Statistical consequences of the Devroye inequality for processes. Applications to a class of non-uniformly hyperbolic dynamical systems



2005 Nonlinearity 18 2341

(http://iopscience.iop.org/0951-7715/18/5/024)

INSTITUTE OF PHYSICS PUBLISHING NONLINEARITY

Nonlinearity **18** (2005) 2341–2364 doi:10.1088/0951-7715/18/5/024

# Statistical consequences of the Devroye inequality for processes. Applications to a class of non-uniformly hyperbolic dynamical systems

**J-R Chazottes**[1], **P Collet**[1] **and B Schmitt**[2]

[1] Centre de Physique Théorique, Ecole Polytechnique, CNRS UMR 7644, F-91128 Palaiseau Cedex, France
[2] Faculté des Sciences Mirande, Département de Mathématiques, Université de Bourgogne, BP 138, 21004 Dijon Cedex, France

E-mail: jeanrene@cpht.polytechnique.fr, collet@cpht.polytechnique.fr and schmittb@u-bourgogne.fr

Received 8 December 2004, in final form 6 June 2005
Published 15 July 2005
Online at stacks.iop.org/Non/18/2341

Recommended by L Bunimovich

**Abstract**
In this paper, we apply the Devroye inequality to study various statistical estimators and fluctuations of observables for processes. Most of these observables are suggested by dynamical systems. These applications concern the co-variance function, the integrated periodogram, the correlation dimension, the kernel density estimator, the speed of convergence of the empirical measure, the shadowing property and the almost-sure central limit theorem. We proved (Chazottes *et al* 2005 *Nonlinearity* **18** 2323–40) that the Devroye inequality holds for a class of non-uniformly hyperbolic dynamical systems introduced in Young (1998 *Ann. Math.* **147** 585–650). In the second appendix we prove that, if the decay of correlations holds with a common rate for all pairs of functions, then it holds uniformly in the function spaces. In the last appendix we prove that, for the subclass of one-dimensional systems studied in Young, the density of the absolutely continuous invariant measure belongs to a Besov space.

Mathematics Subject Classification: 37A50, 37D25, 60E15

## 1. Introduction and set-up

Assume one has a finite sample $x_1, \ldots, x_n$ of a stationary ergodic process taking values in $\mathbb{R}^d$. If we consider an empirical estimator (or an observable) $K(x_1, \ldots, x_n)$ of some statistical properties of the process, we basically wish to determine its fluctuations and its convergence properties as $n$ grows. In statistical terminology, we aim to study the consistency of the estimator $K(x_1, \ldots, x_n)$ and be able to build confidence intervals.

As we shall see in the sequel with various examples, many interesting estimators have a complicated dependence on the sample. In particular, they are not of the form $(u(x_1) + \cdots + u(x_n))/n$, for some function $u$, or cannot be well approximated by such time-averages for which the central limit theorem may apply.

The aim of this paper is to apply what we call the Devroye inequality [8] (see the definition below) to estimate the variance for a general class of estimators $K(x_1, \ldots, x_n)$. For some of them we will further require some weak conditions on the auto-covariance function for functionals of the process.

Our applications concern the empirical auto-covariance function, the integrated periodogram, the correlation dimension, the kernel density estimation of the density of the invariant measure, shadowing properties, the speed of convergence of the empirical measure towards the invariant measure and the almost-sure central limit theorem. Some of these estimators were studied in [6] in the context of piecewise expanding maps on the interval for which a stronger inequality than the Devroye inequality holds.

We shall formulate the results as much as possible in an abstract setting in order to see more clearly what is needed to prove them. As we showed in [5], a class of non-uniformly hyperbolic dynamical systems introduced by Young [18] fits this framework.

Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space and $(X_k)$ be a stationary ergodic sequence of random variables assuming values in $\mathbb{R}^d$.

We will denote the expectation with respect to $\mathbb{P}$ by $\mathbb{E}$, and by $\mu$ the common distribution of the $X_k$. We will assume that the $X_k$ are almost-surely bounded, i.e. there exists a positive constant $A$ such that

$$\|X_k\| \leqslant A \qquad \mathbb{P}\text{—almost-surely.} \tag{1}$$

Let $K$ be a real-valued function on $(\mathbb{R}^d)^n$. We will say that $K$ is separately $\eta$-Hölder in all its variables, if for any $1 \leqslant i \leqslant n$, the following quantities are finite:

$$L_j = L_j(K) := \sup_{x_1,x_2,\ldots,x_{j-1},x_j,x_{j+1},\ldots,x_n} \sup_{\tilde{x}_j \neq x_j} \frac{|K(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n) - K(x_1, \ldots, x_{j-1}, \tilde{x}_j, x_{j+1}, \ldots, x_n)|}{\|x_j - \tilde{x}_j\|^\eta}. \tag{2}$$

We now define what we mean by saying that the process $(X_k)$ satisfies the Devroye inequality.

**Definition 1.1 (Devroye inequality for the variance).** *We will say that the process $(X_k)$ satisfies the Devroye inequality if, for $\eta \in ]0, 1]$, there exists a constant $D = D(\eta) > 0$ such that for any integer $n \geqslant 1$ and for any real-valued separately $\eta$-Hölder function $K$ on $(\mathbb{R}^d)^n$, we have*

$$\mathrm{var}(K) = \mathbb{E}((K - \mathbb{E}(K))^2) \leqslant D \sum_{j=1}^{n} L_j^2. \tag{3}$$

For the case of dynamical systems, $\Omega$ is the phase space on which acts a measurable transformation $f$. We assume that an $f$-invariant ergodic measure $\mu$ is given. One can define a stochastic process $X_k(x) = f^{k-1}(x)$, where $x$ is randomly chosen according to $\mu$. We are interested in observables of the form $K(X_1, \ldots, X_n)(x) = K(x, f(x), \ldots, f^{n-1}(x))$.

One can ask whether there are processes satisfying the Devroye inequality. Indeed, a large class of dynamical systems satisfy the Devroye inequality, as we proved in [5]. Let us recall that this class contains families of piecewise hyperbolic maps, like the Lozi maps; scattering billiards, like the planar periodic Lorentz gas; quadratic and Hénon maps (for parameter sets with positive Lebesgue measure). Let us also briefly mention that such dynamical systems admit an Sinai–Ruelle–Bowen (SRB)-measure, enjoy exponential decay of correlations and a central limit theorem for Hölder continuous observables. Notice that in the sequel we will only need very slow decay of correlations, e.g. $C(\ell) \sim 1/\sqrt{\ell}$ for the integrated periodogram or absolute summability, e.g. for the almost-sure central limit theorem.

## 2. Covariance function

Recall that the auto-covariance of a real-valued, square-integrable function $u$ on $\mathbb{R}^d$ is defined by

$$C(n) = C_u(n) := \mathbb{E}(u(X_1)u(X_n)) - (\mathbb{E}(u(X_1)))^2. \tag{4}$$

An empirical estimator of the auto-covariance is given by

$$\hat{C}_k(n) = \frac{1}{k}\sum_{j=1}^{k} u(X_j)u(X_{j+n}) - \left(\frac{1}{k}\sum_{j=1}^{k} u(X_j)\right)^2.$$

It follows at once from Birkhoff's ergodic theorem that

$$C(n) = \lim_{k\to\infty} \hat{C}_k(n) \qquad \mathbb{P}\text{—almost-surely.}$$

**Theorem 2.1.** *Let $u$ be a real-valued $\eta$-Hölder function on $\mathbb{R}^d$ with Hölder constant denoted by $L_u$. Then, for all integers $k$, $n$, we have*

$$\mathbb{E}((\hat{C}_k(n) - C(n))^2) \leqslant 16DL_u^4A^{2\eta}\frac{n+k}{k^2} + \frac{D^2L_u^4}{k^2}.$$

**Proof.** We have the following identity:

$$\begin{aligned}\mathbb{E}((\hat{C}_k(n) - C(n))^2) &= \mathbb{E}((\hat{C}_k(n) - \mathbb{E}(\hat{C}_k(n)))^2) + (\mathbb{E}(\hat{C}_k(n)) - (C(n)))^2 \\ &= \operatorname{var}(\hat{C}_k(n)) + \left(\operatorname{var}\left(\frac{1}{k}\sum_{j=1}^{k} u(X_j)\right)\right)^2.\end{aligned}$$

The first term is estimated using the Devroye inequality (3) and assuming, without loss of generality, that $u(0) = 0$. We obtain the upper bound of independent interest

$$\operatorname{var}(\hat{C}_k(n)) \leqslant 16DL_u^4A^{2\eta}\frac{n+k}{k^2}.$$

The second term is easily estimated using the Devroye inequality again. This leads immediately to the above estimate. ∎

**Remark.** For the study of $U$-statistics of functionals of $\alpha$- and $\beta$-mixing process, we refer the interested reader to [2] and references therein.

## 3. Integrated periodogram

We recall (see [3]) that if $u$ is a real-valued function, the raw periodogram (of order $n$) of the process $(u(X_k))$ is the function:

$$I_n(\omega) = \frac{1}{n}\left|\sum_{j=1}^{n} \mathrm{e}^{-\mathrm{i}j\omega}(u(X_j) - \mathbb{E}(u(X_1)))\right|^2, \tag{5}$$

where $\omega \in [0, 2\pi]$. The spectral distribution function of order $n$ (integral of the raw periodogram of order $n$) is given by

$$J_n(\omega) = \int_0^{\omega} I_n(s)\,\mathrm{d}s. \tag{6}$$

From a practical point of view, it is worth defining the empirical spectral distribution function of order $n$ as follows:

$$\tilde{J}_n(\omega) = \int_0^{\omega} \frac{1}{n}\left|\sum_{j=1}^{n} \mathrm{e}^{-\mathrm{i}js}\left(u(X_j) - \frac{1}{n}\sum_{\ell=1}^{n} u(X_\ell)\right)\right|^2 \mathrm{d}s.$$

In this section we will make the following assumption.

**Hypothesis 3.1.** *The function $u$ is $\eta$-Hölder continuous and its auto-covariance function $C(\ell) = C_u(\ell)$ satisfies*

$$\sum_{\ell=1}^{\infty} \frac{|C(\ell)|}{\ell} < \infty$$

*(where $C(\ell)$ is defined at (4)).*

Let $\hat{C}(\omega)$ be the Fourier cosine transform of the auto-covariance function, namely

$$\hat{C}(\omega) = \sum_{k=0}^{\infty} \cos(\omega k)\, C(k+1).$$

We will denote by $J(\omega)$ the integral of the following quantity:

$$J(\omega) = \int_0^{\omega} (2\hat{C}(s) - C(1))\,\mathrm{d}s = C(1)\,\omega + 2\sum_{k=1}^{\infty} \frac{\sin(\omega k)}{k} C(k+1). \tag{7}$$

We will use the following convenient quantity:

$$\Delta_n := \frac{2}{n}\sum_{k=1}^{n-1} |C(k+1)| + 2\sum_{k=n}^{\infty} \frac{|C(k+1)|}{k}.$$

Observe that $J(\omega + 2\pi) = J(\omega) + 2\pi C(1) = J(\omega) + J(2\pi)$. In order to estimate $J$, it is therefore enough to restrict to the interval $[0, 2\pi]$.

