Large Deviation Principle for Markov Chains in Continuous Time

Arnaud de la Fortelle

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Large Deviation Principle for Markov Chains in Continuous Time

Arnaud de La Fortelle

N° 3877

Février 2000

THÈME 1

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--3877--FR+ENG

Large Deviation Principle for Markov Chains in Continuous Time

Arnaud de La Fortelle*

Thème 1 — Réseaux et systèmes
Projet MEVAL

Rapport de recherche n° 3877 — Février 2000 — 29 pages

Abstract: Let E be a denumerable state space, Y_t be an homogeneous Markov process on E with generator R. We introduce the *empirical generator* G_t of Y_t, and prove strong LDP local bounds for it. This allows to prove the weak LDP in a very general setting, for irreducible non-explosive Markov processes, not necessarily ergodic. Sanov's theorem is obtained by a contraction argument from the weak LDP for G_t.
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 the empirical generator together with the representation of the rate function as an entropy allow to prove nice properties (uniqueness, continuity, convexity). It also leads to applications in simulation (importance sampling) and in the evaluation of the rate function for sample path LDP in networks. Finally it seems that some technical problems can be reduced to convex programs which can be run with fast algorithms.

Key-words: Large deviation, weak LDP, Markov, countable, empirical measure, Sanov, entropy, information.

(Résumé : tsvp)

The work of the 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

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

Grandes déviations pour les chaînes de Markov en temps continu

Résumé : Soit Y_t un processus de Markov homogène sur un espace d'états dénombrable E, de générateur R. On introduit la notion de *générateur empirique* G_t, qui est une fonctionnelle de $\{Y_s, s \leq t\}$, pour lequel on prouve des bornes locales fortes de type grandes déviations. Un principe faible s'en déduit pour G_t, sous les hypothèses générales que Y_t soit non-explosif et irréductible, mais pas nécessairement ergodique. Une étude fine des propriétés de ces principes *faibles* permet de prouver le théorème de Sanov par contraction.
Nous pensons que ces résultats améliorent les techniques actuelles sous plusieurs aspects. D'abord les conditions de validité ne font pas intervenir l'hypothèse classique d'uniforme ergodicité, ce qui permet d'inclure des applications, comme les réseaux à sauts bornés. Ensuite la fonctionnelle d'action s'exprime explicitement comme une *entropie*, d'où l'on déduit de façon aisée des propriétés comme la continuité ou la convexité. Enfin, les bornes locales sont suffisamment fortes pour montrer que l'entropie est de fait l'unique fonction de Ruelle-Lanford, et donc l'unique fonctionnelle d'action. Sous les hypothèses données, ce résultat n'est pas un corollaire du principe de grandes déviations, comme il l'était dans le cas du principe fort.
La méthode proposée ici semble permettre aussi bien d'obtenir des résultats théoriques que d'améliorer certains algorithmes ("importance sampling" en simulation) ou bien de se ramener à des algorithmes rapides de programmation convexe.

Mots-clé : Grandes déviations, principe faible, Markov, dénombrable, mesure empirique, Sanov, entropie, information.

1 Introduction

Large deviations theory has been widely studied and the range of objects for which an LDP holds is nowadays very large (see [9, 8] and references therein). There are also several approaches [2, 4, 6, 12, 15, 16, 19, 21] by which one can demonstrate these principles, such as subadditivity, convex transforms or change of measures.
The goal of this study is to propose an LDP for a new object, *the empirical generator*, and to show that it is a natural object from which lots of interesting corollaries can be derived. The framework is constituted by continuous time Markov processes on a countable state space, which includes many applications, e.g. queueing networks. In discrete time this would be a "level 2.5" LDP, i.e. Sanov's theorem for the pair empirical measure (see [13] for the definition of the level). But in discrete time (X_i, X_{i+1}) itself is a Markov chain so that the level 2.5 LDP can be derived from the level 2 LDP, the classical Sanov's theorem.
This is no longer true in continuous time since there is no such thing like $(X_t, X_{t+\,\mathrm{d}t})$. Actually, the continuous time entropy[1] H_c has been recently introduced (see [17]), a representation formula of Sanov's rate function has been proved using H_c (see [1]), but there is no LDP. Our results proves that H_c is really a rate function, and that the representation formula is just a contraction of the LDP. Moreover the validity is extended to a countable, instead of finite, state space.
Using the empirical generator has theoretical as well as practical interest. By this mean, Sanov theorem can be derived by contraction for irreducible non-explosive Markov processes and therefore the rate function is easily shown to possess good properties (uniqueness, continuity, convexity), that are not proved in such a general framework. Indeed, in a countable state space, the full LDP does not necessarily hold, and this whole study only requires the weak LDP. It is also an important part of the work to show that, under suitable assumptions, almost all operations usually presented for the sole full LDP (such as contraction) can be performed on weak LDP. Moreover, we prove stronger bounds than LDP bounds, since we are able to prove that the entropy is the Ruelle-Lanford function for the empirical generator. The LDP is then an immediate corollary (see [18]) and this allows additionally to prove the uniqueness of the rate function, which is almost never done for weak LDP. The practical interest is to show how to extend techniques requiring a change of measure, e.g. importance sampling (see [3]). Besides, adapted empirical generators

[1] We shall denote by an index c [resp. d] the entropy function related to the continuous [resp. discrete] time.

can greatly simplify the evaluation of sample path LDP for networks, and even be a (practical) mean to decide between ergodicity and transience.
The organization of the paper is the following: the empirical generator G_t and the entropy function H_c are defined and analyzed in Section 2. H_c is proved to be the Ruelle-Lanford function of G_t in Section 3, which yields the LDP. Properties and operations on weak LDP are discussed in Section 4. Section 5 then derives Sanov's theorem and proves all the usual representations of the rate function to be equal.
In conclusion we wish to emphasize that this method, even though the main object is called a generator, is absolutely not limited to the Markov case. For example, this result can be extended very naturally to semi-Markov processes. The state space can also be generalized, but then more involved analysis is required.

2 Generators and entropy

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.
Because our purpose is to extend its result, we refer the reader to [7] for detailed properties of balanced measures and of the entropy function for *discrete time* Markov chains.

2.1 The empirical generator

Let E be a countable space and $(Y_t)_{t\geq 0}$ be a Markov process on E. Define the following notation

- T_n is the n-th time of jump (with $T_0 = 0$),
- $N_t \stackrel{\text{def}}{=} \sum_{n\geq 0} \mathbb{1}_{\{T_n \leq t\}}$ is the number of jumps at time t ($N_0 = 1$),
- $X_n + 1 \stackrel{\text{def}}{=} Y_{T_n}$ is the embedded chain ($Y_0 = X_1$),
- L_t is the empirical measure of Y_t and $\hat{L}_n$ the pair empirical measure of the embedded chain X_n,

$$L_t \stackrel{\text{def}}{=} \frac{1}{t}\int_0^t \delta_{Y_s}\,\mathrm{d}s, \tag{2.1}$$

$$\hat{L}_n \stackrel{\text{def}}{=} \frac{1}{n}\left(\sum_{i=1}^{n-1} \delta_{X_i X_{i+1}} + \delta_{X_n X_1}\right), \tag{2.2}$$

- $\mathcal{M}_+(E)$ [resp. $\mathcal{M}_1(E)$] is the set of finite positive [resp. probability] measures on E.

Now, the pair empirical measure $\hat{L}_n$ has the particular feature to be *balanced.*

Definition 2.1 (balanced measures) *The set $\mathcal{M}_s(E^2)$ of balanced measures is the subset of $\mathcal{M}_+(E^2)$, with both identical 1–dimensional projections*

$$A \in \mathcal{M}_s(E^2) \iff A(E,.) = A(.,E) \in \mathcal{M}_+(E). \tag{2.3}$$

Let Γ be a graph on E. $\mathcal{M}_s(\Gamma) \subset \mathcal{M}_s(E^2)$ is the set of balanced measures with support included in Γ.

For $A \in M_s(E^2)$, let us put

– $a_{xy} \stackrel{\text{def}}{=} A(\{(x,y)\})$ the 2–dimensional law;

– $a_x \stackrel{\text{def}}{=} A(\{x\} \times E)$, the unique 1–dimensional projection;

– $A_{xy} \stackrel{\text{def}}{=} A(\{(x,y)\} | \{x\} \times E)$, the conditional law, and hence $a_{xy} = a_x A_{xy}$.

Definition 2.2 (empirical generator) *Let Y_t be a Markov process on a countable space E. The empirical generator associated to Y_t is defined by*

$$G_t \stackrel{\text{def}}{=} \left(L_t, \frac{N_t}{t} \hat{L}_{N_t} \right).$$

$\mathcal{G}_s \stackrel{\text{def}}{=} \mathcal{M}_1(E) \times \mathcal{M}_s(E_*^2)$ denotes the set of empirical generators, where E_*^2 is the set of off-diagonal pairs of E. Let $G = (\pi, A)$ be an empirical generator, it is said irreducible if A is irreducible and its support is the support of A.
Let Γ be a subgraph of E_*^2. $\mathcal{G}_s(\Gamma) \subset \mathcal{G}_s(E_*^2)$ is the set of empirical generators with support included in Γ.

