Electronic Journal of Probability

Vol. 7 (2002) Paper no. 23, pages 1–31.

Journal URL
http://www.math.washington.edu/~ejpecp/
Paper URL
http://www.math.washington.edu/~ejpecp/EjpVol7/paper23.abs.html

SELF-NORMALIZED LARGE DEVIATIONS FOR MARKOV CHAINS

Mathieu FAURE
Université de Marne la Vallée, Université de Marne la Vallée
Bat. Copernic, 5 Bd Descartes, 77454 Champs sur Marne, France.
faure@math.univ-mlv.fr

Abstract: We prove a self-normalized large deviation principle for sums of Banach space valued functions of a Markov chain. Self-normalization applies to situations for which a full large deviation principle is not available. We follow the lead of Dembo and Shao [DemSha98b] who state partial large deviations principles for independent and identically distributed random sequences.

Keywords and phrases: Large deviations, Markov chains, partial large deviation principles, self-normalization.

AMS subject classification (2000): Primary: 60F10, 60J10.

Submitted to EJP on July 10, 2000. Final version accepted on November 13, 2002.

1 Introduction

1.1 From Cramér to Shao

Let $(E, \mathcal{E})$ be a Polish space and $(X_n)_n$ be a sequence of E-valued random variables. For a borel-function $g : E \to \mathbb{R}^d$ and $q > 1$, we introduce

$$S_n(g) \stackrel{\triangle}{=} \sum_{i=1}^{n} g(X_i) \quad \text{and} \quad V_{n,q}(g) \stackrel{\triangle}{=} \left(\sum_{i=1}^{n} \|g(X_i)\|^q \right)^{1/q}.$$

If $(X_n)_n$ is an independant and identically distributed (shortened as i.i.d.) sequence with distribution μ and if $g : E \to \mathbb{R}$ is μ-integrable, the classical Cramér-Chernoff large deviation theorem states that

$$\mathbb{P}\left(\frac{S_n(g)}{n} \geq x \right) \leq e^{-n h_g(x)} \quad \text{for all } x > \overline{g} = \int g d\mu,$$

where h_g is the Cramér transform of the i.i.d. sequence $(g(X_i))_i$. This inequality is useful if $h_g(x) > 0$ for all $x > \overline{g}$, i.e. if the "Cramér condition" is satisfied:

$$\text{there exists } \tau > 0 \text{ such that } \int e^{\tau g} d\mu < \infty.$$

Under this condition, we have

$$\frac{1}{n} \log \mathbb{P}\left(\frac{S_n(g)}{n} \geq x \right) \to -h_g(x).$$

However, this assumption is way too strong in many situations. In [Sha97], Shao shows that it is possible to get rid of this exponential moment assumption taking advantage of self-normalization. He considers for instance the self-normalized sequence

$$R_{n,q}(g) \stackrel{\triangle}{=} \frac{S_n(g)}{n^{1-1/q} V_{n,q}(g)}$$

and obtains the following very interesting result (with $\frac{\overline{g}}{\|g\|_{L^q(\mu)}} = 0$ if $\|g\|_{L^q(\mu)} = \infty$ and $\frac{\overline{g}}{\|g\|_{L^q(\mu)}} = 0$ if $\overline{g} = 0$)

$$\forall x > \frac{\overline{g}}{\|g\|_{L^q(\mu)}}, \quad \lim_n \frac{1}{n} \log \mathbb{P}\left(R_{n,q}(g) \geq x \right) \stackrel{\triangle}{=} -K(x) < 0,$$

without any moment assumption on the random variable $(g(X_i))_i$.

In his paper, Shao gives various applications, in particular for the Student statistic:

If $\overline{g} \geq 0$, we note $T_n \stackrel{\triangle}{=} \frac{\sum_{i=1}^n g(X_i)}{n^{1/2}\left(\frac{1}{n-1}\sum_{i=1}^n (g(X_i)-S_n(g))^2\right)^{1/2}}$. For all x such that $x > \frac{\overline{g}}{\|g\|_{L^2(\mu)}}$, we have

$$\lim_n \frac{1}{n} \log \mathbb{P}\left(T_n \geq n^{1/2}x\right) = \lim_n \frac{1}{n} \log \mathbb{P}\left(\frac{S_n(g)}{V_{n,2}(g)} \geq n^{1/2}x\left(\frac{n}{n+nx^2-1}\right)^{1/2}\right) = -K(x) < 0.$$

In this work, we consider the same problematic in the Markovian framework and obtain analog results in section 2.5. (corollary 2).

1.2 Full and partial large deviation principles

Introducing the notion of partial large deviation principle in two articles [DemSha98a] and [DemSha98b], Dembo and Shao give a more general sense to Shao's paper [Sha97] and lighten the tools used to obtain these results.

To help comprehension, we recall the basic vocabulary in large deviation theory. Let E be a metric topological space equiped with its Borel σ-field $\mathcal{E}$. A function $I : E \to [0, \infty]$ is a *good rate function* if its level sets $\{x;\, I(x) \leq t\}$ are compact and it is a *weak rate function* if its level sets are closed (namely, if I is lower semi-continuous, shortened to l.s.c. in the sequel). A sequence of probability measures $(\mu_n)_n$ on $(E, \mathcal{E})$ satisfies a *large deviation principle* (shortened to LDP) with a good rate function I if, for every open subset G and every closed subset F of E,

$$-\inf_{x\in G} I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G), \tag{1}$$

$$-\inf_{x\in F} I(x) \geq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F). \tag{2}$$

We say that the sequence $(\mu_n)_n$ satisfies an upper LDP (resp. a lower LDP) if (2) only (resp. (1) only) holds. Moreover, a weak LDP is said to hold if (2) is satisfied for the compact sets of E only and if I is a weak rate function.

The concept of *partial large deviation principle* (PLDP) has been introduced by Dembo and Shao in [DemSha98a and b] : the sequence $(\mu_n)_n$ satisfies an upper PLDP with weak rate I with respect to a subclass $\mathcal{S}$ of $\mathcal{E}$ if, for every $A \in \mathcal{S}$, we have :

$$-\inf_{x\in \overline{A}} I(x) \geq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(A). \tag{3}$$

The full PLDP is said to hold if (1) is satisfied as well for every open $G \subset E$.

1.3 Plan of the paper

In section 2, we give our main results. A weak large deviation principle for "balanced couples" is stated in section 3 as a preliminary to obtain the main Theorem (the same way as in the i.i.d.

case where the weak Cramér Theorem is the first step to prove self-normalized results). We give some commentaries along with examples in section 4. The proofs of the results are given in sections 5 and 6 : section 5 deals with the weak large deviation principle while section 6 provides partial exponential tightness which is the key to obtain partial large deviation Theorem. At last, section 7 brings some precisions about upper weak large deviations (Theorem 2).

2 Main results, partial LDP

We consider a Markov chain $X = (X_i)_{i\in\mathbb{N}}$ taking values in a Polish space E endowed with its Borel σ-field $\mathcal{E}$. Its transition kernel is denoted by p, $C_b(E)$ is the space of real bounded continuous functions on E and $\mathcal{P}(E)$ the space of probability measures on E equiped with the topology of weak convergence. If ζ belongs to $\mathcal{P}(E^2)$, we denote by ζ^1 and ζ^2 the first and second marginal. If $\xi \in \mathcal{P}(E)$ and $\Gamma \in \mathcal{E}\otimes\mathcal{E}$, then $\xi p(\cdot) = \int \xi(dx)p(x,\cdot)$, $p(f)(x) = \int p(x,dy)f(y)$ and $\xi\otimes p(\Gamma) = \int \mathbb{I}_\Gamma(x,y)\xi(dx)p(x,dy)$.

We work with the canonical form of the Markov chain $(E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}}, (\mathbb{P}_x)_{x\in E}, (X_n)_{n\geq 0})$ and the following notation : for any initial distribution ν, $\mathbb{P}_\nu = \int \nu(dx)\mathbb{P}_x$.

2.1 Assumptions on the Markov chain

These are the assumptions we might have use of in the following, the third one being useless for upper LDP results.

The upper bounds stated in this section require a regularity assumption concerning the Markov chain. Let us recall the classical Feller property and the "almost Fellerian" extension proposed by Dupuis and Ellis [DupEll] and related to a condition introduced by J.G. Attali ([Att]) :

Assumption 1 (Fellerian or almost Fellerian transition).

- *The transition p satisfies the Feller property (or is "Fellerian") if the map : $x \mapsto p(x,\cdot)$ is continuous for the weak convergence topology of $\mathcal{P}(E)$.*
- *More generally, denoting $\mathcal{D}(p)$ the discontinuity set of $x \to p(x,\cdot)$, p is "almost Fellerian" if, for every $x \in E$ and all $\delta > 0$, there exist an open set G_δ of E containing $\mathcal{D}(p)$ and a real number $r(x) > 0$ such that for any $y \in E$, $d(x,y) \leq r(x) \Longrightarrow p(y, G_\delta) \leq \delta$. In particular, for all $x \in E$, $p(x, \mathcal{D}(p)) = 0$.*

We recall that a Lyapunov function is a measurable, non negative borel-function whose level sets are relatively compact. The existence of an invariant probability measure μ for p is guaranteed by the following "stabilization" condition:

$$\text{there exists a Lyapunov function } V,\ a < 1 \text{ and } b \in \mathbb{R} \text{ such that } pV \leq aV + b.$$

If this invariant probability μ is unique, $(X_n)_n$ is μ-stable (almost surely, $L_n \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ converges weakly to μ) and we have the law of large numbers: if $g : E \to \mathbb{R}^d$ is continuous and such that $\|g\| \leq \rho(U)$ with $\frac{\rho(t)}{t} \to 0$ when $t \to \infty$,

$$\frac{S_n(g)}{n} \to \overline{g} = \int g d\mu \text{ almost surely.}$$

The following assumption is a stronger version. It was introduced by Donsker-Varadhan (condition H^* in [DonVar76]) in order to obtain an upper LDP for the empirical distributions of a Markov chain. Our version is taken from Dupuis-Ellis ([DupEll] chap 8) :

Assumption 2 (Criterion of exponential stabilization associated with (U, V)). *There exists a borel-function $U : E \to \mathbb{R}_+$, a Lyapunov function V and a non-negative constant C such that :*

$$V \leq U - \log p(e^U) + C.$$

Remark 1.

a) *Under the assumptions 1 and 2, p always has an invariant probability measure (see Proposition 9.2.6 in [DupEll]). This probability is unique if p is irreducible.*

b) *Assumptions 1 and 2 do not imply irreducibility as we can see considering the example*

$$X_{n+1} = \frac{1}{2}X_n + \varepsilon_{n+1}, \ ((\varepsilon_n)_n \text{ i.i.d. and } \mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = 1/2).$$

c) *U and $W = e^U$ are Lyapunov functions.*

d) *If $E = \mathbb{R}^d$, the stabilization condition is equivalent to $\limsup_{\|x\| \to \infty} \frac{pV(x)}{V(x)} < 1$, while exponential stabilization is equivalent to $\limsup_{\|x\| \to \infty} \frac{pW(x)}{W(x)} = 0$.*

Assumption 3 (Strong irreducibility). *p satisfies the following two conditions :*

1) There exists an integer L such that, for every $(x, y) \in E^2$,

$$\sum_{i=L}^{\infty} \frac{1}{2^i} p^i(x, \cdot) << \sum_{i=1}^{\infty} \frac{1}{2^i} p^i(y, \cdot).$$

2) p has an invariant probability measure μ.

Remark 2. *Assumptions 1,2 and 3 are always satisfied in the i.i.d. case.*

2.2 Particular case of the functional autoregressive models

Let us illustrate our asumptions considering the following model taking values in $\mathbb{R}^d$:

$$X_{n+1} = f(X_n) + \sigma(X_n)\varepsilon_{n+1}.$$

* $(\varepsilon_n)_n$ is an i.i.d. sequence, independent of X_0.
* $f : \mathbb{R}^d \to \mathbb{R}^d$ is a measurable function, bounded on every compact subset of $\mathbb{R}^d$.
* $\sigma : \mathbb{R}^d \to \left[\frac{1}{r}, r\right]$ $(r \geq 1)$ is a measurable function.

