Probab. Theory Relat. Fields 123, 301–322 (2002)
Digital Object Identifier (DOI) 10.1007/s004400200204

P. Collet · S. Martinez · B. Schmitt

Exponential inequalities for dynamical measures of expanding maps of the interval

Received: 14 August 2001 / Revised version: 13 February 2002 /
Published online: 1 July 2002 –

Abstract. We prove an exponential inequality for the absolutely continuous invariant measure of a piecewise expanding map of the interval. As an immediate corollary we obtain a concentration inequality. We apply these results to the estimation of the rate of convergence of the empirical measure in various metrics and also to the efficiency of the shadowing by sets of positive measure.

I. Introduction

Considerable progress has been made recently by Talagrand and others on the concentration properties in product spaces [T1,T2,T3], with striking applications to various areas of Probability theory and Statistics. These results were subsequently developed by several authors ([Mas1.,Mas2.], [Ri1.], [Dem.] among others). We refer to [Le.] for nice reviews and more references. The case of dependent random variables has been investigated more recently. First for Markov chains in [Mar1,Mar2] and then for more general processes in [Mar3.], [Sa.], [Ri2.]. Unfortunately all these papers assume some properties of the correlations which are too strong to be applied to the case of piecewise non Markov expanding maps of the interval. In particular they are neither topologically Markov nor Φ mixing. The main reason is that the forward transition is of course deterministic while the backward transitions are represented by atomic measures. On the other hand these maps have correlations which can be controlled in a suitable topology and we will see below that concentration can be also proven in this case with dynamical applications.

P. Collet: Centre de Physique Théorique, CNRS UMR 7644, Ecole Polytechnique, 91128 Palaiseau Cedex, France. e-mail: collet@cpht.polytechnique.fr

S. Martinez: Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático UMR 2071 UCHILE-CNRS, Casilla 170-3 Correo 3, Santiago, Chile.
e-mail: smartine@dim.uchile.cl

B. Schmitt: Université de Bourgogne, Département de Mathématiques, Faculté des Sciences Mirande, BP-138, 21004 Dijon Cedex, France. e-mail: schmittb@u-bourgogne.fr

Mathematics Subject Classification (2000): 37C30, 60F10, 60E15

Key words or phrases: Empirical measures – Perron-Frobenius operators – Decay of correlations

From now on we will only consider the following situation although we expect the results to be true in more general contexts. Let f be a piecewise regular expanding map of the interval [0, 1] which is topologically mixing. More precisely, we assume there exists a finite partition $\mathcal{A}$ of [0, 1] by intervals where f is regular and monotonous. Moreover, we assume there exists a number $A > 0$ and a number $\rho > 1$ such that for any integer n

$$\inf_{x\in[0,1]} \left|f^{n\prime}(x)\right| \geq A\rho^n .$$

It is well known that there is a unique ergodic absolutely continuous invariant probability measure $d\mu = \varphi dx$ ([L.Y.]). In the sequel, we will assume that the density φ of the invariant measure is bounded below away from zero. We refer to [B.G.R.], appendix B of [Bu.] and [H.] for such statements.

We recall that the transfer operator $\mathcal{L}$ associated to f is given by (see [H.K.] and references therein)

$$\mathcal{L}g(x) = \sum_{z\,,\,f(z)=x} \frac{g(z)}{\left|f'(z)\right|} .$$

We recall that $\mathcal{L}$ is the dual in $L^2([0, 1], dx)$ of the Koopman operator acting on functions by composition with f. We will mostly use the operator L conjugated to $\mathcal{L}$ defined by

$$Lg(x) = \frac{1}{\varphi(x)} \sum_{z\,,\,f(z)=x} \frac{\varphi(z)}{\left|f'(z)\right|} g(z) .$$

L has the following spectral properties in the Banach space BV of functions of bounded variation equipped with the norm

$$\|u\| = \vee u + \int |u(x)|dx .$$

First of all 1 is a simple eigenvalue with eigenvector the constant function and left eigenvector the invariant measure μ. Moreover, the rest of the spectrum is contained in a closed disk of radius strictly smaller than one. In particular, there exists a constant $K > 0$ and a constant $0 \leq \xi < 1$ such that for any function g of bounded variation, we have for any integer n

$$L^n g = \int g d\mu + g_n$$

where the functions of bounded variation g_n satisfies

$$\vee g_n + \|g_n\|_\infty \leq K\xi^n \big(\vee g + \|g\|_\infty \big) . \tag{I.1}$$

This estimate implies the decay of correlations for observables in a suitable function space. Namely there is a constant $K' > 0$ such that for any integrable function g_1 and any function g_2 of bounded variation, we have for any integer k the decay of correlations

$$\left|\int_0^1 d\mu(x) g_1\big(f^k(x)\big) g_2(x) - \int_0^1 d\mu(x) g_1(x) \int_0^1 d\mu(x) g_2(x)\right|$$
$$\leq K' \xi^k \|g_1\|_1 \left(\vee g_2 + \|g_2\|_1\right) \ , \tag{I.2}$$

where ξ is the positive number smaller than one appearing in (I.1). We refer to [H.K.] for the proof of these statements for the operator $\mathcal{L}$. Since the operator of multiplication by φ is bounded in BV together with its inverse (since φ is bounded away from zero), we conclude that the same spectral results hold for L.

We will use later an extension of L (also denoted by L) mapping a function $u\big(x_1, \cdots, x_n\big)$ of n variables to a function of $n-1$ variables and given by

$$Lu\big(x_1, \cdots, x_{n-1}\big) = \frac{1}{\varphi(x_1)} \sum_{z\,,\, f(z)=x_1} \frac{\varphi(z)}{\big|f'(z)\big|} u\big(z, x_1, \cdots, x_{n-1}\big) \ .$$

It is immediate to verify that if the function of one variable v is given by

$$v(x) = u\big(x, f(x), \cdots, f^{n-1}(x)\big) \ ,$$

then

$$Lv(x) = (Lu)\big(x, f(x), \cdots, f^{n-2}(x)\big) \ .$$

Moreover, if u is a function of n variables and $k < n$, we have

$$L^k u\big(x_1, \cdots, x_{n-k}\big)$$
$$= \frac{1}{\varphi(x_1)} \sum_{z\,,\, f^k(z)=x_1} \frac{\varphi(z)}{\big|f^{k'}(z)\big|} u\big(z, f(z), \cdots, f^{k-1}(z), x_1, \cdots, x_{n-k}\big) \ .$$

A real valued function u on $[0, 1]^n$ will be said to be separately Lipschitz if the Lipschitz constants defined for $1 \leq l \leq n$ by

$$\mathrm{Lip}_l(u) = \sup_{x_1,\cdots,x_{l-1},x_l,x_{l+1},\cdots,x_n} \ \sup_{y_l \neq x_l}$$
$$\frac{\big|u(x_1, \cdots, x_{l-1}, x_l, x_{l+1}, \cdots, x_n) - u(x_1, \cdots, x_{l-1}, y_l, x_{l+1}, \cdots, x_n)\big|}{|x_l - y_l|}$$

are all finite.

By abuse of notation, if u is a function of n variables, we will denote by $\mu(u)$ the number

$$\mu(u) = \int u\big(x, f(x), \cdots, f^{n-1}(x)\big)\, d\mu(x) \ .$$

Our first goal is to prove the following result.

Theorem I.1. *There is a constant $C > 0$ such that for any integer n, for any separately Lipschitz real valued function u of n variables, we have*

$$\int_0^1 e^{u\left(x, f(x), \cdots, f^{n-1}(x)\right)} d\mu(x) \leq e^{\mu(u)} e^{C \sum_1^n \left(\mathrm{Lip}_l(u)\right)^2} \ .$$

Using Chebyshev inequality, we can easily derive the following concentration result.

