Optimal speed nonparametric density estimation for one-dimensional dynamical systems

Salim Lardjane

Laboratoire de Statistique et Modélisation, CREST-ENSAI
Campus de Ker-Lann, 35170 Bruz, France
&
Laboratoire SABRES, Université de Bretagne-Sud
Rue Yves Mainguy, Campus de Vannes-Tohannic,
56000 Vannes, France

e-mail : lardjane@ensai.fr

Résumé

L'auteur traite de l'estimation non-paramétrique de la densité invariante pour une famille générale de systèmes dynamiques en dimension un. Il démontre diverses inégalités de mélange et montre que le problème se ramène à celui de l'estimation de la densité marginale d'un processus stochastique stationnaire. Il utilise ce résultat pour démontrer que l'estimateur de Parzen-Rosenblatt de la densité marginale d'un processus dynamique peut converger en moyenne quadratique avec la vitesse optimale du cas i.i.d. Il applique ce résultat à l'estimation de la densité invariante pour des systèmes dynamiques ergodiques et donne une interprétation de l'erreur quadratique moyenne de l'estimateur de Parzen-Rosenblatt dans ce contexte.

Abstract

The author deals with nonparametric invariant density estimation for a general class of one-dimensional dynamical systems. He establishes various mixing inequalities and shows that the problem is equivalent to the estimation of the marginal density of a stationary dynamic process. He uses this equivalence to prove that the Parzen-Rosenblatt estimator of the marginal density of a dynamic process can converge in quadratic mean with the optimal rate of the i.i.d. case. He applies this result to invariant density estimation for ergodic dynamical systems and gives an interpretation of the mean quadratic error of the Parzen-Rosenblatt estimator in this context.

MOTS-CLÉS : Estimation non-paramétrique de densité, Théorie ergodique, Systèmes dynamiques, transformations r-adiques, Processus dynamiques, Chaos.

KEYWORDS : Nonparametric density estimation, Ergodic theory, Dynamical systems, r-adic transformations, Dynamic processes, Chaos.

1 Introduction

Dynamical systems are mathematical objects used to modelize complex deterministic dynamics. As such, they are encountered in physics, economy, biology and meteorology. Formally, a dynamical system is a quadruple $(E, \mathcal{B}, \mu, S)$ where $(E, \mathcal{B}, \mu)$ is a probability space and S is a transformation of E which leaves the probability measure μ invariant, that is to say, such that for all $B \in \mathcal{B}$, $\mu(S^{-1}B) = \mu(B)$. The space E is called the *state* or *phase space* of the system. To each point $x \in E$, one can associate a *trajectory* $(x_t)_{t\in\mathbb{N}}$ by setting $x_0 = x$ and $x_{t+1} = S(x_t)$ for all $t \in \mathbb{N}$. x_0 is termed the *initial condition* or *initial state* of the dynamical system. If $E \subset \mathbb{R}$, the dynamical system $(E, \mathcal{B}, \mu, S)$ is termed *one-dimensional.*

In the 1970's, scientists with different backgrounds recognized the fact that dynamical systems, although showing simple deterministic structures (the map S), could have *chaotic*, *i.e.* random-looking, trajectories [5, 6, 7, 10]. In fact, nonlinearities in S can produce extreme sensitivity with respect to the initial state of the dynamical system, so that the slightest error on x_0 is exponentially amplified as time goes on. As this instability makes predictions impossible, it is necessary, in order to quantify appropriately the behaviour of a chaotic dynamical system, to resort to the *ergodicity property* : the dynamical system $(E, \mathcal{B}, \mu, S)$ and the probability measure μ are termed *ergodic* if every set $B \in \mathcal{B}$ such that $\mu(B \triangle S^{-1}B) = 0$ has μ - measure 0 or 1. Birkhoff's Ergodic Theorem [1] states that if $(E, \mathcal{B}, \mu, S)$ is ergodic, then for all $f \in L^1(E, B, \mu)$ and μ-almost all $x_0 \in E$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} f \circ S^t(x_0) = \int_E f d\mu.$$

An easy consequence of Birkhoff's ergodic theorem is that, given a μ - measurable region $A \subset E$, the system has asymptotically the same probability $\mu(A)$ of showing a state belonging to A, for μ-almost all initial condition x_0. The trajectories of the system asymptotically "fill" the state space, weighting various regions according to the frequency at which they cross them. Ergodicity means that μ-almost all trajectories asymptotically "fill" or "weight" the state space in the same way, namely, according to μ.