**Theorem 3.1.** *There exists a positive constant $\Gamma$ such that for any function $u$ satisfying hypothesis 3.1, and any $n \geqslant 1$, we have:*

$$\mathbb{E}\left(\left(\sup_{\omega\in[0,2\pi]} |\tilde{J}_n(\omega) - J(\omega)|\right)^2\right)$$
$$\leqslant \Gamma \inf_{N\geqslant 1}\left\{N\left[\frac{C(1)^2 + DA^{2\eta}\,L_u^4(1+\log n)^2}{n} + \Delta_n^2\right] + \left[\frac{C(1)}{N} + \Delta_N\right]^2\right\}.$$

**Remark.** If $\Delta_n \leqslant \mathrm{const}/n$, then

$$\mathbb{E}\left(\left(\sup_{\omega\in[0,2\pi]} |\tilde{J}_n(\omega) - J(\omega)|\right)^2\right) \leqslant \mathcal{O}(1)\frac{(1+\log n)^{4/3}}{n^{2/3}}.$$

In particular, if the auto-covariance is absolutely summable, then $\Delta_n \leqslant \mathrm{const}/n$.

For convergence results in distribution sense of the raw periodogram for a class of maps on the interval, we refer to [12].

This theorem is the consequence of two propositions.

**Proposition 3.1.** *For any function u satisfying hypothesis 3.1, and any $n \geqslant 1$, we have:*

$$\mathbb{E}\left(\left(\sup_{\omega\in[0,2\pi]} |J_n(\omega) - J(\omega)|\right)^2\right)$$
$$\leqslant \inf_{N>1}\left\{2(N+1)\left(\frac{(4\pi+1+\log n)^2 L_u^4 D}{n} A^{2\eta} + \Delta_n^2\right) + 8\pi^2\left(\frac{C(1)}{N} + \Delta_N\right)^2\right\}.$$

**Proof.** Let

$$Q_n = \sup_{\omega\in[0,2\pi]} |J_n(\omega) - J(\omega)|. \tag{8}$$

Let $N$ be an integer and define the sequence of numbers $(\omega_p)$ by $\omega_p = 2\pi p/N$ for $p = 0, \dots, N$. It follows at once from the monotonicity of $J$ and $J_n$ (since they are integrals of non-negative functions) that

$$Q_n \leqslant \max\left(\sup_{0\leqslant p\leqslant N-1} |J_n(\omega_{p+1}) - J(\omega_p)|, \sup_{0\leqslant p\leqslant N-1} |J_n(\omega_p) - J(\omega_{p+1})|\right).$$

We now have

$$Q_n \leqslant \sup_{0\leqslant p\leqslant N} |J_n(\omega_p) - J(\omega_p)| + \sup_{0\leqslant p\leqslant N-1} |J(\omega_p) - J(\omega_{p+1})|.$$

Now using definition (7), we get after an easy computation that for all $p = 0, \dots, N-1$

$$|J(\omega_p) - J(\omega_{p+1})| \leqslant 2\pi\left(\frac{C(1)}{N} + \Delta_N\right). \tag{9}$$

It follows that

$$Q_n \leqslant \bar{Q}_n + 2\pi\left(\frac{C(1)}{N} + \Delta_N\right), \tag{10}$$

where

$$\bar{Q}_n = \sup_{0\leqslant p\leqslant N} |J_n(\omega_p) - J(\omega_p)|.$$

We obviously have

$$\mathbb{E}(\bar{Q}_n^2) \leqslant \sum_{p=0}^{N} \mathbb{E}((J(\omega_p) - J_n(\omega_p))^2). \tag{11}$$

We now estimate each term $\mathbb{E}((J(\omega_p) - J_n(\omega_p))^2)$. Observe that for any $\omega$ we have

$$\mathbb{E}((J(\omega) - J_n(\omega))^2) = \mathbb{E}((J_n(\omega) - \mathbb{E}(J_n(\omega)))^2) + (\mathbb{E}(J_n(\omega) - J(\omega)))^2.$$

We have also from the definition of $J_n$:

$$J_n(\omega) = \frac{\omega}{n}\sum_{j=1}^{n}(u(X_j) - \mathbb{E}(u(X_1))^2 + \frac{\mathrm{i}}{n}\sum_{j\neq\ell}^{n}\frac{\mathrm{e}^{-\mathrm{i}(j-\ell)\omega}-1}{j-\ell}(u(X_j) - \mathbb{E}(u(X_1))(u(X_\ell) - \mathbb{E}(u(X_1))$$

$$= \frac{\omega}{n}\sum_{j=1}^{n}(u(X_j) - \mathbb{E}(u(X_1))^2 + \frac{1}{n}\sum_{j\neq\ell}^{n}\frac{\sin((j-\ell)\omega)}{j-\ell}(u(X_j) - \mathbb{E}(u(X_1))(u(X_\ell) - \mathbb{E}(u(X_1)). \quad (12)$$

Using this formula and (7), an easy computation leads to

$$(\mathbb{E}(J_n(\omega)) - J(\omega))^2 \leqslant \Delta_n^2. \quad (13)$$

We now apply the Devroye inequality to $J_n(\omega)$ in the form (12) and get

$$\mathbb{E}((J_n(\omega) - \mathbb{E}(J_n(\omega)))^2) \leqslant \frac{(4\pi + 1 + \log n)^2 L_u^4}{n}\, A^{2\eta} D. \quad (14)$$

Using (13), (14) and (11), it follows that

$$\mathbb{E}(\bar{Q}_n^2) \leqslant (N+1)\left(\frac{(4\pi + 1 + \log n)^2 L_u^4}{n} A^{2\eta} D + \Delta_n^2\right). \quad (15)$$

This completes the proof. ∎

**Proposition 3.2.** *There exists a positive constant $S$ such that for any function $u$ satisfying hypothesis 3.1, and any $n \geqslant 1$, we have*

$$\mathbb{E}\left(\left(\sup_{\omega\in[0,2\pi]} |J_n(\omega) - \tilde{J}_n(\omega)|\right)^2\right) \leqslant S \inf_{N\geqslant 1}\left\{(N+1)\left[\frac{C(1)^2 + DA^{2\eta}L_u^4}{n} + \Delta_n^2\right] + \left[\frac{C(1)}{N} + \Delta_N\right]^2\right\}.$$

The proof is rather similar to the previous one.

**Proof.** Let

$$R_n = \sup_{\omega\in[0,2\pi]} |\tilde{J}_n(\omega) - J_n(\omega)|. \quad (16)$$

Let $N$ be an integer and define as before the sequence of numbers $(\omega_p)$ by $\omega_p = 2\pi p/N$ for $p = 0, \ldots, N$. It follows at once from the monotonicity of $J_n$ and $\tilde{J}_n$ that

$$R_n \leqslant \max\left(\sup_{0\leqslant p\leqslant N-1} |J_n(\omega_{p+1}) - \tilde{J}_n(\omega_p)|, \sup_{0\leqslant p\leqslant N-1} |J_n(\omega_p) - \tilde{J}_n(\omega_{p+1})|\right).$$

We now have

$$R_n \leqslant \bar{R}_n + \sup_{0\leqslant p\leqslant N-1} |J_n(\omega_p) - J_n(\omega_{p+1})|,$$

where

$$\bar{R}_n = \sup_{0\leqslant p\leqslant N} |\tilde{J}_n(\omega_p) - J_n(\omega_p)|. \quad (17)$$

Now we have the estimate

$$\mathbb{E}\left(\sup_{0\leqslant p\leqslant N-1}(J_n(\omega_p)-J_n(\omega_{p+1}))^2\right)$$
$$\leqslant 6\mathbb{E}\left(\sup_{0\leqslant p\leqslant N}(J_n(\omega_p)-J(\omega_p))^2\right)+3\sup_{0\leqslant p\leqslant N-1}(J(\omega_p)-J(\omega_{p+1}))^2.$$

Using proposition 3.1 to estimate the first term and (9) for the second one, we obtain

$$\mathbb{E}\left(\sup_{0\leqslant p\leqslant N-1}(J_n(\omega_p)-J_n(\omega_{p+1}))^2\right)$$
$$\leqslant 12(N+1)\left(\frac{4\pi^2L_u^4D}{n}A^{2\eta}+\Delta_n^2\right)+60\pi^2\left(\frac{C(1)}{N}+\Delta_N\right)^2. \tag{18}$$

We obviously have

$$\mathbb{E}(\bar{R}_n^2)\leqslant\sum_{p=0}^{N-1}\mathbb{E}((J_n(\omega_p)-\tilde{J}_n(\omega_p))^2). \tag{19}$$

We now have to estimate each term $\mathbb{E}((J_n(\omega_p)-\tilde{J}_n(\omega_p))^2)$. Observe that for any $\omega$

$$\mathbb{E}((J_n(\omega)-\tilde{J}_n(\omega))^2)=\mathrm{var}(J_n(\omega)-\tilde{J}_n(\omega))+(\mathbb{E}(J_n(\omega)-\tilde{J}_n(\omega)))^2.$$

Let $S_n := \sum_{j=1}^n u(X_j)$. A simple computation yields

$$J_n(\omega)-\tilde{J}_n(\omega)=\omega\left(\frac{S_n}{n}-\mathbb{E}(u(X_1))\right)^2+\frac{1}{n}\left(\frac{S_n}{n}-\mathbb{E}(u(X_1))\right)$$
$$\times\sum_{j\neq\ell}^{n}\frac{\sin((j-\ell)\omega)}{j-\ell}\left(2u(X_\ell)-\mathbb{E}(u(X_1))-\frac{S_n}{n}\right). \tag{20}$$

An easy computation leads to

$$\mathbb{E}(J_n(\omega)-\tilde{J}_n(\omega))=\left(\omega-\frac{1}{n}\sum_{j\neq\ell}^{n}\frac{\sin((j-\ell)\omega)}{j-\ell}\right)\mathbb{E}\left(\left(\frac{S_n}{n}-\mathbb{E}(u(X_1))\right)^2\right)$$
$$+\frac{2}{n^2}\sum_{r=1}^{n}\sum_{\ell=1}^{n}\mathbb{E}((u(X_r)-\mathbb{E}(u(X_1)))(u(X_\ell)-\mathbb{E}(u(X_1))))\sum_{j\neq\ell}^{n}\frac{\sin((j-\ell)\omega)}{j-\ell}.$$

An easy computation using lemma A.1 shows that there is a constant $c_1>0$ such that for all integer $n$

$$\sup_{\omega\in[0,2\pi]}\left|\left(\omega-\frac{1}{n}\sum_{j\neq\ell}^{n}\frac{\sin((j-\ell)\omega)}{j-\ell}\right)\mathbb{E}\left(\left(\frac{S_n}{n}-\mathbb{E}(u(X_1))\right)^2\right)\right|\leqslant c_1\left(\frac{C(1)}{n}+\Delta_n\right).$$

Similarly, there exists a constant $c_2>0$ such that

$$\sup_{\omega\in[0,2\pi]}\left|\frac{2}{n^2}\sum_{r=1}^{n}\sum_{\ell=1}^{n}C(|\ell-r|+1)\sum_{j\neq\ell}^{n}\frac{\sin((j-\ell)\omega)}{j-\ell}\right|\leqslant c_2\left(\frac{C(1)}{n}+\Delta_n\right).$$

Combining these two estimates, one gets

$$\sup_{\omega\in[0,2\pi]}(\mathbb{E}(J_n(\omega)-\tilde{J}_n(\omega)))^2\leqslant c_3\left(\frac{C(1)}{n}+\Delta_n\right)^2, \tag{21}$$

where $c_3 > 0$ is a constant (independent of $n$). We now apply the Devroye inequality to $J_n(\omega) - \tilde{J}_n(\omega)$ using (20) and lemma A.1. We easily obtain the estimate

$$\sup_{\omega\in[0,2\pi]} \operatorname{var}(J_n(\omega) - \tilde{J}_n(\omega)) \leqslant \frac{c_4 D A^{2\eta} L_u^4}{n}. \tag{22}$$

It follows that

$$\mathbb{E}(\bar{R}_n^2) \leqslant N\left(c_3\left(\frac{C(1)}{n} + \Delta_n\right)^2 + \frac{c_4 D A^{2\eta} L_u^4}{n}\right).$$

The proposition follows by combining this estimate with (18). ■

Theorem 3.1 is proved by combining propositions 3.1 and 3.2.

