead of closed sets (Section 3.1). The combinational bound covers *sub-Markovian* processes, and the trick is to consider finite state processes $(\phi_q(X_n))$, where $\phi_q(x) = \min\{x, q\}$. These processes are sub-Markovian, and hence, for any $x < \inf_{A\in K} H(A\|P)$,

$$\forall A \in K, \qquad \frac{1}{n} \ln \mathbb{P}\left[\phi_q(L_n) = \phi_q(A)\right] \leq -x + 2\frac{\ln n}{n}.$$

Since there is only a finite number of such $\phi_q(A)$ (indeed at most $(n+1)^{q^2}$), the LDP upper bound follows straightly from the inequalities

$$\frac{1}{n} \ln \mathbb{P}\left[L_n \in K\right] \leq \frac{1}{n}\left(q^2 \ln(n+1) - nx + 2\ln n\right) = -x + O\left(\frac{\ln n}{n}\right).$$

The reducible case (Section 3.3) is not conceptually more involved but, since in this case the empirical measures have to be even more restricted, the lower bound hides in a dense impenetrable notational thicket. Nonetheless the result is neither surprising nor more complicated: a weak LDP holds with rate function $H_{|\mathcal{M}}$, where $\mathcal{M}$ is a subset of $M_s(E^2)$ depending only on the ergodic classes of the transition matrix P.

$$H_{|\mathcal{M}}(A\|P) \stackrel{\text{def}}{=} \begin{cases} H(A\|P), & \text{for } A \in \mathcal{M}; \\ \infty, & \text{otherwise.} \end{cases}$$

2 Properties of the relative entropy function

The entropy function being in the heart of the LDP, it is useful to study its properties. On the other hand, this analysis is needed in order to demonstrate the large deviations assertion. Let E be a countable state space identified with the set $\{1, 2, \dots\}$ and P be a Markov kernel on E.

2.1 Entropy and its information theoretic interpretation

Definition 2.1 (balanced measures) *The set $M_s(E^2)$ of balanced measures is the set of measures on E^2 with both identical 1–dimensional projections:*

$$A \in M_s(E^2) \iff A(E,.) = A(.,E) \in M_1(E). \tag{2.1}$$

Let $A \in M_s(E^2)$, we define,

- $a_{ij} \stackrel{\text{def}}{=} A(\{(i,j)\})$ the 2–dimensional law,
- $a_i \stackrel{\text{def}}{=} A(\{i\} \times E) = A(E \times \{i\})$, the 1–dimensional projection,
- $A_{ij} \stackrel{\text{def}}{=} A(E \times \{j\} | \{i\} \times E)$, the conditional law, so that $a_{ij} = a_i A_{ij}$.

Definition 2.2 (relative entropy) *Let $A \in M_s(E^2)$, and P be a Markov kernel. The relative entropy of A with respect to P is defined by*

$$H(A\|P) \stackrel{\text{def}}{=} \sum_{i,j\in E} a_{ij} \ln\left(\frac{a_{ij}}{a_i p_{ij}}\right), \tag{2.2}$$

with the usual convention that $0\ln(0/0) = 0$, $1/0 = +\infty$ and $0 \ln 0 = 0$. $\tilde{H}$ is the extension of the entropy H for finite positive measures A and matrices P verifying $0 \le p_{ij} \le 1$, for all $i,j \in E$.

These results, and the next as well, do have an information significance. In Rényi [22] there is an appendix with an excellent introduction about the link between information theory and probability. For a more complete study see [1], and Hamming [14] for Markov chains. Recall that, by definition, the information brought by a draw Y is $-\ln \mathbb{P}[Y]$, so that $-\ln \mathbb{P}[L_n \in F]/n$ is the mean information brought by the knowledge of L_n being in F.

We think the expression "relative entropy" is not as illuminating as *information gain*. In fact the relative entropy is the *mean information gain*:

$$H(A\|P) = \sum_{i\in E} a_i \sum_{j\in E} A_{ij} \ln\left(\frac{A_{ij}}{p_{ij}}\right) = \mathbb{E}^{A^{(1)}}\left[I(A_{X_1}\|P_{X_1})\right]. \tag{2.3}$$

where $I(A_{X_1}\|P_{X_1}) \stackrel{\text{def}}{=} \sum_j A_{ij} \ln(A_{ij}/p_{ij})$ is the information gain[3] for X_2 conditioning on $X_1 = i$. Thus $H(A\|P)$ is the mean information gain for each sample, taking the

[3]We denote it here by I, but this is actually the relative entropy of measures in $M_1(E)$.

dynamic into account. This is a very suitable expression, from which positivity and lower semi-continuity are easily deduced.

Now, the LDP means, under technical assumptions, that the mean information actually measured tends, on the long run, to the minimum of the mean information over the admissible measures. This is the mathematical formulation of a very general *least information principle*[4].

The process takes the paths giving the minimal information.

2.2 Continuity properties of the entropy function

For the proof of the lower bound of the LDP we need, in close association with the cycles decompositions of Appendix C, some additional special properties of the entropy function. For elementary results see Appendix B.

Definition 2.3 *Let $G \subset E^2$ be a graph on E. $M_s(G) \subset M_s(E^2)$ is the set of balanced measures with support included in G.*

$$A \in M_s(G) \iff A \in M_s(E^2) \text{ and } \mathrm{Supp(A)} \subset \mathrm{G}.$$

Hereafter, the graph G will be assumed to be the graph of the kernel[5] P (see Section C.3). Note that we still do not assume anything about P.

Proposition 2.1 (continuity) *Let G' be a finite subgraph of G. $H(.\|P)$ is continuous on $M_s(G')$.*

Proof : The entropy function is then a finite sum of continuous functions:

$$H(A\|P) = \sum_{(i,j)\in G'} a_{ij} \ln\left(\frac{a_{ij}}{a_i p_{ij}}\right).$$

The sum is finite because G' is finite and the functions $A \mapsto a_{ij}\ln(a_{ij}/a_i p_{ij})$ are continuous (such a function is discontinuous only if $p_{ij} = 0$). ■

[4]Rather than a least action principle, the above principle is in fact a kind of analogue of Fermat's principle, which claims that the light follows the path of extremal propagation time.

[5] Note that $H(A\|P)$ is finite only if $A \in M_s(G)$. Otherwise we would have a term $a_{ij}\ln(a_{ij}/0) = +\infty$ in the entropy sum and the entropy would be infinite.

Proposition 2.2 (exact l.s.c.) *Let $A \in M_s(G)$. There exists a sequence of measures with finite support $A(n) \in M_s(G)$ converging to A, such that their entropy converges to $H(A\|P)$.*

$$\lim_{n\to\infty} A(n) = A \quad \text{and} \quad \lim_{\mathrm{n}\to\infty} \mathrm{H}(\mathrm{A}(\mathrm{n})\|\mathrm{P}) = \mathrm{H}(\mathrm{A}\|\mathrm{P}).$$

Proof : Assume that $H(A\|P) < \infty$. Let consider the decomposition in cycles $A = \sum_{i=1}^{\infty} u_i L(C_i)$ (Proposition C.4). We define

$$\begin{aligned} U_n &\stackrel{\text{def}}{=} \sum_{i=1}^{n} u_i \stackrel{n\to\infty}{\longrightarrow} 1, \\ A(n) &\stackrel{\text{def}}{=} (1-U_n)L(C_1) + \sum_{i=1}^{n} u_i L(C_i) \in M_s(G) \stackrel{n\to\infty}{\longrightarrow} A. \end{aligned}$$

Let us set the ratios

$$\begin{cases} \pi_{ij} \stackrel{\text{def}}{=} a_i p_{ij}, \\ r_{ij} \stackrel{\text{def}}{=} a_{ij}/\pi_{ij}, \\ r_{ij}(n) \stackrel{\text{def}}{=} a_{ij}(n)/\pi_{ij}, \\ r_i(n) \stackrel{\text{def}}{=} a_i(n)/a_i, \end{cases}$$

where $a_{ij}(n)$ is the two-dimensional law of $A(n)$. Let[6] $K = \mathrm{Supp}(C_1) \le q_0$ then[7]

$$\begin{aligned} H(A(n)\|P) &= \sum_{(i,j)\in K} \pi_{ij} r_{ij}(n) \ln r_{ij}(n) - \sum_{i\le q_0} a_i r_i(n) \ln r_i(n) \\ &\quad + \sum_{(i,j)\in K^c} \pi_{ij} r_{ij}(n) \ln r_{ij}(n) - \sum_{i>q_0} a_i r_i(n) \ln r_i(n) \\ &\stackrel{n\to\infty}{\longrightarrow} \sum_{(i,j)\in K} \pi_{ij} r_{ij} \ln r_{ij} + \sum_{(i,j)\in K^c} \pi_{ij} r_{ij} \ln r_{ij} = H(A\|P). \end{aligned}$$

The first finite sums converge because $x \mapsto x \ln x$ is continuous. The infinite sums converge by Lebesgue's theorem: the terms $r_{ij}(n) \ln r_{ij}(n)$ [resp. $r_i(n) \ln r_i(n)$] are

[6] We write $K \le q_0$ for $K \subset \{1, \ldots, q_0\}^2$.

[7] This decomposition is valid if, and only if, $\sum_i a_i r_i(n) \ln r_i(n) < \infty$ to avoid the undetermined form $\infty - \infty$. This is the case because the first q_0 terms are finite, and then $r_i(n) \le 1$ so that $\sum_{i>q_0} a_i r_i(n) \ln r_i(n) \le 0$.

bounded by $r_{ij} \ln r_{ij}$ [resp. $r_i \ln r_i$], terms forming a summable sequence, and all terms converge.

If $H(A\|P) = \infty$, then a finite number of terms are sufficient to prove that $H(A(n)\|P) \geq K$ for any large K, so that the limit is still true. ■

Remark : This allows to prove convexity by Hessian on finite dimensional sets $M_s(G_q) \subset M_s(G)$, and thus to extend the property to infinite support measures. Let $\lambda, \mu \geq 0$ such that $\lambda + \mu = 1$, then

$$\begin{aligned} H(\lambda A + \mu B\|P) &\leq \liminf_{n\to\infty} H(\lambda A(n) + \mu B(n)\|P) \\ &\leq \lim_{n\to\infty} \big(\lambda H(A(n)\|P) + \mu H(B(n)\|P)\big) \\ &= \lambda H(A\|P) + \mu H(B\|P), \end{aligned}$$

where the first inequality comes from the lower semi-continuity, the second from convexity on finite support measures and the third from Proposition 2.2. Thus the entropy function is positive, convex and lower semi-continuous.

Proposition 2.3 (class decomposition) *Let P be a reducible Markov kernel. Let $G' = \bigcup_{i\in I} G_i$ the decomposition in classes of Proposition C.6. The entropy function is infinite outside $M_s(G')$ and, for $A \in M_s(G')$, we have*

$$H(A\|P) = \sum_{i\in I} x_i H_i(A(i)\|P), \tag{2.4}$$

with $A = \sum_{i\in I} x_i A(i)$, $A(i) \in M_s(G_i)$[8] *and $H_i(.\|P)$ the restriction of $H(.\|P)$ to $M_s(E_i^2)$.*

Proof : This follows immediately from the decompositions of A (Proposition C.4) and of $M_s(G)$ (Proposition C.6), gathering by class all minimal cycles together. ■

3 Sanov's theorem

We know that the entropy function H has not necessarily compact level sets (see Section B) thus we aim at proving that a *weak* LDP holds for Markov chains. We shall denote by X the Markov chain with kernel P and initial probability ν.