We do not know many large deviations results for such models. We can mention the LDP for the unidimensional linear autoregressive model with gaussian noise (see [BryDem97], [BerGamRou96]). There also exists a moderate large deviation result for the multidimensional linear autoregressive model or for the kernel estimator with a generalized gaussian noise (see [Wor99]). The study of such models with weaker assumptions for the noise is one of the motivations of self-normalization (following [Sha97]). Let us consider the conditions imposed by the assumptions stated in section 2.1 for this particular model:

- This Markov chain is Fellerian if f and σ are continuous; it is almost Fellerian if f and σ are Lebesgue almost everywhere continuous and ε_1 has a bounded density with respect to the Lebesgue measure on $\mathbb{R}^d$. The almost fellerian assumption allows to study important models in econometry such as threshold autoregressive models for which function f is defined as follows:

  $$f = \sum_{i=1}^{K} f_i \, \mathbb{I}_{\Gamma_i}$$

  with f_i continuous, Γ_i disjoined with boundaries of Lebesgue measure zero (and $\cup\Gamma_i = \mathbb{R}^d$).
- Exponential stabilization is checked for the model if there exists a positive borel-function U such that

  $$\lim_{\|x\|\to\infty} \left(U(x) - \log \mathbb{E}\left(\exp\left(U(f(x) + \sigma(x)\varepsilon_1)\right)\right)\right) = \infty.$$

  Now, for $\rho > 0$,

  $$\log \mathbb{E}\left(\exp\left(\rho\|f(x) + \sigma(x)\varepsilon_1\|\right)\right) \leq \rho\|f(x)\| + \log \mathbb{E}\left(\exp\left(\rho r\|\varepsilon_1\|\right)\right).$$

  Also, if $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing function, for any $s \in]0,1[$, we have

  $$\phi\left(\|f\| + r\|\varepsilon_1\|\right) \leq \max\left\{\phi(s^{-1}\|f\|), \phi((1-s)^{-1}r\|\varepsilon_1\|)\right\} \leq \phi(s^{-1}\|f\|) + \phi((1-s)^{-1}r\|\varepsilon_1\|).$$

  Therefore,

  $$\log \mathbb{E}\left(\exp\left(\phi(\|f(x) + \sigma(x)\varepsilon_1\|)\right)\right) \leq \phi(s^{-1}\|f(x)\|) + \log \mathbb{E}\left(\exp\left(\phi((1-s)^{-1}r\|\varepsilon_1\|)\right)\right).$$

  Hence, we can take $U(x) = \phi(\|x\|)$ under various assumptions. Of course, a less constraining condition on noise will lead to a more restrictive condition on the function f.

a) Under Cramér condition $\mathbb{E}\left(\exp(\tau\|\varepsilon_1\|)\right) < \infty, (\tau > 0)$, we can take $\phi(t) = \frac{\tau}{r}t$ and the condition on f is

$$\lim_{\|x\|\to\infty} (\|x\| - \|f(x)\|) = \infty.$$

b) Under condition $\mathbb{E}\left(\exp(\tau\|\varepsilon_1\|^\beta)\right) < \infty\,(\tau > 0, \beta > 1)$, we can take, for any $s \in]0,1[$, $\phi(t) = \frac{\tau(1-s)^\beta}{r^\beta}t^\beta$ and the condition on f is

$$\lim_{\|x\|\to\infty} \left(\|x\|^\beta - s^{-\beta}\|f(x)\|^\beta\right) = \infty.$$

c) If we only assume that the noise is integrable $\mathbb{E}(\|\varepsilon_1\|) < \infty$, then $\phi(t) = \log(t)$ and the condition on f is

$$\lim_{\|x\|\to\infty} \frac{\|x\|}{\|f(x)\|} = \infty.$$

- At last, strong irreducibility is satisfied as soon as ε_1 has a strictly positive density with respect to Lebesgue measure.

2.3 Donsker-Varadhan rate functions

According to Donsker-Varadhan theory ([DonVar75a] and [DonVar76]), we introduce the relative entropy between two probability measures μ and ν on E :

$$K(\nu \mid \mu) \stackrel{\triangle}{=} \begin{cases} \int \nu(dx)\log\left(\frac{d\nu}{d\mu}(x)\right) & \text{if } \nu << \mu \\ \infty & \text{otherwise} \end{cases} = \sup_{f\in C_b(E)} \left\{ \int f d\nu - \log\int e^f d\mu \right\};$$

and the rates related with the LDP concerning the empirical distributions $L_n \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^n \delta_{X_{i-1},X_i}$ and $L_n^1 \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^n \delta_{X_{i-1}}$ are, if $\zeta \in \mathcal{P}(E^2)$ and $\xi \in \mathcal{P}(E)$,

$$I(\zeta) \stackrel{\triangle}{=} \begin{cases} K(\zeta \mid \zeta^1 \otimes p) = \sup_{f\in C_b(E^2)} \left\{ \int \log\left(\frac{e^f}{p(e^f)}\right) d\zeta \right\} & \text{if } \zeta^1 = \zeta^2 \\ \infty & \text{otherwise} \end{cases}$$

$$I^1(\xi) \stackrel{\triangle}{=} \inf\left\{K\left(\zeta \mid \zeta^1 \otimes p\right),\ \zeta^1 = \zeta^2 = \xi\right\} = \sup_{f\in C_b(E)} \left\{ \int \log\left(\frac{e^f}{p(e^f)}\right) d\xi \right\}.$$

Furthermore, we assume that $(B,\mathcal{B})$ is a separable Banach space endowed with its Borel σ-field. For a B-valued measurable function g on E (resp. G on E^2), we set, for $x \in B$:

$$h_g^0(x) = \inf\left\{ I^1(\xi) \mid \xi \in \mathcal{P}(E), \int \|g\| d\xi < \infty,\ \int g d\xi = x \right\}, \tag{4}$$

$$h_G^0(x) = \inf\left\{ I(\zeta) \mid \zeta \in \mathcal{P}(E^2), \int \|G\| d\zeta < \infty,\ \int G d\zeta = x \right\}. \tag{5}$$

These functions are convex (this statement is proved in Paragraph 5.1) but might not be l.s.c.. Hence, we consider the corresponding l.s.c.-normalized functions h_g and h_G, with

$$h(x) \overset{\triangle}{=} \sup\left\{h^0(\Gamma) \mid \Gamma \text{ neighbourhood of } x\right\} \leq h^0(x).$$

These functions are l.s.c. and convex. We call them "Donsker-Varadhan rates". Finally, the following notations are constantly used in the sequel:

$$\overline{g}_n \overset{\triangle}{=} \frac{1}{n}\sum_{i=1}^{n} g(X_i) \quad \text{and} \quad \overline{G}_n \overset{\triangle}{=} \frac{1}{n}\sum_{i=1}^{n} G(X_{i-1}, X_i).$$

For a function $h : H \to \mathbb{R}_+$ and a subset Γ of H, we note $h(\Gamma) = \inf_{x\in\Gamma} h(x)$.

Remark 3.

- *If $g : E \to B$, the study of the rate h_g and the sequence $\overline{g}_n$ is a particular case of the one involving h_G and $\overline{G}_n$ using the function $G(x, y) \overset{\triangle}{=} g(y)$ (it is not difficult to show that, in this case, $h_g = h_G$). Hence, we shall only work with functions G defined on E^2.*
- *If Γ is an open set, then $h(\Gamma) = h^0(\Gamma)$.*

2.4 Previous LDP for vector valued functionals

With the Donsker-Varadhan rate h_G, we have the following theorem (see [GulLipLot94]) which generalizes known results for bounded functions.

Theorem (Full upper LDP ([GulLipLot94])

Let $G : E^2 \to \mathbb{R}^d$ be a continuous function.

1) *Assume assumptions 1 and 2 are satisfied. Let $V \oplus V$ be defined by $V \oplus V(x, y) \overset{\triangle}{=} V(x) + V(y)$ and*

$$\rho(\|G\|) \leq V \oplus V, \tag{6}$$

where $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ is a function such that $\frac{\rho(t)}{t} \to \infty$ when $t \to \infty$. Then :

- *$h_G = h_G^0$ is a good rate function,*
- *for any compact subset H of $H_M = \{\nu \in \mathcal{P}(E);\ \int U d\nu \leq M\}$ and every closed subset F of $\mathbb{R}^d$*

$$\limsup_n \sup_{\nu\in H} \frac{1}{n} \log \mathbb{P}_\nu(\overline{G}_n \in F) \leq -h_G(F).$$

2) *If assumption 3 is also satisfied, the full LDP with rate h_G is valid.*

As an example, the case studied in section 2.2 (Functional autoregressive models) proves that the domination condition (6) is not easily checked. In section 2.5., we give self-normalized large deviation principles which would be obvious under the assumption (6), as well as [Sha97] and [DemSha98b] who handled self-normalization in the i.i.d. case to get rid of Cramér's condition.

2.5 Partial Large Deviation Principle

We now state our main results extending Dembo-Shao's work ([DemSha98b]) to our Markovian framework.

Theorem 1 (Self-normalized LDP). *Assume that the transition probability p is almost Fellerian and satisfies the criterion of exponential stabilization associated with (U,V). Let $G = (G^{(1)}, G^{(2)}) : E^2 \to \mathbb{R}_+ \times B$ be a continuous function such that $G^{(1)}$ is a Lyapunov function and $G^{(1)} = \rho \circ N(G^{(2)})$ where $N : B \to \mathbb{R}_+$ and $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ are convex continuous functions such that*

- *N has compact level sets*
- *ρ is strictly increasing and $\frac{\rho(st)}{s\rho(t)} \to \infty$ when $\|s\| \to \infty$ and $\|t\| \to \infty$.*

a) Let $h_G^* : \mathbb{R}_+ \times B \to \mathbb{R}_+$ be defined by $h_G^*(x^{(1)}, x^{(2)}) \triangleq h_G([0, x^{(1)}] \times \{x^{(2)}\})$. Then, h_G^* is a convex weak rate function and $((\overline{G}_n^{(1)}, \overline{G}_n^{(2)}))_n$ satisfies an upper PLDP with rate function h_G^* with respect to the sub-class $\mathcal{S}(\rho \circ N) \triangleq \left\{A \in \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B};\ \lim_{x^{(1)} \to \infty} \inf_{(x^{(1)}, x^{(2)}) \in A} \frac{\rho(N(x^{(2)}))}{x^{(1)}} > 0\right\}$.

More precisely, if $H_M \triangleq \left\{\nu \in \mathcal{P}(E);\ \int U d\nu \le M\right\}$, then for any $\Gamma \in \mathcal{S}(\rho \circ N)$ and any compact subset H of H_M,

$$\limsup_{n\to\infty} \sup_{\nu \in H} \frac{1}{n} \log \mathbb{P}_\nu \left((\overline{G}_n^{(1)}, \overline{G}_n^{(2)}) \in \Gamma \right) \le -h_G^*(\overline{\Gamma}).$$

b) Let $\Gamma_r \triangleq \left\{(x^{(1)}, x^{(2)}) \in \mathbb{R}_+ \times B;\ \frac{\rho(N(x^{(2)}))}{x^{(1)}} \ge r\right\}$ and $J(r) \triangleq h_G(\Gamma_r)$. J is increasing, left-continuous and, for every $r > 0$,

- $*$ *$h_G(\Gamma_r) = h_G^*(\Gamma_r)$,*
- $*$ *$h_G(\Gamma_r) = 0$ if and only if there exists an invariant probability measure μ for p such that $G^{(1)}$ is integrable with respect to $\mu \otimes p$ and we have $r \le \left(\rho(N(\int G^{(2)} d(\mu \otimes p)))/\int G^{(1)} d(\mu \otimes p)\right)$.*

In particular, the following Chernoff-like upper bound holds for every compact subset H of H_M :

$$\forall r > 0,\ \limsup_{n\to\infty} \sup_{\nu \in H} \frac{1}{n} \log \mathbb{P}_\nu \left(\frac{\rho(N(\overline{G}_n^{(2)}))}{\overline{G}_n^{(1)}} \ge r \right) \le -h_G(\Gamma_r) = -J(r)$$

and the sequence $(\frac{\rho \circ N(\overline{G}_n^{(2)})}{\overline{G}_n^{(1)}})$ satisfies an upper LDP with rate $J(\cdot)$ on $\mathbb{R}_+$.