Corollary I.2. *Under the above hypothesis, we have for any $t > 0$*

$$\mu\left(\left\{x \mid u\big(x, f(x), \cdots, f^{n-1}(x)\big) > \mu(u) + t\right\}\right) \le e^{-t^2/\left(4C\sum_1^n (\mathrm{Lip}_l(u))^2\right)} .$$

We will below use sometimes the combination of this estimate with the corresponding one for the function $-u$ which leads immediately to

$$\mu\left(\left\{x \mid \left|u\big(x, f(x), \cdots, f^{n-1}(x)\big) - \mu(u)\right| > t\right\}\right) \le 2e^{-t^2/\left(4C\sum_1^n (\mathrm{Lip}_l(u))^2\right)} .$$

Note in particular that in the above results the function u is a function of n independent variables, and not only a function constant on the dynamical cylinders of a finite partition. The above results can also be interpreted in terms of a (sequence of) measures (μ_n) on $[0, 1]^n$ given by

$$d\mu_n(x_1, \cdots, x_n) = d\mu(x_1) \prod_{j=2}^{n} \delta\big(x_j - f^{j-1}(x_1)\big) .$$

By a change of variables, this measure is also given by

$$d\mu_n(x_1, \cdots, x_n) = d\mu(x_n) \sum_{f^{n-1}(z)=x_n} \frac{1}{\left|f^{n-1\prime}(z)\right|} \prod_{j=1}^{n-1} \delta\big(x_j - f^{j-1}(z)\big) .$$

An easy consequence of Theorem I.1 is now the following estimate (see [D.] for the independent case)

$$\mathrm{Var}_{\mu_n}(u) = \int \big(u - \mu_n(u)\big)^2 d\mu_n \le 2C \sum_1^n \big(\mathrm{Lip}_l(u)\big)^2 .$$

This follows at once by replacing u by λu in Theorem I.1, multiplying both sides of the estimate by $e^{-\lambda\mu(u)}$ subtracting 1 to both sides then dividing by λ^2, and letting λ tend to zero.

Another interesting consequence is an information inequality. If ν_1 and ν_2 are two probability measures on $[0, 1]^n$, we recall that their Kantorovich distance $\kappa\big(\nu_1, \nu_2\big)$ is given by

$$\kappa\big(\nu_1, \nu_2\big) = \inf_{\pi} \int \|x_1 - x_2\| \, d\pi(x_1, x_2)$$

where the infimum is over all couplings between ν_1 and ν_2. The information divergence of ν_1 with respect to ν_2 is given by

$$D\big(\nu_1 || \nu_2\big) = \int \log\left(\frac{d\nu_1}{d\nu_2}\right) d\nu_1 .$$

It then follows from Theorem 3.1 in [B.G.] and Theorem I.1 that for any probability measure ν on $[0, 1]^n$

$$\kappa\big(\mu_n, \nu\big) \le \sqrt{2CnD\big(\mu_n||\nu\big)}\ .$$

Concentration results follow from this inequality (see [B.G.]).

The rest of this paper is organized as follows. The proof of Theorem I.1 is given in section II. In section III, we apply the result to the estimation of the rate of convergence of the empirical measure to the invariant measure μ in different metrics. We obtain estimates which are valid for finite samples, not only asymptotically. In Section IV, we give applications to the shadowing by orbits of a given set. In the appendix, a variant of our result is applied to study the rate of convergence of the empirical measure in the Kolmogorov metric.

In the sequel we will sometimes use the same letter to denote different constants.

II. The exponential inequality

In this section we give a proof of Theorem I.1 in the spirit of the so called martingale method of Azuma and Mac Diarmid (see [Dev.] for the case of independent variables and references). We start by recalling the classical Hoeffding inequality (see for example [Dev.]).

Lemma II.1. *Let ν be a probability measure on a space Y, and let g be a real valued bounded measurable function on Y. Then we have*

$$\int_Y e^{g(y)-\nu(g)} d\nu(y) \le e^{(\mathrm{osc}(g))^2/8}$$

where

$$\mathrm{osc}(g) = \sup_{y,\ y'\in Y} \big(g(y) - g(y')\big)\ .$$

Note that we could apply directly this Lemma to try to prove Theorem I.1. We get however a much worse estimate. To get a better estimate, we will use this Lemma recursively through the following result.

Lemma II.2. *If u is a real valued measurable bounded function of n variables, we have*

$$\int_0^1 e^{u\left(x, f(x), \cdots, f^{n-1}(x)\right)} d\mu(x) \le e^{\left(\mathrm{osc}_1(u)\right)^2/8} \int_0^1 e^{Lu\left(x, f(x), \cdots, f^{n-2}(x)\right)} d\mu(x)$$

where

$$\mathrm{osc}_1(u) = \sup_{x_1, x_1'} \sup_{x_2, \cdots, x_{n-1}} u\big(x_1, x_2, \cdots, x_{n-1}\big) - u\big(x_1', x_2, \cdots, x_{n-1}\big)\ .$$

The proof of this Lemma follows easily from the previous result. We observe that

$$\begin{aligned}
&\int_0^1 e^{u\left(x, f(x), \cdots, f^{n-1}(x)\right)} d\mu(x) \\
&= \int_0^1 e^{u\left(x, f(x), \cdots, f^{n-1}(x)\right) - Lu\left(f(x), f^2(x), \cdots, f^{n-1}(x)\right)} e^{Lu\left(f(x), f^2(x), \cdots, f^{n-1}(x)\right)} d\mu(x) \\
&= \int_0^1 L\left(e^{u\left(\cdot, f(\cdot), \cdots, f^{n-1}(\cdot)\right) - Lu\left(f(\cdot), f^2(\cdot), \cdots, f^{n-1}(\cdot)\right)}\right)(x)\, e^{Lu\left(x, f(x), \cdots, f^{n-2}(x)\right)} d\mu(x)
\end{aligned}$$

where the last equality follows from the fact that L is the dual operator to the composition with f with respect to the measure μ.

For a fixed $x \in [0, 1]$ we denote by Y the (finite) set of preimages of x (this set depends of course on x). We now observe that (since $L1 = 1$) the sum over $y \in Y$ of the (non negative) numbers $\varphi(y)/(\varphi(x)|f'(y)|)$ is equal to one, and this defines a probability measure ν on Y (which depends on x). Therefore in the expression

$$\begin{aligned}
&L\left(e^{u\left(\cdot, f(\cdot), \cdots, f^{n-1}(\cdot)\right) - Lu\left(f(\cdot), f^2(\cdot), \cdots, f^{n-1}(\cdot)\right)}\right)(x) \\
&\quad = \frac{1}{\varphi(x)} \sum_{f(y)=x} \frac{\varphi(y)}{|f'(y)|} e^{u\left(y, x, \cdots, f^{n-2}(x)\right) - Lu\left(x, f(x), \cdots, f^{n-2}(x)\right)}
\end{aligned}$$

we can apply Lemma II.1 to the function

$$g(y) = u\big(y, x, \cdots, f^{n-2}(x)\big)$$

observing that

$$\nu(g) = Lu\big(x, f(x), \cdots, f^{n-2}(x)\big)\ .$$

Lemma II.2 follows immediately.

If we apply iteratively $n - 1$ times this estimate we get

$$\begin{aligned}
&\int_0^1 e^{u\left(x, f(x), \cdots, f^{n-1}(x)\right)} d\mu(x) \le e^{(1/8)\sum_{j=0}^{n-2}\left(\mathrm{osc}_1(L^j u)\right)^2} \\
&\quad \int_0^1 e^{\left(L^{n-1}u\right)(x)} d\mu(x) \qquad\qquad \text{(II.1)} \\
&\quad \le e^{(1/8)\sum_{j=0}^{n-1}\left(\mathrm{osc}_1(L^j u)\right)^2} e^{\mu(L^{n-1}u)} = e^{(1/8)\sum_{j=0}^{n-1}\left(\mathrm{osc}_1(L^j u)\right)^2} e^{\mu(u)}\ .
\end{aligned}$$

We have used in last inequalities that $L^{n-1}(u)$ depends only on one variable and we have applied Lemma II.1 with $Y = [0, 1]$, $\nu = \mu$, $g = L^{n-1}(u)$ and the equality $\mu(L^{n-1}u) = \mu(u)$.

In order to prove Theorem I.1, we have to estimate each term on the right hand side. The main tool will be the following Lemma.