[8] If $x_i = 0$ then $A(i)$ is not defined. It can still be chosen with the convention that $0 \times \infty = 0$.

3.1 The upper bound

The sketch of proof is to lump an infinite set of states together in order to handle a finite state process for which a bound is obtained by combinational arguments. The main steps consist, first in proving that the finite state process is sub-Markovian, then, secondly, applying a combinational bound. The third step is to show that this bound is a good approximation on a compact set of measures.

Let us introduce the lumping $\phi_q(x) \stackrel{\text{def}}{=} \min\{x, q\}$.

Lemma 3.1 *The process $\phi_q(X_n)$ is sub-Markovian with pseudo-kernel $\widetilde{P}$ and initial probability $\widetilde{\nu}$,*

$$\widetilde{p}_{ij} = \begin{cases} p_{ij} & \text{if i,j < q,} \\ 1 & \text{otherwise,} \end{cases} \qquad \text{and} \qquad \widetilde{\nu}_{\rm i} = \begin{cases} \nu_i & \text{if i < q,} \\ \nu_q + \nu_{q+1} + \dots & \text{if i = q.} \end{cases}$$

Proof : The probability, beginning at $i < q$, to stay m times consecutively above q and to finish at $j < q$, is easily bounded with $\widetilde{P}$:

$$\sum_{l_1,\dots,l_m \geq q} p_{il_1} p_{l_1 l_2} \dots p_{l_m j} \leq p_{ij}^{(m+1)} \leq 1 = \widetilde{p}_{iq} \widetilde{p}_{qq} \dots \widetilde{p}_{qj}. \tag{3.1}$$

Given a sample path $(k_1, k_2, \dots, k_n)$ with $k_{i_1} = q, \dots\ k_{i_p} = q$, we shall decompose it into transitions below q and paths going above q. Using equation (3.1), we obtain,

$$\begin{aligned} & \mathbb{P}\left[\phi_q(X_1) = k_1, \dots, \phi_q(X_n) = k_n\right] \\ = & \sum_{k_{i_1},\dots,k_{i_p} \geq q} \nu_{k_1} p_{k_1 k_2} \dots p_{k_{n-1} k_n} \leq \widetilde{\nu}_{k_1} \prod_{i=1}^{n-1} \widetilde{p}_{k_i k_{i+1}}, \end{aligned}$$

which exactly expresses the sub-Markovian property of Definition A.9 with pseudo-kernel $\widetilde{P}$. The pair empirical measure is $\phi_q(L_n)$ with[9]

$$\begin{aligned} \phi_q : M_s(E^2) & \mapsto & M_s(\{1, \dots, q\}^2) \\ A & \to & \phi_q(A) = \begin{pmatrix} a_{11} & \dots & a_{1,q-1} & \sum_{j \geq q} a_{1j} \\ \vdots & & & \vdots \\ a_{q-1,1} & \dots & a_{q-1,q-1} & \sum_{j \geq q} a_{q-1,j} \\ \sum_{i \geq q} a_{i1} & \dots & \sum_{i \geq q} a_{i,q-1} & \sum_{i,j \geq q} a_{ij} \end{pmatrix}. \end{aligned}$$

■

[9]We use the same notation for ϕ_q and $\phi_q(x) = \min\{x, q\}$, because their actions are similar.

Theorem 3.2 (LDP upper bound) *Let X be a Markov chain on E with kernel P, and L_n its pair empirical measure. Then, for all compact sets $K \subset M_s(E^2)$, we have*

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left[L_n \in K\right] \leq - \inf_{A\in K} H(A\|P). \tag{3.2}$$

Proof : The Lemma 3.1 allows us to use the combinational bound[10] (Proposition A.7):

$$\forall B \in M_s(\{1,\dots,q\}^2) \cap \mathcal{L}_n, \qquad \ln \mathbb{P}\left[\phi_q(L_n) = B\right] \leq -n\widetilde{H}(\widetilde{B}\|\widetilde{P}) + 2\ln n, \tag{3.3}$$

with $\widetilde{B}$ defined by[11],

$$\widetilde{b}_{ij} \stackrel{\text{def}}{=} \begin{cases} b_{ij} - \dfrac{1}{n} & \text{if } \ \widetilde{p}_{ij} = \displaystyle\min_{k,l:b_{kl}>0} \widetilde{p}_{kl}, \\ b_{ij} & \text{otherwise.} \end{cases}$$

$\widetilde{P}$ has a very special form: $\widetilde{p}_{ij} = 1$ if $i = q$ or $j = q$. In these cases the entropy term $\widetilde{b}_{ij} \ln(\widetilde{b}_{ij}/b_i)$ is negative, so that $\widetilde{H}(\widetilde{B}\|\widetilde{P})$ is easily bounded up by the following truncated entropy,

$$\widetilde{H}(\widetilde{B}\|\widetilde{P}) \leq \widetilde{H}(t_q(\widetilde{B})\|P), \tag{3.4}$$

with the truncation function $t_q(A) \stackrel{\text{def}}{=} (a_{ij})_{i,j<q}$. Let $B = \phi_q(A)$ with $A \in \mathcal{L}_n$; then $t_q(A) = t_q(B)$, but this equality does not hold with $t_q(\widetilde{A})$ and $t_q(\widetilde{B})$. Nevertheless $\widetilde{H}(t_q(\widetilde{B})\|\widetilde{P}) \leq \widetilde{H}(t_q(\widetilde{A})\|P)$, so that,

$$\forall A \in M_s(E^2), \qquad \ln \mathbb{P}\left[\phi_q(L_n) = \phi_q(A)\right] \leq -n\widetilde{H}(t_q(\widetilde{A})\|P) + 2\ln n. \tag{3.5}$$

Let $K \subset M_s(E^2)$ be a compact set, and $x < \inf_{A\in K} H(A\|P)$. Then there exist q and n_0, such that, for all $n \geq n_0$, for all $A \in K \cap \mathcal{L}_n$,

$$\widetilde{H}(t_q(\widetilde{A})\|P) \geq x. \tag{3.6}$$

---

[10] $\widetilde{H}$ and H have the same definition but we keep H for *balanced measures of probability.*

[11] If the minimum p_{ij} is achieved several times, we choose arbitrarily one pair (i,j), and only one. Equation (3.3) makes sense only if $B \in \mathcal{L}_n$, i.e. all b_{ij} are multiples of $1/n$, otherwise $\mathbb{P}\left[\phi_q(L_n) = B\right] = 0$.

Otherwise one could construct a sequence $A(q) \in K \cap \mathcal{L}_{n(q)}$ (with $n(q) \to \infty$) such that $\widetilde{H}(t_q(\widetilde{A}(q))\|P) < x$. Since K is compact, there exists a converging subsequence $A(q) \to A$. Then obviously $t_q(\widetilde{A}(q)) \to A$, and since $\widetilde{H}$ is lower semi-continuous[12],

$$\widetilde{H}(A\|P) = H(A\|P) \leq \liminf_{q\to\infty} \widetilde{H}(t_q(A_q)\|P) \leq x < \inf_{A\in K} H(A\|P),$$

which is a contradiction.

Let us define

$$\mathcal{A}_n \stackrel{\text{def}}{=} \mathcal{L}_n \cap \left\{ A \in M_s(\{1,\dots,q\}^2) : \widetilde{H}(t_q(\widetilde{A})\|P) \geq x \right\}.$$

From the definition of $\mathcal{L}_n$, there are exactly $(n+1)^{q^2}$ measures with all atoms multiples of $1/n$ and support included in $\{1,\dots,q\}^2$, and hence there are at most $(n+1)^{q^2}$ measures in $\mathcal{A}_n$. Note that $L_n \in K \iff L_n \in K \cap \mathcal{L}_n$. Using equation (3.5) and equation (3.6), the LDP upper bound follows from the inequalities

$$\begin{aligned} \frac{1}{n} \ln \mathbb{P}\left[L_n \in K\right] &\leq \frac{1}{n} \ln \mathbb{P}\left[\phi_q(L_n) \in \mathcal{A}_n\right] \\ &\leq \frac{1}{n}\left(q^2 \ln(n+1) - nx + 2\ln n\right). \end{aligned} \tag{3.7}$$

■

3.2 The lower bound

Theorem 3.3 (LDP lower bound) *Let X be an irreducible Markov chain on E with kernel P, and L_n its pair empirical measure. Then, for all open set $O \subset M_s(E^2)$,*

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left[L_n \in O\right] \geq -\inf_{A\in O} H(A\|P). \tag{3.8}$$

Proof : Let P be irreducible and $O \subset M_s(E^2)$ an open set of balanced measures with $A \in O$. We first assume $A \in M_s(G)$ and $\mathrm{Supp}(A)$ is finite.

By Proposition C.5 we get a sequence of G'–cycles $C(n)$ with finite support ($\mathrm{Supp}(G') \leq q < \infty$) such that $L(C(n)) \stackrel{n\to\infty}{\longrightarrow} A$. Note that the length of $C(n)$ always grows to infinity. By Proposition C.1, there exists a global cycle C_g for G' with less than $2q$ transitions.

[12] $\widetilde{H}(.\|P)$ is 1–homogeneous so that $\widetilde{H}(uA\|P) = uH(A\|P)$. Hence $\widetilde{H}(.\|P)$ is lower semi-continuous.

Let $A(m) = L(C_g C(m))$: these balanced measures verify the conditions of Proposition A.7 and

$$\frac{1}{n(m)} \ln \mathbb{P}\left[L_{n(m)} = A(m)\right] \geq -H(A(m)\|P) - R_q(n(m)) + \frac{\ln \nu(S_1(G'))}{n(m)}.$$

with $n(m)$ the length of $C_g C(n)$ and $S_1(G')$ the one–dimensional support of G' (see Definition A.5). If $\nu(S_1(G')) = 0$, we take a larger G' so that this quantity is strictly positive. When $m \to \infty$ we know $n(m) \to \infty$ and

$$\liminf_{m\to\infty} \frac{1}{n(m)} \ln \mathbb{P}\left[L_{n(m)} = A(m)\right] \geq \liminf_{m\to\infty} -H(A(m)\|P)$$

Since H is continuous on $M_s(G')$ by Proposition 2.1, the left side converges and, O being open, for m sufficiently large, $A(m) \in O$ so that

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left[L_n \in O\right] \geq \liminf_{m\to\infty} \frac{1}{n(m)} \ln \mathbb{P}\left[L_{n(m)} = A(m)\right] \geq -H(A\|P) \qquad (3.9)$$

Now we will relax the finiteness assumption: $A \in M_s(G)$. By Proposition 2.2, there exists a sequence $A(q) \in M_s(G)$ with finite support converging to A such that $H(A(q)\|P) \overset{q\to\infty}{\longrightarrow} H(A\|P)$. Equation (3.9) then holds for any $A \in M_s(G)$.

Let relax all assumptions: if $A \notin M_s(G)$ then (see footnote 5) $H(A\|P) = \infty$ and equation (3.9) holds. The proof of the lower bound is concluded. ■

3.3 The reducible case

No concept is really new in this section, which is merely stated for the sake of completeness. Actually, it holds independently of ergodicity or transience, and also for cyclic chains. The only assumption concerns the irreducibility, in order to obtain the lower bound. However, even if the upper bound holds for a reducible kernel P, it can be refined, and the constructive proof of the lower bound has to be done again.