The ergodicity property is of particular importance when the probability measure μ is absolutely continuous with respect to some Lebesgue measure. The asymptotic description provided by μ is then valid for a set of initial conditions with positive Lebesgue measure, that is to say, of physical significance. In this case, the problem of computing an approximation of the density of μ from a sequence of observations is of practical relevance. This approximation is still usually obtained by computing a histogram, but in recent years, there has been a growing research around the use of the Parzen-Rosenblatt estimator for this purpose. An essential difficulty is to go beyond the mere empirical point of view, which means simply plugging our data in the expression of a Parzen-Rosenblatt density estimator without any theoretical result to back its use. A first step in this direction is taken by introducing dynamic stochastic processes (see Bosq [2]). A stochastic process $(X_t)_{t\in\mathbb{N}}$ with base space $(\Omega, \mathcal{A}, \mathbb{P})$ and values in $(E, \mathcal{B})$ is termed a *dynamic process* if there exists a surjective measurable map S on $(E, \mathcal{B})$ such that for all $t \in \mathbb{N}$, $X_{t+1} = S(X_t)$. The map S is termed *the dynamic transformation* of $(X_t)_{t\in\mathbb{N}}$.

The link between dynamic processes and dynamical systems stems from the fact that each trajectory $(X_t(\omega))_{t\in\mathbb{N}}$ of the dynamic process $(X_t)_{t\in\mathbb{N}}$ with dynamic transformation S is such that $X_{t+1}(\omega) = S(X_t(\omega))$ for all $t \in \mathbb{N}$. In other words, each trajectory of the stochastic process $(X_t)_{t\in\mathbb{N}}$ is a trajectory of the dynamical system $(E, \mathcal{B}, \mu, S)$ where μ is the marginal distribution of $(X_t)_{t\in\mathbb{N}}$ (μ is S-invariant by construction).

Now, assume that we have at our disposal n observations $x_1, \ldots, x_n$ from a trajectory of a stochastic dynamic process $(X_t)_{t\in\mathbb{N}}$ with marginal density f and that we use a Parzen-Rosenblatt estimator to estimate $f(x)$. Bosq [2, 3] obtained conditions under which this estimator is consistent in quadratic mean, with a convergence rate of $O((\log n/n)^{4/5})$.

In this paper, we show that it is possible, for a particular class of dynamic processes, to reach the optimal rate of the i.i.d. case, namely $O(n^{-4/5})$, and we apply this result to invariant density estimation for a general class of ergodic dynamical systems.

2 Preliminary results

Let us consider the class $\mathcal{S}$ of transformations

$$\begin{array}{rcccc} S & : & [0,1] & \to & [0,1] \\ & & x & \mapsto & r\,x \mod 1 \end{array}$$

with $r \in \mathbb{N}^*\backslash\{1\}$. $\mathcal{S}$ is the class of *r-adic transformations* of the unit interval. Let us define the intervals $I_1 = [0, \frac{1}{r}]$, $I_r = [\frac{r-1}{r}, 1]$ and $I_k = [\frac{k-1}{r}, \frac{k}{r}[$ for $k = 2, \ldots, r-1$. On each interval I_k, S is given by $S(x) = rx - k + 1$. Let us denote by $\mathcal{B}([0,1])$ the borelian σ-algebra on $[0,1]$. A straightforward calculation shows that the Lebesgue measure λ on $([0,1], \mathcal{B}([0,1]))$ is invariant for any $S \in \mathcal{S}$ and hence that $([0,1], \mathcal{B}([0,1]), \lambda, S)$ is a dynamical system for any $S \in \mathcal{S}$.