The following proposition explain why we still call *generator* the pair $G = (\pi, A)$.

Proposition 2.1 *Let E be a countable state space. There is a one-to-one correspondence between the irreducible empirical generators (π, A) and the generators L_{xy} of irreducible ergodic Markov process on E such that $\sum_{x,y} \pi_x L_{xy} < \infty$, given by*

$$L_{xy} = \frac{a_{xy}}{\pi_x} \quad \textit{if } x \neq y \qquad \textit{and} \quad L_{xx} = -\frac{a_x}{\pi_x}. \tag{2.4}$$

Proof : In this correspondence, L_{xy} can be infinite if $\pi_x = 0$ when $a_x > 0$. There is formally no difficulty to handle this case which corresponds to an instantaneous state x, but we shall not treat it since an instantaneous transition almost never occurs.
Let $G = (\pi, A)$ be an irreducible empirical generator. It is associated a unique generator L_{xy} by (2.4). The transition probabilities for L are $-L_{xy}/L_{xx} = A_{xy}$ for $x \neq y$, hence L is irreducible. Since π is a stationary distribution for L, L is the generator of an irreducible ergodic Markov process.
Reciprocally consider an irreducible ergodic Markov process Y_t with generator L_{xy}, i.e. with intensities $-L_{xx}$ and transition probabilities A_{xy}. There exists a unique stationary distribution π for Y_t. Under the condition $\sum_{x,y} \pi_x L_{xy} < \infty$, $(\pi_x, \pi_x L_{xy})$ is an irreducible empirical generator, since $(\pi_x L_{xy})$ is a *finite* balanced measure.
Obviously both correspondences are inverse of each other, hence the one-to-one correspondence. ■

We shall sometimes denote by l the total mass $\sum_{xy} a_{xy}$ of A. Proposition 2.1 makes it easy to interpreted l, π_x, A_{xy} as the mean intensity of jumps, the stationary distribution and the routing. Therefore a_{xy} can be regarded as the mean number of jumps from x to y per unit time. Obviously these quantities exist if, and only if, the generator is ergodic.
When L is not irreducible, there might exist several stationary distributions which determine several empirical generators. However, if L and π are known, (π, A) is uniquely determined. The worst case is $l = 0$ (i.e. $A = 0$, there is no jump); the mapping $(\pi, A) \to L_{xy}$ is ill-defined for all x such that $\pi_x = 0$, and $L_{xx} = 0$ for all x such that $\pi_x > 0$. Actually the routing is not well defined, but it does not matter as long as it concerns the entropy function. Actually, finding a convenient topology for $\mathcal{G}_s$ is enlightening for all these special cases.

> **Remark : Topological considerations :** The topology for the discrete case is defined by the distance induced by the norm of the underlying Banach space of finite signed measures,
>
> $$d(A, A') = \|A - A'\|_1 \stackrel{\text{def}}{=} \sum_{x,y \in E} |a_{xy} - a'_{xy}|.$$
>
> which yields the same topology on $M_s(E^2)$ as the $\|.\|_\infty$ norm, even though the distances are not equivalent.
>
> The underlying vector space is here the space of signed generators $\mathcal{G}_s^{\text{ev}}(E)$, when π, A or L_{xy} are finite, signed, and where A still verifies the balance equation

(2.3). The generator L has in fact to be defined by a pair (L, π) where π is L-invariant, i.e. $\pi L = 0$. The sum and multiplication by a scalar is defined by $G = \alpha G' + \beta G''$,

$$\begin{cases} \pi = \alpha\pi' + \beta\pi'', \\ a_{xy} = \alpha a'_{xy} + \beta a''_{xy} \end{cases} \quad \text{or} \quad \begin{cases} \pi_x = \alpha\pi'_x + \beta\pi''_x, \\ L_{xy} = \frac{\alpha\pi'_x L'_{xy} + \beta\pi''_x L''_{xy}}{\alpha\pi'_x + \beta\pi''_x} \end{cases}$$

It is easily seen that the real variable is $\pi_x L_{xy}$. It is now natural to define the null element as $A = 0$ *and* $\pi = 0$ or equivalently $\pi = 0$ for the (L_{xy}, π_x) representation.

$\mathcal{G}_s^{\mathrm{ev}}(E)$ is equipped with the norm

$$\|L\|_1 \stackrel{\text{def}}{=} \begin{cases} \displaystyle\sum_{x\in E} |\pi_x| + \sum_{x,y\in E} |\pi_x L_{xy}|, \\ \displaystyle\sum_{x\in E} |\pi_x| + \sum_{x,y\in E} |a_{xy}|, \end{cases}$$

and $\mathcal{G}_s(E)$ by the distance induced by this norm. It is now clear that the two definitions are equivalent, for the representation, the operations and the topology.

♦♦♦

2.2 Entropy

Definition 2.3 (discrete time entropy) *Let P be a transition matrix and $A \in \mathcal{M}_s(E^2)$. The discrete time relative entropy of A with respect to P is defined by*

$$H_d(A\|P) \stackrel{\text{def}}{=} \sum_{x,y\in E} a_{xy} \log\left(\frac{A_{xy}}{P_{xy}}\right), \tag{2.5}$$

with the usual convention

$$0\log(0/0) = 0, \quad 1/0 = +\infty, \quad 0\log 0 = 0.$$

We shall hereafter also denote by P the support of the transition matrix P_{xy}.

Proposition 2.2 *The discrete time relative entropy $H_d(A\|P)$ enjoys the following properties:*

- *H_d is positive, convex and lower semi-continuous with respect to A;*

- *H_d is null if, and only if, one can write $A = P$ as a Markov kernel, i.e. $A_{xy} = P_{xy}$, for all x such that $a_x > 0$ and for all $y \in E$;*

- *Let Γ be a finite subgraph of P. Then $H_d(.\|P)$ is finite and continuous on $\mathcal{M}_s(\Gamma)$;*

- *Let $A \in \mathcal{M}_s(P)$. There exists a sequence of measures $A(n) \in \mathcal{M}_s(P)$, irreducible, with finite support, converging to A, such that their entropy converges to $H_d(A\|P)$.*

$$\lim_{n\to\infty} A(n) = A \quad \textit{and} \quad \lim_{n\to\infty} H_d(A(n)\|P) = H_d(A\|P).$$

Note that the closed level sets of H_d are not necessarily compact, hence a full LDP in discrete time does not hold in general. This remark is still valid in continuous time.

Definition 2.4 (continuous time entropy) *Let R be a generator and $G = (\pi, A)$ an empirical generator. Denote by L_{xy} the corresponding generator (see (2.4)). The relative entropy of G with respect to R is defined by*

$$H_c(G\|R) \stackrel{\text{def}}{=} \sum_{x\neq y\in E} \pi_x I_p(L_{xy}\|R_{xy}), \tag{2.6}$$

$$\textit{where } I_p(\nu\|\lambda) \stackrel{\text{def}}{=} \nu \log\frac{\nu}{\lambda} - \nu + \lambda,$$

with the usual convention of Definition 2.3.

Information theory interpretation The continuous time entropy function H_c has a twofold information theory significance, depending on how the Markov process is considered.

First note that $I_p(\nu\|\lambda)$ is the information gain by unit time when the intensity of a Poisson process is changed from λ to ν. It is linked with the rate function I_e for i.i.d. exponentially distributed random variables by

$$I_p(\nu\|\lambda) = \nu I_e(\nu^{-1}\|\lambda^{-1}) = \nu f(\lambda\nu^{-1}),$$

where $f(u) = u - 1 - \log u$ is the rate function when the mean is 1 and $I_e(.\|x)$ is the rate function when the mean is x. The interpretation is straightforward. The inter-arrival times are exponentially distributed with mean ν^{-1} instead of λ^{-1} so that the

"pay-off" for each jumps is $I_e(\nu^{-1}\|\lambda^{-1})$, multiplied by the number of jumps by unit time ν.
On one hand, the process of jump from x to y occurs as a Poisson process of intensity L_{xy} instead of R_{xy} for a fraction π_x of time, so that the mean information gain is deduced to be (2.6). On the other hand, the jumps at x are exponentially distributed with parameter L_x instead of R_x, so that the change of distribution is "rewarded" $I_e(L_x^{-1}\|R_x^{-1})$ each time, and there is a number of passage a_x per unit time. Moreover the change of routing P to A is rewarded $H_d(A\|P)$ per unit time. Thus it yields the other form

$$H_c(G\|R) = H_d(A\|P) + \sum_{x\in E} a_x I_e(L_x^{-1}\|R_x^{-1}). \tag{2.7}$$

The equality between (2.6) and (2.7) is easily checked. (2.7). This information theory heuristic is exactly the sketch of an alternative proof using conditioning (as in [7]). Since we propose a stronger method, this proof is not detailed here.