We consider hereafter the decomposition into classes of Proposition C.6 and Proposition 2.3. So we have $M_s(G) = M_s(G')$ with $G' = \cup G_i$ where G_i is the graph of P on class E_i. Moreover we assume that for all $x \in E$ there exists a path with positive probability ending at x. Otherwise we could remove all non-accessible elements without changing any property of the Markov chain. Let us introduce the following "good" measures.

Definition 3.1 (serial measures) *A serial measure relatively to a reduc-ible kernel P is a balanced measure $A = \sum x_i A(i)$ with $A(i) \in M_s(G_i)$, such that $E_{i+1} \succ E_i$ for all i, where $\succ$ is the natural partial ordering on the classes. We denote by $\mathcal{M}$ the set of serial measures relatively to P.*

Lemma 3.4 *The serial measures are characterized by the following equivalence.*

$$A \in \mathcal{M} \iff \begin{cases} \exists A(m) \stackrel{m\to\infty}{\longrightarrow} A, \\ \mathbb{P}\left[L_{n(m)} = A(m)\right] > 0. \end{cases} \tag{3.10}$$

Proof : Sufficient condition If $A \notin M_s(G)$ then there exists a forbidden transition $a_{ij} > 0$ (i.e. $p_{ij} = 0$), and for m sufficiently large $n(m)a_{ij}(m) \geq 2$ (with 1 it could have been a ghost transition) so that $\mathbb{P}\left[L_{n(m)} = A(m)\right] = 0$. This is a contradiction. Hence, necessarily, $A \in M_s(G)$ and $A = \sum x_i A(i)$ with $A(i) \in E_i$.

Let $j_0 \in E_j$ and $k_0 \in E_k$ such that $a_{j_0} > 0$ and $a_{k_0} > 0$ (i.e. $x_j > 0$ and $x_k > 0$). As $A(m) \stackrel{m\to\infty}{\longrightarrow} A$, we deduce that there exits m_0 such that for $m \geq m_0$, $a_{j_0}(m) > 0$ and $a_{k_0}(m) > 0$. This means a path $(x_1 \ldots x_n)$ crosses j_0 and k_0 with positive probability, hence $E_j \succ E_k$ or $E_k \succ E_j$. Therefore $\succ$ is a total order relation on the set of E_i for which $x_i > 0$, and A is serial.

Necessary condition Firstly assume A has finite support. Then the decomposition of A as serial measure is finite and its decomposition into cycles also. Hence, we can write $A = \sum_{i=1}^{p} x_i L(C_i)$ with $C_{i+1} \succ C_i$. Therefore there exists a path $T_{12} \ldots T_{p-1,p}$ with positive probability, such that $T_{i,i+1}$ begins on C_i and ends on C_{i+1}. Then we take the same rational approximations as in the proof of Lemma C.2 and the path $T(m) = C_1^{\alpha_1(m)} T_{12} C_2^{\alpha_2(m)} T_{23} \ldots C_p^{\alpha_p(m)}$ has positive probability. Finally $A(m) \stackrel{\text{def}}{=} L(T(m))$ converges toward A with $m \to \infty$ and $\mathbb{P}\left[L_{n(m)} = A(m)\right] > 0$.

Secondly, if A has infinite support, there exists (via the decomposition in cycles used in Proposition 2.2) a sequence of serial $A(n)$ with finite support converging to A. For each $A(n)$ there exists a sequence $A(n,m) \in M_s(G)$ with positive probability converging to $A(n)$. Let $A'(n) = A(n, m_n)$ with m_n such that[13] $d_L(A'(n), A(n)) \leq 1/n$; the sequence $A'(n)$ converges to A and has a positive probability. ■

Theorem 3.5 (reducible case) *Let X be a Markov chain with a reducible kernel P, and $\mathcal{M}$ the set of serial measures relatively to P. The pair empirical measure*

[13]The distance d_L is the Levy distance in $M_1(E^2)$.

of X verifies a weak large deviation principle; for any open set $O \subset M_s(E^2)$ and compact set $K \subset M_s(E^2)$,

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left[L_n \in O\right] \geq -\inf_{A\in O\cap\mathcal{M}} H(A\|P), \tag{3.11}$$

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left[L_n \in K\right] \leq -\inf_{A\in K\cap\mathcal{M}} H(A\|P). \tag{3.12}$$

Proof : Lower bound The lines of the demonstration are similar to those of Theorem 3.3, but the classes must be separated. Let O be an open set of balanced measures and $A \in O \cap \mathcal{M}$ with finite support: $A = \sum_{i=1}^p x_i A(i)$. Let consider the $C_k^{\alpha_k(m)}$ from the previous proof. We define $C(i,m) \stackrel{\text{def}}{=} G_i^g \prod_{k\in E_i} C_k^{\alpha_k(m)}$, adding a global cycle G_i^g (relatively to class i) to each class. The length of $C(i,m)$ is denoted by $n_i(m)$ with $n_i(m)/n(m) \stackrel{m\to\infty}{\longrightarrow} x_i$. Then we link these separate classes with paths T_i $(i<p)$ that begin in e_i, end in e_{i+1}, have length r_i and probability $q_i > 0$. One initial path T_0 is added if $\nu(e_1) = 0$.

As in the previous section this holds only if the classes are finite, but we know that there is no problem to extend the property from finite sets to infinite sets, so we assume hereafter all classes are finite. Finally, define the empirical measure $A(m) \stackrel{\text{def}}{=} L(T_0 C(1,m) T_1 \dots T_{p-1} C(p,m))$. We divide the path into p classes and obtain by Markov conditioning

$$\mathbb{P}\left[L_{n(m)} = A(m)\right] \geq \mathbb{P}\left[T_1\right] \prod_{i=1}^{p} \mathbb{P}_{e_i}\left[L_{n_i(m)} = A(i,m)\right] \prod_{i=1}^{p-1} \mathbb{P}_{e_i}\left[T_i\right],$$

the chain passing successively through T_0, then E_1 with transitions $C(1,m)$, etc. Note that the ghost transitions are known for $L_{n_i(m)}$, $i<p$, because the final states e_i are known.

The lower bound equation (A.20) becomes

$$\begin{aligned}
&\frac{1}{n(m)} \ln \mathbb{P}\left[L_{n(m)} = A(m)\right] \\
&\quad\geq \frac{1}{n(m)} \left(\sum_{i=1}^{p} \frac{n_i(m)}{n(m)} \widetilde{H}_i(A(i,m)\|P) + \sum_{i=1}^{p-1} \ln q_{i,i+1} \right) + o(1) \\
&\quad\stackrel{m\to\infty}{\longrightarrow} -\sum_{i=1}^{p} x_i \widetilde{H}_i(A(i)\|P) = -H(A\|P).
\end{aligned}$$

where $o(1)$ is a sum of rests converging to 0 when m tends to infinity. The identification of the sum of partial entropies to $H(A\|P)$ is due to Proposition 2.3. Of course, for sufficiently large m, $A(m) \in O$. If A has an infinite support, then we use Proposition 2.2 as in the lower bound proof. Hence we have obtained the LDP lower bound equation (3.12).

Upper bound It is much simpler to get. When A is not serial, Lemma 3.4, allows to write

$$A \notin \mathcal{M} \implies \exists \mathcal{V}_A,\ \exists n_0 : \forall n \geq n_0 \quad \mathbb{P}\left[L_n \in \mathcal{V}_A\right] = 0,$$

where $\mathcal{V}_A$ is an open neighborhood of A. Then $\mathbb{P}\left[L_n \in K \setminus \mathcal{V}_A\right]$ equals $\mathbb{P}\left[L_n \in K\right]$ and the upper bound equation (3.2) becomes

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left[L_n \in K\right] \leq - \inf_{A' \in K \setminus \mathcal{V}_A} H(A' \| P).$$

This can be achieved with a finite number of neighborhoods. Finally, using the lower semi-continuity of H,

$$\inf_{A_1,\ldots,A_p \notin \mathcal{M}} \ \inf_{A' \in K \setminus \cup_{i=1}^{p} \mathcal{V}_{A_i}} H(A' \| P) = \inf_{A' \in K \cap \mathcal{M}} H(A' \| P),$$

which gives the upper bound (3.11). ∎

4 Conclusion

There is no way of improving our results toward a full LDP, as shown in remark 1 p. 32. We conjecture that the problem of having a full LDP reduces to the analysis of the entropy function, and this would give a real interest to the weak LDP. Unfortunately, even in our setting, the goodness of the rate function does not imply that a full LDP holds with this good rate function. A basic example is $P_{n,n+1} = 1$ where $H = +\infty$, or small perturbations of this chain: $\inf_A H(A\|P) > 0$ (then a full LDP can not hold) but H is nonetheless a good rate function.

Then, some directions may be explored.

- When is the rate function H a *good* rate function?
- What are the precise relationships between goodness, exponential tightness, and fullness of LDP?
- To which extent is the full LDP necessary?

Partial answers to the first question do exist, e.g. uniformity conditions which can be found in [8, 11, 19, 24]. We are working on *characterization* of goodness, in order to include bounded jumps networks. In fact, for almost all open networks there is no full LDP, because a small amount of homogeneity destroys the fullness.

The second question is open. It is known that the exponential tightness implies a full LDP holds with good rate function H (see [8] p. 8). We conjecture that, if H is good *and* $\inf_A H(A\|P) = 0$, then the family $\{P^{(n)}, n \geq 0\}$ is exponentially tight.

As for the third question, remark that the compactness of the level sets is necessary for the contraction principle. However, this is the case as soon as the function $I(y) \stackrel{\text{def}}{=} \inf_{A \in f^{-1}(y)} H(A\|P)$ is lower semi-continuous and $f^{-1}(K)$ is compact, for any compact K. But is it possible to extend this when $f^{-1}(K)$ is not compact?

Remark 1 : Both conditions are verified when $f(A) = A^{(1)}$. Actually, a LDP holds for the (one–dimensional) empirical measure, with rate function

$$J(\mu\|P) \stackrel{\text{def}}{=} \sup_{u \gg 0} \sum_{j \in E} \mu_j \ln\left(\frac{u_j}{(uP)_j}\right) = \inf_{A^{(1)}=\mu} H(A\|P).$$

It is easy to see that the goodness of J is equivalent to the goodness of H. Therefore both LDP are full or weak together (see also remark 2 p. 32).

This LDP can inversely be extended to k-tuples empirical measure and to the process level, or by considering continuous time or larger state space (like Polish space).

Remark 2 : In the continuous time case, our approach leads to consider the random vector $\left(\hat{L}(N_t), L(t), N_t/t\right)$, where

$$\begin{aligned} \hat{L}(n) &\stackrel{\text{def}}{=} \frac{1}{n+1}\left(\sum_{l=1}^{n} \delta_{X_{T_{l-1}}, X_{T_l}} + \delta_{X_{T_N}, X_{T_0}}\right) \in M_s(E^2), \\ L(t) &\stackrel{\text{def}}{=} \frac{1}{t}\int_0^t \delta_{X_s}\, \mathrm{d}s \in M_1(E), \end{aligned}$$

T_l being the time of l-th jump and N_t the number of jumps before t. Letting $A \in M_s(E^2)$ represent the transitions of the chain, $\eta \in M_1(E)$ the frequency of occupation for each state and $f \geq 0$ the frequency of jumps, a LDP holds with rate function

$$I(A, \eta, f) = fH(A\|P) + \sum_i f a_i l\left(\lambda_i \frac{\eta_i}{f a_i}\right)$$

where $l(u) = u - 1 - \ln u$ (this is the rate function for i.i.d. exponential random variables) and $(\lambda_i p_{ij})$ is the generator of the process.