Now, denote by $\mathcal{C}$ the class of transformations T of an interval $[a,b] \subset \overline{\mathbb{R}}$ for which there exists a transformation $S \in \mathcal{S}$ and an increasing differentiable diffeomorphism $\phi : (a, b) \to (0, 1)$, extended by continuity on $[a, b]$, such that $S \circ \phi = \phi \circ T$. Then, $T = \phi^{-1} \circ S \circ \phi$, S and T are conjugate, and for every transformation $T \in \mathcal{C}$, S is necessarily unique, since any two distinct r-adic transformations are not conjugate. Aditionally, let us note that the probability measure λ_ϕ associated to ϕ is the unique probability measure with support $[a, b]$ which is invariant by T and such that the dynamical system $([a, b], \mathcal{B}([a, b]), \mu_\phi, T)$ is ergodic. To see this, recall that since T and S are conjugate, $\lambda\phi = \lambda_\phi$ is invariant for T. Now, since ϕ is one-to-one, $\phi(A \triangle B) = \phi(A) \triangle \phi(B)$ for any measurable A, B, and this gives for all measurable B, $\lambda\phi(B \triangle T^{-1}B) = \lambda(\phi(B) \triangle \phi \circ T^{-1}(B))$. Using the ergodicity of λ for S, this quantity vanishes iff $\lambda(\phi(B)) \in \{0, 1\}$, which means that λ_ϕ is ergodic for T. Since λ_ϕ is ergodic with support $[a, b]$, it is necessarily unique. In particular, we obtain that if $T \in \mathcal{C}$, with $T = \phi^{-1} \circ S \circ \phi$ and $S \in \mathcal{S}$, then ϕ is necessarily unique.

A well-known example of a chaotic tranformation belonging to the class $\mathcal{C}$ is provided by the *logistic* map on $[0, 1]$, which is given by $T(x) = 4x(1 - x)$. One has $T = \phi^{-1} \circ S \circ \phi$, where S is the 2 - adic map on $[0, 1]$ and

$$\phi(x) = \int_0^x \frac{1}{\pi\sqrt{u(1-u)}}\, du = \frac{1}{2}(1 + \frac{2}{\pi} \arcsin(2x - 1)).$$

ϕ is the distribution function of the $Beta(\frac{1}{2}, \frac{1}{2})$ distribution.

From a practical point of view, transformations belonging to $\mathcal{C}$ can be iterated to simulate samples from given probability distributions and, as such, they are encountered in some non standard Monte-Carlo simulation procedures. This is based on the fact that *any* probability density function f on an interval $[a, b] \subset \overline{\mathbb{R}}$, which is strictly positive on $]a, b[$, is the unique ergodic invariant density of an infinity of transformations $T \in \mathcal{C}$, $T : [a, b] \to [a, b]$ [4].

3 Some inequalities

In this section, we derive some technical inequalities about the speed at which the space is "mixed" by elements of $\mathcal{S}$ and $\mathcal{C}$. We first prove the following lemma.

Lemma 3.1 *For each transformation $S \in \mathcal{S}$, there exists $\rho \in]0, 1[$ such that, for every closed interval $B \subset [0, 1]$ and every $k \in \mathbb{N}$, we have*

$$|\lambda(B \cap S^{-k}B) - \lambda(B)^2| \leq 4\ \lambda(B)\ \rho^k.$$

Proof. Consider a transformation $S \in \mathcal{S}$ and let $\alpha = \{A_1, \ldots, A_r\}$ where $A_k = \mathrm{Int}(I_k)$ for $k = 1, \ldots r$. α is a *measurable partition* of $[0,1]$, that is $\lambda([0,1] \setminus \bigcup_k A_k) = 0$ and $\lambda(A_i \cap A_j) = 0$ for $i \neq j$. Moreover, one can prove by recurrence that α is *independent*, meaning that for all $n \geq 1$ and for all $j_1, \ldots, j_n \in \{1, \ldots, r\}$, one has $\lambda(S^{-n+1} A_{j_n} \cap \cdots \cap A_{j_1}) = \prod_{k=1}^n \lambda(A_{i_k}) = r^{-n}$. The proof is based on the fact that $\lambda(S^{-1} A \cap A_i) = r^{-1} \lambda(A)$ for any measurable A and any $i \in \{1, \ldots, r\}$, which stems from the very definition of S.