We obtain interesting Corollaries :

Corollary 1. *Assume that the assumptions and notations of Theorem 1 hold. In addition, we suppose that the chain satisfies the strong irreducibility hypothesis with invariant probability measure μ and $G^{(1)}$ is integrable with respect to $\mu \otimes p$. Then, for any initial distribution ν, the full partial large deviation principle is valid and we have*

$$-J(r_+) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu \left(\frac{\rho(N(\overline{G}_n^{(2)}))}{\overline{G}_n^{(1)}} > r \right) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu \left(\frac{\rho(N(\overline{G}_n^{(2)}))}{\overline{G}_n^{(1)}} \geq r \right) \leq -J(r).$$

Finally, we give the following more explicit Corollary, applying Theorem 1 to a function $G = (\|F\|^q, F)$. For $q > 1$, we introduce the notation $J_q(r) \stackrel{\triangle}{=} J(r^q)$.

Corollary 2. *Let F be a continuous function from E^2 into $\mathbb{R}^d$. Assume that the transition probability p is almost Fellerian and satisfies the criterion of exponential stabilization associated with (U, V). Then, for any given $q > 1$ and any compact subset H of $H_M = \left\{\nu \in \mathcal{P}(E);\ \int U d\nu \leq M\right\}$, we have*

a) For every $r > 0$, $\limsup_{n\to\infty} \sup_{\nu\in H} \frac{1}{n} \log \mathbb{P}_\nu \left(\frac{\|\sum_{i=1}^n F(X_{i-1},X_i)\|}{n^{1-1/q}(\sum_{i=1}^n \|F(X_{i-1},X_i\|^q)^{1/q}} \geq r \right) \leq -J_q(r)$.

b) $J_q(r) = 0$ if and only if there exists an invariant probability measure μ for p such that

* *F is integrable with respect to $\mu \otimes p$*
* *$r \leq \frac{\|\int F d\mu\otimes p\|}{(\int \|F\|^q d\mu\otimes p)^{1/q}}$.*

$\left(\frac{\|\sum_{i=1}^n F(X_{i-1},X_i)\|}{n^{1-1/q}(\sum_{i=1}^n \|F(X_{i-1},X_i\|^q)^{1/q}} \right)_n$ satisfies an upper LDP with rate function J_q on $\mathbb{R}_+$.

c) If, in addition, the chain satisfies the strong irreducibility hypothesis with the invariant probability measure μ and if $\|F\|^q$ is integrable with respect to $\mu \otimes p$ then, $J_q(r) > 0$ if $\|\int F d\mu \otimes p\| \leq r(\int \|F\|^q d\mu \otimes p)^{1/q}$ and, for any initial distribution ν,

$$-J_q(r_+) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu \left(\frac{\|\sum_{i=1}^n F(X_{i-1}, X_i)\|}{n^{1-1/q}(\sum_{i=1}^n \|F(X_{i-1}, X_i\|^q)^{1/q}} > r \right),$$
$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu \left(\frac{\|\sum_{i=1}^n F(X_{i-1}, X_i)\|}{n^{1-1/q}(\sum_{i=1}^n \|F(X_{i-1}, X_i\|^q)^{1/q}} \geq r \right) \leq -J_q(r)$$

d) If $d = 1$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P} \left(\frac{\sum_{i=1}^n F(X_{i-1}, X_i)}{n^{1-1/q}(\sum_{i=1}^n \|F(X_{i-1}, X_i\|^q)^{1/q}} \geq r \right) \leq -K(r)$$

with $K(r) = 0$ if and only if $\|F\|^q$ is integrable with respect to $\mu \otimes p$ and $\int F d\mu \otimes p \geq r(\int \|F\|^q d\mu \otimes p)^{1/q}$.

Remark 4.

a) If the function U is bounded above on the compact sets of E, the results hold uniformly over the initial states $x \in K$ for any compact subset K of E.

b) If the function U is l.s.c., then H_M is a compact subset of $\mathcal{P}(E)$ and the results hold uniformly over H_M.

2.6 Tests on Markovian models

The results stated in section 2.5 (more particularly corollary 2) are obviously interesting, as in the i.i.d. case, to obtain exponential speed to the law of large numbers. For example, the large deviation upper bounds allow to reproduce Shao's results on Student statistic stated in 1.1. and build tests for assumptions such as "the random variable observed is a Markov chain with transition p", with exponentially decreasing levels. Let us be more specific for a test between two assumptions.

We consider two transition probabilities $(p_i(x,\cdot))(i=0,1)$ on $(E,\mathcal{E})$ satisfying assumptions 1 and 2. Let (μ_i) $(i=0,1)$ be the unique invariant probability measures associated with p_i and let us assume that there exists a measurable, strictly positive function h such that, for any $x \in E$, $p_1(x,dy) = h(x,y)p_0(x,dy)$ (with $\log h$ μ_i-integrable). The distribution of $(X_0, X_1, ..., X_n)$ under $\mathbb{P}^1_\nu$ has density f with respect to its distribution under $\mathbb{P}^0_\nu$, with:

$$(x_0, x_1, ..., x_n) \overset{f}{\mapsto} \prod_{i=1}^n h(x_{i-1}, x_i),$$

$$\int \log h d\mu_1 \otimes p_1 = \int \mu_1(dx) \int \log h(x,y) p_1(x,dy) = \int \mu_1(dx) K\left(p_1(x,\cdot) \mid p_0(x,\cdot)\right) = L(p_1 \mid p_0) \geq 0,$$

$$\int \log \frac{1}{h} d\mu_0 \otimes p_0 = \int \mu_0(dx) K\left(p_0(x,\cdot) \mid p_1(x,\cdot)\right) = L(p_0 \mid p_1) \geq 0.$$

A natural test "p_0 against p_1" will have the rejection region

$$\mathcal{R}_n = \left\{ \sum_{i=1}^n \log h(X_{i-1}, X_i) \geq t \left(n \sum_{i=1}^n \left(\log h(X_{i-1}, X_i)\right)^2 \right)^{1/2} \right\}.$$

The errors are

$$\alpha_n(\nu) = \mathbb{P}^0_\nu\left(\mathcal{R}_n\right) = \mathbb{P}^0_\nu\left(\frac{\sum_{i=1}^n \log h(X_{i-1}, X_i)}{n^{1/2}\left(\sum_{i=1}^n \left(\log h(X_{i-1}, X_i)\right)^2\right)^{1/2}} \geq t \right),$$

$$\beta_n(\nu) = \mathbb{P}^1_\nu\left(\mathcal{R}_n^c\right) = \mathbb{P}^1_\nu\left(\frac{-\sum_{i=1}^n \log h(X_{i-1}, X_i)}{n^{1/2}\left(\sum_{i=1}^n \left(\log h(X_{i-1}, X_i)\right)^2\right)^{1/2}} \geq -t \right).$$

Part d) of corollary 2 leads to the following result. Let us assume that $L(p_1 \mid p_0) > 0$ or $L(p_0 \mid p_1) > 0$: the models are distinguishable. Then, for any t such that

$$\frac{-L(p_0 \mid p_1)}{\left(\int (\log h)^2 \, d\mu_0 \otimes p_0\right)^{1/2}} < t < \frac{L\,(p_1 \mid p_0)}{\left(\int (\log h)^2 \, d\mu_1 \otimes p_1\right)^{1/2}},$$

we have the upper bounds:

$$\limsup_n \frac{1}{n} \log \alpha_n(\nu) < 0,$$
$$\limsup_n \frac{1}{n} \log \beta_n(\nu) < 0.$$

For another application of the self-normalized large deviation, one can look [HeSha96].

3 Weak LDP for vector valued functions with Donsker-Varadhan rates

3.1 Known results concerning empirical distributions

Several upper LDP for the empirical distributions $L_n \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^n \delta_{X_{i-1},X_i}$ or $L_n^1 = \frac{1}{n}\sum_{i=1}^n \delta_{X_{i-1}}$ have been proved by Donsker-Varadhan [DonVar75a], De Acosta [Aco85], [Aco90] and Dupuis-Ellis [DupEll]. The statement that follows is a synthesis of the results we need in our proofs :

Theorem (Donsker-Varadhan, Dupuis-Ellis, De Acosta)
Under Assumptions 1 and 2, I is rate function. For any compact subset H of H_M, we have

$$\limsup_{n\to\infty} \sup_{\nu \in H} \frac{1}{n} \log \mathbb{P}_\nu(L_n \in F) \le -I(F), \ \ \textit{for any closed set } F \tag{7}$$

3.2 About Donsker-Varadhan rate functions

The function $G = (G^{(1)}, G^{(2)})$ considered in Theorem 1 is a particular case of the following class of functions that we will call "balanced couples":

Definition 1. *$G = (G^{(1)}, G^{(2)})$ is a balanced couple on $(E^2, \mathcal{E} \otimes \mathcal{E})$ when*

- *$G^{(1)} : E^2 \to [0, \infty]$ is l.s.c.,*
- *$G^{(2)}$ is continuous from E^2 into a separable Banach space B and, for a continuous function $N : B \to [0, \infty]$ with compact level sets, $\rho \circ N(g^{(2)}) \le g^{(1)}$ where $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and such that $\frac{\rho(t)}{t} \to \infty$.*

Besides, if the function $G^{(1)}$ has compact level sets (i.e. if $G^{(1)}$ is a Lyapunov function), then the couple $(G^{(1)}, G^{(2)})$ will be called "Lyapunov balanced couple".

The following lemma will be proved in section 5.1.

Lemma 1 (Properties of Donsker-Varadhan rate functions).

1) *For any function $G : E^2 \to B$, h_G^0 defined in formula (5) is a convex function. Hence, its l.s.c.-normalized function h_G is a convex weak rate function.*

2) *If G is a Lyapunov balanced couple then h_G^* defined in Theorem 1 is a convex and l.s.c. function.*

Remark 5. *For any $(x^{(1)}, x^{(2)}) \in \mathbb{R}_+ \times B$, $h_G([0, x^{(1)}] \times \{x^{(2)}\}) = h_G^0([0, x^{(1)}] \times \{x^{(2)}\})$.*

3.3 Upper weak LDP for balanced couples

Theorem 2 (Upper weak LDP for balanced couples).
Assume that p is an almost Fellerian transition on $(E, \mathcal{E})$ that satisfies the criterion of exponential stabilization with (U, V). If $G = (G^{(1)}, G^{(2)})$ is a Lyapunov balanced couple on E^2, then $((\overline{G}_n^{(1)}, \overline{G}_n^{(2)}))_n$ satisfies uniform an upper weak LDP for every initial distribution with weak rate function $h_G^*(\cdot)$.
In other words, for any compact subset K of $\mathbb{R}_+ \times B$ and any compact subset H of H_M,

$$\limsup_{n \to \infty} \frac{1}{n} \sup_{\nu \in H} \log \mathbb{P}_\nu \left((\overline{G}_n^{(1)}, \overline{G}_n^{(2)}) \in K \right) \leq -h_G^*(K).$$

In particular, if $R \in]0, \infty[$ and if C is a compact set in B then

$$\limsup_{n \to \infty} \frac{1}{n} \sup_{\nu \in H} \log \mathbb{P}_\nu \left(\overline{G}_n^{(1)} \leq R, \overline{G}_n^{(2)} \in C \right) \leq -h_G([0, R] \times C).$$

3.4 Lower LDP

A general lower LDP relative to the sums of Banach space valued additive functionals of a Markov chain has been proved by De Acosta and Ney ([AcoNey98]) with no other assumptions that the irreducibility of the chain and the measurability of the function. Yet, it seems difficult to compare the "spectral rate" for which their lower LDP holds with h_G.