On the one hand, if $\lambda_i = 1$ for all i, $f = 1$, and $\eta_i = a_i$, i.e. when the process jumps at unit speed, then the rate function is exactly H. On the other hand, if $f \to 0$ (i.e. when it almost never jumps), the rate function tends to $\sum_i \lambda_i \eta_i$. Furthermore, if η is concentrated on a state x, then $I = \lambda_x$: this is exactly the mean (logarithmic) probability of never leaving state x. Thus this rate function is really meaningful in spite of its rather complicated form: the left part concern the transitions and the right part the rate of jump. The problem is to deal with the compact sets.

The entropy function contains indeed some surprises. In the transient case, the global infimum of H can be strictly positive. For instance in the M/M/1 case, when $p_{i,i+1} = \lambda$ and $p_{i,i-1} = \mu$ with $\lambda > \mu$, then $\inf H(.\|P) = -\ln 2\sqrt{\lambda\mu} > 0$. Actually, this infimum characterizes the *transience rate* of the chain. This notion can be applied to transient chains, like the Aloha protocol, to measure the mean explosion time, which is of practical interest. Moreover, since H is convex, numerical calculations are easily performed when explicit solutions are not found.

Other questions are worth considering:

- In which sense does the LDP mean conditioning by a rare event?
- When the rare event F is a H-continuity set[14], and there is a unique minimum $H(A\|P) = \inf_{A\in F} H(A\|P)$, is there a limit process and is this process Markov with kernel A?
- How can the LDP upper bound be extended to closed sets?

This last point faces a serious problem. It appears to be closely related to boundary theory, and more precisely to Martin boundary. The entropy function would be extended to the (minimal) Martin boundary in order to keep the LDP formulation in terms of minima. Note that the ergodic case is simpler, since the Martin space reduces to a single point.

A From combinatorics of paths to LDP

We consider a denumerable state space E, with *a priori* infinite cardinality q. This space is identified with $\{1, 2, \dots, q\}$. Let $(X_i)_{i=1}^{\infty}$ be a Markov chain with transition matrix P and initial distribution ν.

[14] A H-continuity set F is such that the infimum of H is the same on the interior and the closure of F, so that the LDP statement becomes a limit on this set. A lot of practical sets are in fact H-continuity sets.

The reader will note that the demonstration uses some detours in order to avoid the traps. In fact, the thicket of notations hides some intrinsic simplicity.

A.1 Path counting

This section is purely combinational and does not involve any probabilistic structure on the paths. Let $\alpha = (\alpha_{ij})_{i,j\in E}$ be a *finite* integer matrix ($|\alpha| \stackrel{\text{def}}{=} \sum_{i,j\in E} \alpha_{ij} < \infty$).

Definition A.1 $\mathcal{Q}^k_\alpha \subset E^{|\alpha|}$ *[resp. N^k_α] is the set [resp. the number] of paths beginning in k with α_{ij} transitions from state i to state j and $\mathcal{Q}_\alpha \stackrel{\text{def}}{=} \bigcup_{k\in E} \mathcal{Q}^k_\alpha$.*

Definition A.2 (admissible transitions) *$\mathcal{S}$ is the set of α's corresponding at least to one path. $\mathcal{S}^k$ is the restriction of $\mathcal{S}$ to the paths beginning in k.*

$$\mathcal{S}^k \stackrel{\text{def}}{=} \left\{\alpha : N^k_\alpha > 0\right\}, \qquad \text{and} \qquad \mathcal{S} \stackrel{\text{def}}{=} \bigcup_{\text{k}\in\text{E}} \mathcal{S}^{\text{k}}.$$

Our goal is to compute the following quantity

$$\mathbb{P}\left[\mathcal{Q}^1_\alpha\right] = N^1_\alpha \nu_1 \prod_{i,j=1}^{q} p_{ij}^{\alpha_{ij}}.$$

The problem is to evaluate the number N^k_α of paths having the correct number of transitions. Clearly, one can choose $X_1 = k = 1$. In the case of an infinite state space, it must be noticed that sums and products never concern an infinite number of terms. By permutation we can bring all states i such that $\alpha_i > 0$ near 1, so that we can and will assume $q < \infty$ for a given α.

We now wish to describe a complicated set. Let $\mathcal{N}_i(\alpha)$ be the set of all α_i–tuples of elements of E containing the state j exactly α_{ij} times.

$$x \in \mathcal{N}_i(\alpha) \iff x = (e_1, e_2, \ldots, e_{\alpha_i}), \quad \text{with} \sum_{k=1}^{\alpha_i} \mathbb{1}_j(e_k) = \alpha_{ij}, \qquad \forall j \in E. \quad \text{(A.1)}$$

The number of such α_i–tuples is

$$\text{Card}(\mathcal{N}_i(\alpha)) = \frac{\alpha_i!}{\alpha_{i1}!\alpha_{i2}!\ldots\alpha_{iq}!}.$$

Definition A.3 (α–tuples) *Let α be a finite integer matrix. $\mathcal{N}(\alpha)$ [resp. $\xi(\alpha)$] is the set [resp. the number] of α–tuples, where an α–tuple is a sequence of α_i–tuples characterized in equation (A.1).*

By definition $\mathcal{N}(\alpha) \stackrel{\text{def}}{=} \mathcal{N}_1(\alpha) \times \mathcal{N}_2(\alpha) \times \cdots \times \mathcal{N}_q(\alpha)$, thus the cardinality of $\mathcal{N}(\alpha)$ is

$$\xi(\alpha) \stackrel{\text{def}}{=} \mathrm{Card}(\mathcal{N}(\alpha)) = \prod_{i=1}^{q} \mathrm{Card}(\mathcal{N}_i(\alpha)) = \prod_{i=1}^{q} \frac{\alpha_i!}{\prod_{j=1}^{q} \alpha_{ij}!}. \tag{A.2}$$

Lemma A.1 (upper bound) *The number N_α^1 of paths is less than $\xi(\alpha)$.*

Proof : Let $\mathcal{P} \in \mathcal{Q}_\alpha^1$ be a path, which we associate to an α–tuple, representing the successive exits out of each state i, by mean of the following mapping:

$$\begin{array}{ccl} \phi : \mathcal{Q}_\alpha^1 & \mapsto & \mathcal{N}(\alpha) \\ \mathcal{P} & \rightarrow & x = \phi(\mathcal{P}) = \left(x^{(1)}, \ldots, x^{(q)}\right) \\ & & \text{where} \quad x^{(l)} = \left(e_1^{(l)}, e_2^{(l)}, \ldots, e_{\alpha_l}^{(l)}\right), \text{ and } e_n^{(k)} \stackrel{\text{def}}{=} X_{T_n^{(k)}+1}, \end{array}$$

and $T_n^{(k)}$ is the n-th entrance time into state k.

The example of Figure 1 shows that ϕ is a one to one mapping. The paths begin in state 1 ($X_1 = 1$), we get the successor $X_2 = e_1^{(1)} = 4$, then $X_3 = e_1^{(4)} = 3$, etc. Thus the path $\mathcal{P}$ may be reconstructed from $\phi(\mathcal{P})$, which proves the mapping is one to one. Unfortunately, many elements of $\mathcal{N}(\alpha)$ clearly do not correspond to any path, and hence ϕ is not bijective. This yields nevertheless an upper bound for N_α^1,

$$\phi\left(\mathcal{Q}_\alpha^1\right) \subset \mathcal{N}(\alpha) \implies N_\alpha^1 \leq \xi(\alpha).$$

■

The lower bound of N_α is a bit more technical. We need a useful tool:

Definition A.4 (transitions of a cycle) *Let $C = (x_1, \ldots, x_n)$ be a cycle. The transitions of this cycle is the balanced integer matrix*

$$\delta(C) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \delta_{x_i x_{i+1}}, \qquad \textit{with } x_{n+1} = x_1.$$

Lemma A.2 (lower bound) *The number $N_{\alpha'}^1$ of paths is greater than $\xi(\alpha)$ where C is a global cycle*[15] *on $\{1, \ldots, q\}$ and $\alpha' = \alpha + \delta(C)$.*

[15] See Proposition C.1.

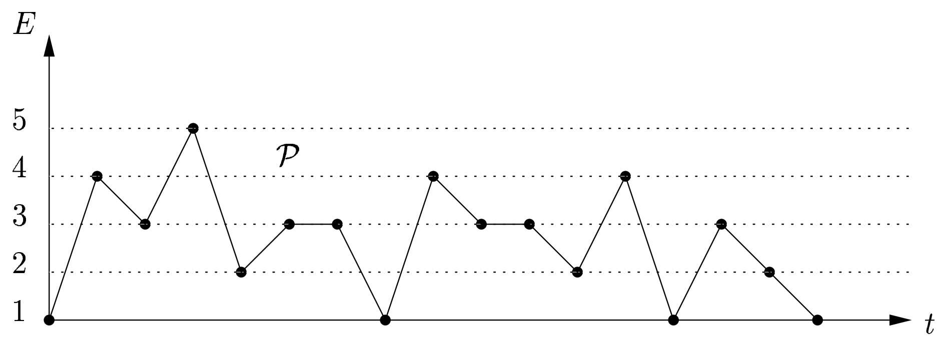

Figure 1: Example of path and hereafter the corresponding objects.

$$\begin{aligned}
\mathcal{P} &= (1,4,3,5,2,3,3,1,4,3,3,2,4,1,3,2,1),\\
\phi(\mathcal{P}) &= \begin{cases} (4,4,3) & \text{(exits of state 1)},\\ (3,4,1) & \text{(exits of state 2)}\\ (5,3,1,3,2,2) & \text{(exits of state 3)},\\ (3,3,1) & \text{(exits of state 4)},\\ (2) & \text{(exits of state 5)}, \end{cases}\\
\alpha &= (\alpha_{ij}) = \begin{pmatrix} 0 & 0 & 1 & 2 & 0\\ 1 & 0 & 1 & 1 & 0\\ 1 & 2 & 2 & 0 & 1\\ 1 & 0 & 2 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}.
\end{aligned}$$

Proof : In order to obtain the lower bound, one must understand more deeply why ϕ is not bijective. Take an element $x \in \mathcal{N}(\alpha)$ and try to reconstruct the chain $\mathcal{P}$ such that $\phi(\mathcal{P}) = x$ (see the proof of Lemma A.1). By convention $\mathcal{P}$ begins with $X_0 = 1$. For instance, if

$$\alpha = (\alpha_{ij}) = \begin{pmatrix} 0 & 1\\ 1 & 1 \end{pmatrix}, \qquad \text{and} \qquad x = \begin{cases} (2),\\ (1,2), \end{cases}$$

the reconstructed chain is $1 \to 2 \to 1$; but it remains one transition $2 \to 2$: the reconstruction process fails. This is not a pathological case and a lot of other examples may be derived from this one.