## 4. Correlation dimension

We recall that the correlation dimension $d_c = d_c(\mu)$ of the measure $\mu$ (recall that $\mu$ is the common distribution of the $X_k$) is defined by

$$\lim_{\epsilon\downarrow 0} \frac{\log \int \mu(B(x',\epsilon))\,\mathrm{d}\mu(x')}{\log \epsilon^{-1}}$$

provided the limit exists (where $B(x',\epsilon)$ is the ball of centre $x'$ and radius $\epsilon$). In practice, one determines for large $n$ the power-law behaviour in $\epsilon$ of $K_{n,\epsilon}^{\vartheta}(x, f(x), \ldots, f^{n-1}(x))$ where

$$K_{n,\epsilon}^{\vartheta}(x_1,\ldots,x_n) = \frac{1}{n^2}\sum_{i\neq j} \vartheta(\epsilon - d(x_i,x_j))$$

and $\vartheta$ is the Heaviside function (i.e. the characteristic function of $\mathbb{R}^+$). It is known that (see, e.g. [14])

$$\lim_{n\to\infty} K_{n,\epsilon}^{\vartheta}(x, f(x),\ldots,f^{n-1}(x)) = \int \mu(B(x',\epsilon))\,\mathrm{d}\mu(x')$$

for $\mu$-almost all $x$ and every continuity point of the non-increasing function $\epsilon \mapsto \int \mu(B(y,\epsilon))\,\mathrm{d}\mu(y)$.

To proceed we need to replace $K_{n,\epsilon}^{\vartheta}(x_1,\ldots,x_n)$ by a component-wise Lipschitz function. For any real-valued Lipschitz function $\phi$, define the sequence of component-wise Lipschitz functions:

$$K_{n,\epsilon}^{\phi}(x_1,\ldots,x_n) := \frac{1}{n^2}\sum_{i\neq j} \phi\left(1 - \frac{d(x_i,x_j)}{\epsilon}\right). \tag{23}$$

**Theorem 4.1.** *For any real-valued Lipschitz function $\phi$, for any $0 < \eta \leqslant 1$, there exists a constant $C = C(\eta) > 0$ such that for any $\epsilon > 0$ and any integer $n$, we have*

$$\operatorname{var}(K_{n,\epsilon}^{\phi}) \leqslant \frac{C}{\epsilon^{2\eta} n}. \tag{24}$$

The proof is a direct application of the Devroye inequality (3).

Several functions $\phi$ are used in the literature. A simple one is given by

$$\phi_0(y) = \begin{cases} 0 & \text{for } y < -\frac{1}{2}, \\ \frac{1}{2} + y & \text{for } -\frac{1}{2} \leqslant y \leqslant \frac{1}{2}, \\ 1 & \text{for } y > \frac{1}{2}. \end{cases}$$

One verifies easily that for all $y \in \mathbb{R}$

$$\vartheta(1-2y) \leqslant \phi_0(1-y) \leqslant \vartheta\left(1 - \frac{y}{2}\right). \tag{25}$$

This implies immediately,

$$K^{\vartheta}_{n,\epsilon/2}(x_1,\ldots,x_n) \leqslant K^{\phi_0}_{n,\epsilon}(x_1,\ldots,x_n) \leqslant K^{\vartheta}_{n,2\epsilon}(x_1,\ldots,x_n) \tag{26}$$

for all $x_1,\ldots,x_n$, $\epsilon > 0$ and $n \geqslant 1$. It follows that, when $d_c > 0$, we have

$$K^{\vartheta}_{n,\epsilon}(x, f(x),\ldots,f^{n-1}(x)) \approx \epsilon^{d_c} \qquad \text{as } \epsilon \to 0$$

is equivalent to

$$K^{\phi_0}_{n,\epsilon}(x, f(x),\ldots,f^{n-1}(x)) \approx \epsilon^{d_c} \qquad \text{as } \epsilon \to 0.$$

Requiring that the typical value is smaller than the size of fluctuations (standard deviations) leads to $\epsilon^{d_c} \gtrsim 1/(\epsilon^{\eta}\sqrt{n})$. In other words,

$$n \gtrsim \epsilon^{-2(d_c+\eta)}.$$

In some iid cases, the optimal estimate has been obtained in [11].

## 5. Empirical measure

We recall that the empirical measure of a sample $X_1,\ldots,X_n$ is a random measure on $\mathbb{R}^d$ defined by

$$\mathcal{E}_n = \frac{1}{n}\sum_{j=1}^{n}\delta_{X_j},$$

where $\delta$ denotes the Dirac measure. We recall that from Birkhoff's ergodic theorem, almost-surely this sequence of random measures weakly converges to the common distribution $\mu$ of the $X_k$'s. It is natural to ask for the speed of this convergence. This of course depends on the distance chosen on the set of probability measures. We will consider the Kantorovich distance [9, 16] defined for two probability measures $\mu_1$ and $\mu_2$ on $\mathbb{R}^d$ by

$$\kappa(\mu_1,\mu_2) = \sup_{g\in\mathscr{L}} \int g(x)\,\mathrm{d}(\mu_1-\mu_2)(x), \tag{27}$$

where $\mathscr{L}$ denotes the set of real-valued Lipschitz functions on $\mathbb{R}^d$ with Lipschitz constant at most one.

We now state the theorems of this section.

**Theorem 5.1.** *By the Devroye inequality (3) we have, for all $n \geqslant 1$,*

$$\operatorname{var}(\kappa(\mathcal{E}_n,\mu)) \leqslant \frac{D(1)}{n}.$$

The proof follows at once from the Devroye inequality (3) using separately the following Lipschitz function of $n$ variables

$$K(x_1,\ldots,x_n) = \sup_{g\in\mathscr{L}}\left[\frac{1}{n}\sum_{j=1}^{n} g(x_j) - \mathbb{E}(g)\right].$$

To get a probability estimate based on this result one needs to give an upper bound for $\mathbb{E}(\kappa(\mathcal{E}_n,\mu))$. The bound we are so far able to obtain in dimension larger than 1 is too pessimistic. We explain below how to obtain a more satisfactory bound in dimension 1. We will require the following property for the auto-covariance. We will denote by $\|u\|_\eta$ the $\eta$-Hölder constant of $u$ (which is bounded by $\mathcal{O}(1)L_1(u)$).

**Hypothesis 5.1.** *For any $\eta \in ]0, 1]$ there is a constant $C_\eta > 0$ such that the auto-covariance $C_u(\ell)$ of any $\eta$-Hölder continuous function $u$ satisfies*

$$\sum_{\ell=1}^{\infty} |C_u(\ell)| \leqslant C_\eta \|u\|_\eta^2.$$

This leads to the following theorem.

**Theorem 5.2.** *Assume that the process $(X_k)$ takes values in $\mathbb{R}$ and that the auto-covariance of $\eta$-Hölder continuous functions satisfies hypothesis 5.1. Then, for any $\eta \in ]0, 1]$, there exists a positive constant $a(\eta)$ such that for all $t > 0$ and $n \geqslant 1$, we have*

$$\mathbb{P}\left(\kappa(\mathcal{E}_n, \mu) > t + \frac{a(\eta)}{n^{1/(2(1+\eta))}}\right) \leqslant \frac{D(1)}{nt^2}.$$

**Remark.** If $a(\eta)$ behaves like $1/\eta$ as $\eta$ tends to zero, then one can optimize by taking $\eta = 1/\log n$.

**Proof.** The theorem of Dall'Aglio [7] states that

$$\kappa(\mu_1, \mu_2) = \int_{\mathbb{R}} |F_{\mu_1}(t) - F_{\mu_2}(t)| \, \mathrm{d}t,$$

where $F_\mu(t)$ is the distribution function of $\mu$.

We wish to estimate the Kantorovich distance between the empirical measure $\mathcal{E}_n$ and $\mu$ (the common distribution of the $X_k$). In this case we have

$$\kappa(\mathcal{E}_n, \mu) = \int_{-A}^{A} \mathrm{d}t \left| \frac{1}{n} \sum_{k=0}^{n-1} \vartheta(t - X_k) - F_\mu(t) \right|,$$

since we assumed from the very beginning that $\|X_k\| \leqslant A$ $\mathbb{P}$-almost-surely, and $\vartheta$ denotes the Heaviside function.

In order to use the decay of correlations, we replace the Heaviside function by a Hölder continuous function $g_\delta$ parametrized by a positive $\delta$ and defined by

$$g_\delta(s) = \begin{cases} 0 & \text{if } s < -\delta, \\ 1 + s/\delta & \text{if } -\delta \leqslant s \leqslant 0, \\ 1 & \text{if } s > 0. \end{cases}$$

We immediately obtain

$$\kappa(\mathcal{E}_n, \mu) \leqslant \delta + \int_{-A}^{A} \mathrm{d}t \left| \frac{1}{n} \sum_{k=0}^{n-1} g_\delta(t - X_k) - F_\mu(t) \right|. \tag{28}$$

We have

$$\begin{aligned} \mathbb{E}(\kappa(\mathcal{E}_n, \mu)) &\leqslant \delta + \mathbb{E}\left( \int_{-A}^{A} \mathrm{d}t \left| \frac{1}{n} \sum_{k=0}^{n-1} g_\delta(t - X_k) - \mathbb{E}(g_\delta(t - X_1)) \right| \right) \\ &\quad + \int_{-A}^{A} \mathrm{d}t \mathbb{E}|g_\delta(t - X_1) - \vartheta(t - X_1)| \\ &\leqslant 2\delta + \mathbb{E}\left( \int_{-A}^{A} \mathrm{d}t \left| \frac{1}{n} \sum_{k=0}^{n-1} g_\delta(t - X_k) - \mathbb{E}(g_\delta(t - X_1)) \right| \right). \end{aligned}$$

Using the Cauchy–Schwarz inequality as in [6], one is led to use the decay of auto-covariance of the functions $g_\delta(t-\cdot)$. Using hypothesis 5.1 we get

$$\mathbb{E}(\kappa(\mathcal{E}_n,\mu)) \leqslant 2\delta + \frac{\mathcal{O}(1)}{\delta^\eta\sqrt{n}}.$$

Using the Chebychev inequality, the above estimate with $\delta = n^{-1/2(1+\eta)}$ and theorem 5.1 we get the theorem. ∎

For the application to dynamical systems satisfying the Devroye inequality (see [5]), moreover we need to verify hypothesis 5.1. It is often proved, see, e.g. [18], that the auto-covariance of observables belonging to a Banach space have a common upper bound for their rate of decay. It turns out that this implies a uniform rate of decay for all functions of norm less than or equal to one, this is the content of theorem B.1 proved in appendix B. So, if this decay is summable then theorem 5.2 holds. For the systems studied in [18], hypothesis 5.1 can be deduced using the estimates provided by approximations #1 and #2 and point 4.2 appearing in that paper.