Lemma II.3. *There is a constant $D > 0$ and a constant $0 \le \sigma < 1$ such that for any integer n, for any separately Lipschitz real valued function u of n variables, we have for any $0 \le k < n$*

$$\mathrm{osc}_1\big(L^k(u)\big) \le D \sum_{j=1}^{k+1} \sigma^{k+1-j} \mathrm{Lip}_j(u)\ .$$

Moreover

$$\mathrm{osc}_1\big(L^n(u)\big) \le D \sum_{j=1}^{n} \sigma^{n-j} \mathrm{Lip}_j(u)\ .$$

Recall that from the definition of L we have

$$\begin{aligned} &L^k u\big(x_1, \cdots, x_{n-k}\big) \\ &\quad = \frac{1}{\varphi(x_1)} \sum_{z\,,\, f^k(z)=x_1} \frac{\varphi(z)}{\big|f^k(z)'\big|} u\big(z, f(z), \cdots, f^{k-1}(z), x_1, \cdots, x_{n-k}\big)\ . \end{aligned}$$

Assuming for the moment Lemma II.3, we finish the proof of Theorem I.1. First of all, we have using Schwartz inequality

$$\begin{aligned} \sum_{k=0}^{n-1} \big(\mathrm{osc}_1(L^k u)\big)^2 &\le D^2 \sum_{k=0}^{n-1} \left(\sum_{j=1}^{k+1} \sqrt{\sigma^{k+1-j}} \sqrt{\sigma^{k+1-j}} \mathrm{Lip}_j(u) \right)^2 \\ &\le D^2 \sum_{k=0}^{n-1} \left(\sum_{j'=1}^{k+1} \sigma^{k+1-j'} \sum_{j=1}^{k+1} \sigma^{k+1-j} \big(\mathrm{Lip}_j(u)\big)^2 \right) \le \frac{D^2}{(1-\sigma)^2} \sum_{j=1}^{n} \mathrm{Lip}_j(u)^2\ . \end{aligned}$$

Theorem I.1 then follows using similarly the last part of Lemma II.3.

In order to prove Lemma II.3, we will need the following result. Recall that $\mathcal{A}$ is a finite partition by intervals of $[0, 1]$ such that f is regular on each atom of $\mathcal{A}$. Denote by $\big(\mathcal{A}_l\big)$ the sequence of partitions given by

$$\mathcal{A}_l = \bigvee_{j=0}^{l} f^{-j}\mathcal{A}\ .$$

Lemma II.4. *There is a finite constant $C > 0$ such that for any integer l we have*

$$\sum_{I \in \mathcal{A}_l} \sup_{x \in I} \frac{1}{|f^{l\prime}(x)|} \le C\ .$$

Proof. If $I \in \mathcal{A}_l$, there is a smallest integer $q_I \le l$ such that $f^{q_I}(I) \cap \partial\mathcal{A} \ne \emptyset$, and denote by $\mathcal{A}_l^p$ the collections of atoms of $\mathcal{A}_l$ with $q_I = p$. We have

$$\sum_{I \in \mathcal{A}_l} \sup_{x \in I} \frac{1}{|f^{l\prime}(x)|} = \sum_{p=0}^{l} \sum_{I \in \mathcal{A}_l^p} \sup_{x \in I} \frac{1}{|f^{l\prime}(x)|}\ .$$

Recall that from the distortion Lemma (see [L.Y.]), there is a constant $C_1 \geq 1$ such that for any integer l, for any $I \in \mathcal{A}_l$, if a is a point in the boundary of I, we have for any $x \in I$, and any integer $k \leq l$

$$\frac{1}{C_1}|f^{k'}(a)| \leq |f^{k'}(x)| \leq C_1|f^{k'}(a)| \,.$$

We now observe that since $\mathcal{A}$ is a partition of monotonicity for f, for any $I \in \mathcal{A}_l^p$ there exists $z \in \partial I$ such that $b = f^p(z) \in \partial\mathcal{A}$. Therefore for any $x \in I$ and $p \leq k \leq l$ we have

$$\left|f^{k'}(x)\right| \geq \frac{1}{C_1}\left|f^{k'}(z)\right| \geq \frac{1}{C_1}\left|(f^{k-p})'(b)\right| \left|f^{p'}(z)\right| \,.$$

Since a preimage of order p of $b \in \partial\mathcal{A}$ is contained in the boundary of at most two elements of $\mathcal{A}_l^p$, we have

$$\sum_{I\in\mathcal{A}_l} \sup_{x\in I} \frac{1}{|f^{l'}(x)|} \leq 2C_1 \sum_{p=0}^{l} \sum_{b\in\partial\mathcal{A}} \frac{1}{\left|(f^{l-p})'(b)\right|} \sum_{f^p(z)=b} \frac{1}{|f^{p'}(z)|}$$
$$\leq \mathcal{O}(1) \sum_{p=0}^{l} \rho^{p-l} \sum_{b\in\partial\mathcal{A}} \mathcal{L}^p 1(b) \,,$$

where $\mathcal{L}$ is the usual transfer operator for the map f. It follows from the spectral theory of $\mathcal{L}$ (see for example [H.K.]) that this quantity is bounded uniformly in l. This finishes the proof of the Lemma.

Proof of Lemma II.3. We first estimate for each fixed $(r_1, \cdots, r_{n-k-1})$ in $[0, 1]^{n-k-1}$ the quantity

$$\sup_{y,y'} \left(u_k(y) - u_k(y')\right) \,,$$

where

$$u_k(y) = L^k u\big(y, r_1, \cdots, r_{n-k-1}\big) =$$
$$\frac{1}{\varphi(y)} \sum_{z\,,\, f^k(z)=y} \frac{\varphi(z)}{\left|f^{k'}(z)\right|} u\big(z, f(z), \cdots, f^k(z), r_1, \cdots, r_{n-k-1}\big) \,.$$

We define for $1 \leq l \leq k$ the sequence of functions of l variables

$$v_l^k\big(x_1, \cdots, x_l\big) = \int u\big(x_1, \cdots, x_l, s, f(s), \cdots, f^{k-l}(s), r_1, \cdots, r_{n-k-1}\big) d\mu(s) \,,$$

and for $k \geq 0$

$$v_0^k = \int u\big(s, f(s), \cdots, f^k(s), r_1, \cdots, r_{n-k-1}\big) d\mu(s) \,.$$

For convenience, we also introduce the notation

$$v_{k+1}^k\big(x_1, \cdots, x_{k+1}\big) = u\big(x_1, \cdots, x_k, x_{k+1}, r_1, \cdots, r_{n-k-1}\big) \,.$$

We observe that

$$u\big(z, f(z), \cdots, f^k(z), r_1, \cdots, r_{n-k-1}\big)$$
$$= v_0^k + \sum_{l=0}^{k} \Big[v_{l+1}^k\big(z, f(z), \cdots, f^l(z)\big) - v_l^k\big(z, f(z), \cdots, f^{l-1}(z)\big)\Big] \ ,$$

and therefore since v_0^k is a constant (hence $Lv_0^k = v_0^k$) we obtain the identity

$$u_k(x) = v_0^k + \sum_{l=0}^{k} \frac{1}{\varphi(x)} \sum_{\substack{z \\ f^k(z)=x}} \frac{\varphi(z)}{\big|f^{k\prime}(z)\big|}$$
$$\Big[v_{l+1}^k\big(z, f(z), \cdots, f^l(z)\big) - v_l^k\big(z, f(z), \cdots, f^{l-1}(z)\big)\Big].$$