We can however force the chain to exhaust all α_i–tuples. Assume the end is state 1 so that the transitions from state 1 are exhausted; we add a transition $1 \to 2$ and finish the reconstruction; thus transitions from state 2 are also exhausted. By adding again a transition into states 3, 4, … n, and so exhaust all transitions. This scheme is formalized by the function

$$
\begin{array}{rcl}
\psi : \mathcal{N}(\alpha) & \mapsto & \mathcal{N}(\alpha') \\
x & \to & x' = \left\{ \begin{array}{l}
(e_1^{(1)}, e_2^{(1)}, \dots, e_{\alpha_1}^{(1)}, 2), \\
(e_1^{(2)}, e_2^{(2)}, \dots, e_{\alpha_2}^{(2)}, 3), \\
\vdots \\
(e_1^{(q-1)}, e_2^{(q-1)}, \dots, e_{\alpha_{q-1}}^{(q-1)}, q), \\
(e_1^{(q)}, e_2^{(q)}, \dots, e_{\alpha_q}^{(q)}, 1),
\end{array} \right.
\end{array}
$$

with $\alpha'_{ij} \stackrel{\text{def}}{=} \alpha_{ij} + \delta_{i,i+1}$ (with the convention that $\delta_{q,q+1} = \delta_{q1}$). This mapping is obviously one to one. The last transition $q \to 1$ is useless but renders the function symmetric, and this will be useful later. The reconstruction process means that, for all α–tuple $x' \in \psi(\mathcal{N}(\alpha))$, there corresponds a path $\mathcal{P} \in \mathcal{Q}^1_{\alpha'}$ such that $\phi(\mathcal{P}) = x'$, i.e. $\psi(\mathcal{N}(\alpha)) \subset \phi(\mathcal{Q}^1_{\alpha'})$. Since ϕ is one to one (see Lemma A.1), this is equivalent to write $\phi^{-1}(\psi(\mathcal{N}(\alpha))) \subset \mathcal{Q}^1_{\alpha'}$. But $\phi^{-1} \circ \psi$ being also one to one, the lower bound is obtained:

$$
\phi^{-1} \circ \psi\,(\mathcal{N}(\alpha)) \subset \mathcal{Q}^1_{\alpha'} \implies \xi(\alpha) \leq N^1_{\alpha'}. \tag{A.3}
$$

This lower bound applies only to paths starting at 1 under the assumption that the end is 1, that is, for balanced α. In case the end is k, it suffices to add termination transitions $k \to k+1$, $k+1 \to k+2$, … $q \to 1$, … $k-2 \to k-1$, and a useless transition $k-1 \to k$, therefore this result applies also for non-balanced α with exactly the same definition for α'. Actually the order of the termination transitions is not important, so that adding any permutation (instead of $(1, 2, \dots, q)$) to α is sufficient. Finally the reconstructed chain may pass several times through each state, provided that it passes through all, then adding any global cycle to α protects the reconstruction against an early ending. ∎

Actually the lower bound is more powerful than it might seem. There is no condition on α except that $\alpha' = \alpha + \delta(C) \in \mathcal{S}^1$ and α be non-negative. Consequently we may apply Lemma A.2 to $\widetilde{\alpha} \stackrel{\text{def}}{=} \alpha - \delta(C)$ with $\alpha \in \mathcal{S}^1$, provided that $\alpha - \delta(C)$ be non-negative. Expressing both the upper bound (Lemma A.1) and the lower bound (Lemma A.2) for any initial state, we obtain

Lemma A.3 (paths multiplicity) *Assuming that $\alpha \in \mathcal{S}$ is null outside $\{1,\ldots,q\}^2$, the number of paths N_α^k is bounded by*

$$\forall k \in \{1,\ldots,q\} \qquad \xi(\widetilde{\alpha}) \leq N_\alpha^k \leq \xi(\alpha). \tag{A.4}$$

The lower bound holds only if $\widetilde{\alpha} \stackrel{\text{def}}{=} \alpha - \delta(C)$ is non-negative, with C a global cycle on $\{1,\ldots,q\}$.

A.2 Asymptotics for paths probabilities

The entropy appears in most of the equations. Let $\alpha \in \mathcal{S}$, we define $n = |\alpha|$ and $A = \alpha/n \in M_1(E^2)$. Using heuristically Stirling's formula, we deduce from Lemma A.3 and equation (A.2)

$$\frac{1}{n}\ln \mathbb{P}\left[\mathcal{Q}_\alpha^k\right] = -\widetilde{H}(A\|P) + O\left(\frac{\ln n}{n}\right). \tag{A.5}$$

This asymptotics gives an idea of the result, but the $O(\ln n/n)$ is uncomfortable because it depends on A and q.

For our purpose we need more precise bounds. We have gathered hereafter some well-known formulæ involving the Γ function, which can be found e.g. in [12].

The logarithmic differential

$$\psi(z) = \frac{\mathrm{d}\ln\Gamma(z)}{\mathrm{d}z},$$

comes into the classical Binet's formulæ,

$$\begin{aligned}
\psi(z) &= \ln(z) + \int_0^\infty \left(\frac{1}{t} - \frac{1}{1-e^t}\right) e^{-tz}\,\mathrm{d}t, && \text{(A.6)}\\
\ln\Gamma(z) &= \left(z - \frac{1}{2}\right)\ln z - z + \frac{1}{2}\ln 2\pi + \int_0^\infty f(t)e^{-tz}\,\mathrm{d}t, && \text{(A.7)}\\
\text{with} \quad & f(t) \stackrel{\text{def}}{=} \left(\frac{1}{e^t - 1} - \frac{1}{t} + \frac{1}{2}\right)\frac{1}{t}.
\end{aligned}$$

Letting $g(t) \stackrel{\text{def}}{=} tf(t)$, one sees that f and g are positive. Moreover

$$g(t) \leq \min\left\{\frac{1}{2}, \frac{t}{12}\right\}, \qquad \forall t \geq 0. \tag{A.8}$$

Instantiating $z = 1$ in equation (A.7) yields immediately

$$\int_0^\infty f(t)e^{-t}\,\mathrm{d}t = 1 - \frac{1}{2}\ln 2\pi > 0. \tag{A.9}$$

Therefore

$$1 - \frac{1}{2}\ln 2\pi - \int_0^\infty f(t)e^{-t(n+1)}\,\mathrm{d}t \geq 0, \quad \forall n \in \mathbb{N}. \tag{A.10}$$

We now aim at bounding $\xi(\alpha)$ as uniformly as possible i.e. independently of A and of q except for the main term.

Lemma A.4 (upper bound) *Let $q \in \mathbb{N}^*$ and α be a non-negative integer matrix,*

$$\ln \xi(\alpha) \leq \sum_{i,j=1}^{q} \alpha_{ij} \ln\left(\frac{\alpha_i}{\alpha_{ij}}\right). \tag{A.11}$$

Proof : Firstly we bound the multinomial coefficient ($A = a_1 + \cdots + a_q$). Binet's formula (A.7) yields

$$\begin{aligned}
\ln\left(\frac{A!}{a_1!\ldots a_q!}\right) &= \sum_{k=1}^{q} a_k \ln\left(\frac{A+1}{a_k+1}\right) + \frac{1}{2}\ln\left(\frac{A+1}{(a_1+1)\ldots(a_q+1)}\right) \\
&\quad + \int_0^\infty f(t)\left(e^{-t(A+1)} - \sum_{k=1}^{q} e^{-t(a_k+1)}\right)\mathrm{d}t \\
&\quad + (q-1)\left(1 - \frac{1}{2}\ln 2\pi\right).
\end{aligned} \tag{A.12}$$

To obtain a uniform upper bound, we need to analyze the rest

$$\begin{aligned}
R(a_1,\ldots,a_q) &\stackrel{\text{def}}{=} \frac{1}{2}\ln\left(\frac{A+1}{(a_1+1)\ldots(a_q+1)}\right) + (q-1)\left(1 - \frac{1}{2}\ln 2\pi\right) \\
&\quad + \int_0^\infty f(t)\left(e^{-t(A+1)} - \sum_{k=1}^{q} e^{-t(a_k+1)}\right)\mathrm{d}t.
\end{aligned}$$

We will show that R is decreasing with respect to each of its variables, so that it is negative. Because of the symmetry of R, let $x \stackrel{\text{def}}{=} a_1 + 1$ and $K \stackrel{\text{def}}{=} a_2 + \cdots + a_q$. Then introducing

$$F(x) \stackrel{\text{def}}{=} \frac{1}{2}\ln\left(\frac{K+x}{x}\right) + \int_0^\infty f(t)\left(e^{-t(K+x)} - e^{-tx}\right)\mathrm{d}t \quad \forall x \geq 1,$$

one can write $R(a_1, \dots, a_q) = F(x) + C^{st}$ where the constant does not depend on a_1. Therefore, for $K \geq 1$,

$$\frac{\partial R}{\partial a_1} = F'(x) = \underbrace{-\frac{1}{2}\frac{K}{x(K+x)}}_{\leq -\frac{1}{2}\frac{1}{x(1+x)}} + \underbrace{\int_0^\infty g(t)e^{-tx}}_{\leq \frac{1}{12}\frac{1-e^{-6x}}{x^2}} \leq -\frac{1}{2x^2} + \frac{1}{12x^2} < 0.$$

One easily checks that differentiation of the integral part is valid. The condition $K \geq 1$ is required by the first term. The bound equation (A.8) is used for the second term. The case $K = 0$ is obvious: $F(x) = 0$. Thus $F(x)$ is decreasing with respect to each of its variables and so does R. Finally, by equation (A.9),

$$R(a_1, \dots, a_q) \leq R(0, \dots, 0) = (q-1)\left(1 - \frac{1}{2}\ln 2\pi - \int_0^\infty f(t)e^{-t}\,\mathrm{d}t\right) = 0.$$

Since $\dfrac{A+1}{a_k+1} \leq \dfrac{A}{a_k}$, we obtain the multinomial upper bound

$$\ln\left(\frac{A!}{a_1! \dots a_q!}\right) \leq \sum_{k=1}^{q} a_k \ln\left(\frac{A}{a_k}\right), \tag{A.13}$$

whence the upper bound for $\xi(\alpha)$. ∎

Lemma A.3 clearly shows that the quantity to be bounded from below is $\xi(\widetilde{\alpha})$ and not $\xi(\alpha)$. The forthcoming relations including $\widetilde{\alpha}$ are of course conditional to its non-negativity. The global cycle defining $\widetilde{\alpha}$ is denoted by C.

Lemma A.5 (lower bound) *Assume that α is null outside $\{1, \dots, q\}^2$, and that $\delta(C)$ has less than $2q$ transitions, then*

$$\ln \xi(\widetilde{\alpha}) \geq \sum_{i,j=1}^{q} \alpha_{ij} \ln\left(\frac{\alpha_i}{\alpha_{ij}}\right) + R_q(n), \qquad \textit{with} \quad \frac{R_q(n)}{n} \overset{n\to\infty}{\longrightarrow} 0. \tag{A.14}$$

with $n = |\alpha| = \alpha_1 + \cdots + \alpha_q$.

Proof : The proof is also based on the study of the multinomial coefficient. Using equation (A.10) in equation (A.12), we obtain,

$$\begin{aligned}
&\ln\left(\frac{A!}{a_1!\dots a_q!}\right) - \sum_{k=1}^{q} a_k \ln\left(\frac{A}{a_k}\right) \\
\geq\ & \sum_{k=1}^{q} a_k \ln\left(\frac{a_k}{a_k+1}\right) + A\ln\left(\frac{A+1}{A}\right) + \frac{1}{2}\ln\left(\frac{A+1}{(a_1+1)\dots(a_q+1)}\right) \\
\geq\ & -q - \frac{q}{2}\ln\left(1+\frac{A}{q}\right).
\end{aligned}$$

The last line comes from the optimization with constraint $(a_1+\cdots+a_q = A)$ of the second line line. One easily checks that the supremum is reached at the symmetric point, i.e. $a_k = A/q$ for all k.