## 6. Kernel density estimation for 1D maps

In this section we assume that $d=1$, namely that the process $(X_k)$ takes values in a bounded interval of $\mathbb{R}$. Moreover, we assume that the common distribution $\mu$ of the $X_k$ is absolutely continuous (with respect to Lebesgue measure) and denote its density by $\Phi$. We consider the random empirical densities $(h_n)$ defined by

$$h_n(X_1,\ldots,X_n;s) = \frac{1}{n\alpha_n}\sum_{j=1}^{n}\psi((s-X_j)/\alpha_n),$$

where $\alpha_n$ is a positive sequence converging to 0 and such that $n\alpha_n$ converges to $+\infty$, and $\psi$ (the kernel) is a bounded, non-negative Lipschitz continuous function with compact support whose integral equals 1. We are interested in the $L^1$ convergence of these empirical densities to the density $\Phi$ of the common distribution $\mu$ of the $X_k$'s.

**Theorem 6.1.** *Assume that the probability density $\Phi$ satisfies*

$$\int |\Phi(s)-\Phi(s-y)|\,\mathrm{d}s \leqslant C|y|^\tau \tag{29}$$

*for some $C>0$, $\tau>0$ and any $y\in\mathbb{R}$. Suppose also that hypothesis 5.1 holds. Then, for any $\eta\in]0,1]$, for any $\psi$ as above, there exists a constant $C'=C'(\eta,\psi)>0$ such that for any integer $n$ and for any $t>C'(\alpha_n^\tau + 1/(\sqrt{n}\alpha_n^{1+\eta}))$, we have*

$$\mathbb{P}\left(\int |h_n(X_1,\ldots,X_n;s)-\Phi(s)|\,\mathrm{d}s > t\right) \leqslant \frac{C'}{t^2 n\alpha_n^{2\eta}}.$$

**Proof.** We define the functions

$$K(x_1,\ldots,x_n) = \int\left|\frac{1}{n\alpha_n}\sum_{j=1}^{n}\psi((s-x_j)/\alpha_n) - \Phi(s)\right|\mathrm{d}s.$$

It is easy to verify that the Hölder constants of this $\eta$-Hölder continuous function satisfy

$$\max_{1\leqslant j\leqslant n} L_j \leqslant \frac{\mathcal{O}(1)}{n\alpha_n^\eta}.$$

Hence, using the Devroye inequality (3), we immediately obtain

$$\mathrm{var}(K) \leqslant \frac{\mathcal{O}(1)}{n\alpha_n^{2\eta}}.$$

The theorem will follow using this and the Chebychev inequality provided we have an upper bound for $\mathbb{E}(K)$. To this purpose we will follow the lines of the proof of theorem III.2 in [6] with the appropriate modifications.

We first estimate the $L^1$-norm of $\Phi - \mathbb{E}(h_n)$. We obtain, using (29) the upper bound

$$\int |\Phi(s) - \mathbb{E}(h_n)(s)|\, \mathrm{d}s \leqslant \alpha_n^{-1} \int \psi\left(\frac{y}{\alpha_n}\right) \mathrm{d}y \int |\Phi(s) - \Phi(s-y)|\, \mathrm{d}s$$
$$\leqslant C\alpha_n^{-1} \int \psi\left(\frac{y}{\alpha_n}\right) |y|^\tau\, \mathrm{d}y \leqslant \mathcal{O}(1)\alpha_n^\tau.$$

We now bound from above the integral

$$\int \mathrm{d}s\ \mathbb{E}(|h_n(X_1,\ldots,X_n;s) - \mathbb{E}(h_n)(s)|).$$

By a well-known computation we have

$$\mathrm{var}(h_n(X_1,\ldots,X_n;s))$$
$$\leqslant \frac{2}{n\alpha_n^2}\sum_{\ell=1}^{n} \mathbb{E}(((\psi((s-X_1)/\alpha_n)) - \tilde{\psi}_n(s))((\psi((s-X_\ell)/\alpha_n)) - \tilde{\psi}_n(s))),$$

where $\tilde{\psi}_n(s) = \mathbb{E}(\psi((s-X_1)/\alpha_n))$. Using the Cauchy–Schwarz inequality and hypothesis 5.1, as in the proof of section 5, we get

$$\int \mathrm{d}s\ \mathbb{E}(|h_n(X_1,\ldots,X_n;s) - \mathbb{E}(h_n)(s)|) \leqslant \frac{\mathcal{O}(1)}{\sqrt{n}\alpha_n^{1+\eta}}.$$

Summarizing, we obtain

$$\mathbb{E}(K) \leqslant \mathcal{O}(1)\left(\alpha_n^\tau + \frac{1}{\sqrt{n}\alpha_n^{1+\eta}}\right).$$

The theorem now follows by the Chebychev inequality and the Devroye inequality. ■

For results on kernel density estimation in the context of piecewise expanding maps on the interval, we refer to [15] and references therein.

We will prove in appendix C that the class of dynamical systems considered in [5, 18], that is the class introduced in [18], satisfies (29) for 1D systems. As explained at the end of the previous section, it also satisfies hypothesis 5.1. Hence the theorem applies. This class includes quadratic maps for a set of parameters of positive Lebesgue measure [18].

## 7. Shadowing and mismatch

For a fixed integer $n$, let $E$ be a measurable subset of $\mathbb{R}^{nd}$. For a trajectory $Y_1,\ldots,Y_n$ of length $n$ of the process $(X_k)$ which is outside $E$, how well can we approximate this trajectory by a trajectory $(X_1,\ldots,X_n)$ of the process belonging to $E$? We first start with a result about the average quality of this 'shadowing'. We will denote by $\mathcal{T}_n$ the set of trajectories of length $n$ of the process.

**Theorem 7.1.** *For any integer $n$, for any measurable subset $E$ of $\mathbb{R}^{nd}$, with $\mathbb{P}((X_1,\ldots,X_n) \in E) > 0$, the function*[3] *defined by*

$$\mathcal{Z}_E(Y_1,\ldots,Y_n) = \frac{1}{n} \inf_{(X_1,\ldots,X_n)\in E\cap\mathcal{T}_n} \sum_{j=1}^{n} \|X_j - Y_j\|$$

[3] The function $\mathcal{Z}_E$ is measurable, see [6].

*satisfies for any* $t > 0$ *the inequality*

$$\mathbb{P}\left(\mathcal{Z}_E \geqslant \frac{1}{n^{1/3}}\left(t + \frac{2^{4/3}D^{1/3}}{\mathbb{P}((X_1,\ldots,X_n)\in E)}\right)\right) \leqslant \frac{D}{n^{1/3}t^2},$$

*where* $D > 0$ *is the constant appearing in (3).*

**Proof.** We first apply the Devroye inequality (3) to the function

$$K(x_1,\ldots,x_n) = \frac{1}{n}\inf_{(X_1,\ldots,X_n)\in E\cap\mathcal{T}_n}\sum_{j=1}^{n}\|X_j - x_j\|.$$

We get

$$\operatorname{var}(\mathcal{Z}_E) \leqslant \frac{D}{n}\cdot \tag{30}$$

The Chebychev inequality yields for any $s > 0$

$$\mathbb{P}\left(\mathcal{Z}_E \geqslant \mathbb{E}(\mathcal{Z}_E) + \frac{s}{n^{1/3}}\right) \leqslant \frac{D}{n^{1/3}s^2}\cdot$$

Proceeding as in [6] and optimizing over $s$ we obtain

$$\mathbb{E}(\mathcal{Z}_E) \leqslant \frac{2^{4/3}D^{1/3}}{n^{1/3}\mathbb{P}((X_1,\ldots,X_n)\in E)}\cdot$$

The theorem follows using again the Chebychev inequality. ∎

**Remark.** There is another way of estimating from above $\mathbb{E}(\mathcal{Z}_E)$. For this observe that $\mathcal{Z}_E$ vanishes in $E$. Therefore it follows from (30) that

$$\mathbb{P}((X_1,\ldots,X_n)\in E)(\mathbb{E}(\mathcal{Z}_E))^2 \leqslant \frac{D}{n}.$$

Hence

$$\mathbb{E}((\mathcal{Z}_E)^2) \leqslant \frac{\sqrt{D}}{\sqrt{n\mathbb{P}((X_1,\ldots,X_n)\in E)}}\cdot$$

We now derive a similar result for the number of mismatch at a given precision.

**Theorem 7.2.** *For any integer* $n$*, for any measurable subset* $E$ *of* $\mathbb{R}^{nd}$*, with* $\mathbb{P}((X_1,\ldots,X_n)\in E) > 0$*, and for any* $\epsilon > 0$*, the function defined by*

$$\mathcal{Z}'_{E,\epsilon}(Y_1,\ldots,Y_n) = \frac{1}{n}\inf_{(X_1,\ldots,X_n)\in E\cap\mathcal{T}_n}\operatorname{Card}\{1 \leqslant j \leqslant n : \|X_j - Y_j\| > \epsilon\}$$

*satisfies for any* $t > 0$ *the following*

$$\mathbb{P}\left(\mathcal{Z}'_{E,\epsilon} \geqslant \frac{1}{\epsilon^{2/3}n^{1/3}}\left(t + \frac{2^{4/3}D^{1/3}}{\mathbb{P}((X_1,\ldots,X_n)\in E)}\right)\right) \leqslant \frac{D}{\epsilon^{2/3}n^{1/3}t^2},$$

*where* $D > 0$ *is the constant appearing in (3).*

The industrious reader can follow the lines of the proof of theorem IV.2 in [6]. Using Hölder estimates instead of Lipschitz estimates yields the same formula with $\epsilon^{2/3}$ replaced with $\epsilon^{2\eta/3}$, for any $0 < \eta \leqslant 1$. However the constant $D$ depends on $\eta$ in an implicit way, so it is not clear how to optimize over $\eta$.

For the case of dynamical systems, given an initial condition $y$ outside a measurable subset $S$ of the phase space with positive measure, the questions considered above mean that we look at how good is the shadowing of the orbit of $y$ by an orbit starting from $S$ (in that case $E = S \times \mathbb{R}^{(n-1)d}$).

## 8. Almost-sure central limit theorem

We say that the process $(u(X_k))$, where $u$ is a real-valued function, satisfies the central limit theorem if

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sum_{j=1}^{n} u(X_j) - n\mathbb{E}(u)}{\sigma\sqrt{n}} \leqslant t\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} \mathrm{e}^{-\xi^2/2}\,\mathrm{d}\xi, \tag{31}$$

where $\sigma^2 = \sigma^2(u)$ is assumed to be strictly positive and is defined by

$$\sigma^2 = C(1) + 2\sum_{\ell=2}^{\infty} C(\ell), \tag{32}$$

where we assume that the series is finite (see (4) for the definition of $C(\ell)$).