We now define for $0 \le l \le k$

$$w_l^k(y) = \frac{1}{\varphi(y)} \sum_{z\,,\, f^l(z)=y} \frac{\varphi(z)}{\big|f^{l\prime}(z)\big|}$$
$$\Big[v_{l+1}^k\big(z, f(z), \cdots, f^l(z)\big) - v_l^k\big(z, f(z), \cdots, f^{l-1}(z)\big)\Big] \ .$$

and using the chain rule, we get

$$u_k(x) = v_0^k + \sum_{l=0}^{k} \Big(L^{k-l} w_l^k\Big)(x) \ . \tag{II.2}$$

In order to be able to exploit the spectral properties of L, we have to estimate the L^∞ norm of w_l^k and also its variation. We first observe that since μ is an invariant measure, we have

$$v_{l+1}^k\big(x_1, \cdots, x_{l+1}\big)$$
$$= \int u\big(x_1, \cdots, x_{l+1}, s, f(s), \cdots, f^{k-l-1}(s), r_1, \cdots, r_{n-k-1}\big) d\mu(s)$$
$$= \int u\big(x_1, \cdots, x_{l+1}, f(s), f^2(s), \cdots, f^{k-l}(s), r_1, \cdots, r_{n-k-1}\big) d\mu(s) \ .$$

Therefore

$$v_{l+1}^k\big(z, \cdots, f^l(z)\big) - v_l^k\big(z, \cdots, f^{l-1}(z)\big)$$
$$= \int \Big(u\big(z, \cdots, f^{l-1}(z)\ , f^l(z), f(s), f^2(s), \cdots, f^{k-l}(s), r_1, \cdots, r_{n-k-1}\big)$$
$$- u\big(z, \cdots, f^{l-1}(z), s,\ f(s), f^2(s), \cdots, f^{k-l}(s), r_1, \cdots, r_{n-k-1}\big)\Big)\, d\mu(s)$$

and by the Lipschitz hypothesis the modulus of this quantity is bounded by

$$\Big| v_{l+1}^k\big(z, \cdots, f^l(z)\big) - v_l^k\big(z, \cdots, f^{l-1}(z)\big)\Big|$$
$$\le \mathrm{Lip}_{l+1}(u) \sup_{z\in[0,1]} \int \big|s - f^l(z)\big| d\mu(s) \le \mathrm{Lip}_{l+1}(u) \ , \tag{II.3}$$

since μ is a probability measure on [0, 1]. It follows immediately that

$$\left\| w_l^k \right\|_\infty \le \mathrm{Lip}_{l+1}(u)\ .$$

We now come to the estimate on the variation. This estimate is reminiscent of the estimate of Lasota and Yorke but there are some major differences. First of all we have

$$\begin{aligned}\vee w_l^k &\le \bigvee\left(\frac{1}{\varphi}\right)\|\varphi\, w_l^k\|_\infty + \left\|\frac{1}{\varphi}\right\|_\infty \vee(\varphi w_l^k)\\ &\le \vee\varphi \left\|\frac{1}{\varphi^2}\right\|_\infty \|\varphi\, w_l^k\|_\infty + \left\|\frac{1}{\varphi}\right\|_\infty \vee(\varphi w_l^k)\ .\end{aligned}$$

By our previous bound and since φ is of bounded variation and bounded below, we have

$$\vee w_l^k \le \mathcal{O}(1)\mathrm{Lip}_{l+1}(u) + \mathcal{O}(1)\vee(\varphi w_l^k)\ .$$

Recall that

$$\varphi(x) w_l^k(x) = \sum_{z\,,\, f^l(z)=x} \frac{\varphi(z)}{\left|f^{l\prime}(z)\right|} S_l^k(z)$$

where

$$S_l^k(z) = \left[v_{l+1}^k\big(z, f(z), \cdots, f^l(z)\big) - v_l^k\big(z, f(z), \cdots, f^{l-1}(z)\big)\right]\ .$$

Since f^l is injective on each atom I of $\mathcal{A}_l$ we can introduce the inverse function ψ_I from $f^l(I)$ to I. We have

$$\varphi(x) w_l^k(x) = \sum_{I\in\mathcal{A}_l} \frac{\varphi(\psi_I(x))\chi_{f^l(I)}(x)}{\left|f^{l\prime}(\psi_I(x))\right|} S_l^k(\psi_I(x))\ .$$

We now have

$$\vee\big(\varphi w_l^k\big) \le T_1 + T_2 + T_3 + T_4$$

where

$$\begin{aligned}T_1 &= \sum_{I\in\mathcal{A}_l} \bigvee_{f^l(I)} (\varphi\circ\psi_I) \left\|\frac{1}{f^{l\prime}\circ\psi_I}\right\|_{L^\infty(I)} \left\|S_l^k\circ\psi_I\right\|_{L^\infty(I)}\\ T_2 &= \sum_{I\in\mathcal{A}_l} \|\varphi\circ\psi_I\|_{L^\infty(I)} \bigvee_{f^l(I)} \left(\frac{1}{f^{l\prime}\circ\psi_I}\right) \left\|S_l^k\circ\psi_I\right\|_{L^\infty(I)}\\ T_3 &= \sum_{I\in\mathcal{A}_l} \|\varphi\circ\psi_I\|_{L^\infty(I)} \left\|\frac{1}{f^{l\prime}\circ\psi_I}\right\|_{L^\infty(I)} \bigvee_{f^l(I)} \big(S_l^k\circ\psi_I\big)\\ T_4 &= 2\sum_{a\in\partial\mathcal{A}_l} \frac{\varphi(a)}{\left|f^{l\prime}(a)\right|}\left|S_l^k(a)\right|\ .\end{aligned}$$

The factor 2 in the last term comes from the fact that each boundary point of a segment in $\mathcal{A}_l$ appears twice, once as a left end point and another time as a right end point. We now control each term separately. Using the previous estimates we have easily

$$T_1 \leq \mathcal{O}(1)\mathrm{Lip}_{l+1}(u) \sum_{I\in\mathcal{A}_l} \vee_I \varphi \leq \mathcal{O}(1)\mathrm{Lip}_{l+1}(u) \ .$$

We now come to the estimation of the term T_2. A classical estimate using distortion (see for example [L.Y.]) shows that

$$\bigvee_I \left(\frac{1}{f^{l'}}\right) \leq \mathcal{O}(1) \sup_{x\in I} \left|\frac{1}{f^{l'}(x)}\right| \ .$$

Therefore

$$T_2 \leq \mathcal{O}(1)\mathrm{Lip}_{l+1}(u) \sum_{I\in\mathcal{A}_l} \sup_{x\in I} \frac{1}{|f^{l'}(x)|} \leq \mathcal{O}(1)\mathrm{Lip}_{l+1}(u)$$

by Lemma II.4. In order to estimate the third term T_3, it is enough to estimate the variation of v^k_{l+1} (and of v^k_l but the argument is similar) on an interval I. This is where our Lipschitz hypothesis is crucial. For an increasing sequence of points $a_1, \cdots, a_r$ in $I \in \mathcal{A}_l$ we have

$$\begin{aligned}
&\sum_{s=1}^{r-1} \left|v^k_{l+1}(a_{s+1}, f(a_{s+1}), \cdots, f^l(a_{s+1})) - v^k_{l+1}(a_s, f(a_s), \cdots, f^l(a_s))\right| \\
&\leq \sum_{s=1}^{r-1}\sum_{j=0}^{l} \Big|v^k_{l+1}(a_s, f(a_s), \cdots, f^{j-1}(a_s), f^j(a_{s+1}), f^{j+1}(a_{s+1}), \cdots, f^l(a_{s+1})) \\
&\quad -v^k_{l+1}(a_s, f(a_s), \cdots, f^j(a_s), f^{j+1}(a_{s+1}), f^{j+2}(a_{s+1}) \cdots, f^l(a_{s+1}))\Big| \\
&\leq \sum_{j=0}^{l} \mathrm{Lip}_{j+1}(u) \sum_{s=1}^{r-1} \left|f^j(a_{s+1}) - f^j(a_s)\right| \leq \sum_{j=0}^{l} \mathrm{Lip}_{j+1}(u)\left|f^j(I)\right| \leq \mathcal{O}(1) \\
&\quad \sum_{j=0}^{l} \mathrm{Lip}_{j+1}(u)\rho^{-l+j} \ .
\end{aligned}$$