Let ζ and η be two measurable partitions of $[0,1]$. Their *joint* $\zeta \vee \eta$ is defined to be the measurable partition $\zeta \vee \eta = \{M \cap N : M \in \zeta, N \in \eta\}$.

Set $\alpha_0^k = \bigvee_{i=0}^k S^{-i}\alpha$. Then, α_0^k is the collection of the interiors of the intervals on which the r^k bijective branches of S^k are defined. Thus, $\alpha_0^k = \big\{\big(ir^{-k}, (i+1)r^{-k}\big) : i = 0, \ldots, r^k - 1\big\}$. Now, let B be a closed interval in $[0,1]$. Then,

$$\begin{aligned}
\lambda(B \cap S^{-k}B) - \lambda(B)^2 &= \lambda\Big(\bigcup_{A \in \alpha_0^k} B \cap S^{-k}B \cap A\Big) - \lambda(B)^2 \\
&= \sum_{A \in \alpha_0^k} \lambda\left(B \cap S^{-k}B \cap A\right) - \lambda(B)^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\lambda(B \cap S^{-k}B) - \lambda(B)^2 \\
&\quad = \sum_{\substack{A \in \alpha_0^k \\ A \subset B}} \lambda(S^{-k}B \cap A) + \sum_{\substack{A \in \alpha_0^k \\ A \cap B \neq \emptyset, A \cap B \neq A}} \lambda(B \cap S^{-k}B \cap A) \\
&\qquad + \sum_{\substack{A \in \alpha_0^k \\ A \cap B = \emptyset}} \lambda(B \cap S^{-k}B \cap A) - \lambda(B)^2 \\
&\quad = \sum_{\substack{A \in \alpha_0^k \\ A \subset B}} \frac{\lambda(B)}{r^k} + \sum_{\substack{A \in \alpha_0^k \\ A \cap B \neq \emptyset, A \cap B \neq A}} \lambda(B \cap S^{-k}B \cap A) - \lambda(B)^2.
\end{aligned}$$

But, since B is an interval, one can find at most *two* sets $A_{1,2} \in \alpha_0^k$ such that $A_{1,2} \cap B \neq \emptyset$ and $A_{1,2} \cap B \neq A_{1,2}$, each of them covering an extremity of B. Hence,

$$|\lambda(B \cap S^{-k}B) - \lambda(B)^2| \leq \Big|\frac{\lambda(B)}{r^k} \sum_{\substack{A \in \alpha_0^k \\ A \subset B}} 1 - \lambda(B)^2\Big| + 2\frac{\lambda(B)}{r^k}.$$

Since $\lambda(B) \geq \sum_{A \in \alpha_0^k, A \subset B} r^{-k} = \sum_{A \in \alpha_0^k, A \subset B} \lambda(A)$, the above sum is less than $4\lambda(B) r^{-k}$, so we can set $\rho = r^{-1}$ to obtain the result. ■

Using lemma 3.1, we can prove the following corollary.

Corollary 3.1 *Let $T \in \mathcal{C}$, $T : [a,b] \to [a,b]$, such that $T = \phi^{-1} \circ S \circ \phi$, $S \in \mathcal{S}$. Then, there exists $\rho \in]0,1[$ such that, for every closed interval $B \subset [a,b]$ and every $k \in \mathbb{N}$, we have*

$$|\mu(B \cap T^{-k}B) - \mu(B)^2| \leq 4\ \mu(B)\ \rho^k$$

where $\mu = \lambda\phi = \mu_\phi$ is the probability measure associated to the distribution function ϕ.