We will prove an almost-sure central limit theorem, see, e.g. [1], for a review of this field. Our result is slightly stronger since it asserts the convergence in the Kantorovich distance $\kappa$ already used above, see formula (27). We shall use it for measures on $\mathbb{R}$ and real-valued Lipschitz functions on $\mathbb{R}$. Note that we can replace $g$ by $g - g(0)$ in (27) since $\mu_1$ and $\mu_2$ are probability measures. In other words, there is no loss of generality in assuming

$$g \in \mathscr{L}_0 := \{g \in \mathscr{L} \mid g(0) = 0\}.$$

It is convenient to define the sequence of weighted empirical (random) measures of the normalized partial sum $S_k = u(X_1) + \cdots + u(X_k)$ by

$$\mathcal{A}_n = \frac{1}{D_n}\sum_{k=1}^{n}\frac{1}{k}\,\delta_{S_k/\sqrt{k}},$$

where $D_n = \sum_{k=1}^{n} 1/k$. We shall investigate the convergence of this sequence of weighted empirical measures to the Gaussian measure in the Kantorovich metric.

We now state the result of this section.

**Theorem 8.1.** *Consider the process $(u(X_k))$ where $u$ is a Hölder continuous function with zero $\mu$ average (recall that $\mu$ is the common law of the $X_k$). Assume that $\sigma^2 \neq 0$ (see (32)), that the auto-covariance of $(u(X_k))$ is absolutely summable and that (31) holds (central limit theorem). Then $\mathbb{P}$-almost-surely*

$$\lim_{n\to\infty} \kappa(\mathcal{A}_n, \mathcal{N}(0,\sigma^2)) = 0, \tag{33}$$

*where $\mathcal{N}(0,\sigma^2)$ is the Gaussian measure with mean zero and variance $\sigma^2$.*

The assumptions of the theorem hold for the class of dynamical systems discussed in [5,18]. For piecewise expanding maps of the interval, a stronger result was proved in [4]. Notice that this theorem immediately implies that almost-surely $\mathcal{A}_n$ converges weakly to the Gaussian measure.

**Proof.** We first prove that

$$\lim_{n\to\infty} \mathbb{E}(\kappa(\mathcal{A}_n, \mathcal{N}(0,\sigma^2))) = 0.$$

Let $B$ be a positive constant to be chosen large enough later on. We have for any $g \in \mathscr{L}_0$ vanishing at 0 and any $x$

$$|g(x)| \leqslant |x|.$$

Therefore

$$\kappa(\mathcal{A}_n, \mathcal{N}(0,\sigma^2)) \leqslant \sup_{g\in\mathscr{L}_0} \int_{-B}^{B} g(\mathrm{d}\mathcal{A}_n - \mathrm{d}\mathcal{N}(0,\sigma^2)) + \int_{|y|>B} |y|\,\mathrm{d}\mathcal{A}_n(y) + \int_{|y|>B} |y|\,\mathrm{d}\mathcal{N}(0,\sigma^2)(y). \tag{34}$$

We first estimate the expectation of the second term uniformly in $n$. Since the correlations are absolutely summable, we get for any $j$

$$\mathbb{E}(S_j^2)^{1/2} \leqslant \mathcal{O}(1)\sqrt{j}. \tag{35}$$

Therefore, using the Cauchy–Schwarz and Bienaymé–Chebychev inequalities we get

$$\mathbb{E}\left(\int_{|y|>B} |y|\,\mathrm{d}\mathcal{A}_n(y)\right) = \frac{1}{D_n}\sum_{k=1}^{n}\frac{1}{k}\mathbb{E}\left(\chi_{[B,\infty[}\left(\frac{|S_k|}{\sqrt{k}}\right)\frac{|S_k|}{\sqrt{k}}\right) \leqslant \frac{\mathcal{O}(1)}{B}. \tag{36}$$

In order to estimate the first term on the rhs of (34), we observe that since $[-B, B]$ is compact, we can apply Ascoli–Arzela theorem to conclude that for any $\epsilon > 0$ there is a number $r = r(\epsilon)$ and a finite sequence $g_1, \ldots, g_r$ of functions in $\mathscr{L}_0$ such that for any $g \in \mathscr{L}_0$, there is at least one integer $1 \leqslant j \leqslant r$ such that

$$\sup_{|y|\leqslant B} |g(y) - g_j(y)| \leqslant \epsilon.$$

Therefore

$$\sup_{g\in\mathscr{L}_0} \int_{-B}^{B} g(\mathrm{d}\mathcal{A}_n - \mathrm{d}\mathcal{N}(0,\sigma^2)) \leqslant \sup_{1\leqslant j\leqslant r(\epsilon)} \int_{-B}^{B} g_j(\mathrm{d}\mathcal{A}_n - \mathrm{d}\mathcal{N}(0,\sigma^2)) + 2\epsilon. \tag{37}$$

We now consider the $r$ sequences of random variables

$$Y_{n,j} = \int_{-B}^{B} g_j(\mathrm{d}\mathcal{A}_n - \mathrm{d}\mathcal{N}(0,\sigma^2))$$

with $1 \leqslant j \leqslant r$.

We first estimate the variance of $Y_{n,j}$. Let the sequence of functions $(K_{n,j})$ of $n$ variables $x_1, \ldots, x_n$ and $1 \leqslant j \leqslant r$ be defined by

$$K_{n,j}(x_1,\ldots,x_n) = \frac{1}{D_n}\sum_{k=1}^{n}\frac{1}{k}\left[g_j\left(\frac{\sum_{l=1}^{k} u(x_l)}{\sqrt{k}}\right) - \mathcal{N}(0,\sigma^2)(g_j)\right],$$

where $\mathcal{N}(0,\sigma^2)(\cdot)$ denotes the integration against the Gaussian measure. It is easy to verify that all these functions are separately Lipschitz with respect to all their variables, and that the Lipschitz constant with respect to the $q$th variable is bounded by $\mathcal{O}(1)/(\sqrt{q}D_n)$ uniformly in $n$. Applying the Devroye inequality (3), we get

$$\mathrm{var}(Y_{n,j}) = \mathrm{var}(K_{n,j}) \leqslant \frac{\mathcal{O}(1)}{D_n^2}\sum_{q=1}^{n}\frac{1}{q} \leqslant \frac{\mathcal{O}(1)}{D_n}.$$

We now have, using the Cauchy–Schwarz inequality,

$$\mathbb{E}\left(\sup_{1\leqslant j\leqslant r} Y_{n,j}\right) \leqslant \mathbb{E}\left(\sum_{j=1}^{r}|Y_{n,j}|\right) \leqslant \sum_{j=1}^{r}\mathbb{E}(|Y_{n,j} - \mathbb{E}(Y_{n,j})|) + \sum_{j=1}^{r}|\mathbb{E}(Y_{n,j})|$$
$$\leqslant \sum_{j=1}^{r}\mathrm{var}(Y_{n,j})^{1/2} + \sum_{j=1}^{r}|\mathbb{E}(Y_{n,j})| \leqslant \frac{r\mathcal{O}(1)}{\sqrt{D_n}} + \sum_{j=1}^{r}|\mathbb{E}(Y_{n,j})|.$$

By the central limit theorem (31), we have for each $j$

$$\lim_{n\to\infty} \mathbb{E}(Y_{n,j}) = 0$$

and therefore from the above estimates for a fixed $r(\epsilon)$ we have

$$\limsup_{n\to\infty} \mathbb{E}\left(\sup_{1\leqslant j\leqslant r(\epsilon)} \int_{-B}^{B} g_j(\mathrm{d}\mathcal{A}_n - \mathrm{d}\mathcal{N}(0,\sigma^2))\right) \leqslant 0.$$

It now follows from (36) and (37) that for any $\epsilon > 0$ and any $B > 0$,

$$0 \leqslant \limsup_{n\to\infty} \mathbb{E}(\kappa(\mathcal{A}_n, \mathcal{N}(0,\sigma^2))) \leqslant 2\epsilon + \frac{\mathcal{O}(1)}{B} + \int_{|y|>B} |y| \, \mathrm{d}\mathcal{N}(0,\sigma^2)(y).$$

Letting $B$ tend to infinity and $\epsilon$ to zero we get

$$\lim_{n\to\infty} \mathbb{E}(\kappa(\mathcal{A}_n, \mathcal{N}(0,\sigma^2))) = 0.$$

We now estimate the variance of $\kappa(\mathcal{A}_n, \mathcal{N}(0,\sigma^2))$. Applying the Devroye inequality (3) as above to the function $K_n$ of $n$ variables $x_1, \ldots, x_n$:

$$K_n(x_1,\ldots,x_n) = \sup_{g\in\mathcal{L}_0} \frac{1}{D_n} \sum_{j=1}^{n} \frac{1}{j}\left[g\left(\frac{\sum_{l=1}^{j} u(x_l)}{\sqrt{j}}\right) - \mathcal{N}(0,\sigma^2)(g)\right] \tag{38}$$

we get

$$\mathbb{E}([\kappa(\mathcal{A}_n, \mathcal{N}(0,\sigma^2)) - \mathbb{E}(\kappa(\mathcal{A}_n, \mathcal{N}(0,\sigma^2)))]^2) = \mathrm{var}(K_n) \leqslant \frac{\mathcal{O}(1)}{D_n^2} \sum_{j=1}^{n} \frac{1}{j} \leqslant \frac{\mathcal{O}(1)}{D_n}.$$

If for $0 < \rho < 1$ we define

$$n_k = \mathrm{e}^{k^{1+\rho}}$$

we conclude that

$$\sum_k \mathbb{E}([\kappa(\mathcal{A}_{n_k}, \mathcal{N}(0,\sigma^2)) - \mathbb{E}(\kappa(\mathcal{A}_{n_k}, \mathcal{N}(0,\sigma^2)))]^2) < \infty,$$

which implies by B Lévi's theorem that

$$\lim_{k\to\infty} (\kappa(\mathcal{A}_{n_k}, \mathcal{N}(0,\sigma^2)) - \mathbb{E}[\kappa(\mathcal{A}_{n_k}, \mathcal{N}(0,\sigma^2))]) = 0 \qquad \mathbb{P}\text{—almost-surely.}$$

We now observe that if $n_k < n \leqslant n_{k+1}$, we have

$$|\kappa(\mathcal{A}_n, \mathcal{N}(0,\sigma^2)) - \kappa(\mathcal{A}_{n_k}, \mathcal{N}(0,\sigma^2))| \leqslant \frac{D_n - D_{n_k}}{D_n} \kappa(\mathcal{A}_{n_k}, \mathcal{N}(0,\sigma^2)) + \sup_{g\in\mathcal{L}_0} \frac{1}{D_n} \sum_{j=n_k+1}^{n} \frac{1}{j}\left[g\left(\frac{S_j}{\sqrt{j}}\right) - \mathcal{N}(0,\sigma^2)(g)\right].$$

The first term tends to zero almost-surely by our previous estimates. We now prove that the second term tends to zero almost-surely. We have

$$\sup_{g\in\mathcal{L}} \frac{1}{D_n} \sum_{j=n_k+1}^{n} \frac{1}{j}\left[g\left(\frac{S_j}{\sqrt{j}}\right) - \mathcal{N}(0,\sigma^2)(g)\right] \leqslant \frac{1}{D_n} \sum_{j=n_k+1}^{n} \frac{1}{j}\left[\frac{|S_j|}{\sqrt{j}} + \mathcal{N}(0,\sigma^2)(|x|)\right]$$
$$\leqslant \frac{1}{D_{n_k}} \sum_{j=n_k+1}^{n_{k+1}} \frac{1}{j}\left[\frac{|S_j|}{\sqrt{j}} + \mathcal{N}(0,\sigma^2)(|x|)\right].$$