Proposition 2.3 *The relative entropy $H_c(G\|R)$ is positive, lower semi-continuous and convex with respect to $G = (\pi, A)$. It is null if, and only if, $L_{xy} = R_{xy}$ for all $x, y \in E$.*

Proof : All those properties are immediately derived from the same properties of I_p, by (2.6). ∎

As in the discrete time, the proof of the LDP requires continuity properties of the entropy function. Hereafter R will be assumed to be an irreducible generator, and for the sake of simplicity $\mathcal{G}_s(R)$ will denote the empirical generators on the graph of R. Note that $H_c(G\|R)$ is finite only if $G \in \mathcal{G}_s(R)$.

Proposition 2.4 *Let Γ be a finite subgraph of* $\mathrm{Supp}(R)$. *Then $H_c(.\|R)$ is finite continuous on $\mathcal{G}_s(\Gamma)$. It has compact level sets.*

Proof : By (2.6), the entropy function is a finite sum, for $(x, y) \in \Gamma$, of continuous finite functions with compact level sets. ∎

Proposition 2.5 (exact l.s.c.) *Let $G \in \mathcal{G}_s(R)$. There exists a sequence of generators $G(n) \in \mathcal{G}_s(R)$, irreducible, with finite support, converging to G, such that their entropy converge to $H_c(G\|R)$.*

$$\lim_{n\to\infty} G(n) = G \quad \textit{and} \quad \lim_{n\to\infty} H_c(G(n)\|R) = H_c(G\|R).$$

Proof : By (2.7), this is an extension of the same property for the discrete entropy function H_d (Proposition 2.2); there exist $A(n)$ converging to A such that $A(n)$ is irreducible, with finite support and $H_d(A(n)\|P)$ converges to $H_d(A\|P)$. The proof is completed by taking $G(n) = (\pi(n), A(n))$ where $\pi(n)$ is an approximation of π that has the same support as $A(n)$. ■

3 LDP at fixed time

Let E be a countable state space, and Y_t be a Markov process with generator R. We prove here that the entropy H_c is a Ruelle-Lanford function for the empirical generator G_t, which is a stronger result than a weak LDP. For the notion of Ruelle-Lanford functions and the related theorems, we refer the reader to Lewis and Pfister [18] (or see Section 4.2).

Theorem 3.1 *Let Y_t be a non-explosive Markov process on a countable state space E with generator R and initial distribution ν. The entropy function $H_c(.\|R)$ is a Ruelle-Lanford function for the empirical generator G_t.*

The proof relies on an exponential change of measure, from which strong upper and lower bounds are derived. Using the continuity properties of the entropy function, these bounds are extended to all empirical generators. Finally $H_c(.\|R)$ is deduced to be a Ruelle-Lanford function for the empirical generator G_t.

Change of measure Let $G = (\pi, A)$ be a generator and L_{xy} be the corresponding generator[2] by (2.4). In order to define properly h (see below), the support of G is restricted to Supp(R), i.e. $G \in \mathcal{G}_s(R)$,

- $h(x,y) \stackrel{\text{def}}{=} \begin{cases} \log(L_{xy}/R_{xy}) \in \mathbb{R} \cup \{-\infty\} & \text{if } x \in \text{Supp}(G), \\ 0 & \text{otherwise,} \end{cases}$
- $\Lambda(x) \stackrel{\text{def}}{=} \sum_{y \neq x} R_{xy}(e^{h(x,y)} - 1)$,
- and the process,

$$M_t \stackrel{\text{def}}{=} \exp\left\{ \sum_{i=1}^{N_t - 1} h(X_i, X_{i+1}) - \int_0^t \Lambda(Y_s)\,\mathrm{d}s \right\}. \tag{3.1}$$

[2]Note that the relation (2.4) allows to define L_{xy} even though Proposition 2.1 is valid only for irreducible G with finite support.

To prove that M_t is a martingale, one first calculate the derivative in 0

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_x\left[M_t\right]\Big|_{t=0} = \sum_{y\neq x} R_{xy}(e^{h(x,y)}-1) - \Lambda(x) = 0, \quad \forall x \in E. \tag{3.2}$$

Using the Markov property, this means that the derivative is null for all $t \geq 0$, so that $\mathbb{E}\left[M_t\right] = M_0 = 1$. Then, using again the Markov property, this proves that $\mathbb{E}\left[M_s|\mathcal{F}_t\right] = M_t$ for all $s \geq t \geq 0$, i.e. M_t is a martingale with respect to $\mathcal{F}_t$.
A new probability measure is defined by

$$\mathbb{P}^*\left[B\right] \stackrel{\text{def}}{=} \mathbb{E}\left[\mathbb{1}_{\{B\}} M_t\right], \qquad \forall B \in \mathcal{F}_t.$$

It is again classical that under the probability $\mathbb{P}^*$, Y_t is a Markov process with intensity L_{xy},

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_x^*[f(Y_t)]\Big|_{t=0} = \begin{cases} \sum_{y\neq x} L_{xy}(f(y)-f(x)), & \text{if } x \in \mathrm{Supp}(G), \\ \sum_{y\neq x} R_{xy}(f(y)-f(x)), & \text{if } x \notin \mathrm{Supp}(G). \end{cases}$$

We proved that Y_t is a Markov process with generator L_{xy} on $\mathrm{Supp}(G)$ which is absorbing under $\mathbb{P}^*$, and generator R_{xy} outside $\mathrm{Supp}(L)$. We shall denote by $\mathrm{Supp}(\mathbb{P}^*)$ the graph of all possible transitions under $\mathbb{P}^*$, i.e. the pairs (x,y) of $\mathrm{Supp}(G)$ when $x \in \mathrm{Supp}(G)$ and of $\mathrm{Supp}(R)$ when $x \notin \mathrm{Supp}(G)$.
To this change of measure is associated the linear mapping

$$\phi(\pi', A') \stackrel{\text{def}}{=} \sum_{x,y\in E} a'_{xy} h(x,y) + \sum_{x\in E} \pi'_x \Lambda(x),$$

where $\pi' \in \mathcal{M}_1(E)$ and $A' \in \mathcal{M}_+(\mathrm{Supp}(\mathbb{P}^*))$. ϕ is finite and continuous on $\mathcal{G}(E) \stackrel{\text{def}}{=} \mathcal{M}_1(E) \times \mathcal{M}_+(\mathrm{Supp}(\mathbb{P}^*))$ and a simple calculation shows that

$$\phi(G) = H_c(G\|R). \tag{3.3}$$

Strong bounds This change of measure is particularly well suited for the following additive version of the empirical generator G_t,

$$\widetilde{G}(t_1,t_2) \stackrel{\text{def}}{=} \left(\frac{1}{t_2-t_1}\int_{t_1}^{t_2} \delta_{Y_u}\,\mathrm{d}u,\ \frac{1}{t_2-t_1}\sum_{i=N_{t_1}}^{N_{t_2}-1} \delta_{X_i X_{i+1}} \right) \in \mathcal{G}(E),$$

since

$$\log M_t = t\phi\left(\widetilde{G}(0,t)\right). \tag{3.4}$$

The difference between $\widetilde{G}(0,t)$ and G_t [resp. $\mathcal{G}(E)$ and $\mathcal{G}_s(E)$] is that the first one is not necessarily balanced. Moreover, they are related by

$$G_t = \widetilde{G}(0,t) + \frac{1}{t}(0, \delta_{Y_t Y_0}). \tag{3.5}$$

For the purpose of Ruelle-Lanford function, it is necessary to obtain much stronger bounds than for proving the LDP. Here, the lower bound holds for compact sets (instead of open sets) and the upper bound holds for open sets (instead of compact sets). Now, it is easier to prove these bounds for $\widetilde{G}(0,t)$ than for G_t.

Lemma 3.2 *Let Y_t be a Markov process with irreducible generator R_{xy}, then*

$$\lim_{r\to 0} \sup_{K\subset B(G,R)} \liminf_{t\to\infty} \mathbb{P}\left[G_t \in K\right] \geq -H_c(G\|R), \tag{3.6}$$

$$\lim_{r\to 0} \limsup_{t\to\infty} \mathbb{P}\left[G_t \in B(G,r)\right] \leq -H_c(G\|R), \tag{3.7}$$

where K denotes compact sets. (3.6) holds for any irreducible generator $G \in \mathcal{G}_s(R)$ with finite support while (3.7) holds for any $G \in \mathcal{G}_s(R)$.