Using Lemma II.4, the third term T_3 is therefore bounded by

$$T_3 \leq \mathcal{O}(1) \sum_{j=1}^{l+1} \mathrm{Lip}_j(u)\rho^{-l+j} \ .$$

The last term T_4 is then bounded using again Lemma II.4, namely

$$T_4 \leq \mathcal{O}(1)\mathrm{Lip}_{l+1}(u) \sum_{I\in\mathcal{A}_l} \sup_{x\in I} \frac{1}{|f^{l'}(x)|} \leq \mathcal{O}(1)\mathrm{Lip}_{l+1}(u) \ .$$

We now have proven the estimate

$$\vee w_l^k \le \mathcal{O}(1) \sum_{j=0}^{l} \mathrm{Lip}_{j+1}(u) \rho^{-l+j} \ .$$

We can now come back to the expression

$$u_k(x) = v_0^k + \sum_{l=0}^{k} \left(L^{k-l} w_l^k \right)(x) \ .$$

From the spectral theory of L, using (I.1) we obtain

$$\vee u_k(x) \le \mathcal{O}(1) \sum_{l=0}^{k} \xi^{k-l} \sum_{j=0}^{l} \rho^{-l+j} \mathrm{Lip}_{j+1}(u) \le \mathcal{O}(1) \sum_{j=1}^{k+1} \sigma^{k+1-j} \mathrm{Lip}_j(u)$$

where $\sigma = (1 + \sup(\xi, \rho^{-1}))/2 < 1$. As a consequence, we obtain

$$\sup_{y,y'} \left(u_k(y) - u_k(y') \right) \le \mathcal{O}(1) \sum_{j=1}^{k+1} \sigma^{k+1-j} \mathrm{Lip}_j(u) \ ,$$

which is the first part of Lemma II.3. The second part follows in a similar way.

III. Rate of convergence of the empirical measure

We recall that the empirical measure of n samples is the random measure defined by

$$\mathcal{E}_n(x) = \frac{1}{n} \sum_{j=1}^{n-1} \delta_{f^j(x)}$$

where δ denotes the Dirac measure. Birkhoff's ergodic theorem tells us that since the measure μ is ergodic, for almost every x this sequence of random probability measures converges weakly when n tends to infinity to the (non random) probability measure μ. For statistical purposes it is important to know the speed of this convergence. To do this we first have to select a metric between probability measures (see [Ra.] for some examples). This leads to several famous statistical tests whose asymptotic speed is well known for the case of sequences of independent samples (see for example [Bo.]). Some results have recently been obtained in dependent cases, see for example [Ri2.], and [Mae.] for results on fluctuations for maps of the interval.

We will consider below some examples of distances which are Lipschitz functions of the samples in order to apply our previous estimates. We first start with a result on the Kantorovich distance κ which for probability measures on the unit interval is also given by

$$\kappa(\mu, \nu) = \int_0^1 \left| F_\mu(s) - F_\nu(s) \right| ds \ ,$$

where $F_\mu(s) = \mu([0,s])$ is the distribution function of μ. We refer to [Ra.] for equivalent definitions of this distance.

Theorem III.1. *There exists a number $t_0 > 0$ and a constant $R > 0$ such that for any $t > t_0$ and any integer n,*

$$\mu\left(\left\{x \mid \kappa(\mathcal{E}_n(x), \mu) > tn^{-1/2}\right\}\right) \le e^{-Rt^2} .$$

Note that this result is not just asymptotic, indeed we have an estimate valid for any integer n. We also remark that from the results in [B.G.R.], [Bu.], [C.], [Li.], [Mau.], [Sc.], it is possible to give constructive estimates for the above constants t_0 and R in terms of quantities depending on f, namely estimates on the finite regularity and information about topological mixing (this is unavoidable as already shown by the case of Markov chains). In other words, one can give uniform estimates for bounded sets of transformations in a suitable topology.

In order to apply the results of the previous section to prove Theorem III.1, we will consider the sequence of functions of n variables

$$u_n(x_1, \cdots, x_n) = \int_0^1 |F_n(x_1, \cdots, x_n, t) - F(t)|\, dt ,$$

where F_n is the empirical distribution of the sequence $x_1, \cdots, x_n$, namely

$$F_n(x_1, \cdots, x_n, t) = \frac{1}{n}\text{Card}\left(\left\{1 \le i \le n \mid x_i \le t\right\}\right) .$$

We point out that when x is chosen with probability μ, $F_n(x, \cdots, f^{n-1}(x), t)$ is the empirical distribution, namely $\mathcal{E}_n(x)\big([0,t]\big)$.

We first have to show that u_n is Lipschitz and to estimate the Lipschitz constants. For this purpose, we consider an index $1 \le k \le n$ and change the value of x_k to x'_k. For definiteness we will assume $x'_k > x_k$, the other case being similar. It is easy to verify from the definition that $F_n(x_1, \cdots, x_k, \cdots, x_n, t)$ and $F_n(x_1, \cdots, x'_k, \cdots, x_n, t)$ differ only for $x_k \le t \le x'_k$, and the difference is bounded in modulus by $1/n$. Therefore

$$\sup_{x_1,\cdots,\hat{x}_k,\cdots,x_n} \left|u_n(x_1, \cdots, x_k, \cdots, x_n) - u_n(x_1, \cdots, x'_k, \cdots, x_n)\right| \le \frac{|x_k - x'_k|}{n} .$$

In other words

$$\sup_{1 \le j \le n} \text{Lip}_j\big(u_n\big) \le \frac{1}{n} .$$

Before applying corollary I.2, we give an estimate on the average $\mu\big(u_n\big)$ of u_n. We have by Schwartz inequality

$$\mu\big(u_n\big) = \int_0^1 \varphi(x)\, dx \int_0^1 dt \left|\frac{1}{n}\sum_{j=0}^{n-1} \chi_{[0,t]}\big(f^j(x)\big) - F(t)\right|$$

$$\le \left[\int_0^1 dt \int_0^1 d\mu(x) \left(\frac{1}{n}\sum_{j=0}^{n-1} \chi_{[0,t]}\big(f^j(x)\big) - F(t)\right)^2\right]^{1/2} .$$

Expanding the sum in the square and using the invariance of μ, we get

$$\int_0^1 d\mu(x)\left(\frac{1}{n}\sum_{j=0}^{n-1}\chi_{[0,t]}\big(f^j(x)\big)-F(t)\right)^2 = \frac{1}{n}\int_0^1 d\mu(x)\left(\chi_{[0,t]}(x)-F(t)\right)^2$$
$$+\frac{2}{n}\sum_{j=1}^{n-1}(1-\frac{j}{n})\int_0^1 d\mu(x)\left(\chi_{[0,t]}(x)-F(t)\right)\Big(\chi_{[0,t]}\big(f^j(x)\big)-F(t)\Big)\,.$$

We now observe that for each $t\in[0,1]$, $F(t)$ is the average of $\chi_{[0,t]}\big(f^j(x)\big)$ with respect to μ. We can now use the decay of correlations (I.2) and we obtain

$$\int_0^1 d\mu(x)\left(\frac{1}{n}\sum_{j=0}^{n-1}\chi_{[0,t]}\big(f^j(x)\big)-F(t)\right)^2 \le \frac{\mathcal{O}(1)}{n}\,.$$

It then follows at once that

$$\mu\big(u_n\big)\le\mathcal{O}(1)n^{-1/2}\,.$$

We now apply Corollary I.2 to conclude the proof.