It follows easily from our choice of $(n_k)$ that

$$\lim_{k\to\infty} \frac{1}{D_{n_k}} \sum_{j=n_k+1}^{n_{k+1}} \frac{1}{j} \mathcal{N}(0,\sigma^2)(|x|) = 0.$$

We now prove the almost-sure convergence to zero of the sequence,

$$T_k = \frac{1}{D_{n_k}} \sum_{j=n_k+1}^{n_{k+1}} \frac{|S_j|}{j^{3/2}}.$$

For this purpose we estimate the expectation of the square of $T_k$. Using the Cauchy–Schwarz inequality and (35) we obtain

$$\mathbb{E}(T_k^2) \leqslant \frac{1}{D_{n_k}^2} \sum_{p,q=n_k+1}^{n_{k+1}} \frac{\mathbb{E}(S_p^2)^{1/2}}{p^{3/2}} \frac{\mathbb{E}(S_q^2)^{1/2}}{q^{3/2}} \leqslant \frac{(\log n_{k+1} - \log n_k + \mathcal{O}(1))^2}{D_{n_k}^2} \leqslant \frac{\mathcal{O}(1)}{k^2}.$$

It follows at once that $\mathbb{E}(T_k^2)$ is summable in $k$. The result now follows using B Lévi's theorem. The theorem is proved. ■

**Remarks.** We note that the above proof also leads to an estimate on the probability that $\kappa(\mathcal{A}_n, \mathcal{N}(0, \sigma^2))$ is larger than some given number $\epsilon > 0$.

For a dynamical system $(\Omega, f)$ it often occurs that the invariant measure is supported on an attractor which is a small subset of the phase space $\Omega$. When there exists a SRB measure, one would like to have theorem 8.1 almost-surely with respect to Lebesgue measure on $\Omega$. Assuming that the stable foliation is absolutely continuous, and the forward contraction is uniform and exponential along local stable manifolds (see [18] for several examples), it is sufficient to prove that

$$\lim_{n\to\infty} |K_n(x, f(x), \ldots, f^{n-1}(x)) - K_n(\tilde{x}, f(\tilde{x}), \ldots, f^{n-1}(\tilde{x}))| = 0,$$

where $K_n$ is defined by (38) and $x, \tilde{x}$ belong to the same local stable manifold. This follows at once from the definition of $K_n$ and the uniform exponential contraction along local stable manifolds.

## Acknowledgments

JRC and PC acknowledge the CIC, Cuernavaca, México, for its warm hospitality, where part of this work was done as well as the Institut de Mathématiques de Bourgogne in Dijon. BS acknowledges the kind hospitality of CPhT at Ecole Polytechnique. The authors acknowledge the anonymous referee for a very careful reading of the paper.

## Appendix A. About trigonometric series

For the convenience of the reader we prove in this appendix the following (probably well known) result on trigonometric series for which we have not been able to locate a reference.

**Lemma A.1.** *We have the following*

$$\sup_{m\in\mathbb{N},\omega\in[0,2\pi]} \left| \sum_{k=1}^{m} \frac{\sin k\omega}{k} \right| < \infty\,.$$

**Proof.** First observe that it is enough to assume that $\omega \in [0, \pi]$. Now we have

$$\sum_{k=1}^{m} \frac{\sin k\omega}{k} = \frac{1}{2} \int_0^{\omega} \mathrm{e}^{\mathrm{i}s} \frac{1 - \mathrm{e}^{\mathrm{i}ms}}{1 - \mathrm{e}^{\mathrm{i}s}} \mathrm{d}s + \text{c.c.}$$

By an easy estimate, one gets

$$\sup_{m\in\mathbb{N},\omega\in[0,\pi]}\left|\int_0^\omega \mathrm{e}^{\mathrm{i}s}\,\frac{1-\mathrm{e}^{\mathrm{i}ms}}{1-\mathrm{e}^{\mathrm{i}s}}\,\mathrm{d}s-\int_0^\omega \mathrm{e}^{\mathrm{i}s}\,\frac{1-\mathrm{e}^{\mathrm{i}ms}}{\mathrm{i}s}\,\mathrm{d}s\right|<\infty.$$

Finally,

$$\int_0^\omega \mathrm{e}^{\mathrm{i}s}\,\frac{1-\mathrm{e}^{\mathrm{i}ms}}{2\mathrm{i}s}\,\mathrm{d}s+\text{c.c.}=\int_0^\omega\frac{\sin s}{s}\mathrm{d}s-\int_0^\omega\frac{\sin(m+1)s}{s}\mathrm{d}s=-\int_\omega^{(m+1)\omega}\frac{\sin s}{s}\,\mathrm{d}s.$$

It is well known that the modulus of this quantity is uniformly bounded in $\omega$ and $m$. ■

## Appendix B. On the uniform decay of correlations

In this appendix we prove a general result on decay of correlations which may be useful in other contexts. Consider a dynamical system on a phase space $\Omega$ given by a measurable map $f$ from $\Omega$ to itself. Let $\mu$ be an ergodic invariant measure. The decay of correlations is often proved in the following form: there is a non-increasing sequence $(\gamma_n)$ and two Banach spaces $\mathscr{B}_1$ and $\mathscr{B}_2$ of measurable functions on $\Omega$ such that for any functions $\psi_1\in\mathscr{B}_1$ and $\psi_2\in\mathscr{B}_2$, there is a constant $C_{\psi_1,\psi_2}$ such that for any integer $n$

$$\left|\int\psi_1\circ f^n\psi_2\,\mathrm{d}\mu-\int\psi_1\,\mathrm{d}\mu\int\psi_2\,\mathrm{d}\mu\right|\leqslant C_{\psi_1,\psi_2}\gamma_n. \tag{39}$$

It is often useful to have some information on the constant $C_{\psi_1,\psi_2}$, in particular if it can be bounded by a product of norms of the two functions (and a uniform constant). It turns out that this apparently stronger result follows from the previous estimate under the following natural assumptions.

(i) The constant functions belong to $\mathscr{B}_1$.
(ii) The integration with respect to $\mu$ defines a continuous linear functional on $\mathscr{B}_1$.
(iii) The Koopman operator $U$ (of composition with $f$) is continuous in $\mathscr{B}_1$.
(iv) $\mathscr{B}_2$ is contained in the dual of $\mathscr{B}_1$ (duality with respect to the integration by $\mu$) with a topology at least as fine as the dual norm topology.

As will become clear from the proof, the result below is due to the special form of the correlation integral.

**Theorem B.1.** *Assume the above properties (i)–(iv), and inequality (39) hold. Then there exists a constant $K$ such that for any integer $n$ and any $\psi_1\in\mathscr{B}_1$, $\psi_2\in\mathscr{B}_2$, we have*

$$\left|\int\psi_1\circ f^n\psi_2\,\mathrm{d}\mu-\int\psi_1\,\mathrm{d}\mu\int\psi_2\,\mathrm{d}\mu\right|\leqslant K\|\psi_1\|_{\mathscr{B}_1}\|\psi_2\|_{\mathscr{B}_2}\gamma_n. \tag{40}$$

A frequent example is $\mathscr{B}_1=L^\infty(\Omega,\,\mathrm{d}\mu)$, while $\mathscr{B}_2$ is a space of more regular functions (functions of bounded variation, Lipschitz or Hölder functions). In [18], $\mathscr{B}_1=\mathscr{B}_2$ is the space of Hölder continuous functions. We give below a proof based on the principle of uniform boundedness.

**Proof.** We first deal with the easy case where for some integer $n_0$ we have $\gamma_{n_0}=0$. For any $n>n_0$, using the identity $\psi_1\circ f^n=(\psi_1\circ f^{n-n_0})\circ f^{n_0}$ and (iii), we conclude that the correlation integral (lhs of (39)) is equal to zero for any $\psi_1\in\mathscr{B}_1$ and $\psi_2\in\mathscr{B}_2$. On the other hand, it follows from (iv) that there is a positive number $K_0$ such that

$$\sup_{\|\psi_1\|_{\mathscr{B}_1}\leqslant1,\|\psi_2\|_{\mathscr{B}_2}\leqslant1}\left|\int\psi_1\psi_2\,\mathrm{d}\mu\right|\leqslant K_0$$

and (40) follows immediately with

$$K = K_0 \sup_{n,\gamma_n>0} \|U^n\|_{\mathscr{B}_1} \gamma_n^{-1}.$$

We now assume $\gamma_n > 0$ for any integer $n$. We first control the dependence on $\psi_1$ and for this purpose we first fix $\psi_2 \in \mathscr{B}_2$. We then define a sequence of non-negative continuous functions $(p_n^{\psi_2})$ on $\mathscr{B}_1$ by

$$p_n^{\psi_2}(\psi_1) = \gamma_n^{-1} \left| \int \psi_1 \circ f^n \psi_2 \, \mathrm{d}\mu - \int \psi_1 \, \mathrm{d}\mu \int \psi_2 \, \mathrm{d}\mu \right|.$$

We have obviously for any integer $n$ and any $\psi_1$, $\psi_1'$ and $\psi_1''$ belonging to $\mathscr{B}_1$:

$$p_n^{\psi_2}(\psi_1' + \psi_1'') \leqslant p_n^{\psi_2}(\psi_1') + p_n^{\psi_2}(\psi_1'') \qquad \text{and} \qquad p_n^{\psi_2}(\psi_1) = p_n^{\psi_2}(-\psi_1).$$

It follows immediately from (39) that for each $\psi_1 \in \mathscr{B}_1$ we have

$$\sup_n p_n^{\psi_2}(\psi_1) \leqslant C_{\psi_1,\psi_2} < \infty.$$

Therefore, we can apply the principle of uniform boundedness [10, theorem 1.29, section III, p 136] to conclude that there is a finite constant $D_{\psi_2}$ such that

$$\sup_{n,\|\psi_1\|_{\mathscr{B}_1}\leqslant 1} p_n^{\psi_2}(\psi_1) \leqslant D_{\psi_2}.$$

In other words, for any integer $n$, for any $\psi_1 \in \mathscr{B}_1$ and any $\psi_2 \in \mathscr{B}_2$ we have

$$\left| \int \psi_1 \circ f^n \psi_2 \, \mathrm{d}\mu - \int \psi_1 \, \mathrm{d}\mu \int \psi_2 \, \mathrm{d}\mu \right| \leqslant D_{\psi_2} \|\psi_1\|_{\mathscr{B}_1} \gamma_n. \tag{41}$$

We shall now control the dependence in $\psi_2$. Let $\Lambda = \mathbb{N} \times B_1$ where $B_1$ is the closed unit ball of $\mathscr{B}_1$. We define a family $(q_\lambda)_{\lambda\in\Lambda}$ of continuous, non-negative functions of $\mathscr{B}_2$ by

$$q_{(n,\psi_1)}(\psi_2) = \gamma_n^{-1} \left| \int \psi_1 \circ f^n \psi_2 \, \mathrm{d}\mu - \int \psi_1 \, \mathrm{d}\mu \int \psi_2 \, \mathrm{d}\mu \right|.$$

We have immediately for any $\lambda \in \Lambda$ and for any $\psi_2$, $\psi_2'$ and $\psi_2''$ in $\mathscr{B}_2$:

$$q_\lambda(\psi_2' + \psi_2'') \leqslant q_\lambda(\psi_2') + q_\lambda(\psi_2'') \qquad \text{and} \qquad q_\lambda(\psi_2) = q_\lambda(-\psi_2).$$

Moreover it follows from (41) that for any $\psi_2 \in \mathscr{B}_2$,

$$\sup_{\lambda\in\Lambda} q_\lambda(\psi_2) \leqslant D_{\psi_2} < \infty.$$

We can apply as above the principle of uniform boundedness to conclude that there is a finite constant $K$ such that

$$\sup_{\lambda\in\Lambda,\|\psi_2\|_{\mathscr{B}_2}\leqslant 1} q_\lambda(\psi_2) \leqslant K,$$

which immediately implies (40). ■

In the case where $\gamma_n$ in (39) is summable and assumptions (i)–(iv) hold, theorem B.1 implies hypothesis 5.1 with $\mathscr{B}_1 = \mathscr{B}_2$ being the space of $\eta$-Hölder continuous functions ($0 < \eta \leqslant 1$).