The multinomial coefficient lower bound is not uniform but it suffices for the large deviation lower bound. The lower bound for ξ is obtained again by optimization. Omitting the details, we have

$$\ln \xi(\alpha) \geq \sum_{i,j=1}^{q} \alpha_{ij} \ln\left(\frac{\alpha_i}{\alpha_{ij}}\right) - q^2 - \frac{q^2}{2}\ln\left(1+\frac{n}{q^2}\right). \tag{A.15}$$

Similar arguments gives the lower bound for $\xi(\widetilde{\alpha})$, but with a slightly more complicated rest that we summarize in $R_q(n)$. ■

Definition A.5 (one-dimensional support) *The one-dimensional support of an integer matrix α is the set (with a similar definition for a balanced measure),*

$$S_1(\alpha) = \{i \in E : \alpha_i > 0\}.$$

Lemma A.6 (paths probability) *Under the assumption of Lemma A.3, the probability of passing exactly α_{ij} times from state i to state j, $\forall i,j$, is bounded by*

$$-\widetilde{H}(A\|P) - R_q(n) + \frac{\ln \nu(S_1(\alpha))}{n} \leq \frac{1}{n}\ln \mathbb{P}\left[\mathcal{Q}_\alpha\right] \leq -\widetilde{H}(A\|P),$$

with $A \stackrel{\text{def}}{=} \alpha/|\alpha|$ and R_q satisfying $R_q(n) \stackrel{n\to\infty}{\longrightarrow} 0$.

Proof : This is the precise bounds for the heuristic equation (A.5). Gathering the results of Lemma A.3 (bound of N_α^k), improved by Lemma A.4 (upper bound for

$\xi(\alpha)$) and Lemma A.5 (lower bound for $\xi(\widetilde{\alpha})$), we obtain,

$$-\widetilde{H}(A\|P) - R_q(n) + \frac{\ln \nu_k}{n} \le \frac{1}{n} \ln \mathbb{P}\left[\mathcal{Q}_\alpha^k\right] \le -\widetilde{H}(A\|P) + \frac{\ln \nu_k}{n}, \tag{A.16}$$

with R_q verifying $R_q(n) \overset{n\to\infty}{\longrightarrow} 0$. In the present lemma, we sum up the $\mathbb{P}\left[\mathcal{Q}_\alpha^k\right]$ over the initial state k. Since the bounds are uniform with respect to the initial state, the result follows. The upper bound does not change, because $\ln \nu(S_1(\alpha)) \le 0$.

Eventually all states less than q may be an initial state, because of the global cycle. We introduced $S_1(\alpha)$ to encapsulate with the case when α does not pass through every state less than q and when the "global" cycle just covers theses states. ∎

A.3 Bounds for the empirical measure probabilities

Definition A.6 (empirical measure) *The pair empirical measure for a chain X is the random variable*

$$L_n \overset{\text{def}}{=} \frac{1}{n}\left(\sum_{i=1}^{n-1} \delta_{X_i,X_{i+1}} + \delta_{X_n,X_1}\right) \in M_s(E^2). \tag{A.17}$$

The transition $X_n \to X_1$ is called a ghost transition*. The pair empirical measure for a path, or a cycle, is similarly defined.*

Definition A.7 (predecessors of balanced path) *Let α be balanced. $\mathcal{S}_\alpha$ is the set of integer matrices of length n for which the empirical measure is $\alpha/|\alpha|$, and $\mathcal{S}_\alpha^k$ is the set of those which start in k*

$$\mathcal{S}_\alpha^k \overset{\text{def}}{=} \left\{\beta \in \mathcal{S}^k : \exists i \in E, \quad \beta + \delta_{ik} = \alpha\right\} \qquad \text{with} \qquad \mathcal{S}_\alpha \overset{\text{def}}{=} \bigcup_{\mathrm{k}\in\mathrm{E}} \mathcal{S}_\alpha^{\mathrm{k}}.$$

For a given α, there are many empirical measures $\beta \in \mathcal{S}_\alpha^k$: which transition is the ghost transition? Since there is at most n transitions, $\text{card}(\mathcal{S}_\alpha^k) \le n$, and therefore $\text{card}(\mathcal{S}_\alpha) \le n^2$ (for the upper bound it is not desirable that q appear at all).

Definition A.8 *Let A be a balanced measure. A is an empirical measure of order n if it is the empirical measure of a path of length n. The set of these measures is (with $\mathcal{L}$ for* lattice*)*

$$\mathcal{L}_n \overset{\text{def}}{=} \left\{A \in M_s(E^2) : nA \in \mathcal{S}\right\}.$$

Definition A.9 (Sub-Markovian process) *A random process $X_n \in E$ is sub-Markovian if there exists a probability $\nu \in M_1(E)$ and $p_{ij} \leq 1$ for $i, j \in E$ such that*

$$\mathbb{P}\left[X_1 = k_1, \dots, X_n = k_n\right] \leq \nu_{k_1} \prod_{i=1}^{n-1} p_{k_i k_{i+1}}, \tag{A.18}$$

where ν is the initial probability and P a pseudo-kernel[16].

Proposition A.7 (probability of an empirical measure) *Let X be a sub-Markovian process with pseudo-kernel P. Let L_n be its pair empirical measure and $A \in \mathcal{L}_n$, then*

$$\frac{1}{n} \ln \mathbb{P}\left[L_n = A\right] \leq -\widetilde{H}(\widetilde{A} \| P) + \frac{2 \ln n}{n}, \tag{A.19}$$

$$\text{where,} \qquad \forall A \in \mathcal{L}_n, \qquad \widetilde{a}_{ij} \stackrel{\text{def}}{=} \begin{cases} a_{ij} - \dfrac{1}{n} & \text{if } \mathrm{p_{ij}} = \min\limits_{\mathrm{k,l}\,:\,\mathrm{a_{kl}}>0} \mathrm{p_{kl}}, \\ a_{ij} & \text{otherwise.} \end{cases}$$

Moreover if X is a Markov chain with transition probabilities P, initial probability ν, if $\operatorname{Supp}(A) \subset \{1, \dots, q\}^2$, and if there exists a global cycle $(x_1, \dots, x_l)$ with $l \leq 2q$ such that $a_{x_i, x_{i+1}} > 0$ for $i \leq l$, then, for $n > 2q$, we have

$$\frac{1}{n} \ln \mathbb{P}\left[L_n = A\right] \geq -H(A \| P) - R_q(n) + \frac{\ln \nu(S_1(A))}{n}, \tag{A.20}$$

with the rest satisfying $\lim_{n \to \infty} R_q(n) = 0$.

Proof : The proof is based on the equality $\mathbb{P}\left[L_n = A\right] = \sum_{\beta \in \mathcal{S}_\alpha} \mathbb{P}\left[\mathcal{Q}_\beta\right]$, where $A \in \mathcal{L}_n$ and $\alpha = nA$. From Lemma A.5, for $\beta \in \mathcal{S}_\alpha$, we obtain, provided that $\widetilde{\beta}$ exists,

$$\ln \xi(\alpha) + R_q(n) \leq \ln \xi(\widetilde{\alpha}) - (\ln n + 1) \leq \ln \xi(\widetilde{\beta}) \leq \ln \xi(\beta) \leq \ln \xi(\alpha),$$

therefore the rest R_q is different for $\widetilde{\beta}$ but still verifies $\lim_{n \to \infty} R_q(n)/n = 0$.

The problem now is the ghost transition. All paths do not have the same transitions β, but we need this lost transition $p_{X_n X_1}$ for the entropy expression. Taking

[16] Of course there is no uniqueness of P; indeed taking $p_{ij} = 1$, every process is sub-Markovian!

the worst case for $\beta \in \mathcal{S}_\alpha$ the bound of Lemma A.6 turns to be

$$\begin{aligned}
\frac{1}{n}\ln \mathbb{P}\left[\mathcal{Q}_\beta\right] &\geq -H(A\|P) - R_q(n) + \frac{\ln \nu(S_1(A))}{n} && \text{(A.21)}\\
\frac{1}{n}\ln \mathbb{P}\left[\mathcal{Q}_\beta\right] &\leq -H(A\|P) - \frac{1}{n}\min_{i,j\,:\,\alpha_{ij}\geq 1} \ln p_{ij}. && \text{(A.22)}
\end{aligned}$$

As $-\max_{i,j:\alpha_{ij}\geq 1} \ln p_{ij} \geq 0$ this term does not change the lower bound. The problem of the lower bound is the existence of at least one positive $\widetilde{\beta}$ for $\beta \in \mathcal{S}_\alpha$. In fact a sufficient condition (see Lemma A.3) is the existence of a global cycle $C = (x_1, \ldots, x_n)$ such that

$$\forall i, \qquad \alpha_{x_i x_{i+1}} \geq 1,$$

and the existence of at least one more transition (this is the case if $n > 2q$).

The upper bound is more practical in an other format. Let define $\widetilde{A}$

$$\forall A \in \mathcal{L}_n, \qquad \widetilde{a}_{ij} \stackrel{\text{def}}{=} \begin{cases} a_{ij} - \dfrac{1}{n} & \text{if } \; p_{ij} = \min\limits_{k,l\,:\,a_{kl}>0} p_{kl}, \\ a_{ij} & \text{otherwise.} \end{cases}$$

If there are several pairs (i,j) with $a_{ij} > 0$ such that p_{ij} is minimum, then we choose arbitrarily one; in this case, all $\widetilde{A}$ have the same entropy. Note that all a_{ij} are multiple of $1/n$ so that all $\widetilde{a}_{ij}$ are non negative. Then the upper bound for a given β writes

$$\frac{1}{n}\ln \mathbb{P}\left[\mathcal{Q}_\beta\right] \leq -\widetilde{H}(\widetilde{A}\|P).$$

Moreover the upper bound of equation (A.21) may have the unpleasant form $-\infty + \infty$, which is not the case for the above equation.

In the whole reasoning X does not need to be a Markov chain. The upper bound just requires the probability of any path to be bounded above by the product of p_{ij}. Therefore the upper bound still holds for a sub-Markovian process. Taking into account the multiplicity of predecessors for an empirical measure ($\operatorname{Card} \mathcal{S}_\alpha \leq n^2$) we obtain the upper bound. ■

Remark : The conditions of Proposition A.7 are extremely weak. Moreover, the result resembles what can be obtained by heuristic calculations, so that there is no big surprise. The point is just to know what is obvious?

B Elementary properties of the entropy function

Proposition B.1 (positivity) *The relative entropy $H(A\|P)$ is positive. It is null if, and only if, one can write $A = P$ as a Markov kernel, i.e. $A_{ij} = p_{ij}$, $\forall i, j \in E$.*

Proof : We derive the proof from the information theoretic approach. Let consider the equation (2.3). The information gain I is strictly positive for $A_i \neq P_i$, writing A_i and P_i for the i-conditional law (this a classical result derived from Jensen's inequality).