We now consider the convergence in total variation. However we have to use a smoothed empirical measure. We fix once for all a non negative regular function ψ defined on $\mathbf{R}^+$, equal to one on $[0,1/4]$, vanishing on $[1,\infty)$, and with integral $1/2$. Note that all the above results also hold for the circle under the same hypothesis. In order to avoid to treat boundary terms we will from now on consider maps of the circle S^1. Let (α_n) be a positive sequence converging to zero. As usual (see [Bo.]) we will assume that $n\alpha_n$ tends to infinity with n. We consider the sequence of regularized (random) empirical measures $\mathcal{H}_n(x)$ with densities (h_n) defined by

$$h_n(x,s)=\frac{1}{n\alpha_n}\sum_{j=1}^{n}\psi\big(|s-f^j(x)|/\alpha_n\big)\,,$$

where $|\ |$ denotes the Riemaniann distance on the circle. This is also known as a Parzen non parametric estimate of the density. The distance in total variation is the (random) quantity

$$d_{\mathrm{TV}}(\mathcal{H}_n(x),\mu)=\int_{S^1}|h_n(x,s)-\varphi(s)|\,ds\,.$$

Theorem III.2. *There exists a constant $R'>0$ such that for any integer n, and for any $t>\alpha_n+1/\sqrt{n\alpha_n}$ we have*

$$\mu\left(\big\{x\mid d_{TV}\big(\mathcal{H}_n(x),\mu\big)>t\big\}\right)\le e^{-R't^2n\alpha_n^2}\,.$$

In order to prove this theorem, we define a sequence of functions (u_n) of n variables in S^1 by

$$u_n\big(x_1, \cdots, x_n, s\big) = \int_{S^1} \left| \frac{1}{n\alpha_n} \sum_{j=1}^{n} \psi\big(|s - x_j|/\alpha_n\big) - \varphi(s) \right| ds \ .$$

As before, our first goal is to prove that this function is separately Lipschitz and to estimate the Lipschitz constant. For each integer k between 1 and n we have to study the variation of the function with x_k. We obtain

$$\begin{aligned}\sup_{x_1,\cdots,\hat{x}_k,\cdots,x_n} &\left|u_n(x_1, \cdots, x_k, \cdots, x_n) - u_n(x_1, \cdots, x_k', \cdots, x_n)\right| \\ &\le \tfrac{1}{n\alpha_n} \textstyle\int_{S^1} \left|\psi\big(\alpha_n^{-1}|s - x_k|\big) - \psi\big(\alpha_n^{-1}|s - x_k'|\big)\right| ds \ .\end{aligned}$$

We now distinguish two cases. We also assume α_n small since we are only interested in this situation. First if $|x_k - x_k'| > 4\alpha_n$, then the above quantity is equal to 2 which is smaller than $|x_k - x_k'|/\alpha_n$. On the other hand, if $|x_k - x_k'| \le 4\alpha_n$ the above quantity is bounded by

$$= \frac{1}{n\alpha_n^2} \int_{|s-x_k|\le 6\alpha_n} ds \ \|\psi'\|_\infty \ \big||s - x_k| - |s - x_k'|\big| \le \mathcal{O}(1) \frac{|x_k - x_k'|}{n\alpha_n} \ .$$

Therefore for all $1 \le j \le n$

$$\mathrm{Lip}_j\big(u_n\big) \le \frac{\mathcal{O}(1)}{n\alpha_n} \ .$$

In order to estimate the average of the total variation, it will be convenient to replace φ by a regularized function $\tilde{h}_n$ given by

$$\tilde{h}_n(s) = \alpha_n^{-1} \int_{S^1} d\mu(x)\psi\left(\frac{|x - s|}{\alpha_n}\right) \ .$$

We now have to estimate the L^1 norm of the error. We have

$$\begin{aligned}\int_{S^1} ds \left|\varphi(s) - \tilde{h}_n(s)\right| &\le \int_{S^1} ds \ \alpha_n^{-1} \int_{S^1} \psi\big(|y|\alpha_n^{-1}\big)|\varphi(s) - \varphi(s - y)|dy \\ &\le \sum_{1\le k\le\alpha_n^{-1}+1} \int_{(k-1)\alpha_n}^{k\alpha_n} ds \ \alpha_n^{-1} \int_{S^1} \psi\big(|y|\alpha_n^{-1}\big)|\varphi(s) - \varphi(s - y)|dy \\ &\le \sum_{1\le k\le\alpha_n^{-1}+1} \vee_{[(k-\beta)\alpha_n,(k+\beta+1)\alpha_n]}\varphi \int_{(k-1)\alpha_n}^{k\alpha_n} ds \ \alpha_n^{-1} \int_{S^1} \psi\big(|y|\alpha_n^{-1}\big)dy \ ,\end{aligned}$$

where β denotes the diameter of the support of ψ. By the normalization of ψ, the integral over y is equal to α_n. The intervals $[(k - \beta)\alpha_n, (k + 1 + \beta)\alpha_n]$ form a covering of the circle with multiplicity smaller than $2\beta + 2$, therefore

$$\int_{S^1} ds \left|\varphi(s) - \tilde{h}_n(s)\right| \le \mathcal{O}(1)\alpha_n \vee \varphi \ .$$

We now observe that for α_n small enough

$$\sup_s \bigvee \psi\left(\frac{|\cdot - s|}{\alpha_n}\right) \le \mathcal{O}(1) \ ,$$

while

$$\sup_s \int_{S^1} d\mu(x) \psi\left(\frac{|x-s|}{\alpha_n}\right) \le \mathcal{O}(1)\alpha_n \ .$$

Therefore using again the decay of correlations (I.2), we get for any s

$$\left| \int_{S^1} d\mu(x) \Big(\alpha_n^{-1}\psi\big(\alpha_n^{-1}|f^j(x)-s|\big) - \tilde{h}_n(s)\Big)\Big(\alpha_n^{-1}\psi\big(\alpha_n^{-1}|x-s|\big) - \tilde{h}_n(s)\Big)\right| \\ \le \mathcal{O}(1)\xi^j(\alpha_n^{-1}+1) \ .$$

This implies

$$\int_{S^1} d\mu(x) \left(h_n(x,s) - \tilde{h}_n(s)\right)^2 \le \frac{\mathcal{O}(1)}{n\alpha_n} \ .$$

Therefore

$$\int_{S^1} d_{TV}\big(\mathcal{H}_n(x), \mu\big) d\mu(x) \le \int_{S^1} d\mu(x) \int_{S^1} ds \left|\tilde{h}_n(s) - \varphi(s)\right| \\ + \left(\int_{S^1} d\mu(x) \big(h_n(x,s) - \tilde{h}_n(s)\big)^2\right)^{1/2} \le \mathcal{O}(1)\left(\alpha_n + \frac{1}{\sqrt{n\alpha_n}}\right) \ .$$

The result follows by a direct application of Corollary I.2.

IV. Application to the shadowing

In this section we apply the results of section I to the shadowing properties of some subsets of trajectories. The basic problem can be formulated as follows. Let A be a set of initial conditions, if x is an initial condition not in A, how well can we approximate the trajectory of x by a trajectory from an element of A. This is in some sense the analog of the well known consequence of concentration for independent random variables which states that sets of measure one half are big. We first start with a result about the average quality of shadowing.

Theorem IV.1. *There are positive constants C_1 and C_2 such that if A is a set of positive measure, for any integer n the sequence (Z_n) of functions defined by*

$$Z_n(x) = \frac{1}{n} \inf_{y\in A} \sum_{j=0}^{n-1} \left|f^j(y) - f^j(x)\right| \ ,$$

satisfies for any $t \ge 0$

$$\mu\left(\left\{x \;\middle|\; Z_n(x) \ge C_1 \frac{\sqrt{\log n}}{\mu(A)\sqrt{n}} + \frac{t}{\sqrt{n}}\right\}\right) \le e^{-C_2 t^2} \ .$$

We remark that by restricting the infimum over a countable dense subset it is easy to verify that Z_n is measurable. Moreover, $0 \leq Z_n \leq 1$, and $Z_n(x) = 0$ if $x \in A$. We now give a proof of this Theorem.