Our demonstration relies on the dynamic programming method developped by Dupuis and Ellis ([DupEll]) for proving the lower LDP which needs a stronger assumption than standard irreducibility (Condition 8.4.1. in [DupEll]). Therefore, we achieve a less general result than that of De Acosta and Ney but it holds with the rate h_G as well as the upper LDP.

The following Theorem requires strong irreducibility but no assumption about the regularity of p or G.

Theorem 3. *If p fulfills the strong irreducibility assumption and if $G : E^2 \to B$ is measurable, integrable with respect to $\mu \otimes p$, then, for every initial distribution ν, the sequence $(\overline{G}_n)_n$ satisfies for any open set U of B*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu \left(\overline{G}_n \in U\right) \geq -h_G(U).$$

4 Comments and examples

4.1 Cramér and Donsker-Varadhan for i.i.d. random vectors

We consider a Polish space E and an i.i.d. E-valued random sequence $(X_n)_n$ with distribution μ.

- If $g : E \to B$ is a measurable function (where B is a separable Banach space), the sequence $(g(X_n))_n$ is i.i.d., B-valued, and $(\overline{g}_n)_n$ satisfies a weak-convex LDP with the Cramér rate :
$$h_g^{Cramer}(x) = \sup_{\theta\in B^*} \left\{ < \theta, x > - \log \int e^{<\theta, g(y)>} \mu(dy) \right\}$$
(where B^* denotes the topological dual of B). This result is due to Bahadur and Zabell [BahZab79] (see also [DemZei] and [Aze]).

  Under the Cramér condition $\mathbb{E}(\exp(t\|g(X)\|)) < \infty$ (for any t if B is a separable Banach space and for at least a $t > 0$ in the particular situation $B = \mathbb{R}^d$), h_g^{Cramer} is a good rate function and the full LDP holds (see [DonVar76] or [Aco88]).

- On the other hand, we are in our context with $p(x, \cdot) = \mu(\cdot)\ \forall x \in E$. There always exists a Lyapunov function V such that $\int e^V d\mu < \infty$. Hence, the criterion of exponential stabilization associated with (V, V) is satisfied. The strong irreducibility hypothesis is satisfied and Theorem 3 holds for any measurable function $g : E \to B$, integrable with respect to μ.

The convex large deviation principle allows us to write for any $\nu \in \mathcal{P}(E)$, any $x \in B$ and $B(x, r) \stackrel{\triangle}{=} \{y \in B;\ d_B(x, y) < r\}$

$$-h_g(x) \leq -\lim_{r\to 0} h_g(B(x, r)) \leq \lim_{r\to 0}\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu \left[\overline{g}_n \in B(x, r)\right] = -h_g^{Cramer}(x).$$

Hence, $h_g \geq h_g^{Cramer}$ and all our upper bound results stated in section 2 involving rate h_g are still valid with h_g^{Cramer} (without assuming Cramér condition). The lower bound results obtained in Theorem 1 and in corollaries 1 and 2 hold with the rate h_g^{Cramer} according to the weak-convex Theorem.

As a direct consequence of full upper LDP Theorem in i.i.d. case, $h_g = h_g^{Cramer}$ whenever we have $\mathbb{E}(\exp \rho(\|g(X)\|)) < \infty$, (with $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\frac{\rho(t)}{t} \to \infty$). Moreover, for

$B = \mathbb{R}^d$, if x is the gradient of $\theta \to \log \int \exp(\langle \theta, g(y) \rangle)\mu(dy)$ at a point $\theta(x)$ belonging to the interior of $\left\{\theta;\ \int \exp(\langle \theta, g(y) \rangle)\mu(dy) < \infty\right\}$ and if $\gamma_{\theta(x)}$ is the probability measure proportional to $\exp(\langle \theta(x), g \rangle)\mu$, we have $\int g d\gamma_{\theta(x)} = x$ and

$$h_g^{Cramer}(x) = K\left(\gamma_{\theta(x)} \mid \mu\right) \geq h_g(x).$$

Taking up these two facts, one might ask whether $h_g = h_g^{Cramer}$ is always true. At this point, we cannot answer to this question. But, we show in the following that it is true for our partial large deviations bounds.

In order to avoid situations in which the Cramér rate is senseless (for example, when the Laplace transform is infinite everywhere except in 0), it is natural to consider the weak LDP associated to a balanced couple $(g^{(1)}, g^{(2)})$ for which the domain of the Laplace transform contains the set $]-\infty, 0[\times B^*$. This is the idea of Dembo and Shao [DemSha98a and b], [Sha97]. Our paper follows the steps of [DemSha98b] where the authors consider an i.i.d. sequence $(Y_n)_n$ taking values in B and the balanced couple $(\rho \circ N(x), x)$ (ρ and N defined as in Theorem 1). Therefore, when $\rho \circ N(Y)$ is integrable, corollaries 1 and 2 yield the same self-normalized results than [DemSha98b] with the same rate (namely the Cramér rate). Without assuming that $\rho \circ N(Y)$ is integrable, Theorem 1 and parts a,b of Corollary 2 remain valid.

4.2 About the full upper LDP

Can we extend the full upper LDP stated in 2.4 for functions g such that $\|g\| = O(V)$? Answer is no as we can see with the following counter-example (inspired by [AcoNey98]).

- **Description of the Markov chain**

  Denoting $E(\cdot)$ the integer part, the following sets form a partition of $\mathbb{N}$:

  $$\{u_n \triangleq 9m;\ m \in \mathbb{N}\}, \left\{v_n \triangleq 3\left(m + E\left(\frac{m-1}{2}\right)\right);\ m \in \mathbb{N}\right\} \text{ and } \left\{w_n \triangleq m + E\left(\frac{m-1}{2}\right);\ m \in \mathbb{N}\right\}.$$

  We consider the $\mathbb{N}$-valued Markov chain defined by the following transition kernel :

  $*$ $p(0, u_m) = p_m > 0$, $p(u_m, v_m) = p(v_m, w_m) = p(w_m, 0) = 1$.

  This Markov chain is recurrent and aperiodic with stationary distribution μ such that :

  $*$ $\mu(0) = \frac{1}{4-3p_0} = c$, $\mu(u_m) = \mu(v_m) = \mu(w_m) = cp_m$ for every m.

  In order to compute the Donsker-Varadhan rate I^1, we must determine which transition kernels q are absolutely continuous with respect to p. They are necessarily of the following form : for each $m \in \mathbb{N}$,

  $*$ $q(0, u_m) = q_m > 0$, $q(u_m, v_m) = q(v_m, w_m) = q(w_m, 0) = 1$.

  The invariant probabilities for q are defined by :

$*$ $\mu_q(0) = \frac{1}{4-3q_0} = c(q)$, $\mu_q(u_m) = \mu_q(v_m) = \mu_q(w_m) = c(q)q_m$.

Therefore, $I(\mu_q \otimes q) = \mu_q(0)K(q(0,\cdot) \mid p(0,\cdot)) = \mu_q(0)\sum_{m=0}^{\infty} q_m \log \frac{q_m}{p_m}$, and, for any probability measure on $\mathbb{N}^2$ which cannot be written $\mu_q \otimes q$, the rate I is infinite.

- **A function** g

  Let g be the function defined by $g(u_m) = m$, $g(v_m) = -2m$ and $g(w_m) = m$.

  For any probability measure ν such that $I^1(\nu) < \infty$, we have $\int g d\nu = 0$ and $h_g(x) = 0$ if $x = 0$, $h_g(x) = \infty$ elsewhere. On the other hand, if we set $r(t) = \sum_{j \geq t} p_j$ and $a > 0$, then we have, for this function,

$$\mathbb{P}_x\left(\overline{g}_n \geq a\right) = \mathbb{P}(X_{n-1} = 0)r(an),$$

$$\mathbb{P}_x\left(\overline{g}_n \leq -a\right) = \mathbb{P}(X_{n-2} = 0)r(an).$$

  Moreover, $\mathbb{P}(X_n = 0) \to c = \frac{1}{4-3p_0}$.

  Therefore, if the sequence $(p_m)_m$ is such that $\frac{1}{n}\log r(an) \to R(a) > -\infty$ and if R is continuous, increasing to infinity, then $(\overline{g}_n)_n$ satisfies a LDP with rate $J(x) = -R(|x|)$ for every initial state (see the Lemma 9). The upper weak LDP cannot possibly hold with rate h_g. We now check if these results are compatible with the upper LDP Theorem given in section 2.4.

- **A criterion of exponential stabilization**

  Assume that p satisfies the criterion of exponential stabilization associated with (U,V), V non-negative, increasing and such that :

$$\begin{aligned} V(0) &= U(0) - \log\textstyle\sum_{m\geq 0} e^{U(9m)}p_m > -\infty,\ V(u_m) = U(u_m) - U(v_m) \to \infty, \\ V(v_m) &= U(v_m) - U(w_m) \to \infty,\ V(w_m) = U(w_m) - 1 \to \infty. \end{aligned}$$

  The first point implies that there exists $M > 0$ such that $\sum_{m\geq 0} e^{U(9m)}p_m \leq M$, hence $e^{U(9am)}r(am) \leq M$. Therefore, if $\frac{1}{n}\log r(an) \to R(a) > -\infty$, then $\limsup_{n\to\infty}\frac{1}{n}U(9an)$ is bounded above by a constant *i.e.* $U(n) = O(n)$.

  Let us consider, for example, the sequence $p_n = \frac{1}{2^{n+1}}$. We then have $\frac{1}{n}\log r(an) \to -[a]\log 2$. If we take the function $U(m) = \frac{m\log 2}{18}$, then all four points are satisfied. The chain checks the criterion of exponential stabilization associated with (U,V) and $\|g\| = O(V)$. Yet, the upper LDP does not hold with rate h_g.

Unlike in the i.i.d. case, in a Markovian framework, the condition (6) leading to the upper LDP cannot be replaced by $\|g\| = O(V)$.

4.3 Large deviation upper bound and regularity of the function

This example shows that the regularity of the function G is necessary to get an upper LDP (unlike the lower LDP).

We consider the model

$$X_0 = x\ ,\ X_{n+1} = \frac{1}{2}X_n + \varepsilon_{n+1}$$

where $(\varepsilon_i)_i$ is an i.i.d. sequence having the distribution $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = \frac{1}{2}$ and we take $E = [-2, 2]$. The transition kernel of this Markov chain is

$$p(x, \cdot) = \frac{1}{2}\delta_{x/2+1} + \frac{1}{2}\delta_{x/2-1}$$

and the corresponding invariant probability is the Lebesgue measure on $[-2, 2]$. The only transition kernels which are absolutely continuous with respect to $p(x, \cdot)$ on $[-2, 2]$ have the following form :

$$q_{t(\cdot)}(x, \cdot) = t(x)\delta_{x/2+1} + (1 - t(x))\delta_{x/2-1} \text{ (where } t \text{ is some [0,1]-valued function).}$$

Let ζ be an invariant probability for $q_{t(\cdot)}$ and $\mathcal{D}$ be the following subset of $[-2, 2]$:

$$\left\{0, \sum_{j=0}^{n} \frac{i_j}{2^j},\ n \in \mathbb{N},\ \{i_0, i_1, ..., i_n\} \in \{-1, 1\}^{n+1}\right\}.$$

We can prove by induction that, if $q_{t(\cdot)}$ is absolutely continuous with respect to p, then necessarily its invariant probability measure ζ is such that $\zeta(\mathcal{D}) = 0$. As a consequence, the rate $h_{\mathbb{I}_\mathcal{D}}$ is infinite everywhere except in 0 where it is nul. But, starting from 0, the chain remains in $\mathcal{D}$: therefore, we have $\frac{1}{n}\log \mathbb{P}_0((\overline{\mathbb{I}_\mathcal{D}})_n \in \{1\}) = 0$. According to these two observations, the upper LDP cannot hold for the sequence $(\overline{\mathbb{I}_\mathcal{D}})_n$ with rate $h_{\mathbb{I}_\mathcal{D}}$.

Similarly, our results about Lyapunov balanced couples are no longer valid when the function is not regular enough : for instance, if $G(x) = (|x|, \mathbb{I}_\mathcal{D}(x))$, then we have $\mathbb{P}_0(\overline{G}_n \in [0, 2] \times \{1\}) = 1$, while $h_G([0, 2] \times \{1\}) = \infty$.