Proof. First note, that according to our preliminary results, the measure μ_ϕ is T-invariant and ergodic. Now, consider a closed interval $B \subset [a,b]$. Since ϕ is bijective, we have, for every $k \in \mathbb{N}$

$$\begin{aligned} |\mu(B \cap T^{-k}B) - \mu(B)^2| &= |\lambda(\phi B \cap \phi T^{-k}\phi^{-1}\phi B) - \lambda(\phi B)^2| \\ &= |\lambda(\phi B \cap S^{-k}\phi B) - \lambda(\phi B)^2|. \end{aligned}$$

The continuity of ϕ implies that the image under ϕ of every closed interval is a closed interval. Hence, according to lemma 3.1, we obtain

$$|\mu(B \cap T^{-k}B) - \mu(B)^2| \leq 4\ \lambda(\phi B)\ r^{-k} = 4\ \mu(B)\ \rho^k,$$

where $0 < \rho = r^{-1} < 1$. ■

We are now able to derive a bound on the rate at which the space is "mixed" by transformations belonging to $\mathcal{C}$.

Theorem 3.1 *Let $T \in \mathcal{C}$, $T : [a,b] \to [a,b]$, such that $T = \phi^{-1} \circ S \circ \phi$, $S \in \mathcal{S}$. Then, for every closed interval $B \subset [a,b]$ and every $n \in \mathbb{N}^*$, we have*

$$\frac{1}{\binom{n}{2}} \sum_{\substack{0 \leq j,k \leq n-1 \\ j \neq k}} |\mu(T^{-j}B \cap T^{-k}B) - \mu(B)^2| \leq \frac{16\ \mu(B)}{n}$$

where $\mu = \lambda\phi = \mu_\phi$ is the probability measure associated to the distribution function ϕ.

Proof. Since μ is invariant by S, we have

$$\begin{aligned} &\frac{1}{\binom{n}{2}} \sum_{\substack{0 \leq j,k \leq n-1 \\ j \neq k}} |\mu(S^{-j}B \cap S^{-k}B) - \mu(B)^2| \\ &= \frac{4}{n-1} \sum_{1 \leq k \leq n-1} (1 - \frac{k}{n})\ |\mu(B \cap S^{-k}B) - \mu(B)^2| \\ &\leq \frac{16\ \mu(B)}{n-1} \sum_{1 \leq k \leq n-1} (1 - \frac{k}{n}) \cdot r^{-k} \\ &\leq \frac{16\ \mu(B)}{n} \cdot \frac{1}{r-1} \cdot \frac{(1-r^{-1})(n-1) - r^{-1} + r^{-n}}{(1-r^{-1})(n-1)} \leq \frac{16\ \mu(B)}{n}. \ ■ \end{aligned}$$

Now, consider $T \in \mathcal{C}$, $T : [a,b] \to [a,b]$, such that $T = \phi^{-1} \circ S \circ \phi$, $S \in \mathcal{S}$. Define $\mu = \lambda\phi$, and let $(X_t)_{t\in\mathbb{N}}$ denote the sequence of transformations $(T^t)_{t\in\mathbb{N}}$, where $X_0 = T^0$ denotes the identity map on $[a,b]$.

Corollary 3.1 *$(X_t)_{t\in\mathbb{N}}$ is a stationary dynamic process with dynamic transformation T and marginal probability measure μ, and is such that, for every closed interval $B \subset [a,b]$ and every $n \in \mathbb{N}^*$, we have*

$$\frac{1}{\binom{n}{2}} \sum_{\substack{0\le j,k\le n-1 \\ j\neq k}} |\mu_{j,k}(B\times B) - \mu(B)^2| \le \frac{16\ \mu(B)}{n}$$

where $\mu_{j,k}$ is the joint probability measure of (X_j, X_k) for all $j,k \in \mathbb{N}$ such that $j \neq k$.

Proof. The identity map of $[0,1]$ and the maps $\{T^t : t \in \mathbb{N}^*\}$ are all $\mathcal{B}([a,b])/\mathcal{B}([a,b])$ - measurable by construction. Thus, the sequence $(X_t)_{t\in\mathbb{N}}$ is a stochastic process with base space $([a,b], \mathcal{B}([a,b]), \mu)$ and is obviously a dynamic process, with dynamic transformation T. Moreover, for all $s \in \mathbb{N}^*$, $k \in \mathbb{N}$, $t_1, \ldots, t_s \in \mathbb{N}$ such that $t_1 < t_2 < \cdots < t_s$, and all borelian sets $B_1, \ldots, B_s \in \mathcal{B}([a,b])$, one has

$$\begin{aligned}\mu(T^{-t_1-k}B_1 \cap \cdots \cap T^{-t_s-k}B_s) &= \mu(T^{-k}(T^{-t_1}B_1 \cap T^{-t_2}B_2 \cap \cdots \cap T^{-t_s}B_k)) \\ &= \mu(T^{-t_1}B_1 \cap T^{-t_2}B_2 \cap \cdots \cap T^{-t_s}B_k).\end{aligned}$$