Proof : The technical proof that follows does not express anything else than: *under $\mathbb{P}^*$, Y_{t_0} reaches* Supp(G) *with non-null probability. Then Y_t remains in* Supp(G) *because it is absorbing, and by ergodicity $\widetilde{G}(t_0,t)$ converges to G.* The problem is to exhibit compact sets for the lower bound.

Let $G \in \mathcal{G}_s(R)$ be an irreducible generator with finite support. Choose x_0 such that $\nu(x_0) > 0$ and $x_1 \in$ Supp(G). From the irreducibility of R, there exists t_1 with $\mathbb{P}^*\left[Y_0 = x_0,\ Y_{t_1} = x_1\right] > 0$. Since the space of càdlàg functions $\mathbb{D}([0,t_1],E)$ is Polish, there exists a compact set K_1 of trajectories on $[0,t_1]$ such that

$$K_0 \subset \{Y_0 = x_0\} \cap \left\{Y_{t_1} = x_1\right\}, \quad \text{and} \quad \mathbb{P}^*\left[K_0\right] = C > 0. \tag{3.8}$$

$\widetilde{G}(0,t_1)$ is a continuous functional of the trajectory $\{Y_t, t \in [0,t_1]\}$, so that

$$K_1 = \left\{\widetilde{G}(0,t_1)(\omega),\ \forall \omega \in K_0\right\}$$

is compact. But Supp(G) is absorbing for $\mathbb{P}^*$, hence the set of states is now restricted to a finite set.

Define K_2 to be the closed ball of center G and radius $\varepsilon > 0$, in $\mathcal{G}(\mathrm{Supp}(G))$. It is a compact set since G has finite support and $l = \sum_{x,y} a_{xy}$ is bounded. G is irreducible, thus the Markov process is ergodic, once in $\mathrm{Supp}(G)$, hence $\widetilde{G}(0,t)$ tends to G and there exists t_2 such that

$$\mathbb{P}^*_{x_1}\left[\widetilde{G}(0,t) \in K_2\right] \geq \frac{1}{2}, \qquad \forall t \geq t_2. \tag{3.9}$$

Now, the two parts have to be joined. By the additivity of $\widetilde{G}$,

$$\widetilde{G}(0,t) = \frac{t_1}{t}\widetilde{G}(0,t_1) + \left(1 - \frac{t_1}{t}\right)\widetilde{G}(t_1,t).$$

Fix $\varepsilon \leq 1$ and define the compact set

$$K \stackrel{\text{def}}{=} \bigcup_{0 \leq u \leq \varepsilon} \{uK_1 + (1-u)K_2\} \ \subset \mathcal{G}(E).$$

Note that K is included in a ball of center G and radius 2ε in $\mathcal{G}(E)$. Now, using the Markov property, by (3.8) and (3.9)

$$\mathbb{P}^*\left[\widetilde{G}_t \in K\right] \geq \mathbb{P}^*\left[\widetilde{G}(0,t_1) \in K_1\right] \mathbb{P}^*_{x_1}\left[\widetilde{G}(0,t-t_1) \in K_2\right] \geq \frac{C}{2},$$

for all $t \geq t_0/\varepsilon$ and $t - t_1 \geq t_2$.
Using the definition of $\mathbb{P}^*$ and by (3.4),

$$\begin{aligned} \frac{1}{t}\log \mathbb{P}\left[\widetilde{G}_t \in K\right] &= \frac{1}{t}\log \mathbb{E}^*\left[\mathbb{I}_{\{\widetilde{G}_t \in K\}} M_t^{-1}\right] \\ &\geq -\sup_{G' \in K} \phi(G') + \frac{1}{t}\log \mathbb{P}^*\left[\widetilde{G}_t \in K\right]. \end{aligned} \tag{3.10}$$

Since the probability for $\widetilde{G}_t$ to belongs to K is greater than $C/2 > 0$ for t large enough,

$$\liminf_{t\to\infty} \frac{1}{t}\log \mathbb{P}\left[\widetilde{G}_t \in K\right] \geq -\sup_{G' \in K} \phi(G').$$

Now ϕ is continuous and its value at G is $H_c(G\|R)$. Moreover K is included in a small neighborhood of G, so that the strong lower bound holds,

$$\lim_{r\to 0} \sup_{K \subset B(G,r)} \liminf_{t\to\infty} \frac{1}{t}\log \mathbb{P}\left[\widetilde{G}_t \in K\right] \geq -H_c(G\|R), \tag{3.11}$$

where K denotes compact sets.
To convert this result for G_t is easily done by (3.5). Note that the second term $t^{-1}(0, \delta_{Y_tY_0})$ belongs to a compact K_3 when Y_t lies in the finite support of G and $Y_0 = x_0$. In addition it is very small when t is large. Formally, define

$$K' \stackrel{\text{def}}{=} \bigcup_{0 \le u \le \varepsilon} \{uK_1 + (1-u)K_2 + ut_1^{-1}K_3\} \ \subset \mathcal{G}(E).$$

Then

$$\widetilde{G}(0,t) \in K \implies G_t \in K'.$$

Moreover K' is included in a ball of center G and radius 3ε as soon as $t \ge t_1/\varepsilon$. Finally, by (3.11), the lower bound (3.6) is valid for irreducible $G \in \mathcal{G}_s(R)$ with finite support.

The upper bound is much easier to obtain. Reversing the inequality of (3.10) yields, for any measurable set $O \subset \mathcal{G}(E)$,

$$\frac{1}{t} \log \mathbb{P}\left[\widetilde{G}(0,t) \in O\right] \le - \inf_{G' \in O} \phi(G'), \tag{3.12}$$

where the probability has been bounded by 1. By (3.5), the difference between G_t and $\widetilde{G}(0,t)$ is bounded by $1/t$ so that

$$G_t \in B(G,\varepsilon) \implies \widetilde{G}(0,t) \in B(G, 2\varepsilon), \qquad \forall t > 1/\varepsilon,$$

hence

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left[G_t \in B(G,\varepsilon)\right] \le - \inf_{G' \in B(G,2\varepsilon)} \phi(G').$$

Finally, by continuity of ϕ and using (3.3), the upper bound (3.7) is proved for any $G \in \mathcal{G}_s(R)$. ■

proof of Theorem 3.1 : Ruelle-Lanford function In a first step, the lower bound (3.6) has to be extended to all empirical generators.
Let $G \in \mathcal{G}_s(E)$, not necessarily irreducible nor with finite support and $r > 0$. By Proposition 2.5, for all $\varepsilon > 0$, there exists an irreducible finite support generator $G' \in B(G,r)$ such that $|H_c(G'\|R) - H_c(G\|R)| < \varepsilon$. Since $B(G,r)$ is open, there exists $r' > 0$ such that $B(G',r')$ is included in $B(G,r)$, and, by (3.6) for all $\eta > 0$ there exists a compact set $K \subset B(G',r')$ such that

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left[G_t \in K\right] \ge -H_c(G'\|R) - \eta \ge -H_c(G\|R) - \eta - \varepsilon.$$

Since this is true for all $\varepsilon > 0$ and all $\eta > 0$, (3.6) is valid for G, hence the extension of Lemma 3.2.
Remind that the upper and lower rate functions $\overline{I}$ and $\underline{I}$ are defined by (see Definition 4.2),

$$\underline{I}(G) \stackrel{\text{def}}{=} -\lim_{r\to 0}\liminf_{t\to\infty}\frac{1}{t}\log \mathbb{P}\left[G_t \in B(G,r)\right], \tag{3.13}$$

$$\overline{I}(G) \stackrel{\text{def}}{=} -\lim_{r\to 0}\sup_{K\subset B(G,r)}\limsup_{t\to\infty}\frac{1}{t}\log \mathbb{P}\left[G_t \in K\right], \tag{3.14}$$

This extension of Lemma 3.2 yields

$$\overline{I}(G) \leq \underline{I}(G) \leq H_c(G\|R) \leq \overline{I}(G),$$

for all generators $G \in \mathcal{G}_s(R)$. This means $\underline{I}(G) = \overline{I}(G) = H_c(G\|R)$ for these generators. Both bounds are proved: H_c is a Ruelle-Lanford function for G_t. ∎

Theorem 3.3 (LDP) *Let Y_t be a non-explosive Markov process on a countable state space E. The empirical generator at fixed time G_t verifies a weak LDP with rate function H_c. If E is finite, G_t verifies a full LDP with good rate function H_c.*