The upper large deviations Theorem given in 2.4 does not apply to every measurable function G, even when it is bounded. This remark is also true for our weak upper LDP (Theorem 2), hence for our PLDP.

5 Proof of the results stated in section 3

5.1 Proof of Donsker-Varadhan rate functions properties

Proof of Remark 5

Let $(x^{(1)}, x^{(2)}) \in \mathbb{R}_+ \times B$. For any $x \in [0, x^{(1)}]$ and $n \in \mathbb{N}$,

$$h_G(x, x^{(2)}) \geq h_G^0 \left( ]x - 1/n, x + 1/n[\times B(x^{(2)}, 1/n) \right).$$

Let ζ_n be a sequence of $\mathcal{P}(E^2)$ such that $\int G d\zeta_n \in]x - 1/n, x + 1/n[\times B(x^{(2)}, 1/n)$ and $I(\zeta_n) \leq h_G^0 \left( ]x - 1/n, x + 1/n[\times B(x^{(2)}, 1/n) \right) + 1/n$.

Then, since $G^{(1)}$ is a Lyapunov function, $(\zeta_n)_n$ is relatively compact and there exists a subsequence $(\zeta_{u(n)})_n$ which converges weakly to ζ.

Thanks to an argument of uniform integrability ($G^{(2)}$ is uniformly integrable with respect to the sequence $(\zeta_{u(n)})_n$), $\int G^{(1)} d\zeta \leq \liminf_n \int G^{(1)} d\zeta_{u(n)} = x$, $\int G^{(2)} d\zeta = \lim_n \int G^{(2)} d\zeta_{u(n)} = x^{(2)}$ and, I being l.s.c., $I(\zeta) \leq \liminf I(\zeta_n)$.

$$h_G^0 \left([0, x^{(1)}] \times \{x^{(2)}\} \right) \leq \liminf_n h_G^0 \left( ]x - 1/n, x + 1/n[\times B(x^{(2)}, 1/n) \right) \leq h_G(x, x^{(2)}).$$

Finally, $h_G^0 \left([0, x^{(1)}] \times \{x^{(2)}\} \right) \leq h_G \left([0, x^{(1)}] \times \{x^{(2)}\} \right)$ which proves Remark 5. $\diamondsuit$

Proof of lemma 1

1) Let $x, y \in B$, $t \in [0, 1]$ and $\varepsilon > 0$. We assume that $h_G(x)$ and $h_G(y)$ are finite (otherwise, there is nothing to prove). Then :

   $\exists \zeta_x \in \mathcal{P}(E^2)$ such that $\int \|G\| d\zeta_x < \infty$, $\int G d\zeta_x = x$ and $I(\zeta_x) \leq \varepsilon + h_G(x)$,

   $\exists \zeta_y \in \mathcal{P}(E^2)$ such that $\int \|G\| d\zeta_y < \infty$, $\int G d\zeta_y = y$ and $I(\zeta_y) \leq \varepsilon + h_G(y)$,

   $\int \|G\| d(t\zeta_x + (1-t)\zeta_y) < \infty$ and $\int G d(t\zeta_x + (1-t)\zeta_y) = tx + (1-t)y$.

   Consequently, for every $t \in [0, 1]$, the convexity of I yields

   $h_G(tx + (1-t)y) \leq I(t\zeta_x + (1-t)y) \leq tI(\zeta_x) + (1-t)I(\zeta_y) \leq \varepsilon + th_G(x) + (1-t)h_G(y)$.

   Hence, h_G is convex.

2) The convexity of h_G^* follows by a similar argument. Let us prove the lower semi-continuity of h_G^*.

   Let $(x_n)_n = (x_n^{(1)}, x_n^{(2)})_n$ be a sequence of $\mathbb{R}_+ \times B$ converging to $x = (x^{(1)}, x^{(2)})$ and $\varepsilon > 0$. Assume that $\liminf h_G^*(x_n) < \infty$ (otherwise there is nothing to prove). Let $(\zeta_n)_n$ be a sequence of $\mathcal{P}(E^2)$ such that :

   $$\varepsilon + h_G^*(x_n) \geq I(\zeta_n), \quad \int G^{(1)} d\zeta_n \leq x_n^{(1)}, \quad \int G^{(2)} d\zeta_n = x_n^{(2)}.$$

   Therefore, as the function $\zeta \mapsto \int G^{(1)} d\zeta$ has compact level sets (because $G^{(1)}$ is a l.s.c. Lyapunov function) and as $(x_n^{(1)})_n$ is bounded, the sequence $(\zeta_n)_n$ is relatively compact.

   Then let $(x_{u(n)})_n$ be a subsequence such that $\lim_n h_G^*(x_{u(n)}) = \liminf_n h_G^*(x_n)$ and $(\zeta_{u(n)})_n$ converges weakly to some probability measure ζ. For the same reasons of uniform integrability, we have

   $$\int G^{(1)} d\zeta \leq \liminf \int G^{(1)} d\zeta_{u(n)} = x^{(1)} \quad \text{and} \quad \lim_n \int G^{(2)} d\zeta_{u(n)} = x^{(2)}.$$

Consequently, $h_G^*(x) \le I(\zeta) \le \liminf_{n\to\infty} I(\zeta_n) \le \varepsilon + \liminf_{n\to\infty} h_G^*(x_n)$ and h_G^* is l.s.c..
$\diamondsuit$

5.2 Proof of the weak upper LDP

Let $G = (G^{(1)}, G^{(2)})$ be a balanced couple. Let K be a compact subset of $\mathbb{R}_+ \times B$ and $\Gamma_K \triangleq \{\zeta \in \mathcal{P}(E^2);\ \int \|G\| d\zeta < \infty,\ \int G d\zeta \in K\}$. We have $\{\overline{G}_n \in K\} = \{L_n \in \Gamma_K\} \subset \{L_n \in \overline{\Gamma_K}\}$.
Since $G^{(1)}$ is a Lyapunov function, then Γ_K is a relatively compact subset of $\mathcal{P}(E^2)$. According to Donsker-Varadhan Theorem given in section 3.1, we have:

$$\limsup_{n\to\infty} \sup_{\nu\in H} \frac{1}{n} \log \mathbb{P}_\nu(\overline{G}_n \in K) = \limsup_{n\to\infty} \sup_{\nu\in H} \frac{1}{n} \log \mathbb{P}_\nu\left(L_n \in \Gamma_K\right) \le -I(\overline{\Gamma_K}).$$

Theorem 2 follows from this Lemma :

Lemma 2. $I(\overline{\Gamma_K}) = h_G^*(K)$.

Proof
First, $h_G^*(K) = I(\Gamma_K) \ge I(\overline{\Gamma_K})$. Let us prove that $I(\Gamma_K) \le I(\overline{\Gamma_K})$. For every given $\zeta \in \overline{\Gamma_K}$, there exists a sequence $(\zeta_n)_n$ of Γ_K which converges weakly to some probability measure ζ.

For each n, let $x_n = \left(x_n^{(1)}, x_n^{(2)}\right) = \left(\int G^{(1)} d\zeta_n, \int G^{(2)} d\zeta_n\right) \in K$. Exists a subsequence $(x_{u(n)})_n$ which converges to $x = (x^{(1)}, x^{(2)}) \in K$. $\int G^{(1)} d\zeta \le x^{(1)}$ and $\int G^{(2)} d\zeta = x^{(2)}$. Therefore, by definition of h_G^*, the following inequality holds : $h_G^*(K) \le h_G^*(x) \le I(\zeta)$ for every given $\zeta \in \overline{\Gamma_K}$. Consequently, $h_G^*(K) \le I(\overline{\Gamma_K})$.$\diamondsuit$

5.3 Proof of the lower LDP

5.3.1 Properties of the strong irreducibility

We note :

$$U(x,.) \triangleq \sum_{n\ge 1} p^n(x,\cdot) \ \text{ and } U_a(x,.) \triangleq \sum_{n\ge 1} a^n p^n(x,\cdot)\ (0<a<1).$$

If the strong irreducibility hypothesis holds with the invariant probability measure μ, we have, for any x :

$$\sum_{i=L}^{\infty} \frac{1}{2^i} p^i(x,\cdot) << \mu(\cdot) << \sum_{i=1}^{\infty} \frac{1}{2^i} p^i(x,\cdot).$$

This property implies the μ-irreducibility of the Markov chain. According to Nummelin ([Num]), the chain has a small set C. In other words, there exists a probability measure ξ on C, a real number $h > 0$ and $a > 0$ such that :

$$\forall x \in E,\ U_a(x, \cdot) \geq h\mathbb{I}_C(x)\xi(\cdot).$$

In particular, $\xi << \mu$. If we note $\xi U(C) \stackrel{\triangle}{=} \int \xi(dx)U(x, C)$ and $\mathcal{R}_C \stackrel{\triangle}{=} \{\omega;\ \sum_{i\geq 0} \mathbb{I}_C(X_i(\omega)) = \infty\}$, two situations can occur :

* C is a transient small set : $\xi U(C) < \infty$ and $\mathbb{P}_\xi(\mathcal{R}_C) = 0$;

* C is a recurrent small set : $\xi U(C) = \infty$ and $\mathbb{P}_\xi(\mathcal{R}_C) = 1$.

The transient case is incoherent here because it would imply that $\mu(C) = \xi(C) = 0$. Consequently, C is a recurrent small set. We set $\Gamma_C \stackrel{\triangle}{=} \{x;\ \mathbb{P}_x(\mathcal{R}_C) = 1\}$ and note that $1 = \mathbb{P}_\xi(\mathcal{R}_C) = \mathbb{P}_\xi(\sum_{i=n}^{\infty} \mathbb{I}_C(X_i) = \infty) = \mathbb{E}_\xi(\mathbb{P}_{X_n}(\mathcal{R}_C))$ for any n.
Moreover, $\mathbb{P}_{X_n}(\mathcal{R}_C) = 1$ if and only if $X_n \in \Gamma_C$, hence $\xi p^n(\Gamma_C) = 1$ for all n. Therefore, for any x of E and any integer $l \geq L$, we have $p^n(x, \Gamma_C^c) = 0$.

Every point of E leads to the recurrent small set C almost surely and the transition kernel has an invariant probability measure. Therefore, the chain is positive recurrent (see, for example, Theorem 8.2.16 in [Duf]).

Lemma 3. *Under assumption 2, for any probability measure $\zeta \in \mathcal{P}(E^2)$ such that $I(\zeta) < \infty$, there exists a transition kernel q with respect to which ζ^1 is invariant and $q^n(x, \cdot) << p^n(x, \cdot)$ for every $n \in \mathbb{N}$ and $x \in E$. Moreover, $\zeta^1 << \mu$.*

Proof

Since $K\left(\zeta \mid \zeta^1 \otimes p\right) < \infty$, according to a result obtained by Donsker and Varadhan, there exists a transition kernel q with stationary distribution ζ^1 such that $\zeta = \zeta^1 \otimes q$. Consequently, $q(x, \cdot) << p(x, \cdot)$ ζ^1 almost surely. We can modify q setting $q(x, \cdot) = p(x, \cdot)$ for points $x \in E$ such that $q(x, \cdot) << p(x, \cdot)$ is not satisfied. ζ^1 remains invariant for this modified q.

By an iterative argument, $q^n(x, \cdot) << p^n(x, \cdot)$ for every $n \geq 1$. Therefore, for all x in E and all $l \geq L$, $q^l(x, \cdot) << \mu(\cdot)$ and, while $\sum_{i\geq L} q^i(x, \cdot) << \sum_{i\geq L} p^i(x, \cdot) << \mu$, we have $\zeta^1 << \mu$. $\diamondsuit$

5.3.2 Lower Laplace principle

We follow the method developped by Dupuis-Ellis ([DupEll]) to prove the lower LDP for empirical distributions.