Hence, the process $(X_t)_{t\in\mathbb{N}}$ is stationary. Its marginal distribution is μ since for all $B \in \mathcal{B}([0,1])$, $\mu(X_0^{-1}B) = \mu(B)$. Now, for all $j,k \in \mathbb{N}$ such that $j \neq k$ on has $\mu_{j,k}(B\times B) = \mu(X_j^{-1}B \cap X_k^{-1}B) = \mu(T^{-j}B \cap T^{-k}B)$. The result then stems from theorem 3.1. ■

Let us emphasize the specificity of the results obtained in this section. By restricting B to be a *closed interval*, we have managed to obtain a bound on $|\lambda(B\cap S^{-k}B) - \lambda(B)^2|$ which is stronger than the bounds which are derived for *all* measurable sets B in ergodic theory. These bounds, which are obtained either by spectral or symbolic representation methods, have the form $K \cdot \rho^k$, where $\rho \in (0,1)$ and K is a positive constant. Accordingly, they are not sufficiently sharp to be used in the context of nonparametric statistics, as it will appear in the sequel (see the bound on the term $C_{nt}(x)$ in the proof of theorem 4.1).

Note that the bound obtained in corollary 3.1 obviously holds for ψ, ϕ^* and ϕ - mixing stochastic processes. However, this could not have led us to the result since dynamic processes are not even strongly mixing (α - mixing). To see this, denote by $(X_t)_{t\in\mathbb{N}}$ a dynamic process with dynamic transformation F. Then, for every $s, t \in \mathbb{N}$ such that $s < t$, on has $X_t = F^{t-s}X_s$ and hence, $\sigma(X_t) \subset \sigma(X_s)$. Therefore,

$$\alpha(k) \geq \alpha^{(2)}(k) := \sup_{t\in\mathbb{N}} \sup_{\substack{A\in\sigma(X_t)\\ B\in\sigma(X_{t+k})}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| = 1/4$$

for every $k \in \mathbb{N}$, where $\alpha(k)$ denotes the strong mixing coefficient of order k, and $\alpha^{(2)}(k)$ the 2-α mixing coefficient of order k [3]. Since $\alpha(k) \leq 1/4$ by construction, we obtain that $\alpha(k) = 1/4$ for every $k \in \mathbb{N}$. The above calculation also shows that dynamic processes are not 2-α mixing, although this property is less restrictive than strong mixing (see Bosq [3]).

4 Invariant density estimation

Consider $T \in \mathcal{C}$, $T : [a, b] \to [a, b]$, such that $T = \phi^{-1} \circ S \circ \phi$, $S \in \mathcal{S}$. Define $\mu = \lambda\phi$, and let $(X_t)_{t\in\mathbb{N}}$ be the dynamic stochastic process associated to T. Denote by μ' the probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given by $\mu'(B) = \mu(B \cap [a, b])$ for all $B \in \mathcal{B}(\mathbb{R})$. Then, $f := d\mu'/d\lambda_{\mathbb{R}} = \phi' \cdot \mathbb{I}_{(a,b)}$, where $\lambda_{\mathbb{R}}$ is the Lebesgue measure on $\mathbb{R}$ and $\mathbb{I}_A$ denotes the indicator function of a set A. Denote by $\mathcal{C}(f)$ the set of continuity points of f and by $\mathcal{C}_{2,1}(B)$ the set of twice differentiable real valued functions on $\mathbb{R}$, such that $||f||_\infty < B$ and $||f''||_\infty < B$, with $B \in \mathbb{R}^+$.