The demonstration would be over if the conditional laws of A were always defined. This is the case only if $a_i > 0$. If $a_i = 0$, $I(A_i\|P_i)$ need not be defined because it must just be defined $A^{(1)}$-a.s. In this case we can choose $A_i = P_i$. Therefore we have proved the proposition. ■

> **Remark :** We must emphasize that the equation $H(A\|P) = 0$ may have no solution, or many. If P is irreducible, there exists a unique solution if, and only if, P is ergodic. If P possesses two ergodic classes, with respective stationnary distribution $\pi(1)$ and $\pi(2)$. $(\pi_i(1)p_{ij})$ and $(\pi_i(2)p_{ij})$ are both balanced measures and their entropies are null (and any barycentric combination of these measures is also a solution).

Lemma B.2 (lower semi-continuity) *The entropy function is lower semi-continuous with respect to the the pair (A, P),*

$$\forall P_n \to P, \quad \forall A_n \to A \qquad H(A\|P) \leq \liminf_{n\to\infty} H(A_n\|P_n).$$

Proof : The lower semi-continuity is also a consequence of equation (2.3). The information $I(A_i\|P_i)$ is lower semi-continuous for all i with respect to both A_i and P_i, and nonnegative. (this is another classical result derived from the convexity of $(x, y) \to x\ln(x/y)$). As sum of nonnegative lower semi-continuous functions, the relative entropy $H(A\|P)$ is lower semi-continuous. ■

Lemma B.3 (convexity) *The entropy function is a convex*[17] *function.*

$$\begin{aligned} &\forall A, A', \forall x, x' \geq 0 : \ x + x' = 1, \\ &\qquad\qquad H(xA + x'A'\|P) \leq xH(A\|P) + x'H(A'\|P). \end{aligned} \tag{B.1}$$

[17] In fact H is convex with A or P. Unfortunately the relative entropy function is *not* convex with both variables as I in the one–dimensional case. Finding a counter-example is a good way to understand the behavior of H.

Proof : Let us perform the heuristic calculation of the Hessian.

$$\frac{\partial^2 H}{\partial a_{ij}\partial a_{kl}} = \frac{\delta_{jl}\delta_{ik}}{a_{ij}} - \frac{\delta_{ik}}{a_i}, \tag{B.2}$$

$$Q(A,P).X.X \stackrel{\text{def}}{=} \sum_{i,j,k,l=1}^{q} \frac{\partial^2 H}{\partial a_{ij}\partial a_{kl}} x_{ij}x_{kl}$$

$$= \sum_{i,j=1}^{q} a_{ij}\left(\frac{x_{ij}}{a_{ij}}\right)^2 - \sum_{i=1}^{q} a_i \left(\frac{x_i}{a_i}\right)^2 \geq 0, \tag{B.3}$$

with $x_i \stackrel{\text{def}}{=} \sum_j x_{ij}$. The positivity of the Hessian results from Jensen's inequality applied to every line i (the convexity of $x \to x^2$).

This method is correct, first, when A belongs to $M(G)$, the set of positive finite measures on the graph of P, because the second derivative is defined only if all concerned p_{ij} are non-null; secondly, if all concerned a_{ij} are non-null. Then, the Hessian being not a continuous function, the convexity can be deduced only for finite support measures, where the Hessian is continuous. Nonetheless we give the sketch of a rigorous proof after Proposition 2.2. ■

At the end of this appendix, we would like to state an important "negative" property. Letting $\Psi(K) \stackrel{\text{def}}{=} \{A \in M_s(E^2) : H(A\|P) \leq K\}$ denotes the level set of the entropy function.

The closed level set $\Psi(K)$ is not necessarily compact.

Proof : Consider the irreducible ergodic kernel P defined on $\mathbb{N}$ by

$$p_{01} = 1, \qquad \text{and} \qquad p_{ij} = \begin{cases} 3/8 & \text{if } \; j = i-1, \\ 1/2 & \text{if } \; j = i, \\ 1/8 & \text{if } \; j = i+1, \end{cases} \qquad \text{when } i \geq 1.$$

For $q > 0$ let $A(q) \in M_s(\mathbb{N}^2)$, with the mass concentrated on $q \to q$ ($a_{qq}(q) = 1$). Then,

$$H(A(q)\|P) = a_{qq}(q) \ln \frac{a_{qq}(q)}{a_q(q)p_{qq}} = -\ln p_{qq} = \ln 2.$$

The set $\{A(q) : q > 0\} \subset \Psi(\ln 2)$ is not tight, therefore $\Psi(\ln 2)$ is not compact. ■

Remark 1 : This shows that we cannot reach a full LDP which would require the entropy function to have compact level sets. Nevertheless we can have either a weak LDP or a full LDP for a restrained class of chains. This paper prove that a weak LDP holds for all Markov chains.

We have been dealing with *pair* empirical measures, but the example proves that the same phenomenon is true for *single* empirical measures, with its appropriate entropy function.

Remark 2 : We do not resist to prove that the condition (1.1) implies the goodness of the entropy function, hence a full LDP. Take a very weak version of (1.1), i.e. there exists $M \geq 1$ such that $P(x, F) \leq MP(y, F)$ for all $x, y \in E$ and $F \subset E$. Then, taking $x = i$, $y = 0$ and $F = \{j\}$, one obtain, as in equation (2.3), $H(A \| P) \geq I(A^{(1)} \| P_0)$. It is easy then to prove the compactness of the level sets. Note that, in the irreducible case, considering the LDP statement, the goodness of $H(. \| P^{(n)})$ implies the goodness of $H(. \| P)$. Hence the condition $P^{(n)}(x, .) \leq MP^{(n)}(y, .)$ is sufficient. The condition (1.1) is a bit more technical.

C Cycles and decomposition of balanced measures

Our purpose is to exhibit empirical measures and their corresponding paths converging to a given balanced measure in order to prove the LDP lower bound. A good setup for this is the theory of graph which is available e.g. in ([2]).

C.1 Elementary definitions and properties

We shortly remain some useful definitions and properties. All graphs are assumed to be oriented 1–graphs on a finite set E. In the following paragraphs $(x_1, \dots, x_n)$ will denote both the cycle and the path. There may exist several such paths, but here it suffices to prove the existence of at least one. We denote by $\mathcal{G}$ the set of all cycles with support on a graph $G \subset E^2$.

A cycle is minimal if, and only if, it passes through each vertex at most once, which implies that the number of minimal cycles is finite. We denote these cycles by $C_1, \dots, C_p$. They play an important role in this section, because any cycle is a product of minimal cycles.

Proposition C.1 *A global cycle is a cycle visiting every point of E. Let q be the cardinality of E. If G is strongly connected then there exists a global cycle with less than $2q$ transitions.*

Definition C.1 (composition of cycles) *Let* $C_1 \stackrel{\text{def}}{=} (x_1, \ldots, x_n)$ *and* $C_2 \stackrel{\text{def}}{=} (y_1, \ldots, y_m)$ *be two cycles possessing at least one common vertex (say* $x_1 = y_1$*). The product* C_1C_2 *is any of the possible cycles*

$$C_1C_2 \stackrel{\text{def}}{=} (x_1, \ldots, x_n, y_1, \ldots, y_n).$$

We define similarly the product of paths, or of a path and a cycle. C^n *denotes the product* $CC\ldots C$*,* n *times.*

The empirical measure $L(C)$ of a cycle C (see Definition A.6) belongs to $M_s(G)$. An interesting property of this measure is that $(l_1+l_2)L(C_1C_2) = l_1L(C_1)+l_2L(C_2)$, with l_1 [resp. l_2] the length of C_1 [resp. C_2], independently of the representative of the class C_1C_2.

C.2 Cycles and balanced measures

Lemma C.2 *If* G *is finite and strongly connected, the closure of the image of* $\mathcal{G}$ *by* L *is the convex hull generated by the empirical measure of the minimal cycles,*

$$\overline{L(\mathcal{G})} = \text{co}(\{L(C_1), \ldots, L(C_p)\}). \tag{C.1}$$

Proof : By the decomposition into minimal cycles, $L(\mathcal{G})$ is included in the convex hull generated by $L(C_1), \ldots, L(C_p)$, and so is the closure (the convex hull is closed).

Let $A \in \text{co}(\{L(C_1), \ldots, L(C_p)\})$. By definition, $A = \sum_{i=1}^p u_iL(C_i)$, with $\sum_i u_i = 1$ and $0 \leq u_i \leq 1$. Let $l_i > 0$ be the length of cycle C_i and $v_i = u_i/l_i$. Rational approximations of v, $v_i(n) = \alpha_i(n)k(n)$, $\alpha_i(n), k(n) \in \mathbb{N}$, can be constructed with $\sum_{i=1}^p v_i(n) = 1$ and $k(n) \to \infty$. Since G is strongly connected, there exists a cycle $C_t = T_{12}T_{23}\ldots T_{p1}$ of length r, where T_{ij} is a path beginning in $\text{Supp}(C_i)$ and ending in $\text{Supp}(C_j)$. Let define

$$C(n) \stackrel{\text{def}}{=} C_1^{\alpha_1(n)}T_{12}C_2^{\alpha_2(n)}T_{23}\ldots C_p^{\alpha_p(n)}T_{p1} \in \mathcal{G}.$$

Noting that $k(n) = \sum_i l_i\alpha_i(n)$, we can write

$$\begin{aligned} A(n) \stackrel{\text{def}}{=} L(C(n)) &= \frac{1}{r+k(n)}\left(\sum_i \alpha_i(n)l_iL(C_i) + rL(C_t)\right) \\ &\stackrel{n\to\infty}{\longrightarrow} \sum_i u_iL(C_i) = A, \end{aligned}$$

i.e. $\text{co}(\{L(C_1), \ldots, L(C_p)\}) \subset \overline{L(\mathcal{G})}$. ∎

Proposition C.3 (empirical and balanced measure) *If G is finite and strongly connected, any balanced measure with support in G is the limit of empirical measures of G–cycles, that is,*

$$M_s(G) = \overline{L(\mathcal{G})} = \mathrm{co}(\{L(C_1), \ldots, L(C_p)\}). \tag{C.2}$$

Proof : By definition $L(\mathcal{G})$ is included in $M_s(G)$ and so is its closure ($M_s(G)$ is closed).

Inversely let $A \in M_s(G)$. First, we decompose A on $\{L(C_1), \ldots, L(C_p)\}$ and consider $A - uL(C_1)$. This is a balanced measure, positive for $u = 0$. Define

$$u_1 \stackrel{\text{def}}{=} \max \{u : A - uL(C_1) \geq 0\}, \tag{C.3}$$

and $B_1 = A - u_1 L(C_1)$. We iterate this operation on B_1 with C_2, and so on. Finally we get a decomposition $(u_1, \ldots, u_p)$ with all nonnegative coefficients and $B = A - \sum_i u_i L(C_i)$ is positive. Moreover $\mathrm{Supp}(B)$ does not possess any cycle; otherwise it would possess a minimal cycle C_i and u_i would not satisfy equation (C.3). Let assume $B \neq 0$, then there exists $b_{x_1 x_2} > 0$. Since B is positive and balanced, $\sum_j b_{x_2 j} = \sum_i b_{i x_2} > 0$ and there exists x_3 such that $b_{x_2 x_3} > 0$ and so on. Since E is finite, we may extract a cycle from $x_1, x_2, \ldots$, which is a contradiction. Then $B = 0$ and $A = \sum_i u_i L(C_i) \in \mathrm{co}(\{L(C_1), \ldots, L(C_p)\})$. ■

In the following propositions we consider an infinite graph G on a countable set E. We need some notion about convex hulls and closure in infinite dimensional spaces (see, for instance, [3] or[16]).