Proof : The first part of the theorem is theorem 3.1 in [18] (or Theorem 4.2 here). For the second part, it is only necessary to extend the upper bound from compact to closed sets. The rate function is known to be good (Proposition 2.4) but anyway this is implied by the fullness of the LDP.
The upper bound is valid for closed sets F such that there exists an upper bound l_0 for $l = \sum_{x,y} a_{xy}$; they are compact since E is finite so that $\mathcal{M}_1(E)$ and $\mathcal{M}_s(E_*^2)$ are compact. The extension of this bound is obtained by considering the asymptotics of $\mathbb{P}\left[N_t/t \geq l_0\right]$.
In fact, for a Poisson process with intensity r, for all $l_0 > r$

$$\mathbb{P}\left[N_t/t > l_0\right] = e^{-rt}\sum_{n>l_0 t}\frac{(rt)^n}{n!} \leq e^{-rt}\frac{(rt)^{l_0 t}}{(l_0 t)!}\frac{l_0}{l_0 - r},$$

because $n! \leq n_0! n_0^{n-n_0}$ for $n \geq n_0$. This bound also holds for the Markov process Y_t with $\bar{r} \stackrel{\text{def}}{=} \sup\{R_x : x \in E\} < \infty$, so that

$$\limsup_{t\to\infty}\frac{1}{t}\log \mathbb{P}\left[N_t/t > l_0\right] \leq -\bar{r} + l_0 - l_0\log\frac{l_0}{\bar{r}} \stackrel{l_0\to\infty}{\longrightarrow} -\infty$$

Hence, writing $\Psi(l_0)$ the set of all generators (π, A) verifying $l \leq l_0$, for all $C > 0$, there exists $l_0 < \infty$ such that

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left[G_t \in \Psi^c(l_0)\right] \leq -C.$$

By the lower bound (3.6), this implies that H_c is greater than C on $\Psi^c(l_0)$.
Now consider $C > \inf_F H_c(.\|R)$ and a corresponding l_0. F is contained in the union of $F \cap \Psi(l_0)$ and $\Psi^c(l_0)$, so that

$$\begin{aligned} &\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left[G_t \in F\right] \\ \leq\quad & -\min\left\{\inf_{G\in F\cap\Psi(l_0)} H_c(G\|R), C\right\} = -\inf_{G\in F} H_c(G\|R) \end{aligned} \tag{3.15}$$

because the infimum of H_c is reached on $F \cap \Psi(l_0)$. The exponential tightness for G_t is proved, hence the extension of the weak LDP to a full LDP. ■

Remark explosive process : If the process Y_t can jump an infinite number of times in a finite duration, an LDP can not be derived this way, yet Lemma 3.2 is almost true. The stumbling block is the new beginning of the process after a time of explosion: it can create a new non-balanced state, and finally (3.5) no more holds. This means that $\widetilde{G}_t$ does not tend necessarily toward $\mathcal{G}_s(E)$. The rate function is not necessarily infinite outside $\mathcal{G}_s(E)$ and can be less than H_c on $\mathcal{G}_s(E)$. What can be proved with this method is an LDP for

$$\mathbb{E}\left[\mathbb{1}_{\{G_t\in .\}}\mathbb{1}_{\{Y_t \text{ does not explode before } t\}}\right].$$

Even for an explosive process, this substochastic process is well-defined (see Feller [14]) and unique.

♦♦♦

4 Operations on weak LDP

A full LDP is often required in the literature for important corollaries of LDP, such as Varadhan's integral lemma or the contraction principle. In fact, under some natural restrictions, the same theorems applies for weak LDP.

4.1 Definition and extension

Definition 4.1 (Weak LDP) *Let $\mu_\varepsilon \subset \mathcal{M}_1(E)$ be a family of probability measures on a Hausdorff topological space E. It satisfies a weak LDP with rate function I if*

- *(W1) $I: \ \mathcal{M}_1(E) \mapsto \mathrm{I\!R}_+$ is a lower semi-continuous function;*
- *(W2) for all open sets $O \subset \mathcal{M}_1(E)$,*
$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(O) \geq -I(O); \tag{4.1}$$
- *(W3) for all compact sets $K \subset \mathcal{M}_1(E)$,*
$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K) \leq -I(K); \tag{4.2}$$

denoting by $I(F)$ the infimum of I over a set $F \subset \mathcal{M}_1(E)$.

To handle closed sets an auxiliary function θ is required.

Assumption (H1) *There exists a lower semi-continuous mapping $\theta: \ E \to F$ on the Hausdorff topological space F such that μ_ε is exponentially tight in θ, i.e. for all finite $C \in \mathrm{I\!R}$ there exists a compact $K \subset F$ such that*

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\theta^{-1}(K^c)) \leq -C. \tag{4.3}$$

Such a function θ is called an auxiliary function.

Theorem 4.1 *Let $\mu_\varepsilon \subset \mathcal{M}_1(E)$ be a family of probability measures on a Hausdorff topological space E satisfying a weak LDP with rate function I. If there exists an auxiliary function $\theta: \ E \to F$ (i.e. assumption (H1)), then the upper bound (4.2) holds for all closed sets F that have compact θ-level sets, i.e. such that $F \cap \theta^{-1}(K)$ is compact, for all compact sets $K \subset F$.*

Proof : Let $F \subset \mathcal{M}_1(E)$ be a closed set with compact θ-level sets, and $C > I(F)$. By (H1) there exists a compact set $K \subset F$ such that (4.3) is satisfied. By the lower bound, this implies that $I(x) \geq C$ for all $x \in \theta^{-1}(K^c)$, therefore $I(F) = I(F \cap \theta^{-1}(K))$.
Then divide F into $F \cap \theta^{-1}(K)$, which is compact by assumption and into $F \cap \theta^{-1}(K^c)$ which is included in $\theta^{-1}(K^c)$. Then,

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(F) \leq -\min\{I(F \cap \theta^{-1}(K)); C\} = -I(F),$$

where the first term of the min is due to the the upper bound (4.2) and the second to the exponential tightness in θ. Hence the upper bound. ■

Note that θ needs only to be lower semi-continuous so as $\theta^{-1}(K)$ to be closed, and the Hausdorff assumption ensures that compact sets are closed. In the classical theory I is required to have compact level sets and μ_ε to be exponentially tight. One easily checks that these assumptions means, in our framework, that $\theta = I$ in which case (4.3) is the usual exponential tightness assumption. But there are many other interesting cases where the exponential tightness does not hold but where this theorem has interesting consequences.

4.2 Ruelle-Lanford functions and uniqueness

A first step to the identification is the classical uniqueness, which is not not usually given in the framework of weak LDP. The following results are contained in [18].

Definition 4.2 *Let $\mu_\varepsilon \subset \mathcal{M}_1(E)$ be a family of probability measures on a Hausdorff topological space E. A Ruelle-Lanford function is uniquely defined by $\overline{I} = \underline{I}$ with*

$$\underline{I}(x) \stackrel{\text{def}}{=} -\lim_{r\to 0} \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\mathcal{V}_r), \tag{4.4}$$

$$\overline{I}(x) \stackrel{\text{def}}{=} -\lim_{r\to 0} \sup_{K\subset\mathcal{V}_r} \limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K), \tag{4.5}$$

where $\mathcal{V}_r$ is a decreasing (when $r \to 0$) base of open neighborhoods of $x \in E$ and K are compact sets. Both $\overline{I}$ and $\underline{I}$ are lower semi-continuous.

Theorem 4.2 (Existence) *Let $\mu_\varepsilon \subset \mathcal{M}_1(E)$ be a family of probability measures on a Hausdorff topological space E. If $\underline{I} \le \overline{I}$ in Definition 4.2, then μ_ε verifies a weak LDP with rate function I, for any lower semi-continuous I satisfying $\underline{I} \le I \le \overline{I}$.*

Theorem 4.3 (Uniqueness) *Let $\mu_\varepsilon \subset \mathcal{M}_1(E)$ be a family of probability measures on a Hausdorff topological space E. If it satisfies a weak LDP with rate function I, either there is a Ruelle-Lanford function which is the unique rate function I, or there is no uniqueness.*

Proof : A necessary and sufficient condition for an LDP to hold is

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K) \le \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(O),$$

for all open sets O and all compact sets $K \subset O$, which ensures that $\overline{I} \ge \underline{I}$ in Definition 4.2.

Since it governs a LDP, I must satisfy $\underline{I} \leq I \leq \overline{I}$. Moreover, any lower semi-continuous function I verifying $\underline{I} \leq I \leq \overline{I}$ is a rate function (see Theorem 4.2).
Now, either I is a Ruelle-Lanford function and the uniqueness is proved, or there is a strict inequality $\overline{I} > \underline{I}$ and there are several rate functions. ■

4.3 Contraction

Hereafter, we prove a somewhat restricted contraction principle, that needs (H1) and a further assumption.

Assumption (H2) *Let $\theta: E \mapsto F'$ be an auxiliary function and $f: E \to F$ be a continuous mapping on the Hausdorff topological space F. Is is assumed that for all compact sets $K \subset F$ and $K' \subset F'$,*

$$f^{-1}(K) \cap \theta^{-1}(K') \quad \textit{is compact.}$$

Then f is said to have compact θ-level sets.