Now, assume that we observe n successive states: $x_0, \ldots, x_{n-1}$ of the dynamical system $([a, b], \mathcal{B}([a, b]), \mu, T)$ and that we want to estimate the value $f(x)$ with $x \in \mathcal{C}(f)$. According to corollary 3.1, our observations are realizations of the random variables $X_0, \ldots, X_{n-1}$, which are the first n terms of the dynamic process $(X_t)_{t\in\mathbb{N}}$, the marginal density of which is f.
Thus, the value $f(x)$ can be estimated by the Parzen-Rosenblatt estimator [8, 9] :

$$f_n(x) = \frac{1}{nh_n} \sum_{t=0}^{n-1} \mathbb{I}_{[-1/2,1/2]}(\frac{x - X_t}{h_n}),$$

where the bandwidths $(h_n)_{n\in\mathbb{N}}$ are strictly positive and decrease to 0.

Theorem 4.1 *With the above notations, the conditions*

$$h_n \underset{n\to\infty}{\longrightarrow} 0, \; nh_n \underset{n\to\infty}{\longrightarrow} \infty$$

ensure that $\mathbb{E}[(f_n(x) - f(x))^2] \underset{n\to\infty}{\longrightarrow} 0$.
If there exists $B \in \mathbb{R}_+^*$ *such that* $f \in \mathcal{C}_{2,1}(B)$, *then for any bandwidths given by* $h_n = c \cdot n^{-1/5}$ *with* $c \in \mathbb{R}_+^*$, *we have*

$$\mathbb{E}[(f_n(x) - f(x))^2] = O(n^{-4/5}).$$

Proof. Set $K = \mathbb{I}_{[-1/2,1/2]}$. We have

$$\mathbb{E}[(f_n(x) - f(x))^2] = B_n^2(x) + V_n(x) = B_n^2(x) + V_n^{\perp}(x) + C_n(x),$$

where $B_n(x) = \mathbb{E}f_n(x) - f(x)$, $V_n(x) = Var(f_n(x))$,

$$V_n^{\perp}(x) = \frac{1}{n}Var(\frac{1}{h_n}K(\frac{x - X_0}{h_n})),\ C_n(x) = \frac{2}{nh_n^2}\sum_{t=1}^{n-1}(1 - \frac{t}{n})C_{n,t}(x),$$

with $C_{n,t}(x) = Cov(K(\frac{x-X_0}{h_n}), K(\frac{x-X_t}{h_n}))$.
The bias term $B_n(x)$ is the same as in the i.i.d. case and thus converges to 0 [8, 9]. $V_n^{\perp}(x)$ is the variance of the estimator in the i.i.d. case and thus, $nh_n^d V_n^{\perp}(x) \sim f(x)$ [8, 9]. Now, let $W_n(x) = [x - h_n/2, x + h_n/2]$. From corollary 3.1, we can write

$$\begin{aligned} |C_n(x)| &\leq \frac{n-1}{2nh_n^2} \cdot |\frac{1}{\binom{n}{2}} \sum_{\substack{0 \leq s,t \leq n-1 \\ |s \neq t}} [\mu(S^{-s}W_n(x) \cap S^{-t}W_n(x)) - \mu(W_n(x))^2]| \\ &\leq \frac{8}{h_n^2} \cdot \frac{\mu(W_n(x))}{n}. \end{aligned}$$

Thus $|C_n(x)| = O(1/nh_n)$, $Var f_n(x) = O(1/nh_n)$ and hence $\mathbb{E}(f_n(x) - f(x))^2 = O(h_n^4) + O(\frac{1}{nh_n})$. Thus, if $h_n \to 0$, $nh_n \to \infty$, our estimator is consistent in quadratic mean. If there exists $B \in \mathbb{R}_+^*$ such that $f \in \mathcal{C}_{2,1}(B)$, set $h_n = c \cdot n^{-1/5}$. Then, a straightforward calculation gives $n^{4/5}\mathbb{E}(f_n(x) - f(x))^2 = O(1)$. ■

Theorem 4.1 states that the Parzen-Rosenblatt estimator of the marginal density of the dynamic process associated to any map $T \in \mathcal{C}$ by the construction of section 3 converges in quadratic mean for the same bandwidths as in the i.i.d. case and that the optimal rate of the i.i.d. case can be achieved.