* *Representation formula*

For any initial distribution ν and every $j \leq n$, we introduce the following notations:

$$L_{j,n} \triangleq \frac{1}{n}\sum_{i=1}^{j} \delta_{X_{i-1},X_i} \quad \text{and} \quad L_{n,n} \triangleq L_n;$$

$$\overline{G}_n \triangleq \int G dL_n = \frac{1}{n}\sum_{i=1}^{n} G(X_{i-1},X_i); \;\; \overline{G}_{j,n} \triangleq \int G dL_{j,n} = \frac{1}{n}\sum_{i=1}^{j} G(X_{i-1},X_i).$$

Let f be a bounded Lipschitz function $B \to \mathbb{R}$; $\mathcal{F}_j$ is the σ-field generated by $X_0, ..., X_j$.

$$H_{0,n}(x,0) = -\frac{1}{n}\log \mathbb{E}\left(\exp(-nf(\overline{G}_n))\right) = -\Lambda_n(f),\; H_{j,n}(X_j, \overline{G}_{j,n}) \triangleq -\frac{1}{n}\log \mathbb{E}\left(\exp(-nf(\overline{G}_n))/\mathcal{F}_j\right).$$

By Markov property, we obtain :

$$H_{n,n}(X_n, \overline{G}_{n,n}) = f(\overline{G}_n),$$
$$H_{j,n}(y,r) = -\frac{1}{n}\log \int \exp\left(-nH_{j+1,n}\left(z, r+\frac{1}{n}G(u,z)\right)\right)\delta_y \otimes p(du,dz).$$

The Donsker-Varadhan variational identity gives :

$$H_{j,n}(y,r) = \inf_{\zeta \in \mathcal{P}(E^2)} \left\{\frac{1}{n}K(\zeta \mid \delta_y \otimes p) + \int H_{j+1,n}\left(z, r+\frac{1}{n}G(u,z)\right)\zeta(du,dz)\right\}.$$

This is the dynamic programming equation for the controlled Markov chain with state space $E \times B$, control space $\mathcal{P}(E^2)$ and the transition kernel at time j is $Q_{j,n}$:

$$Q_{j,n}(y,r,\zeta;\bullet) = \text{ distribution of } (Z^2, r+\frac{1}{n}G(Z^1,Z^2)) \text{ where } Z=(Z^1,Z^2) \text{ has the distribution } \zeta.$$

The final cost is $(y,r) \to f(r)$ and the running cost is $c_j(y,r,\zeta) = \frac{1}{n}K(\zeta \mid \delta_y \otimes p)$.

$*$ *Lower Laplace principle*

Let q be a transition kernel of a recurrent Markov chain with stationary distribution α such that $I(\alpha \otimes q) < \infty$ and $\int \|G\| d\alpha \otimes q > \infty$.

For a Markov chain $(Y_j)_{0 \leq j \leq n}$ with initial distribution ν and transition kernel q, $(T_j)_{0 \leq j \leq n}$, defined by $T_0 = (Y_0, 0)$ and $T_j = \left(Y_j, \sum_{i=1}^{j} G(Y_{i-1}, Y_i)\right)$ for $1 \leq j \leq n$, is the controlled Markov chain for the control $\zeta = \delta_y \otimes q$. Consequently,

$$-\Lambda_n(f) \leq \mathbb{E}\left(\sum_{i=0}^{n-1} K\left(\delta_{Y_i} \otimes q \mid \delta_{Y_i} \otimes p\right) + f\left(\sum_{i=1}^{j} G(Y_{i-1}, Y_i)\right)\right),$$

$$-\Lambda_n(f) \leq \mathbb{E}\left(\sum_{i=0}^{n-1} K\left(q(Z_i,\cdot) \mid p(Z_i,\cdot)\right) + f\left(\frac{1}{n}\sum_{i=1}^{n} G(Y_{i-1},Y_i)\right)\right).$$

$$\liminf \Lambda_n(f) \geq -K\left(\alpha\otimes q \mid \alpha\otimes p\right) - f\left(\int G d\alpha\otimes q\right) = -\left(I(\alpha\otimes q) + f\left(\int G d\alpha\otimes q\right)\right).$$

Following the demonstration of Dupuis-Ellis ([DupEll], chapitre 8), we introduce $\mathcal{N}$ which is the set of all probability measures $\zeta \in \mathcal{P}(E^2)$ such that :

- $I(\zeta) < \infty$;
- $\zeta = \zeta^1 \otimes q$ where q is the transition kernel of a recurrent Markov chain.

We have :

$$\liminf \Lambda_n(f) \geq -inf\left\{I(\zeta) + f\left(\int G d\zeta\right) \mid \zeta \in \mathcal{N}, \int \|G\| d\zeta < \infty\right\}.$$

On the other hand :

$$\begin{aligned}
\inf_{x\in B}\{f(x) + h_G(x)\} &= \inf_{x\in B}\left\{f(x) + inf\left\{I(\zeta) \mid \zeta \in \mathcal{P}(E^2), \int \|G\| d\zeta < \infty, \int G d\zeta = x\right\}\right\} \\
&= inf\left\{I(\zeta) + f\left(\int G d\zeta\right) \mid \zeta \in \mathcal{P}(E^2), \int \|G\| d\zeta < \infty\right\}.
\end{aligned}$$

Since f is a bounded Lipschitz function, the following Lemma will prove the identity (8):

$$\begin{aligned}
&\inf\left\{I(\zeta) + f\left(\int G d\zeta\right) \mid \zeta \in \mathcal{P}(E^2), \int \|G\| d\zeta < \infty\right\} = \\
&\inf\left\{I(\zeta) + f\left(\int G d\zeta\right) \mid \zeta \in \mathcal{N}, \int \|G\| d\zeta < \infty\right\}. \qquad (8)
\end{aligned}$$

Lemma 4. *Let G be a measurable function from E^2 into B. If p satisfies the strong irreducibility hypothesis and if $\int \|G\| d\mu\otimes p < \infty$, then, for any $\zeta \in \mathcal{P}(E^2)$ such that $I(\zeta) < \infty$, $\int \|G\| d\zeta < \infty$ and all $\delta > 0$, there exists a probability measure $\nu_\delta \in \mathcal{N}$ such that :*

$$\int \|G\| d\nu_\delta < \infty;\ \ I(\nu_\delta) \leq I(\zeta) \ \ \textit{and} \ \ \left\|\int G d\zeta - \int G d\nu_\delta\right\| \leq \delta.$$

Proof

For all $t \in]0,1[$, we define $\zeta_t \stackrel{\triangle}{=} (1-t)\zeta + t\mu\otimes p$. Clearly,

$$\left\|\int G d\zeta - \int G d\zeta_t\right\| \leq t\left(\int \|G\| d\zeta + \int \|G\| d\mu\otimes p\right) = tM$$

and we will take $\nu_\delta = \zeta_{\delta/M}$.

By convexity of the relative entropy,

$$I(\zeta_t) = K\left(\zeta_t \mid \zeta_t^1 \otimes p\right) \leq (1-t)K\left(\zeta \mid \zeta^1 \otimes p\right) < \infty,$$

$$\int \|G\| d\zeta_t = (1-t)\int \|G\| d\zeta + t\int \|G\| d\mu \otimes p < \infty.$$

We now prove that ζ_t belongs to $\mathcal{N}$. According to Lemma 4, there exists a transition kernel q_t with stationary distribution ζ_t^1 such that, for all n and any $x \in E$, $q_t^n(x,\cdot) << p_t^n(x,\cdot)$. Moreover, $\zeta_t^1 << \mu$. Let us show that q_t satisfies the strong irreducibility hypothesis.

Obviously, $\zeta_t^1(\cdot) \geq t\mu(\cdot)$. The probability measures ζ_t^1 and μ are equivalent. We denote h the density of ζ_t^1 with respect to μ ; $h \geq t$. Let A and B belong to E ;

$$\int_A q_t(x,B)h(x)\mu(dx) = \zeta_t(A \times B) \geq t\mu \otimes p(A \times B) = t\int_A p(x,B)\mu(dx).$$

Consequently , $q_t(x,\cdot) \geq \frac{t}{f(x)}p(x,\cdot)$ μ almost surely. We modify q_t as we did in Lemma 4 for this inequality to be true on E. This modification raises no problem since $\zeta_t^1 \sim \mu$: we change q_t on a set N such that $\zeta_t^1(N) = 0$. Therefore, ζ_t^1 remains invariant for q_t and $q_t^n(x,\cdot) \sim p^n(x,\cdot)$ for every $n \in \mathbb{N}$ and all $x \in E$. $\diamondsuit$

The lower "Laplace principle" is proved : for any bounded Lipschitz function f,

$$\liminf_{n\to\infty} \Lambda_n(f) \geq -\inf_{x\in B}\{h_G^0(x) + f(x)\}.$$

The lower part of Bryc Theorem ("lower Laplace principle" $\Longrightarrow$ "lower LDP") is proved only considering bounded Lipschitz functions and without using the lower semi-continuity of the "rate" (see [DupEll]). Consequently, we have the lower LDP for any initial distribution ν :

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_\nu\left(\overline{G}_n \in \Gamma\right) \geq -h_G^0(U) = -h_G(U) \quad \text{for any open subset } U \text{ of } B\ .\diamondsuit$$

6 Proof of the partial LDP

6.1 An exponential tightness result

The following result allows us to take advantage of exponential tightness criteria stated in [DemSha98b]:

Lemma 5. *Suppose that assumptions 1 and 2 are satisfied and let F be a continuous function from E^2 into $\mathbb{R}_+$. Then, there exists a continuous non-negative function T increasing to infinity such that $h_{T\circ F}$ is a good rate function and, for every compact subset H of H_M and $r > 0$,*

$$\limsup_{n\to\infty} \sup_{\nu\in H} \frac{1}{n}\log \mathbb{P}_\nu\left(\overline{(T\circ F)}_n \geq r\right) \leq -h_{T\circ F}([r,\infty[).$$

Proof

The result is clear if the function F is bounded. Let us assume that F is unbounded and consider

$$\alpha(t) \stackrel{\triangle}{=} \sup\{F(x,y);\ (x,y) \in E^2,\ V \oplus V(x,y) \leq t\}.$$

This function increases to infinity. Let β be a continuous function, strictly increasing to infinity and such that $\beta \geq \alpha$. According to the definition of α, we have $F \leq \alpha \circ (V \oplus V) \leq \beta \circ (V \oplus V)$ and thus $\beta^{-1} \circ F \leq V \oplus V$. We consider a continuous increasing function k such that $k(t)/t \to 0$ and $k(t) \to \infty$. Setting $T \stackrel{\triangle}{=} k \circ \beta^{-1}$, we have $T \circ F \leq k(V \oplus V)$.