Theorem 4.4 (Contraction) *Let $\mu_\varepsilon \subset \mathcal{M}_1(E)$ be a family of probability measures on a Hausdorff topological space E which satisfies a weak LDP with rate function I. If $\theta: E \mapsto F_2$ verifies assumption (H1) and $f: E \to F$ satisfies (H2), then $\mu_\varepsilon \circ f^{-1}$ verifies a weak LDP with rate function J,*

$$J(y) \stackrel{\text{def}}{=} I(f^{-1}(y)) = \inf_{x \in f^{-1}(y)} I(x). \tag{4.6}$$

Proof : There are three points (W1)–(W3) to verify (see Definition 4.1). The lower bound (W2) is immediate, the upper bound (W3) is an application of Theorem 4.1, there is only the lower semi-continuity left (W1).
Let $y_n \in F$ be a sequence converging to y. We shall denote by K the compact set $\{y_1, y_2, \ldots, y\}$. It is possible to assume

$$\liminf_{n \to \infty} J(y_n) = C_1 < \infty,$$

otherwise the lower semi-continuity is obvious. Then extract a sequence such that $J(y_n)$ converges to C_1. By (H1) there exists $K_2 \subset F_2$ such that (4.3) is satisfied with $C > C_1$. Let $\varepsilon > 0$, there exists $x_n \in f^{-1}(y_n)$ such that $I(x_n) \leq J(y_n) + \varepsilon$. For n large enough, $J(y_n)$ is near C_1 so that x_n belongs to $f^{-1}(K) \cap \theta^{-1}(K_2)$ which is compact by (H2). Thus it is possible to extract a sub-sequence x_n converging to x. Now f is continuous, so that $f(x) = y$ and by the lower semi-continuity of I,

$$\lim_{n \to \infty} J(y_n) + \varepsilon \geq \liminf_{n \to \infty} I(x_n) \geq I(x) \geq J(y).$$

Since this is true for all $\varepsilon > 0$, the lower semi-continuity is proved. ■

This idea to use an auxiliary function θ contains the two cases where this theorem has been proved; first the full LDP, where $\theta = I$ which is lower semi-continuous and has compact level sets. In this case (H1) is the usual exponential tightness assumption; and second the weak contraction principle as in [18] where $f^{-1}(K)$ is required to be compact for compact K, in which case e.g. $\theta = 0$. In both cases the assumption (H2) is obvious but this method allows further results such as the contraction of the empirical generator to the empirical measure in the Markov case. The following properties will be useful in the Markov case, but their interest are not restrained to this framework.

Lemma 4.5 *Assume (H1) and (H2) as in Theorem 4.4. For all $y \in F$ such that $J(y) < \infty$, the infimum $J(y) = I(f^{-1}(y))$ is reached on a non-empty and compact set.*

Proof : Let $C > J(y)$ be finite. By (H1), there exists K such that (4.3) holds. The infimum $J(y)$ is equal to $I(f^{-1}(y) \cap \theta^{-1}(K))$ (see the proof of Theorem 4.1). By (H2), this is a compact set. Since I is lower semi-continuous, the infimum is reached on this compact. Now the set of infima is exactly $I^{-1}(J(y)) \cap f^{-1}(y)$ which is closed, and the proof is completed. ■

Proposition 4.6 *Assume (H1) and (H2) as in Theorem 4.4. If I is a Ruelle-Lanford function for μ_ε then J is a Ruelle-Lanford function for $\mu_\varepsilon \circ f^{-1}$.*

Proof : Let $\varepsilon > 0$. For each $x \in E$, let $\mathcal{V}_x$ be a small open neighborhood of x, such that $I(\mathcal{V}_x) \geq I(x) - \varepsilon$. Let $y \in F$ such that $J(y) < \infty$; otherwise the lower bound is trivial. By Lemma 4.5, $K = f^{-1}(y) \cap I^{-1}(J(y))$ is a non-empty compact set. Therefore it is possible to find a finite covering $\bigcup_{i=1}^n \mathcal{V}_{x_i}$ of K. By assumption,

$$I(\bigcup_{i=1}^n \mathcal{V}_{x_i}) \geq \min_{i \leq n} I(x_i) - \varepsilon \geq I(K) - \varepsilon.$$

Let $\mathcal{V}'_r$ be a decreasing base of open neighborhoods of y. By continuity of f, when r tends to 0, $f^{-1}(\mathcal{V}_r)$ converges to K. Thus there exists $r > 0$ such that $f^{-1}(\mathcal{V}'_r)$ is included in $\bigcup_{i=1}^n \mathcal{V}_{x_i}$. Then, defining $\underline{J}$ and $\overline{J}$ by (4.4)–(4.5),

$$\underline{J}(y) \geq \lim_{r \to 0} I\big(f^{-1}(\mathcal{V}'_r)\big) \geq I\big(\bigcup_{i=1}^n \mathcal{V}_{x_i}\big) \geq I(x) - \varepsilon \geq J(y) - \varepsilon.$$

Since this is true for all $\varepsilon > 0$, $\underline{J}(y) \geq J(y)$.
Similarly,

$$\begin{aligned}\overline{J}(y) &\leq \lim_{r\to 0} \sup_{K\subset \mathcal{V}'_r} I\big(f^{-1}(K)\big) \\ &\leq \lim_{r\to 0} I\big(f^{-1}(\mathcal{V}'_r)\big) \leq I(x) + \varepsilon \leq J(y) + \varepsilon,\end{aligned}$$

using the lower semi-continuity of I to bound $I\big(f^{-1}(\mathcal{V}'_r)\big)$ for r small enough. This implies $\overline{J}(y) \leq J(y)$. But $\overline{J} \geq \underline{J}$ since $\mu_\varepsilon \circ f^{-1}$ verifies a weak LDP, so that $J = \overline{J} = \underline{J}$ and the proof is completed. ■

Proposition 4.7 (exact l.s.c.) *Assume (H1) and (H2) as in Theorem 4.4 and additionally assume that $F = f(E)$. If $\mathcal{B} \subset E$ is a dense subset of E such that I is the least lower semi-continuous function coinciding with I on $\mathcal{B}$, then $f(\mathcal{B})$ is a dense subset of F such that J is the least lower semi-continuous function coinciding with J on $f(\mathcal{B})$.*

Proof : The set $f(\mathcal{B})$ is dense in $f(E)$ by the continuity of f.
Let $y \in F$ such that $J(y) < \infty$. By Lemma 4.5 there exists $x \in E$ such that $y = f(x)$ and $I(x) = J(y)$. By assumption there exists a sequence x_n converging to x such that $I(x_n)$ converges to $I(x)$. Then

$$\limsup_{n\to\infty} J(f(x_n)) \leq \lim_{n\to\infty} I(x_n) = I(x) = J(y),$$

hence $J(f(x_n))$ converges to $J(y)$ by the lower semi-continuity of J. Now f is continuous, thus $f(x_n)$ converges to $y = f(x)$.
If $J(y) = \infty$, by the lower semi-continuity of J, the limit of $J(y_n)$ is infinite for all y_n converging to y. The proof is completed. ■

Proposition 4.8 (convexity) *Assume (H1) and (H2) as in Theorem 4.4. If f is affine and I is convex then J is convex.*

Proof : The assumption of an affine f implies that E and F are closed convex sets of locally convex topological vector spaces, where barycentration operates.
Let $y_1, y_2 \in F$, $\alpha_1, \alpha_2 \geq 0$ such that $\alpha_1 + \alpha_2 = 1$. Since f is affine, the following sets are equal,

$$f^{-1}(\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 f^{-1}(y_1) + \alpha_2 f^{-1}(y_2),$$

hence

$$\begin{aligned}J(\alpha_1 y_1 + \alpha_2 y_2) &= I(\alpha_1 f^{-1}(y_1) + \alpha_2 f^{-1}(y_2)) \\ &\leq \alpha_1 I(f^{-1}(y_1)) + \alpha_2 I(f^{-1}(y_2)) = \alpha_1 J(y_1) + \alpha_2 J(y_2).\end{aligned}$$

The inequality is true for general convex functions. Hence the convexity of J. ■

4.4 Varadhan's integral lemma

The general result that has to be extended is

Theorem 4.9 *Let $\mu_\varepsilon \subset \mathcal{M}_1(E)$ be a family of probability measures on a Hausdorff topological space E which satisfies a weak LDP with rate function I. Let $g : E \to \mathbb{R}$ be a continuous function. Then, for all open sets O and compact sets K in E,*

$$\liminf_{\varepsilon\to 0} \varepsilon \log \int_O e^{g(x)/\varepsilon} \mu_\varepsilon(\mathrm{d}x) \geq -\inf_{x\in O}\{I(x) - g(x)\}, \tag{4.7}$$

$$\limsup_{\varepsilon\to K} \varepsilon \log \int_K e^{g(x)/\varepsilon} \mu_\varepsilon(\mathrm{d}x) \leq -\inf_{x\in K}\{I(x) - g(x)\}. \tag{4.8}$$

Proof : Since g is finite and continuous, g is locally bounded and the proof is straightforward. ■

To extend this result to the usual Varadhan's integral lemma, one needs severe restriction.