## Appendix C. Property of the density of the invariant measure for a class of 1D maps

The purpose of this section is to prove that property (29) in theorem 6.1 is indeed valid for maps on the interval satisfying the axioms of [18]. In other words, the density of the absolutely continuous invariant measure belongs to a Besov space (see [17] for definitions). In particular, quadratic maps for a set of parameters of positive Lebesgue measure [18] are included. We refer the reader to [18] (and [5]) for notation and properties of such dynamical systems and their associated tower maps.

Recall that the density $\Phi$ of the SRB measure $\mu$ reads [13, 18] as follows:

$$\Phi(y) = \sum_{j\geqslant 1}\sum_{k=0}^{R_j-1} a_{kj}(y)\chi_{f^k(\Lambda_j)}(y), \tag{42}$$

where we set for any $k < R_j$ and for any $y \in f^k(\Lambda_j)$,

$$a_{kj}(y) = \frac{\varphi(y_{kj})}{f'^k(y_{kj})},$$

where $y_{kj}$ is the unique point in $\Lambda_j$ satisfying $f^k(y_{kj}) = y$, and $\varphi$ is the density of the $f^R$-invariant measure. It is convenient to assume that $a_{kj}$ vanishes outside $f^k(\Lambda_j)$.

We will use repeatedly the following properties from [18].

(i) There exists $\theta > 0$ such that $\sum_j \mathrm{e}^{\theta R_j}|\Lambda_j| < \infty$.

(ii) There are constants $C > 0$ and $\alpha \in ]0, 1[$ such that for all $j$ and all $k < R_j$ and any $y$, $y'$ in $f^k(\Lambda_j)$:

$$\left|\frac{a_{kj}(y)}{a_{kj}(y')} - 1\right| \leqslant C\alpha^{s(y,y')}.$$

We recall that $s(y, y')$ is the separation time of the orbits of $y$ and $y'$, see [18].

(iii) There exists a constant $C > 1$ such that for all $j$ and all $k < R_j$ and any $y$ in $f^k(\Lambda_j)$

$$C^{-1}\,|\Lambda_j| \leqslant a_{kj}(y)|f^k(\Lambda_j)| \leqslant C\,|\Lambda_j|.$$

(iv) Let $B := \|f'\|_\infty > 1$. For all $j$ and all $k < R_j$ and any $y$ in $f^k(\Lambda_j)$

$$a_{kj}(y) \leqslant C B^{R_j-k}\,\frac{|\Lambda_j|}{|\Lambda|}.$$

Property (i) follows from the exponential tail for Markovian return times. Property (ii) follows from the distortion bound in [18]. Property (iii) follows from (ii) and the fact that $f^k|\Lambda_j$ is a diffeomorphism and $\varphi$ is bounded. Finally, property (iv) follows from (iii) and the fact that $f^{R_j-k}(f^k(\Lambda_j)) = \Lambda$ and $f^{R_j-k}|f^k(\Lambda_j)$ is a diffeomorphism.

We will use the following lemma.

**Lemma C.1.** *There exists a constant $C > 0$ such that for any measurable set $A \in \mathbb{R}$ we have*

$$\mu(A) \leqslant C\, m(A)^{\varrho},$$

*where $m$ is Lebesgue measure and $\varrho = \min\{\theta/\log B, 1\} > 0$ ($\theta$ and $B$ are defined in (i) and (iv), respectively).*

**Proof.** We have using Hölder inequality with $p = \log B/(\log B - \min\{\theta, \log B\})$ and $q = p/(p-1) = \varrho^{-1}$,

$$\mu(A) = \sum_{j\geqslant 1}\sum_{k=0}^{R_j-1}\int \mathrm{d}y a_{kj}(y)\chi_A(y)\chi_{f^k(\Lambda_j)}(y) \leqslant m(A)^{1/q}\sum_{j\geqslant 1}\sum_{k=0}^{R_j-1}\left(\int \mathrm{d}y a_{kj}^p(y)\chi_{f^k(\Lambda_j)}(y)\right)^{1/p}.$$

Using (iii), (iv) and (i), this is bounded above by

$$\mathcal{O}(1)\, m(A)^{1/q} \sum_{j\geqslant 1} \sum_{k=0}^{R_j-1} |\Lambda_j|\; B^{(R_j-k)(p-1)/p} \leqslant \mathcal{O}(1) m(A)^{\varrho}.$$

The lemma is proved. ■

The main result of this section is the following theorem.

**Theorem C.1.** *For an interval map satisfying hypotheses of [18], for any positive* $\tau < \min\{\log\alpha^{-1}/(2\log B), \frac{1}{4}(\min\{1, \theta/\log B\})^3\}$, *there exists* $C > 0$ *such that*

$$\int |\Phi(y) - \Phi(y-\delta)|\, \mathrm{d}y \leqslant C|\delta|^{\tau}$$

*for any* $\delta \in \mathbb{R}$. *In other words,* $\Phi$ *belongs to the Besov space* $\Lambda_{\tau}^{1,\infty}$ *(see [17]).*

**Proof.** It is enough to consider $0 < \delta < 1$. We have

$$\int |\Phi(y) - \Phi(y-\delta)|\, \mathrm{d}y \leqslant \sum_{j\geqslant 1} \sum_{k=0}^{R_j-1} \int \mathrm{d}y |a_{kj}(y)\chi_{f^k(\Lambda_j)}(y) - a_{kj}(y-\delta)\chi_{f^k(\Lambda_j)}(y-\delta)|. \tag{43}$$

For a fixed $\delta > 0$, we split the sum over $j$ and $k$ in (43) according to the condition $\delta > |f^k(\Lambda_j)|/2$ and the complementary condition. The first sum is bounded above by

$$\begin{aligned}
&2 \sum_{j\geqslant 1} \sum_{\substack{k=0 \\ |f^k(\Lambda_j)|<2\delta}}^{R_j-1} \sup_{y\in f^k(\Lambda_j)} a_{kj}(y)\, |f^k(\Lambda_j)| \\
&\quad = 2 \sum_{j\geqslant 1} \sum_{\substack{k=0 \\ |f^k(\Lambda_j)|<2\delta}}^{R_j-1} \left( \sup_{y\in f^k(\Lambda_j)} a_{kj}(y)\, |f^k(\Lambda_j)| \right)^{1-\tau} \left( \sup_{y\in f^k(\Lambda_j)} a_{kj}(y) \right)^{\tau} |f^k(\Lambda_j)|^{\tau} \\
&\quad \leqslant \mathcal{O}(1)\, \delta^{\tau} \sum_{j\geqslant 1} \sum_{k=0}^{R_j-1} |\Lambda_j|\; B^{\tau(R_j-k)} \leqslant \mathcal{O}(1)\, \delta^{\tau},
\end{aligned} \tag{44}$$

where this last inequality follows from (i), (iii) and (iv).

Now we turn to the second sum, namely the sum over the indices $j$, $k$ satisfying $|f^k(\Lambda_j)| \geqslant 2\delta$. This sum is bounded above by

$$\begin{aligned}
&\sum_{j\geqslant 1} \sum_{\substack{k=0 \\ |f^k(\Lambda_j)|\geqslant 2\delta}}^{R_j-1} \int \mathrm{d}y a_{kj}(y) |\chi_{f^k(\Lambda_j)}(y) - \chi_{f^k(\Lambda_j)}(y-\delta)| \\
&\quad + \sum_{j\geqslant 1} \sum_{\substack{k=0 \\ |f^k(\Lambda_j)|\geqslant 2\delta}}^{R_j-1} \int \mathrm{d}y |a_{kj}(y) - a_{kj}(y-\delta)| \chi_{f^k(\Lambda_j)}(y-\delta).
\end{aligned}$$

Since $f^k(\Lambda_j)$ is an interval and $|f^k(\Lambda_j)| \geqslant 2\delta$, we have

$$\int \mathrm{d}y a_{kj}(y)|\chi_{f^k(\Lambda_j)}(y) - \chi_{f^k(\Lambda_j)}(y-\delta)|$$
$$\leqslant \mathcal{O}(1) \sup_{y\in f^k(\Lambda_j)} a_{kj}(y) \int \mathrm{d}y|\chi_{f^k(\Lambda_j)}(y) - \chi_{f^k(\Lambda_j)}(y-\delta)|$$
$$\leqslant \mathcal{O}(1) \sup_{y\in f^k(\Lambda_j)} a_{kj}(y)\delta. \tag{45}$$

On the other hand, using (iii), the same integral is bounded above by $2\sup_{y\in f^k(\Lambda_j)} a_{kj}(y)$ $|f^k(\Lambda_j)| \leqslant \mathcal{O}(1)\ |\Lambda_j|$. Proceeding as in (44), we obtain for the sum over $j$ and $k$ the upper bound $\mathcal{O}(1)\ \delta^\tau$. We now estimate for each $j$ and $k$ the integral

$$\int \mathrm{d}y|a_{kj}(y) - a_{kj}(y-\delta)|\chi_{f^k(\Lambda_j)}(y-\delta)$$
$$= \int_{(f^k(\Lambda_j))^c} \mathrm{d}y|a_{kj}(y) - a_{kj}(y-\delta)|\chi_{f^k(\Lambda_j)}(y-\delta) \tag{46}$$
$$+ \int_{f^k(\Lambda_j)} \mathrm{d}y|a_{kj}(y) - a_{kj}(y-\delta)|\chi_{f^k(\Lambda_j)}(y-\delta). \tag{47}$$

It is easy to verify that the integral (46) can be bounded above like the integral (45). For the integral (47) we have the obvious upper bound

$$2 \sup_{y\in f^k(\Lambda_j)} a_{kj}(y)\ |f^k(\Lambda_j)| \leqslant \mathcal{O}(1)\ |\Lambda_j|. \tag{48}$$