The conditions required to apply the full upper LDP given in section 2.4 are satisfied with $\rho = k^{-1}(\cdot)$, hence $h_{T\circ F}$ is a good rate function and we have, for every $r > 0$,

$$\limsup_{n\to\infty} \sup_{\nu\in H} \frac{1}{n} \log \mathbb{P}_\nu \left(\overline{(T \circ F)}_n \geq r \right) \leq -h_{T\circ F}([r, \infty[).\ \diamondsuit$$

An immediate consequence of this result is that, for any $\delta > 0$,

$$\lim_{r\to\infty} \limsup_{n\to\infty} \sup_{\nu\in H} \frac{1}{n} \log \mathbb{P}_\nu \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{F(X_i-1, X_i)\geq r} \geq \delta \right) = -\infty. \tag{9}$$

Indeed, we have

$$\left\{\tfrac{1}{n} \textstyle\sum_{i=1}^n \mathbb{I}_{F(X_{i-1},X_i)\geq r} \geq \delta\right\} \quad \subset \quad \left\{\tfrac{1}{n} \textstyle\sum_{i=1}^n T \circ F(X_{i-1}, X_i) \geq \delta T(r)\right\}$$

and $h_{T\circ F}([\delta T(r), \infty[) \to \infty$ as $r \to \infty$, for every given $\delta > 0$. $\diamondsuit$

6.2 Proof of Theorem 1, part a.

Lemma 6. *The function $(G^{(1)}, G^{(2)})$ considered in Theorem 1 is a Lyapunov balanced couple (see section 3.2).*

Proof

We only need to check that $\rho(t)/t \to \infty$ as $t \to \infty$. There exists $B > 0$ such that, if we introduce $u = \rho(B)$, $v = \rho(B+1)$, then $a \stackrel{\triangle}{=} v - u > 0$. Let $b = u - aB$. For every $t \geq B + 1$, we have $\rho(t) \geq at + b$ thanks to the convexity of $\rho(\cdot)$. If we set $L(t) = \rho(t)/t$ then $L(t)/L(t^{1/2}) = \rho(t)/(t^{1/2}\rho(t^{1/2}))$ and therefore $L(t) \to \infty$. As a matter of fact, if $(t_n)_n \uparrow \infty$ was a sequence such that $\limsup L(t_n) < \infty$, then we would have $L(t_n)/L(t_n^{1/2}) \leq \limsup L(t_n)/a < \infty$, which contradicts our hypothesis on ρ. $\diamondsuit$

Now, we can apply Theorem 2 to the function G : the sequence $(\overline{G}_n^{(1)}, \overline{G}_n^{(2)})_n$ satisfies a weak upper LDP with rate h_G^*. According to [DemSha98b], the partial LDP stated in Theorem 1 holds as soon as $(\overline{G}_n^{(1)}, \overline{G}_n^{(2)})_n$ is exponentially tight with respect to the class $\mathcal{S}(\rho \circ N)$; in other words if, for any positive number R and any set $A \in \mathcal{S}(\rho \circ N)$, there exists a compact subset K_R of $\mathbb{R}_+ \times B$ such that $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left((\overline{G}_n^{(1)}, \overline{G}_n^{(2)}) \in A \cap K_R^c\right) \leq -R$.

To prove such a statement, we apply formula (9) proved in the previous Paragraph to the function $F = N \circ G^{(2)}$: this yields, for every $\delta > 0$,

$$\lim_{r\to\infty} \limsup_{n\to\infty} \sup_{\nu \in H} \frac{1}{n} \log \mathbb{P}_\nu \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{N\circ G^{(2)}(X_i-1, X_i) \geq r} \geq \delta \right) = -\infty.$$

On the other hand, $\rho \circ N$ belongs to the class of functions introduced in definition 0.1.1. of [DemSha98b]. Therefore, the proof of Lemma 0.1.1. in [DemSha98b] applies to the function $(G^{(1)}, G^{(2)}) = (\rho \circ N(G^{(2)}), G^{(2)})$, and this entails, for any $\varepsilon > 0$

$$\lim_{r\to\infty} \limsup_{n\to\infty} \sup_{\nu \in H} \frac{1}{n} \log \mathbb{P}_\nu \left(\rho \circ N(\overline{G}_n^{(2)}) \geq \varepsilon \overline{G}_n^{(1)} \geq \varepsilon r \right) = -\infty.$$

This property allows us to apply Proposition 0.1.2. of [DemSha98b] to prove that the distributions of $(\overline{G}_n^{(1)}, \overline{G}_n^{(2)})$ are exponentially tight with respect to $\mathcal{S}(\rho \circ N)$. Part $a)$ of Theorem 1 is then proved $\diamondsuit$.

6.3 Proof of Theorem 1 part b.

Lemma 7. *If* $\Gamma_r \triangleq \left\{ (x^{(1)}, x^{(2)}) \in \mathbb{R}_+ \times B;\ \frac{\rho(N(x^{(2)}))}{x^{(1)}} \geq r \right\}$, *then :*

a) $h_G^*(\Gamma_r) = h_G(\Gamma_r)$.

b) $h_G(\Gamma_r) = 0$ *if and only if* p *has an invariant probability measure* μ *such that*

- $G^{(1)}$ *is integrable with respect to* $\mu \otimes p$
- $r \leq \frac{\rho(N(\int G^{(2)} d\mu \otimes p))}{\int G^{(1)} d\mu \otimes p}$.

Proof

a) Let $(x^{(1)}, x^{(2)}) \in \Gamma_r$ and $\varepsilon > 0$ be given. Let $\zeta \in \mathcal{P}(E^2)$ be such that

- $y^{(1)} \triangleq \int G^{(1)} d\zeta \leq x^{(1)}$,
- $\int G^{(2)} d\zeta = x^{(2)}$
- $I(\zeta) \leq \varepsilon + h_G^*(x)$.

The couple $(y^{(1)}, x^{(2)})$ belongs to Γ_r and

$$h_G^*(\Gamma_r) \leq h_G(\Gamma_r) \leq h_G((y^{(1)}, x^{(2)})) \leq I(\zeta) \leq \varepsilon + h_G^*(x^{(1)}, x^{(2)}).$$

b) A part of the proof is trivial because $I(\mu \otimes p) = 0$ when μ is a p-invariant measure.

Assume that $h_G(\Gamma_r) = 0$. This implies the existence, for any integer n, of a probability measure ζ_n such that $\int G d\zeta_n \in \Gamma_r$ and $I(\zeta_n) \leq \frac{1}{n}$. As I is a good rate function, we can consider a subsequence $(\zeta_n)_n$ that converges weakly to ζ. By construction of $(\zeta_n)_n$, we have $I(\zeta) = 0$ and $\int G d\zeta \in \Gamma_r$. The first assertion entails that $\zeta^1 = \zeta^2$ (which we note μ in the following) and that μ is an invariant probability measure for p. The second assertion implies that $G^{(1)}$ is integrable with respect to $\zeta = \zeta^1 \otimes p$, and that $r \leq \rho \circ N(\int G^{(2)} d(\mu \otimes p)) / \int G^{(1)} d(\mu \otimes p)$. $\diamondsuit$

For any $r > 0$, the set $\Gamma_r \triangleq \left\{(x^{(1)}, x^{(2)});\ \frac{\rho \circ N(x^{(2)})}{x^{(1)}} \geq r\right\}$ belongs to the class $\mathcal{S}(\rho \circ N)$. The Chernoff upper bound

$$\limsup \frac{1}{n} \log \mathbb{P}\left[\frac{\rho \circ N(\overline{G}_n^{(2)})}{\overline{G}_n^{(1)}} \geq r\right] \leq -J(r).$$

is an obvious consequence of part *a*). $\diamondsuit$

We now prove the upper LDP with rate J.

Lemma 8. *J is a left continuous non-decreasing function. In addition, J is nul in zero and infinite on $]1, \infty[$.*

Proof

If $r' > r$ then $\Gamma_{r'}$ is a subset of Γ_r, hence $h_G(\Gamma_{r'}) \geq h_G(\Gamma_r)$ and J is non-decreasing. Moreover, for $r > 1$, the set of probability measures ζ such that $\int G d\zeta \in \Gamma_r$ is empty (since the function $\rho \circ N$ is convex) and $J(r) = \infty$.

We now prove the left-continuity of J (and consequently its lower semi-continuity). Let $r \in]0, 1]$ and let $(r_n)_n$ be an increasing sequence with limit r such that $\sup_n J(r_n) \leq J(r)$. Let then $(x_n)_n$ be a $\mathbb{R}_+ \times B$ valued sequence such that $x_n \in \Gamma_{r_n}$ and ζ_n be a sequence of $\mathcal{P}(E^2)$ such that $\int G d\zeta_n = x_n$ and

$$I(\zeta_n) < h_G(x_n) + 1/n < h_G(\Gamma_{r_n}) + 2/n = J(r_n) + 2/n.$$

Therefore $\sup_n I(\zeta_n) < \infty$ and, since I has compact level sets, there exists a subsequence $\zeta_{u(n)}$ which converges weakly to some $\zeta \in \mathcal{P}(E^2)$. The uniform integrability argument we used in 5.1 leads to the following identities:

$$\int G^{(1)} d\zeta \leq \liminf \int G^{(1)} d\zeta_{u(n)} \ \text{ and } \ \int G^{(2)} d\zeta = \lim \int G^{(2)} d\zeta_{u(n)},$$

and thus

$$\frac{\rho \circ N(\int G^{(2)} d\zeta)}{\int G^{(1)} d\zeta} \geq \limsup_n \frac{\rho \circ N(\int G^{(2)} d\zeta_{u(n)})}{\int G^{(1)} d\zeta_{u(n)}} = \limsup_n \frac{\rho \circ N(x_{u(n)}^{(2)})}{x_{u(n)}^{(1)}} \geq r.$$

i.e. $(\int G^{(1)} d\zeta, \int G^{(2)} d\zeta) \in \Gamma_r$. Consequently, since I is l.s.c., we have

$$h_G(\Gamma_r) \leq I(\zeta) \leq \liminf I(\zeta_n) \leq \limsup_{n \to \infty} h_G(\Gamma_{r_n}) = \limsup J(r_n).$$

We reach a contradiction ; J is left-continuous therefore l.s.c.. $\diamondsuit$

Since J is l.s.c. and infinite on $]1,\infty[$, it is a good rate function. The upper LDP of rate J comes via the following Lemma taken from [Wor00].

Lemma 9. *Let $(\mu_n)_n$ be a sequence of probability measures on $\mathbb{R}_+$.*
If there exists a $\mathbb{R}_+$-valued function J, increasing to infinity, nul in zero, left continuous and such that, for any $r>0$,

$$\limsup \frac{1}{n} log\mu_n([r,\infty[) \leq -J(r),$$

then $(\mu_n)_n$ satisfies an upper LDP with rate $J(\cdot)$ on $\mathbb{R}_+$.

We apply this result to our situation and we obtain that $(\frac{\rho\circ N(\overline{G}_n^{(2)})}{\overline{G}_n^{(1)}})_n$ satisfies an upper LDP with rate J on $\mathbb{R}_+$.

Proof (taken from [Wor00])
Since J converges to infinity, the sequence $(\mu_n)_n$ is exponentially tight. If we consider the rate H associated with a LDP satisfied by a sub-sequence, we have, for any $r>0$,

$$-H(]r,\infty[) \leq -\liminf_{n\to\infty} \frac{1}{u(n)} \log \mu_{u(n)}(]r,\infty[) \leq \limsup_{n\to\infty} \frac{1}{u(n)} \log \mu_{u(n)}(]r,\infty[) \leq -J(r),$$

and $H(r) \geq H(]r_n,\infty[) \geq J(r_n)$ for a sub-sequence $(r_n)_n$ increasing to r. $H(r) \geq J(r)$ by the lower semi-continuity of J. $\diamondsuit$

6.4 Proof of Corollary 2, part d.

Applying Theorem 1 to the borelian set $A_r \in \mathcal{S}(|.|^q)$ defined by

$$A_r \stackrel{\triangle}{=} \left\{ (x^{(1)}, x^{(2)}) \in \mathbb{R}_+ \times \mathbb{R} \mid \ \frac{x^{(2)}}{(x^{(1)})^{1/q}} \geq r \right\},$$

we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P} \left(\frac{\sum_{i=1}^n F(X_{i-1},X_i)}{n^{1-1/q}(\sum_{i=1}^n \|F(X_{i-1},X_i\|^q)^{1/q}} \geq r \right) \leq -K(r),$$

with $K(r) = h_G^*(A_r) = h_G(A_r)$. Using the same arguments as in the proof of lemma 7, we have $K(r) = 0$ if and only if $\|F\|^q$ is integrable with respect to $\mu \otimes p$ and $\int F d\mu \otimes p \geq r(\int \|F\|^q d\mu \otimes p)^{1/q}$. $\diamondsuit$

7 Appendix

As the weak Cramér Theorem in i.i.d. case, the weak upper LDP stated in Theorem 2 might have its own interest. We show in this section that it can be easily checked, without assuming the criterion of exponential stabilization, if the transition is fellerian.

7.1 Weak LDP for the empirical distributions

The following upper LDP result has been essentially proved by Donsker and Varadhan [DonVar75a] and [DonVar76].