Theorem 4.10 *Let $\mu_\varepsilon \subset \mathcal{M}_1(E)$ be a family of probability measures on a Hausdorff topological space E which satisfies a weak LDP with rate function I. Let $g : E \to \mathbb{R}$ be a continuous function. Assume (H1) with the auxiliary function $\theta : E \mapsto F$ and (H2); assume that $\mu_\varepsilon^g = e^{g/\varepsilon}\mu_\varepsilon$ is exponentially tight in g, then*

$$\lim_{\varepsilon\to 0} \varepsilon \log \int e^{g(x)/\varepsilon} \mu_\varepsilon(\mathrm{d}x) = -\inf_{x\in E}\{I(x) - g(x)\}. \tag{4.9}$$

Proof : By assumptions, for all finite C there exist compact sets $K \subset F$ and $[a,b] \in \mathbb{R}$ such that $K_0 = \theta^{-1}(K) \cap g^{-1}([a,b])$ is compact and

$$\limsup_{\varepsilon\to 0} \varepsilon \log \int_{K_0^c} e^{g(x)/\varepsilon} \mu_\varepsilon(\mathrm{d}x) \leq -C.$$

Applying (4.8), this means that

$$\limsup_{\varepsilon\to 0} \varepsilon \log \int e^{g(x)/\varepsilon} \mu_\varepsilon(\,\mathrm{d}x) \le -\min\Big\{ \inf_{x\in K_0} \{I(x)-g(x)\}, C \Big\}. \tag{4.10}$$

If I is uniformly infinite, (4.10) means that the limit (4.9) is verified and infinite. Otherwise, let $C > \inf_{x\in E}\{I(x)-g(x)\}$. The lower bound (4.7) proves that the infimum is reached on K_0, so that (4.10) implies the upper bound

$$\limsup_{\varepsilon\to 0} \varepsilon \log \int e^{g(x)/\varepsilon} \mu_\varepsilon(\,\mathrm{d}x) \le -\inf_{x\in E} \{I(x)-g(x)\}.$$

Combined with the lower bound (4.7) for $O = E$, this means that (4.9) holds. ■

5 One-dimensional Sanov's theorem

Note that all operations of Section 4 can be performed on the LDP for empirical generators G_t, for example the analog of Varadhan's integral lemma Theorem 4.9 or Theorem 4.10. But we shall further only explore the contraction so as to obtain the one-dimensional Sanov's theorem.

5.1 Contraction

Instantiating Theorem 4.4 for $f(G) = \pi$, one can checks in this case that the compactness assumption ($f^{-1}(K)$ compact for K compact) is *not* generally satisfied, so that the LDP for the one–dimensional empirical measure can not be directly derived. In the remark of page 16, we have shown that the real feature is to ensure the finiteness of the speed of the jumps. However, when the only measured parameter of the process is the empirical measure L_t, it is impossible to guarantee a finite number of jumps in a finite interval of time.

There must therefore be a guarantee that R corresponds to a non-explosive process. Unfortunately it is not easy to give conditions for non-explosion and the lack of structure on E prevents us to treat easily systems with unbounded intensities, like the $M/M/\infty$ queue. A good way for ensuring both the uniqueness of the process and the contraction principle is to bound the intensities. Otherwise the sketch of the proof has to be adapted case by case.

Theorem 5.1 *Let Y_t be an irreducible Markov process with generator R and assume that the intensities R_i are bounded by $\bar{r}$. Then*

$$I_0(\pi\|R) \stackrel{\text{def}}{=} \inf_{G=(\pi,A)} H_c(G\|R), \tag{5.1}$$

is the Ruelle-Lanford function for the empirical measures L_t. Therefore G_t satisfies a weak LDP with the unique, convex rate function I_0 that is the least lower semi-continuous function equal to $I_0(\pi\|R)$ on the finite support irreducible measures $\pi \in M_1(E)$.

Proof : Hereafter f refers to the continuous projection $f(G) = \pi$. The heart of the proof is that $\theta(G) = l = \sum a_{xy}$ is an auxiliary function such that (H1) and (H2) are valid. (H1) has already been proved in Theorem 3.3 under the assumption of bounded intensities. The main difficulty is to prove (H2).
Let $K \subset \mathcal{M}_1(E)$ be compact and $l_0 \in \mathbb{R}_+$. Both $f^{-1}(K)$ and $\theta^{-1}([0, l_0])$ are closed, but none is compact.
Since $K \subset \mathcal{M}_1(E)$ is compact, for all $\eta > 0$ there exists a compact $K_1 \subset E$ such that

$$\pi(K_1) = \sum_{x \in K_1} \pi_x \leq \eta, \qquad \forall \pi \in K.$$

Let $G = (\pi, A)$ be an element of both $f^{-1}(K)$ and $\theta^{-1}([0, l_0])$, and denote by $a(F) = \sum_{x\in F} a_x$.

$$\begin{aligned} H_c(G\|R) &\geq \sum_{x \notin K_1} \pi_x I_p\Big(\frac{a_x}{\pi_x}\Big\| R_x\Big) \\ &\geq \pi(K_1^c) I_p\Big(\frac{a(K_1^c)}{\pi(K_1^c)}\Big\| \sum_{x\in K_1^c} \pi_x R_x\Big) \\ &= I_p\Big(a(K_1^c)\Big\|\pi(K_1^c) \sum_{x \in K_1^c} \pi_x R_x\Big). \end{aligned} \tag{5.2}$$

The second line is a consequence of the convexity of $I_p(.\|.)$ with respect to both variables and the last of its homogeneity. Let $\varepsilon > 0$. Fix $\eta \leq \varepsilon$ and the corresponding K_1 such that

$$I_p(\varepsilon\|\eta^2\bar{r}) > l_0 \text{ and } \varepsilon > \eta^2\bar{r}.$$

The term $\pi(K_1^c)\sum_{x\in K_1^c} \pi_x R_x$ in (5.2) is less than $\eta^2\bar{r}$, so that, since $I_p(x\|y)$ is increasing w.r.t. x and decreasing w.r.t. t when $x > y$,

$$H_c(G\|R) \geq I_p(\varepsilon\|\eta^2\bar{r}) > l_0, \quad \forall \pi \in K,$$

for all G such that $a(K_1^c) > \varepsilon$. This means that, for all $\varepsilon > 0$, there exists a compact set $K_1 \subset E$ such that,

$$G \in f^{-1}(K) \cap \theta^{-1}([0, l_0]) \implies \begin{cases} a(K_1^c) \leq \varepsilon, \\ \pi(K_1^c) \leq \varepsilon, \end{cases}$$

i.e. $f^{-1}(K) \cap \theta^{-1}([0, l_0])$ is tight. Since it is closed, it is compact. (H2) is proved.

Here comes the arsenal developed in Section 4. By Proposition 4.6, $I_0(.\|R)$ is a Ruelle-Lanford function for $L_t = f(G_t)$. By Theorem 4.2 and Theorem 4.3, L_t verifies a weak LDP with unique rate function $I_0(.\|R)$. Since f is linear and $H_c(.\|R)$ is convex, by Proposition 4.8, $I_0(.\|R)$ is also convex.
By Proposition 2.5 and Proposition 4.7, the finite support irreducible measures are a dense subset of $\mathcal{M}_1(E)$ and $I_0(.\|R)$ is the least lower semi-continuous function equal to $I_0(\pi\|R)$ on the finite support irreducible measures $\pi \in \mathcal{M}_1(E)$. It is an easy matter to characterize the finite support irreducible measures as the finite support measures such that $\mathrm{Supp}(\pi)$ is connected in the graph $\mathrm{Supp}(R)$.
Note that, by Lemma 4.5, $I_0(\pi\|R) = H_c(G\|R)$ for a non-empty compact set of generators G. ■

5.2 Identification of the 1-dimensional rate function

Exactly as in the discrete case (see [7]) there are several ways of representing the rate function in Sanov's theorem for the one-dimensional empirical measures

$$L_t \stackrel{\text{def}}{=} \frac{1}{t} \int_0^t \delta_{Y_u} \, \mathrm{d}u.$$