Using (ii), this integral is also bounded above by

$$\mathcal{O}(1) \int_{f^k(\Lambda_j)} \mathrm{d}y a_{kj}(y)\alpha^{s(y,y-\delta)} \chi_{f^k(\Lambda_j)}(y-\delta),$$

where $s(y, y-\delta)$ is the separation time of the orbits of $y$ and $y-\delta$. In order to estimate this integral we introduce a partition of $f^k(\Lambda_j)$ into four subsets defined by

$$\boldsymbol{B}^1_{kj} = \left\{ y \in f^k(\Lambda_j) | s(y, y-\delta) > \frac{\tau \log\delta}{\log\alpha} \right\}$$
$$\boldsymbol{B}^2_{kj} = \{y \in f^k(\Lambda_j) \cap (\boldsymbol{B}^1_{kj})^c | R(f^{s(y,y-\delta)}(y)) > \sigma\ \log\delta^{-1}\},$$

where $R(\cdot)$ is the Markovian return-time function defined in [18], and $\sigma := 1/4 \log B$.

$$\boldsymbol{B}^3_{kj} = \{y \in f^k(\Lambda_j) \cap (\boldsymbol{B}^1_{kj})^c \cap \boldsymbol{B}^2_{kj} |\ |f^{s(y,y-\delta)}(y) - f^{s(y,y-\delta)}(y-\delta)| < \sqrt{\delta}\}$$
$$\boldsymbol{B}^4_{kj} = \{y \in f^k(\Lambda_j) \cap (\boldsymbol{B}^1_{kj})^c \cap \boldsymbol{B}^2_{kj} |\ |f^{s(y,y-\delta)}(y) - f^{s(y,y-\delta)}(y-\delta)| \geqslant \sqrt{\delta}\}.$$

We will estimate the contribution of these four sequences of sets separately. We have obviously

$$\mathcal{O}(1) \int_{\boldsymbol{B}^1_{kj}} a_{kj}(y)\,\mathrm{d}y\alpha^{s(y,y-\delta)} \leqslant \delta^\tau\ |f^k(\Lambda_j)| \sup_{y\in f^k(\Lambda_j)} a_{kj}(y)$$

and therefore we can bound the sum over $k$ and $j$ using (iii) and (i).

To estimate the contribution of $\boldsymbol{B}^2_{kj}$ we introduce the set

$$\mathcal{C} := \bigcup_{\ell=0}^{\lfloor \tau\log\delta/\log\alpha \rfloor} \{y : f^\ell(y) \in \{R > \sigma\ \log\delta^{-1}\}\}.$$

From the invariance of the SRB measure $\mu$ and lemma C.1 we have

$$\mu(\mathcal{C}) \leqslant \left\lfloor \frac{\tau\log\delta}{\log\alpha} \right\rfloor \mu\{R > \sigma\log\delta^{-1}\} \leqslant C\frac{\tau\log\delta}{\log\alpha} m(R > \sigma\log\delta^{-1})^\varrho.$$

Now observe that

$$\boldsymbol{B}^2_{kj} \subset \mathcal{C} \cap f^k(\Lambda_j),$$

which implies using (i) and Chebychev inequality that

$$\begin{aligned}\int_{\boldsymbol{B}^2_{kj}} a_{kj}(y)\,\mathrm{d}y &\leqslant \sup_{y\in f^k(\Lambda_j)} a_{kj}(y)\mu(\mathcal{C})\\ &\leqslant \mathcal{O}(1)\sup_{y\in f^k(\Lambda_j)} a_{kj}(y)\frac{\tau\log\delta}{\log\alpha} m(R>\sigma\log\delta^{-1})^{\varrho}\\ &\leqslant \mathcal{O}(1)\log(\delta^{-1})\sup_{y\in f^k(\Lambda_j)} a_{kj}(y)\delta^{\varrho\sigma\theta}.\end{aligned}$$

Using (iv) and interpolating with the bound (48), we get

$$\sum_{j\geqslant 1}\sum_{k=0}^{R_j-1}\int_{\boldsymbol{B}^2_{kj}} a_{kj}(y)\,\mathrm{d}y \leqslant \mathcal{O}(1)\log(\delta^{-1})\delta^{\varrho^2\sigma\theta} = \mathcal{O}(1)\log(\delta^{-1})\delta^{\varrho^3/4}.$$

We now treat the integral over the set $\boldsymbol{B}^3_{kj}$. We define the sets

$$\mathcal{D}_0 = \left\{ y : y\in\Lambda, \mathrm{d}\left( y, \bigcup_{j, R_j<\sigma\log\delta^{-1}} \partial\Lambda_j \right) < \sqrt{\delta} \right\}$$

and

$$\mathcal{D} = \bigcup_{\ell=0}^{\lfloor\sigma\log\delta^{-1}\rfloor} f^{-\ell}(\mathcal{D}_0).$$

From the invariance of the SRB measure $\mu$ we get

$$\mu(\mathcal{D}) \leqslant \sigma\log\delta^{-1}\,\mu(\mathcal{D}_0).$$

We now estimate

$$\mu(\mathcal{D}_0) = \sum_{j\geqslant 1}\sum_{k=0}^{R_j-1}\int_{f^k(\Lambda_j)\cap\mathcal{D}_0} a_{kj}(y)\,\mathrm{d}y.$$

As we have done several times above, each integral in these sums has two bounds. From the definition of $\mathcal{D}_0$ we have

$$\int_{f^k(\Lambda_j)\cap\mathcal{D}_0} a_{kj}(y)\,\mathrm{d}y \leqslant 2\sqrt{\delta}\,\#\{j : R_j<\sigma\log\delta^{-1}\}\sup_{y\in f^k(\Lambda_j)} a_{kj}(y).$$

Since $f^{R_j}(\Lambda_j)=\Lambda$ for all $j\geqslant 1$, we have for any $j\geqslant 1$ that

$$B^{R_j}|\Lambda_j| \geqslant |\Lambda|.$$

Therefore since the $\Lambda_j$ are disjoint and their union is $\Lambda$ we obtain for any integer $q\geqslant 1$

$$\#\{j : R_j\leqslant q\} \leqslant B^q.$$

It follows using (iv) that

$$\int_{f^k(\Lambda_j)\cap\mathcal{D}_0} a_{kj}(y)\,\mathrm{d}y \leqslant \mathcal{O}(1)\delta^{1/4}B^{R_j-k}|\Lambda_j|.$$

Therefore interpolating with the trivial bound (48) as before, one gets

$$\mu(\mathcal{D}_0) \leqslant \mathcal{O}(1)\delta^{\theta/(4\log B)}.$$

We now observe that

$$\boldsymbol{B}^3_{kj} \subset \mathcal{D}\cap f^k(\Lambda_j),$$

which implies, using (ii), that

$$\int_{\boldsymbol{B}^3_{kj}} a_{kj}(y)\,\mathrm{d}y \leqslant \sup_{y\in f^k(\Lambda_j)} a_{kj}(y)\mu(\mathcal{D}) \leqslant \mathcal{O}(1) \sup_{y\in f^k(\Lambda_j)} a_{kj}(y)\ \log(\delta^{-1})\delta^{\theta/4\log B}.$$

Using (i) and (iv) and interpolating with the bound (48), we get

$$\sum_{j\geqslant 1}\sum_{k=0}^{R_j-1}\int_{\boldsymbol{B}^3_{kj}} a_{kj}(y)\,\mathrm{d}y \leqslant \mathcal{O}(1)(\log(\delta^{-1}))^{\theta/\log B}\delta^{\theta^2/4(\log B)^2}.$$

Finally, if $y\in\boldsymbol{B}^4_{kj}$ we have using $B=\|f'\|_\infty$

$$\sqrt{\delta}\leqslant|f^{s(y,y-\delta)}(y)-f^{s(y,y-\delta)}(y-\delta)|\leqslant\delta\ B^{s(y,y-\delta)}.$$

This immediately implies that

$$s(y,y-\delta)\geqslant-\frac{\log\delta}{2\log B}.$$

Using this bound, properties (iii) and (i), we obtain

$$\sum_{j\geqslant 1}\sum_{k=0}^{R_j-1}\int_{\boldsymbol{B}^4_{kj}} a_{kj}(y)\alpha^{s(y,y-\delta)}\,\mathrm{d}y \leqslant \mathcal{O}(1)\delta^{\log\alpha^{-1}/(2\log B)}\sum_{j\geqslant 1}\sum_{k=0}^{R_j-1}|\Lambda_j| \leqslant \mathcal{O}(1)\delta^{\log\alpha^{-1}/(2\log B)}.$$

This ends the proof of the theorem. ■

## References


[1] Berkes I 1998 Results and problems related to the pointwise central limit theorem *Asymptotic Methods in Probability and Statistics (Ottawa, ON, 1997)* ed B Szyszkowicz (Amsterdam: North-Holland) pp 59–96

[2] Borovkova S, Burton R and Dehling H 2001 Limit theorems for functionals of mixing processes with applications to $U$-statistics and dimension estimation *Trans. Am. Math. Soc.* **353** 4261–318

[3] Brockwell P J and Davis R A 1991 *Time Series: Theory and Methods* 2nd edn (Berlin: Springer)

[4] Chazottes J-R and Collet P 2005 Almost sure central limit theorems and Erdos-Rényi type law for expanding maps of the interval *Ergod. Theory Dynam. Syst.* **25** 419–41

[5] Chazottes J-R, Collet P and Schmitt B 2005 Devroye inequality for a class of non-uniformly hyperbolic dynamical systems *Nonlinearity* **18** 2323–40

[6] Collet P, Martínez S and Schmitt B 2002 Exponential inequalities for dynamical measures of expanding maps of the interval *Probab. Theory Relat. Fields* **123** 301–22

[7] Dall'Aglio G 1956 Sugli estremi dei momenti delle funzioni di ripartizione doppia *Ann. Scuola Norm. Sup. Pisa* **10** 35–74

[8] Devroye L 1991 Exponential inequalities in nonparametric estimation *Nonparametric Functional Estimation and Related Topics (Spetses, 1990)* (*NATO Advanced Science Institute Series C, Mathematical Physical Sciences* vol 335) (Dordrecht: Kluwer) pp 31–44

[9] Dudley R M 2002 Real analysis and probability *Cambridge Studies in Advanced Mathematics* vol 74 (Cambridge: Cambridge University Press)

[10] Kato T 1966 *Perturbation Theory for Linear Operators* (New York: Springer)

[11] Kawaguchi A and Yanagawa T 2001 Estimating correlation dimension in chaotic time series *Bull. Inform. Cybernet.* **33** 63–71

[12] Lopes A O and Lopes S R C 2002 Convergence in distribution of the periodogram of chaotic processes *Stochastics Dyn.* **2** 609–24

[13] Luzzatto S 2004 Stochastic-like behaviour in nonuniformly expanding maps *Preprint* math.DS/0409085

[14] Manning A and Simon K 1998 A short existence proof for correlation dimension *J. Stat. Phys.* **90** 1047–9

[15] Prieur C 2001 Density estimation for one-dimensional dynamical systems *ESAIM Probab. Stat.* **5** 51–76

[16] Rachev S T 1991 *Probability Metrics and Stability of Stochastic Processes* (New York: Wiley)

[17] Stein E M 1970 *Singular Integrals and Differentiability Properties of Functions* (*Princeton Mathematical Series* vol 30) (Princeton, NJ: Princeton University Press)

[18] Young L-S 1998 Statistical properties of dynamical systems with some hyperbolicity *Ann. Math.* **147** 585–650