In order to apply Corollary I.2, we define a sequence of functions (u_n) by

$$u_n(x_1, \cdots, x_n) = \frac{1}{n} \inf_{y \in A} \sum_{j=0}^{n-1} \left| f^j(y) - x_{j+1} \right| .$$

It is easy to verify that these functions are separately Lipschitz and moreover, for any $1 \leq l \leq n$ we have

$$\mathrm{Lip}_l(u_n) \leq n^{-1} .$$

Applying Corollary I.2, there is a constant $C_2 > 0$ such that for any A and any $t > 0$ we have

$$\mu\left(\left\{x \;\middle|\; Z_n(x) \geq \mu\big(Z_n\big) + \frac{t}{\sqrt{n}}\right\}\right) \leq e^{-C_2 t^2} .$$

We now estimate $\mu\big(Z_n\big)$ by the usual trick (see for example [T1]). For a fixed $s > 0$, let

$$B_s = \left\{Z_n(x) \geq \mu\big(Z_n\big) + \frac{s}{\sqrt{n}}\right\} .$$

We have

$$\mu\big(Z_n\big) = \int Z_n d\mu = \int_A Z_n d\mu + \int_{A^c \cap B_s^c} Z_n d\mu + \int_{B_s} Z_n d\mu .$$

The first integral is equal to zero, the second integral is bounded by

$$\int_{A^c \cap B_s^c} Z_n d\mu \leq \left(\mu\big(Z_n\big) + \frac{s}{\sqrt{n}}\right) \mu(A^c) ,$$

and the third integral is estimated by $\mu(B_s)$ using $Z_n \leq 1$. We obtain

$$\mu\big(Z_n\big) \leq \left(\mu\big(Z_n\big) + \frac{s}{\sqrt{n}}\right) \mu(A^c) + e^{-C_2 s^2} ,$$

which implies

$$\mu\big(Z_n\big) \leq \mu(A)^{-1} \left(\frac{s}{\sqrt{n}} + e^{-C_2 s^2}\right) ,$$

and the result follows by choosing s adequately.

We now derive a similar result for the number of mismatch at a given precision. Again in order to avoid unessential complications at the boundary we assume we are working on the circle. For a measurable set A of positive measure, we define for any integer n and any $\epsilon > 0$

$$Z'_{n,\epsilon}(x) = \frac{1}{n} \inf_{y \in A} \mathrm{Card}\big\{0 \leq j < n \;\big|\; |f^j(y) - f^j(x)| > \epsilon\big\}.$$

This a measurable function of x because instead of taking the infimum over all points in A, it is enough to take the infimum over a countable dense subset in A containing also all the preimages (in A) of order up to n of the points of discontinuity of f.

Theorem IV.2. *There are positive constants C_1 and C_2 such that if A is a set of positive measure, for any integer n, for any $0 < \epsilon < 1/2$ and any $t \geq 0$*

$$\mu\left(\left\{x \;\middle|\; Z'_{n,\epsilon}(x) \geq C_1 \epsilon^{-1} \mu(A)^{-1} \sqrt{\frac{\log n}{n}} + \frac{t\epsilon^{-1}}{\sqrt{n}}\right\}\right) \leq e^{-C_2 t^2} .$$

The new difficulty in applying Corollary I.2 is that the Hamming function

$$u_{n,\epsilon}\big(x_1, \cdots, x_n\big) = \frac{1}{n} \inf_{y \in A} \sum_{j=0}^{n-1} \big(1 - \chi_{[f^j(y)-\epsilon, f^j(y)+\epsilon]}(x_{j+1})\big)$$

is not Lipschitz. We therefore replace the characteristic function of the complement of the interval $[-\epsilon, \epsilon]$ by a piecewise linear approximation. Let

$$g_\epsilon(s) = \begin{cases} |s|/\epsilon & \text{if } |s| < \epsilon \\ 1 & \text{otherwise.} \end{cases}$$

We now define a sequence of functions

$$v_{n,\epsilon}\big(x_1, \cdots, x_n\big) = \frac{1}{n} \inf_{y \in A} \sum_{j=0}^{n-1} g_\epsilon\left(f^j(y) - x_{j+1}\right) .$$

These functions are separately Lipschitz with Lipschitz constants ϵ^{-1}/n and $u_n \leq v_n$. Let

$$\tilde{Z}_{n,\epsilon}(x) = v_{n,\epsilon}\big(x, f(x), \cdots, f^{n-1}(x)\big) .$$

Since

$$Z'_{n,\epsilon}(x) = \frac{1}{n} \inf_{y \in A} \sum_{j=0}^{n-1} \left(1 - \chi_{[f^j(y)-\epsilon, f^j(y)+\epsilon]}(f^j(x))\right) \leq \tilde{Z}_{n,\epsilon}(x)$$

we have using Corollary I.2

$$\begin{aligned} &\mu\left(\left\{x \;\middle|\; Z_{n,\epsilon}(x) \geq \mu\big(\tilde{Z}_{n,\epsilon}\big) + \frac{t}{\sqrt{n}}\right\}\right) \\ &\quad \leq \mu\left(\left\{x \;\middle|\; \tilde{Z}_{n,\epsilon}(x) \geq \mu\big(\tilde{Z}_{n,\epsilon}\big) + \frac{t}{\sqrt{n}}\right\}\right) \leq e^{-C_2 t^2} . \end{aligned}$$

The result follows by an estimate of $\mu\big(\tilde{Z}_{n,\epsilon}\big)$ as above.

Appendix

In this appendix we prove a variant of the above results for the case of the Kolmogorov metric. We recall that if ν_1 and ν_2 are two measures, the Kolmogorov metric $\rho(\nu_1, \nu_2)$ is defined by

$$\rho(\nu_1, \nu_2) = \sup_t \left|F_{\nu_1}(t) - F_{\nu_2}(t)\right| .$$

As before, if $\mathcal{E}_n$ denotes the empirical measure for n successive samples, and μ the invariant measure, we have

$$\rho(\mathcal{E}_n(x), \mu) = \sup_t \left|\frac{1}{n}\text{Card}\left(\left\{0 \le j \le n-1 \mid f^j(x) \le t\right\}\right) - F(t)\right| .$$

The Kolmogorov metric can be compared to the Kantorovich metric. We have of course the trivial inequality

$$\kappa(\mathcal{E}_n(x), \mu) \le \rho(\mathcal{E}_n(x), \mu)$$

and since μ has a bounded density φ, we have also

$$\rho(\mathcal{E}_n(x), \mu) \le \sqrt{2\|\varphi\|_\infty \kappa(\mathcal{E}_n(x), \mu)} .$$

However using this inequality together with Theorem III.1 gives only a poor estimate for the rate of convergence of the Kolmogorov distance. We will derive below a better estimate.

It will be useful to introduce the function of $n+1$ variables defined on $[0, 1]^{n+1}$ by

$$U_n(x_1, \cdots, x_n, t) = \text{Card}\left(\left\{1 \le j \le n \mid x_j \le t\right\}\right) - nF(t) .$$

Unfortunately the supremum over t of the absolute value of this function is not separately Lipschitz in $x_1, \cdots, x_n$, and we cannot apply Theorem I.1.

We will now prove the following result.

Theorem A.1. *There is a constant $C > 0$ such that for any integer n, we have for any $s > 0$*

$$\mu\left(\{\rho(\mathcal{E}_n, \mu) > s + \|\varphi\|_\infty/n\}\right) \le 2ne^{-Cns^2} .$$

We remark that except for the $2n$ prefactor, which can be absorbed by modifying the constant C for s larger than $\mathcal{O}(1)\sqrt{\log n/n}$, we have the same scaling as in the well known Kolmogorov theorem which holds in the independent case (see [Bo.]). As mentioned above, we observe that this result holds for any n. Moreover, C and $\|\varphi\|_\infty$ can eventually be explicitly estimated.

We first prove an exponential estimate with a bound uniform in t where we use the convenient notation

$$N_n(x, t) = \text{Card}\left(\left\{0 \le j \le n-1 \mid f^j(x) \le t\right\}\right) .$$

Note that the expectation of $N_n(x, t)$ with respect to μ is equal to $nF(t)$.