Proposition C.4 (balanced measure decomposition) *The set of balanced measures with support in G is the closed convex hull generated by the empirical measures of the minimal cycles of G,*

$$M_s(G) = \overline{\mathrm{co}}(\{L(C_i)\}_{i \in \mathbb{N}}). \tag{C.4}$$

Proof : Note that the set of minimal cycles is countable, thus it is possible to use a countable index i. Firstly, by definition, $\mathrm{co}(\{L(C_i)\}_{i\in\mathbb{N}}) \subset M_s(G)$. Since $M_s(G)$ is closed (in $M(E^2)$ which is a topological vector space with norm L_1), $\overline{\mathrm{co}}(\{C_i\}_{i\in\mathbb{N}}) \subset M_s(G)$ (the closed convex hull is the closure of the convex hull).

Secondly, the decomposition process of Proposition C.3 still works and yields

$$A = \sum_{i \in \mathbb{N}} u_i L(C_i) \in \overline{\mathrm{co}}(\{L(C_i)\}_{i\in\mathbb{N}}).$$

The sum $\sum u_i = 1$ converges because all u_i are nonnegative and A is a probability measure. Since $M_1(E^2)$ is complete for the norm L_1, so is $M_s(G)$ and the sum $\sum_{i\in\mathbb{N}} u_i L(C_i)$ converges in $\overline{\mathrm{co}}(\{L(C_i)\}_{i\in\mathbb{N}})$. ■

C.3 Markov chains and cycles

In this section E stand for a denumerable state space, and P is an irreducible Markov kernel on E. The graph G of P is the set of all strictly positive transitions. Obviously P is irreducible if, and only if, G is strongly connected.

Proposition C.5 *Let $A \in M_s(G)$ with a finite support*[18]*. There exists a sequence $C(n)$ of G'–cycles, where G' is a finite strongly connected subgraph of G, such that $L(C(n)) \to A$.*

Proof : Let $\mathrm{Supp}(A) \leq q$. Firstly, we construct G', verifying

$$\forall i, j \leq q, \qquad (i,j) \in G \implies (i,j) \in G'. \tag{C.5}$$

Since G is strongly connected, there exists a path $T(i,j)$ connecting any i to any j. Let G' be the union of all the transitions of $T(i,j)$ and all transitions (i,j) for $i,j \leq q$. G' is finite and satisfies (C.5). Let $i,j \in S_1(G')$[19],

- if $i,j \leq q$, the path $T(i,j)$ is a G-path.
- If $i \leq q$ and $j > q$, j belongs to a path $T(x_1,x_2)$ with $x_1,x_2 \leq q$. Then the path $T(i,x_1)T(x_1,x_2)$ contains a path linking i and j.
- If $i,j > q$ we construct a path $T(x_1,x_2)T(x_2,x_3)T(x_3,x_4)$ with i in $T(x_1,x_2)$ and j in $T(x_3,x_4)$ and $x_1,x_2,x_3,x_4 \leq q$. This path contains a path linking i and j.

Thus G' is strongly connected.

Secondly, by (C.5), $A \in M_s(G')$, and since G' is strongly connected, using Proposition C.3, there exists a sequence of G'–cycles $C(n)$, such that $L(C(n)) \to A$. Note the length of the $C(n)$ grows to infinity. ■

[18] For the sake of shortness we denote by $G' \leq q$ the property $G' \subset \{1,\ldots,q\}^2$.

[19] $S_1(G')$ is the one-dimensional projection of $\mathrm{Supp}(G')$ (see Definition A.5).

Proposition C.6 (class decomposition) *Let P be a reducible Markov kernel and let*
$\{E_i\}_{i\in I}$ be the classes of this kernel. There exists a subgraph G' and strongly connected subgraphs G_i with $G_i \subset E_i^2$ of G such that

$$G' = \bigcup_{i\in I} G_i \qquad \text{and} \quad M_s(G') = M_s(G). \tag{C.6}$$

Proof : This is a corollary of Proposition C.4. Let G_i be the set of all allowed transitions ($p_{ij} > 0$) from E_i to E_i. By definition of a class the graphs G_i are strongly connected and there is no cycle crossing two different classes therefore the class graphs G_i are unconnected. Moreover $\bigcup_{i\in I} E_i = E$, and thus any cycle belongs to some G_i. ∎

D Notation

We use the following notation:

- E is a finite or denumerable set. Its cardinality is q, so that E will be identified with $\{1,\ldots,q\}$.
- $(X_i)_{i=1}^{\infty} \in E^{\mathbb{N}}$ is a Markov chain with transition matrix $P = (p_{ij})$.
- $M_1(E)$ is the set of probability measures on E. Let $\mu \in M_1(E)$, μ_i denotes $\mu(\{i\})$.
- $M_s(E^2)$ is the set of balanced measures on E^2 (Definition 2.1). Let $A \in M_s(E^2)$ then:
  - $a_{ij} \stackrel{\text{def}}{=} A(\{(i,j)\})$ is the 2–dimensional law,
  - $A^{(1)}$ the 1–dimensional projection
    with $A^{(1)}(\{i\}) \stackrel{\text{def}}{=} a_i \stackrel{\text{def}}{=} A(\{i\}\times E) = A(E\times\{i\})$,
  - $A_{ij} \stackrel{\text{def}}{=} A(E\times\{j\}|\{i\}\times E)$, the conditional law, so that $a_{ij} = a_i A_{ij}$.
- G is a graph, $M_s(G)$ is the set of pair empirical measures on this graph.
- Let M be a balanced measure, a strongly connected graph or a balanced integer matrix
  - Supp(M) is the support of M,

  - $S_1(M)$ its 1–dimensional projection,
  - $\mathrm{Supp}(M) \leq q$ expresses that $\mathrm{Supp}(M) \subset \{1, \ldots, q\}^2$.

- $L_n(\omega)$ is the pair empirical measure (Definition A.6).
- $\mathcal{L}_n$ is the set of empirical measures produced in n steps (Definition A.8).
- $H(.\|.)$ is the relative entropy (Definition 2.2). We denote by $\tilde{H}$ the extension of H.
- Letting $\alpha = (\alpha_{ij})_{i,j \in E} \in \mathbb{N}^{E^2}$ be a *finite* integer matrix, we introduce the following quantities:
  - $|\alpha| \stackrel{\text{def}}{=} \sum_{i,j} \alpha_{ij} < \infty$,
  - $\alpha_i \stackrel{\text{def}}{=} \sum_j \alpha_{ij}$,
  - $\mathcal{Q}_\alpha^k \subset E^{|\alpha|}$ the set of paths beginning in k with α_{ij} transitions from i to j,
  - N_α^k the cardinality of $\mathcal{Q}_\alpha^k$,
  - $\mathcal{S}^k$ and $\mathcal{S}$ the admissible paths (Definition A.2),
  - $\mathcal{S}_\alpha^k$ and $\mathcal{S}_\alpha$ the predecessors of balanced paths (Definition A.7),
  - $\xi(\alpha)$, the number of α–tuples (Definition A.3),

References

[1] Ash, R. *Information theory.* Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1965. Interscience Tracts in Pure and Applied Mathematics, No. 19.

[2] Berge, C. *Graphs*, vol. 6 of *North-Holland Mathematical Library.* North-Holland Publishing Co., Amsterdam, 1985. Second revised edition of part 1 of the 1973 English version.

[3] Bourbaki, N. *Éléments de mathématique. Fasc. XV. Livre V: Espaces vectoriels topologiques. Chapitre I: Espaces vectoriels topologiques sur un corps valué. Chapitre II: Ensembles convexes et espaces localement convexes.* Hermann, Paris, 1966. Actualités Scientifiques et Industrielles, No. 1189. Deuxième édition revue et corrigée.

[4] Cattiaux, P., and Léonard, C. Minimization of the Kullback information for some Markov processes. In *Séminaire de Probabilités, XXX*, vol. 1626 of *Lecture Notes in Math.* Springer, Berlin, 1996, pp. 288–311.

[5] Csiszár, I., Cover, T. M., and Choi, B. S. Conditional limit theorems under Markov conditioning. *IEEE Trans. Inform. Theory 33*, 6 (1987), 788–801.

[6] Csiszár, I., and Körner, J. *Coding theorems for discrete memoryless systems.* Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1981.

[7] de Acosta, A. Large deviations for empirical measures of Markov chains. *J. Theoret. Probab. 3*, 3 (1990), 395–431.

[8] Dembo, A., and Zeitouni, O. *Large deviations techniques and applications.* Jones and Bartlett Publishers, Boston, MA, 1993.

[9] Deuschel, J.-D., and Stroock, D. W. *Large deviations*, vol. 137 of *Pure and Applied Mathematics.* Academic Press Inc., Boston, MA, 1989.

[10] Ellis, R. S. *Entropy, large deviations, and statistical mechanics*, vol. 271 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science].* Springer-Verlag, New York, 1985.

[11] Ellis, R. S. Large deviations for the empirical measure of a Markov chain with an application to the multivariate empirical measure. *Ann. Probab. 16*, 4 (1988), 1496–1508.

[12] Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F. G. *Higher transcendental functions. Vol. I.* Robert E. Krieger Publishing Co. Inc., Melbourne, Fla., 1981. Based on notes left by Harry Bateman, With a preface by Mina Rees, With a foreword by E. C. Watson, Reprint of the 1953 original.

[13] Gallager, R. *Information Theory and Reliable Communication.* Wiley, New York, 1968.

[14] Hamming, R. W. *Coding and information theory.* Prentice-Hall Inc., Englewood Cliffs, N.J., 1980.

[15] Jain, N. C. Large deviation lower bounds for additive functionals of Markov processes. *Ann. Probab. 18*, 3 (1990), 1071–1098.

[16] LAY, S. R. *Convex sets and their applications.* Robert E. Krieger Publishing Co. Inc., Malabar, FL, 1992. Revised reprint of the 1982 original.

[17] LINDQVIST, B. On the loss of information incurred by lumping states of a Markov chain. *Scand. J. Statist. 5*, 2 (1978), 92–98.

[18] LINNIK, J. V. An information-theoretic proof of the central limit theorem with Lindeberg conditions. *Teor. Verojatnost. i Primenen 4* (1959), 311–321.

[19] LIPTSER, R. S. Large deviations for empirical measures of Markov processes (discrete time, the noncompact case). *Teor. Veroyatnost. i Primenen. 41*, 1 (1996), 65–88.

[20] MICLO, L. Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies. In *Séminaire de Probabilités, XXXI*, vol. 1655 of *Lecture Notes in Math.* Springer, Berlin, 1997, pp. 136–167.

[21] MILGRAM, M. Barycentric entropy of Markov systems. *Cybernet. Systems 12*, 1-2 (1981), 141–178.

[22] RÉNYI, A. *Calcul des probabilités. Avec un appendice sur la théorie de l'information.* Dunod, Paris, 1966. Traduit de l'allemand par C. Bloch. Collection Universitaire de Mathématiques, No. 21.

[23] SHWARTZ, A., AND WEISS, A. *Large deviations for performance analysis.* Stochastic Modeling Series. Chapman & Hall, London, 1995. Queues, communications, and computing, With an appendix by Robert J. Vanderbei.

[24] WU, L. M. Grandes déviations pour les processus de Markov essentiellement irréductibles. I. Temps discret. *C. R. Acad. Sci. Paris Sér. I Math. 312*, 8 (1991), 609–614.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399