Actually, the formulations depend on the spaces of $E \mapsto \mathbb{R}$ functions usually considered: $C_b(E)$ (bounded continuous functions) and $U(E)$, the space of continuous bounded functions h with $\inf h > 0$. Note that in discrete spaces, all functions are continuous. The other expressions of the rate function are given in terms of transforms of the quantity

$$\Lambda(f) \stackrel{\text{def}}{=} \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E} \left[\exp \int_0^t f(Y_u) \, \mathrm{d}u \right],$$

and of the kernel K^f satisfying

$$K_t^f = P_t + \int_0^t P_{t-s}(f K_s^f) \, \mathrm{d}s,$$

i.e. formally the kernel with generator $L^f = L + f$. The Feynman-Kac formula

$$K_t^f \phi(x) = \mathbb{E}_x \left[\phi(Y_t) \exp\left\{ \int_0^t f(Y_u)\, \mathrm{d}u \right\} \right]$$

connects both quantities.
An almost exhaustive list of the main rate functions is given below.

$$\begin{cases} I_0(\pi) = \inf_{G=(\pi,A)} H_c(G\|R), \\ I_1(\pi) = \sup_{h\in U(E)} -\sum_{x\in E} \pi_x \dfrac{Rh(x)}{h(x)}, \\ I_2(\pi) = \sup_{f\in C_b(E)} \left[\langle f,\pi\rangle - \Lambda(f))\right], \\ I_3(\pi) = \sup_{f\in C_b(E)} \left[\langle f,\pi\rangle + \log \mathcal{R}(K_1^f)\right], \\ I_4(\pi) = \lim_{\varepsilon\to 0} \dfrac{1}{\varepsilon} I_d(\pi\|P^\varepsilon), \end{cases}$$

where $\mathcal{R}$ denotes the convergence parameter ($-\log \mathcal{R}$ is sometimes called logarithmic spectral radius), $P_{xy}^\varepsilon = \mathbb{P}_x[Y_\varepsilon = y]$ the transition matrix at time ε for the process Y_t and I_d the rate function in the discrete time.
Inequalities under very general conditions between the four last expressions are proved in [9]. Actually there are equalities when the process is Feller continuous. However, these expressions are known to be rate functions only for restricted classes of Markov processes. Apart from the classical uniform condition, there is not as much work as in the discrete time for which the identification can be deduced from [11] (see also [4, 5, 10, 19, 20]).

Remark : The problem is to compare these expressions to I_0. It is not possible to use the uniqueness of I_0 as a rate function because the rate functions I_1–I_4 are not proved to be rate functions in a general enough framework. It is hereafter assumed that the intensities are bounded which is the condition of Theorem 5.1 but this restriction can be somewhat relaxed.

Let π be a finite support irreducible measure. Optimizing I_0 and I_1 yields

$$\sum_{y\neq x} \pi_y \frac{R_{yx}h(x)}{h(y)} = \sum_{y\neq x} \pi_x \frac{R_{xy}h(y)}{h(x)}, \tag{5.3}$$

$$\sum_{y\neq x} \pi_y L_{yx} = \sum_{y\neq x} \pi_x L_{xy}, \tag{5.4}$$

which are equivalent since $L_{xy} = R_{xy}h(y)/h(x)$ through Lagrange's multipliers; in fact the multipliers are $\log h(x)$ and correspond to the linear constraints (5.4) for L. Since H_c is strictly convex, since π is irreducible and since the constraints (5.4) are linear (so that the set where H_c is to be minimized is convex), there is a unique solution L_{xy}, from which a unique solution for h is derived, up to a constant. Finally it is easily checked that $I_0 = I_1$, and since I_1—I_4 are equal, all rate functions are equal for the finite support irreducible measures.

Note that the definition of I_4 also works for the generators. Formally, if $P^\varepsilon = I + \varepsilon R + o(\varepsilon)$ and $A^\varepsilon = I + \varepsilon G + o(\varepsilon)$ for $G = (\pi, A)$ then π is invariant for all A^ε and

$$\begin{aligned} H_d(A^\varepsilon \| P^\varepsilon) &= \sum_{x,y\in E} \pi_x A^\varepsilon_{xy} \log \frac{A^\varepsilon_{xy}}{P^\varepsilon_{xy}} \\ &= \sum_{x\in E} \pi_x \left(\sum_{y\neq x} \varepsilon L_{xy} \log \frac{L_{xy}}{R_{xy}} + \log \frac{1-L_x}{1-R_x} \right) + o(\varepsilon) \\ &= \varepsilon H_c(G\|R) + o(\varepsilon). \end{aligned}$$

This calculation is valid for irreducible finite support generators. Indeed, for such a generator there exists a stochastic semi-group of transition corresponding to the generator, and A^ε can be defined similarly to P^ε. Eventually

$$H_c(G\|R) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} H_d(A^\varepsilon \| P^\varepsilon).$$

It is now necessary to prove that I_4 is not too large for the other generators. Let π be a measure and choose a generator G that has the same entropy (by Lemma 4.5). By Proposition 2.5 choose irreducible finite support $G(n)$ converging to G such that their entropy converges to $H_c(G\|R)$. $\pi(n)$ denotes the invariant measure of $G(n)$. As above, for such an irreducible finite support generator $G(n)$ it is possible to define $A^\varepsilon(n)$. This ensures that $\pi(n)$ is the invariant probability of $A^\varepsilon(n)$. Since $I_d(\pi\|P) = \inf_{A:\ A(.,E)=\pi} H_d(A\|P)$,

$$I_d(\pi(n)\|P^\varepsilon) \le H_d(A^\varepsilon(n)\|P^\varepsilon) \le \varepsilon H_c(G(n)\|R) + o(\varepsilon).$$

$H_c(G(n)\|R)$ converges to $H_c(G\|R)$ and $\pi(n)$ to π, so that

$$I_4(\pi) \le \limsup_{n\to\infty} I_4(\pi(n)) \le H_c(G\|R) = I_0(\pi\|R),$$

by the lower semi-continuity of I_4. This proves $I_4 \le I_0$. Now, by the continuity property in Theorem 5.1, I_0 is known to be the least lower semi-continuous function such that $I = I_0$ for all finite support irreducible measures. But on these measures, $I_0 = I_1 = I_4$ so that finally $I_4 = I_0$ for all measures.

♦♦♦

References

[1] Baldi, P., and Piccioni, M. A representation formula for the large deviation rate function for the empirical law of a continuous time Markov chain. *Statist. Probab. Lett. 41*, 2 (1999), 107–115.

[2] Bolthausen, E. Markov process large deviations in τ-topology. *Stochastic Process. Appl. 25*, 1 (1987), 95–108.

[3] Bucklew, J. A., Ney, P., and Sadowsky, J. S. Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Probab. 27*, 1 (1990), 44–59.

[4] de Acosta, A. Upper bounds for large deviations of dependent random vectors. *Z. Wahrsch. Verw. Gebiete 69*, 4 (1985), 551–565.

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

[6] de Acosta, A., and Ney, P. Large deviation lower bounds for arbitrary additive functionals of a Markov chain. *Ann. Probab. 26*, 4 (1998), 1660–1682.

[7] de La Fortelle, A., and Fayolle, G. Large deviation principle for markov chains in discrete time. In *Rapport de recherche*, vol. 3791. INRIA, 1999, p. 34.

[8] Dembo, A., and Zeitouni, O. *Large deviations techniques and applications*, second ed. Springer-Verlag, New York, 1998.

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

[10] Dinwoodie, I. H. Identifying a large deviation rate function. *Ann. Probab. 21*, 1 (1993), 216–231.

[11] Dinwoodie, I. H., and Ney, P. Occupation measures for Markov chains. *J. Theoret. Probab. 8*, 3 (1995), 679–691.

[12] Donsker, M. D., and Varadhan, S. R. S. On some problems of large deviations for Markov processes. In *Proceedings of the 40th Session of the International Statistical Institute (Warsaw, 1975), Vol. 1. Invited papers* (1975), vol. 46, pp. 409–416, 417–418 (1976). With discussion.

[13] 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.

[14] Feller, W. On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc. 48* (1940), 488–515.

[15] Iscoe, I., Ney, P., and Nummelin, E. Large deviations of uniformly recurrent Markov additive processes. *Adv. in Appl. Math. 6*, 4 (1985), 373–412.

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

[17] Kesidis, G., and Walrand, J. Relative entropy between Markov transition rate matrices. *IEEE Trans. Inform. Theory 39*, 3 (1993), 1056–1057.

[18] Lewis, J., and Pfister, C. Thermodynamic probability theory: some aspects of large deviations. *Uspekhi Mat. Nauk 50*, 2(302) (1995), 47–88.

[19] Ney, P., and Nummelin, E. Markov additive processes: large deviations for the continuous time case. In *Probability theory and mathematical statistics, Vol. II (Vilnius, 1985).* VNU Sci. Press, Utrecht, 1987, pp. 377–389.

[20] Nummelin, E. *General irreducible Markov chains and nonnegative operators.* Cambridge University Press, Cambridge, 1984.

[21] Sanov, I. N. On the probability of large deviations of random variables. In *Select. Transl. Math. Statist. and Probability, Vol. 1.* Inst. Math. Statist. and Amer. Math. Soc., Providence, R.I., 1961, pp. 213–244.

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