Let us translate this result in deterministic terms. Let $x_0, \ldots, x_{n-1}$ denote the observed states of the dynamical system $([a, b], \mathcal{B}([a, b]), \mu, T)$. Let f denote the density of μ, and x a continuity point of f. The Parzen-Rosenblatt estimator of $f(x)$ writes

$$\begin{aligned} f_n(x) &= \frac{1}{nh_n}\sum_{t=0}^{n-1}\mathbb{I}_{[-1/2,1/2]}(\frac{x - x_t}{h_n}) = \frac{1}{nh_n}\sum_{t=0}^{n-1}\mathbb{I}_{[-1/2,1/2]}(\frac{x - S^t x_0}{h_n}) \\ &=: f_n(x_0, x). \end{aligned}$$

The notation $f_n(x_0, x)$ is introduced to emphasize the fact that the estimated value of $f(x)$ at time n depends only on the first observed state x_0.

Now, theorem 4.1 states that if $h_n \to 0$, $nh_n \to \infty$, then

$$\int_{[a,b]} (f_n(x_0, x) - f(x))^2 \, d\mu(x_0) \underset{n\to\infty}{\longrightarrow} 0.$$

If there exists $B \in \mathbb{R}_+^*$ such that $f \in \mathcal{C}_{2,1}(B)$, theorem 4.1 states that for sequences of bandwidths given by $h_n = c \cdot n^{-1/5}$ with $c \in \mathbb{R}_+^*$, we have :

$$\int_{[a,b]} (f_n(x_0, x) - f(x))^2 \, d\mu(x_0) = O(n^{-4/5}).$$

The importance of these results stems from the fact that the mean quadratic error $\int_{[a,b]} (f_n(x_0, x) - f(x))^2 \, d\mu(x_0)$ is the quadratic error one makes on the average when using repeatedly various sequences of observations of length n from a typical trajectory. To see this, remember that in the ergodic case, the frequency at which values in the range $[x_0 - \epsilon, x_0 + \epsilon]$ appear in a typical trajectory is given by $\mu([x_0 - \epsilon, x_0 + \epsilon])$ ($\epsilon \approx 0$ denotes the precision of our observations).
Accordingly, the frequency at which our first observation takes the value x_0, when repeatedly using sequences of n observations from the same trajectory, is approximately equal to $\mu([x_0 - \epsilon, x_0 + \epsilon])$, and roughly speaking, to $d\mu(x_0)$.

Now, in most practical cases, one can watch only the "real-world" trajectory of a dynamical system and has no control over its initial condition. The mean quadratic error then reflects the dispersion of the estimated values obtained by practitians who observe the same phenomenon at various stages of its evolution and for the same duration.

Accordingly, convergence in quadratic error means that the highest the number of observations each practitian records, the less the dispersion of the estimated values, that is to say, the less important the moments at which practicians start recording their observations.

References

[1] G. D. Birkhoff. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. USA*, 17:656–660, 1931.

[2] D. Bosq. Optimal asymptotic quadratic error of density estimators for strong mixing or chaotic data. *Statistics and Probability Letters*, 22:339–347, 1995.

[3] D. Bosq. *Nonparametric Statistics for Stochastic Processes : Estimation and Prediction.* Number 110 in Lecture Notes in Statistics. Springer, 2nd edition, 1998.

[4] A. Lasota and M. Mackey. *Chaos, Fractals and Noise : Stochastic Aspects of Dynamics.* Springer-Verlag, 1994.

[5] R. May. Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467, 1976.

[6] A. Medio. *Chaotic Dynamics : Theory and applications to Economics.* Cambridge, 1992.

[7] Y. Neimark and P. Landa. *Stochastic and Chaotic Oscillations.* Kluwer Academic Publishers, 1992.

[8] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33:1065–1076, 1962.

[9] M. Rosenblatt. Remarks on some non parametric estimates of a density function. *Ann. Math. Stat.*, (27):832–837, 1956.

[10] D. Ruelle. *Chaotic evolution and strange attractors.* Cambridge, 1989.