Theorem

Assume that p is fellerian. Then $(L_n)_n$ satisfies a uniform upper weak LDP over all initial distributions:

$$\limsup_n \sup_{\nu\in\mathcal{P}(E)} \frac{1}{n}\log\mathbb{P}_\nu\left(L_n\in\Gamma\right)\leq -I(\Gamma)\ \text{ for every compact subset } \Gamma \text{ of } \mathcal{P}(E^2)$$

Proof

- Denoting $\mathcal{LB}(E^2)$ the space of Lipschitz bounded functions on E^2 provided with the norm $\|.\|_{\mathcal{LB}}=\|.\|_\infty+r(\cdot)$, $r(\cdot)$ being the Lipschitz constant of the function, let J be the weak rate defined on $\mathcal{P}(E^2)$ by

$$J(\zeta)=K\left(\zeta\mid\zeta^1\otimes p\right)=\sup_{G\in\mathcal{LB}(E^2)}\left\{\int G(x,y)\zeta(dx,dy)-\log\int e^{G(x,y)}\zeta^1(dx)p(x,dy)\right\}.$$

As a matter of fact, J is l.s.c. because p is Fellerian hence $\nu\to\nu\otimes p$ is a continuous map. Setting $p^*G(x)\stackrel{\triangle}{=}\log\int e^{G(x,y)}p(x,dy)$, we have $p^*G\in C_b(E)$ when $G\in\mathcal{LB}(E^2)$ and the following identity is obtained with conditional expectations, for every initial distribution ν :

$$1=\mathbb{E}_\nu\left(e^{\sum_{j=1}^n G(X_{j-1},X_j)-p^*G(X_{j-1})}\right)=\mathbb{E}_\nu\left(e^{n\int(G(x,y)-p^*G(x))L_n(dx,dy)}\right).$$

For any Borel subset Γ of $\mathcal{P}(E^2)$ and any probability measure ν on $\mathcal{P}(E)$, the latter identity yields the following key equation

$$\begin{aligned}\mathbb{P}_\nu(L_n\in\Gamma) &\leq \mathbb{E}_\nu\left(\mathbb{I}_{L_n\in\Gamma}\,e^{\sup_{\zeta\in\Gamma}(-n\int(G(x,y)-p^*G(x))d\zeta}\,e^{n\int(G(x,y)-p^*G(x))dL_n}\right)\ \forall G\in\mathcal{LB}(E^2)\\ &\leq e^{-n\sup_{G\in\mathcal{LB}(E^2)}\inf_{\zeta\in\Gamma}\left\{\int(G(x,y)-p^*G(x))\zeta(dx,dy)\right\}}.\end{aligned}$$

Assume now that Γ is compact. If $\lambda<J(\Gamma)$ then, for any $\zeta\in\Gamma$, there exists a function G in $\mathcal{LB}(E^2)$ such that $\int(G(x,y)-p^*G(x))\zeta(dx,dy)>\lambda$. Hence, Γ can be recovered by the family $(\Gamma_G)_{G\in\mathcal{LB}(E^2)}$ where $\Gamma_G\stackrel{\triangle}{=}\left\{\zeta;\ \int(G(x,y)-p^*G(x))\zeta(dx,dy)>\lambda\right\}$.
But for every given $G\in\mathcal{LB}(E^2)$, the map $\zeta\to\int(G(x,y)-p^*G(x))\zeta(dx,dy)$ is continuous, hence Γ_G is an open subset of $\mathcal{P}(E^2)$. Consequently, there exists $N\in\mathbb{N}$ and functions $(G_j)_{j=1..N}$ in $\mathcal{LB}(E^2)$ such that

$$\Gamma\subset\cup_{j=1}^N\left\{\zeta;\ \int(G_j(x,y)-p^*G_j(x))\zeta(dx,dy)>\lambda\right\}=\cup_{j=1}^N\Gamma_{G_j}.$$

For each $j \leq N$, $\sup_{G\in\mathcal{LB}(E^2)} \inf_{\zeta\in\Gamma_{G_j}} \left\{\int (G(x,y) - p^*G(x))\zeta(dx,dy)\right\} \geq \lambda$. Hence, $\mathbb{P}_\nu(L_n \in \Gamma_{G_j}) \leq e^{-n\lambda}$, and it comes

$$\sup_{\nu\in\mathcal{P}(E)} \frac{1}{n} \log \mathbb{P}_\nu(L_n \in \Gamma) \leq -\lambda + \log N/n.$$

Letting $n \to \infty$ and choosing λ arbitrarily close to $J(\Gamma)$ yields the following upper bound for every compact subset $\Gamma \in \mathcal{P}(E^2)$

$$\limsup_{n\to\infty} \sup_{\nu\in\mathcal{P}(E)} \frac{1}{n} \log \mathbb{P}_\nu(L_n \in \Gamma) \leq -J(\Gamma). \tag{10}$$

- The proof will be over as soon as we prove that (10) holds with the rate I instead of J. The weak topology of $\mathcal{P}(E)$ can be defined via Dudley's distance : for μ and ν in $\mathcal{P}(E)$,

$$d_{Dudley}(\mu,\nu) \stackrel{\triangle}{=} \inf\left\{\left|\int f d\mu - \int f d\nu\right|;\ f \in \mathcal{LB}(E) \text{ and } \|f\|_{\mathcal{LB}} \leq 1\right\}.$$

The function $\zeta \in \mathcal{P}(E^2) \mapsto (\zeta^1,\zeta^2) \in \mathcal{P}(E)\times\mathcal{P}(E)$ is continuous and so is $\zeta \mapsto d_{Dudley}(\zeta^1,\zeta^2)$. We then introduce the following closed subsets :

$$\Delta_n \stackrel{\triangle}{=} \{\zeta \in \mathcal{P}(E^2);\ d_{Dudley}(\zeta^1,\zeta^2) \leq \frac{2}{n}\} \ \text{ and } \ \Delta \stackrel{\triangle}{=} \{\zeta \in \mathcal{P}(E^2);\ \zeta^1 = \zeta^2\}.$$

The sequence $(\Gamma\cap\Delta_n)_n$ of compact sets decreases to the compact set $\Gamma\cap\Delta$. Since J is l.s.c., we have $I(\Gamma) = J(\Gamma\cap\Delta) = \lim_n \uparrow J(\Gamma\cap\Delta_n)$. On the other hand, for any $n\in\mathbb{N}$, $L_n \in \Delta_n$. Consequently, for any integer m, formula (10) yields

$$\limsup_{n\to\infty} \sup_{\nu\in\mathcal{P}(E)} \frac{1}{n} \log \mathbb{P}_\nu(L_n \in \Gamma) = \limsup_{n\to\infty} \sup_{\nu\in\mathcal{P}(E)} \frac{1}{n} \log \mathbb{P}_\nu(L_n \in \Gamma\cap\Delta_m) \leq -J(\Delta\cap\Gamma_m),$$

and the proof is over by letting $m \to \infty$. $\diamondsuit$

7.2 Weak LDP for balanced couples

Taking advantage of the result proved in section 7.1, we can easily check this altered version of Theorem 2:

Theorem 4.
Assume that p is fellerian. If $G = (G^{(1)}, G^{(2)})$ is a Lyapunov balanced couple on E^2, then $((\overline{G}_n^{(1)}, \overline{G}_n^{(2)}))_n$ satisfies a uniform weak upper LDP for every initial distribution with the weak rate function $h_G^*(\cdot)$.
In other words, for any compact subset K of $\mathbb{R}_+ \times B$,

$$\limsup_{n\to\infty} \sup_{\nu\in\mathcal{P}(E)} \frac{1}{n} \log \mathbb{P}_\nu\left((\overline{G}_n^{(1)}, \overline{G}_n^{(2)}) \in K\right) \leq -h_G^*(K).$$

In particular, if $R \in]0,\infty[$ and if C is a compact set in B then

$$\limsup_{n\to\infty} \sup_{\nu\in\mathcal{P}(E)} \frac{1}{n} \log \mathbb{P}_\nu \left(\overline{G}_n^{(1)} \le R, \overline{G}_n^{(2)} \in C \right) \le -h_G([0,R]\times C).$$

Acknowledgments I would like to thank Professor Marie Duflo and Julien Worms for their very useful help. I am grateful to the referees Amir Dembo and Ted Cox for their helping and valuable comments. The author also thanks Rémi Leclercq for his careful reading of the paper.

References

[Aco85] A. De Acosta. Upper bounds for large deviations of dependent random vectors. *Z. Wahrscheinlichkeitstheorie*, 69:551–565, 1985.

[Aco88] A. De Acosta. Large deviations for vector-valued Functionals of a Markov chain: Lower bounds. *Ann. Prob.*, 16:925–960, 1988.

[Aco90] A. De Acosta. Large deviations for empirical measures of Markov chains. *J. of Theoretical Prob.*, 3:395–431, 1990.

[AcoNey98] A. De Acosta and P. Ney. Large deviation lower bounds for arbitrary additive functionals of a Markov chain. *Ann. Prob.*, 26:1660–1682, 1998.

[Att] J. G. Attali. *1. Méthodes de stabilité pour des chaînes de Markov non fellériennes, 2. Sur quelques autres problèmes issus des réseaux de neurones, Thèse,* Université Paris 1, 1999

[Aze] R. Azencott. Grandes déviations et applications. *Ecole de Proba. de Saint-Flour VIII,* number 774 in lectures notes in math., pages 1-176. Springer, 1980.

[BahZab79] R. Bahadur and S. Zabell. Large deviations for the sample mean in general vector spaces. *Ann. Prob.*, 7:587-621, 1979.

[BJV91] J.R. Baxter, N.C. Jain and S.R.S. Varadhan. Some familiar examples for which the large deviation principle does not hold. *Commun. Pure Appl. Math.*, 911–923, 1991.

[BerGamRou96] B. Bercu, F. Gamboa and A. Rouault. Large deviations for quadratic forms of stationary gaussian processes *Stoch. Proc. and their Appl.*, 71:75–90, 1997.

[BryDem97] W. Bryc and A. Dembo. Large deviations for quadratic Functionals of gaussian processes. *J. of Theoretical Prob.*, 10:307–322, 1997.

[DemSha98a] A. Dembo and Q. M. Shao. Self-normalized moderate deviations and LILs. *Stoch. Proc. and their Appl.*, 75:51–65, 1998.

[DemSha98b] A. Dembo and Q. M. Shao. Self-normalized large deviations in vector spaces. *Progress in probability proceeding Oberwolfach*, 43:27–32, 1998.

[DemZei] A. Dembo and O. Zeitouni *Large deviations Techniques and Applications.* Springer, 1998.

[Din93] I. H. Dinwoodie. Identifying a large deviation rate function. *Ann. Prob.*, 23 : 216–231, 1993.

[DinZab92] I. H. Dinwoodie and S. L. Zabell. Large deviations for exchangeable random vectors. *Ann. Prob.*, 3:1147–1166, 1992.

[DonVar75a] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time I. *Commun. Pure Appl. Math.*, 28:1–47, 1975.

[DonVar76] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time III. *Commun. Pure Appl. Math.*, 29:389–461, 1976.

[Duf] M. Duflo. *Random Iterative Models*, Volume 38 de *Applications of mathematics.* Springer, 1997.

[DupEll] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations.* Wiley, 1997.

[GulLipLot94] O. V. Gulinskii, R. S. Liptser and S. V. Lototskii. Large deviations for unbounded additive Functionals of a Markov process with discrete time (non compact case). *J. Appl. Math. Stoch. Anal.*, 7(3):423–436, 1994.

[HeSha96] X. He and Q. M. Shao. Bahadur efficiency and robustness of Studentized score tests. *Ann Inst. Statist. Math.*, 48:295–314, 1996.

[Jai90] N. Jain. Large deviations for additive Functionals of Markov processes: discrete time, non compact case. *Ann Prob.*, 18:1071–1098, 1990.

[NeyNum87] P. Ney and E. Nummelin. Markov additive processes (ii): Large deviations. *Ann. Prob.*, 15:593–609, 1987.

[Sha97] Q. M. Shao. Self-normalized large deviations. *Ann. Prob.*, 25:285–328, 1997.

[Wor99] J. Worms. Moderate deviations for stable Markov chains and regression models. *Electronic journal of Probability* , 4:1–28, 1999

[Wor00] J. Worms. *Principes de déviations modérées pour des martingales et applications statistiques, Thèse* Université de Marne-la-Vallée, France, 2000