Lemma A.2. *There is a constant $D' > 0$ such that for any integer n, we have for any real λ*

$$\sup_{0\le t\le 1}\int e^{\lambda\left(N_n(x,t)-nF(t)\right)}\,d\mu(x)\le e^{nD'\lambda^2}$$

Proof. For a fixed t, we use recursively Lemma II.2 as in the proof of Theorem I.1 and we obtain

$$\int e^{\lambda\left(N_n(x,t)-nF(t)\right)}\,d\mu(x)\le e^{(\lambda^2/8)\sum_{j=0}^{n-1}\left(\mathrm{osc}_1 L^j U_n\right)^2}\ .$$

Instead of using Lemma II.3, we will estimate directly $\mathrm{osc}_1\left(L^j U_n\right)$. We observe that

$$U_n(x_1,\cdots,x_n,t)=\sum_{j=0}^{n-1}\chi_{[0,t]}(x_j)-nF(t)\ .$$

Therefore for any integer $0\le k\le n-1$ and fixed t, we have

$$L^k(U_n)(y,r_1,\cdots,r_{n-k-1})=\frac{1}{\varphi(y)}\sum_{z\ ,\ f^k(z)=y}\frac{\varphi(z)}{|f^{k\prime}(z)|}\sum_{j=0}^{k}\left(\chi_{[0,t]}\left(f^j(z)\right)-F(t)\right)$$
$$+\sum_{j=1}^{n-k-1}\left(\chi_{[0,t]}(r_j)-F(t)\right)=\sum_{l=0}^{k}L^l\left(\chi_{[0,t]}-F(t)\right)(y)$$
$$+\sum_{j=1}^{n-k-1}\left(\chi_{[0,t]}(r_j)-F(t)\right)\ .$$

Using (I.1) and observing that the variation and L^∞ norms of $\chi_{[0,t]}$ are independent of t, we conclude that there are two constants $C''>0$ and $0<\sigma<1$ such that for any $t>0$

$$\bigvee L^l\left(\chi_{[0,t]}-F(t)\right)\le C''\sigma^l$$

which implies

$$\mathrm{osc}_1\left(L^k U_n\right)\le\frac{C''}{1-\sigma}$$

and the result follows.
By a simple application of Markov's inequality, we obtain the following result.

Corollary A.3. *There is a constant $D''>0$ such that for any integer n, we have for any $s>0$*

$$\sup_{0\le t\le 1}\mu\left(\left\{\left|N_n(\cdot,t)-nF(t)\right|>s\right\}\right)\le 2e^{-D''s^2/n}\ .$$

We now finish the proof of Theorem A.1. We first observe that if $t \in [k/n, (k+1)/n]$, with $0 \leq k \leq n-1$, we have for any x

$$N_n(x, k/n) \leq N_n(x, t) \leq N_n(x, (k+1)/n)$$

and similarly

$$F(k/n) \leq F(t) \leq F((k+1)/n) \ .$$

Therefore,

$$\begin{aligned} &\left|N_n(x,t) - nF(t)\right| \\ &\quad \leq \max\left\{\left|N_n(x,k/n) - nF((k+1)/n)\right| \ , \ \left|N_n(x,(k+1)/n) - nF(k/n)\right|\right\} \\ &\quad \leq \max\left\{\left|N_n(x,k/n) - nF(k/n)\right| \ , \ \left|N_n(x,(k+1)/n) - nF((k+1)/n)\right|\right\} \\ &\qquad + \|\varphi\|_\infty \ . \end{aligned}$$

In particular, if x is not a preimage of 0, taking also into account that $N_n(x,0) = F(0) = 0$ and $N_n(x,1) = nF(1) = n$, we get

$$\sup_{0 \leq t \leq 1} \left|N_n(x,t) - nF(t)\right| \leq \sup_{0<k<n} \left|N_n(x,k/n) - nF(k/n)\right| + \|\varphi\|_\infty \ .$$

Therefore, if

$$\sup_{0 \leq t \leq 1} \left|N_n(x,t) - nF(t)\right| \geq \|\varphi\|_\infty + sn$$

there is at least one integer $1 \leq k \leq n-1$ such that

$$\left|N_n(x,k/n) - nF(k/n)\right| \geq sn$$

and the result follows from Corollary A.3 since the countable set of preimages of the origin is of measure zero.

Acknowledgments. The authors thank J.Buzzi for pointing out references [Bu.] and [H.]. The authors are grateful to the ECOS-CONYCIT grants CE99E09 and CE99E10, DIM and CMM of the Universidad de Chile, Fondap in applied Mathematics, the CNRS, University of Dijon and Ecole Polytechnique for support and hospitality.

References

[B.G.] Bobkov, S.G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. **163**, 1–28 (1999)

[B.G.R.] Barbour, A.D., Gerrard, R.M., Reinert, G.: Iterates of expanding maps. Probab. Theory Related Fields **116**, 151–180 (2000)

[Bo.] Borovkov, A.A.: *Mathematical statistics*. Gordon and Breach Science Publishers, Amsterdam, 1998

[Bu.] Buzzi, J.: Specification on the interval. Trans. Amer. Math. Soc. **349**, 2737–2754 (1997)

[C.] Collet, P.: An estimate of the decay of correlations for mixing non-Markov expanding maps of the interval. Unpublished.

[D.] Devroye, L.: Exponential inequalities in nonparametric estimation. Nonparametric functional estimation and related topics (Spetses, 1990), 31–44, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 335, Kluwer Acad. Publ., Dordrecht, 1991

[H.] Hofbauer, F.: Piecewise invertible dynamical systems. Probab. Theory Relat. Fields **72**, 359–386 (1986)

[H.K.] Hofbauer, F., Keller, G.: Ergodic properties of invariant measures for piecewise monotonic transformations. Math. Z. **180**, 119–140 (1982)

[Le.] Ledoux, M.: *Concentration of measures and logarithmic Sobolev inequalities*. Séminaire de Probabilités XXXIII. lecture notes in Mathematics **1709**, 120–216, Springer 1999. The concentration of measure phenomenon. Proceedings of the CIMPA summer school "From Classical to Modern Probability", Temuco 2001. P. Picco & J. San Martín editors, to appear.

[Li.] Liverani, C.: Decay of correlations. Ann. of Math. (2) 142 239–301 (1995)

[L.Y.] Lasota, A., Yorke, J.A.: On the existence of invariant measures for piecewise monotonic transformations. Trans. Amer. Math. Soc. **186**, 481–488 (1974)

[Mae.] Maes, J.: Estimation non paramétrique pour des processus dynamiques dilatants. C. R. Acad. Sci. Paris Sér. I Math. **330**, 831–834 (2000)

[Mas1.] Massart, P.: About the constants in Talagrand's concentration inequalities for empirical processes. Ann. Probab. **28**, 863–884 (2000)

[Mas2.] Massart, P.: Optimal constants for Hoeffding type inequalities. Preprint.

[Mar1.] Marton, K.: Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration. Ann. Probab. **24**, 857–866 (1996)

[Mar2.] Marton, K.: A measure concentration inequality for contracting Markov chains. Geom. Funct. Anal. **6**, 556–571 (1996)

[Mar3.] Marton, K.: Measure concentration for a class of random processes. Probab. Theory Relat. Fields **110**, 427–439 (1998)

[Mau.] Maume-Deschamps, V.: Correlation decay for Markov maps on a countable state space. Ergod. Theor. & Dyn. Syst. **21**, 165–196 (2000)

[Ra.] Rachev, S.T.: *Probability metrics and the stability of stochastic models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley and Sons, Ltd., Chichester, 1991

[Ri1.] Rio, E.: Inégalités exponentielles pour les processus empiriques. (Exponential inequalities for empirical processes) C. R. Acad. Sci. Paris Sér. I Math. **330**, 597–600 (2000)

[Ri2.] Rio, E.: Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. (Hoeffding inequalities for Lipschitz functions of dependent sequences) C. R. Acad. Sci. Paris Sér. I Math. **330**, 905–908 (2000)

[Sa.] Samson, P.: Concentration of measure inequalities for Markov chains and Φ mixing processes. Ann. Probab. **28**, 416–461 (2000)

[Sc.] Schmitt, B.: Spectral gap, Poincaré inequality and mass concentration for g measures. Proceedings of the school "Dynamics and Randomness", A. Mass, S. Martínez & J. San Martín editors. Santiago 2000, to appear.

[T1.] Talagrand, M.: Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Etudes Sci. Publ. Math. **81**, 73–205 (1995)

[T2.] Talagrand, M.: New concentration inequalities in product spaces. Invent. Math. **126**, 505–563 (1996)

[T3.] Talagrand, M.: A new look at independence. Ann. Probab. **24**, 1–34 (1996)