Determinism is not a curse

Examples of stationary strongly dependent processes whose density can be estimated with optimal i.i.d. rate

Salim Lardjane

Université de Bretagne Sud & CREST
Laboratoire de Statistique & Modélisation
ENSAI, Rue Blaise Pascal
35170 Bruz, France

e-mail : lardjane@ensai.fr

Abstract

The author introduces a class of ergodic dynamical systems for which nonparametric estimation of the invariant density can be done with optimal i.i.d. rate.

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

Published in the Journal of Nonparametric Statistics Volume 15, Numbers 4-5/August-October 2003, pp. 479-483.

The purpose of this paper is to give an example of a class of ergodic dynamical systems for which nonparametric estimation of the invariant density can be done at the same rate as in the situation where the data are independent and identically distributed.

To specify these systems, for any integer r larger or equal to 2, define the transform

$$S_r \quad : \quad x \in [0,1] \quad \mapsto \quad r\,x \mod 1 \in [0,1].$$

The Lebesgue measure is ergodic for any such transform (Lasota and Mackey [4], p.77).
A transformation T is called conjugated with S_r if there exists an interval $[a,b]$ and a diffeomorphism ϕ from $[a,b]$ to $[0,1]$ such that $S_r \circ \phi = \phi \circ T$. Sometimes, we will say that T is ϕ-conjugated with S_r. We will say that T and S_r are positively conjugated if ϕ is increasing. In particular, if (y_n) is a trajectory for S_r, that is $y_n = S_r(y_{n-1})$, then $x_n = \phi^{-1}(y_n)$ satisfies $x_n = T(x_{n-1})$. Thus, the dynamics of one of those two sequences can be recovered from that of the other one.

We claim that such a transformation T, positively conjugated to some S_r, identifies ϕ, a, b and S_r. Indeed, since ϕ is assumed to be increasing, it can be viewed as the cumulative distribution of a probability measure λ_ϕ on $[a,b]$. Since ϕ is one-to-one, for every set $B \subset [0,1]$,

$$\phi(B \triangle T^{-1}B) = (\phi B) \triangle (\phi T^{-1} B) = (\phi B) \triangle (S_r^{-1} \phi B).$$

Thus, λ being ergodic for S_r, for any Borel set B in $[0,1]$, $\lambda(B \triangle T^{-1}B)$ vanishes if and only if $\lambda(\phi B)$ is either 0 or 1. This proves that λ_ϕ is ergodic for T. Since λ_ϕ is ergodic, absolutely continuous with support $[a,b]$, it is necessarily unique (see Boyarsky & Góra [2], p. 37). This proves our claim.

Let us now consider a map T, positively ϕ-conjugated to some S_r. Let X_0 be a point drawn randomly from the distribution with c.d.f. ϕ. Define the sequence $X_i = T(X_{i-1})$, $i \geq 1$. When conditioned on X_0, this sequence is purely deterministic. Its marginal density is the derivative ϕ'. Based on $X_0, \ldots, X_{n-1}$, the Parzen-Rosenblatt estimator of the density ϕ' at point x with a uniform kernel is

$$\phi_n'(x) = \frac{1}{nh_n} \sum_{i=0}^{n-1} \chi_{[-1/2,1/2]}\left((x - X_i)/h_n\right)$$

where (h_n) is a sequence of positive numbers.

It has been established that, under suitable conditions, this estimator is consistent in quadratic mean, with a convergence rate which is the same as in the i.i.d. case up to a logarithm (Bosq [1], p. 57).
The following result shows that despite the strong dependence involved in the process, the estimation of the density ϕ_n' can actually be made at the *same* pointwise rate as in the i.i.d. case.

Theorem 1 *For the process (X_n) above, if ϕ' is continuous at point x, and if both $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n = \infty$, then $\phi_n'(x)$ is a pointwise consistent estimator of $\phi'(x)$ in quadratic mean. Moreover, if ϕ' is twice continuously differentiable at point x and $h_n = c\, n^{-1/5}$, then*

$$\mathbb{E}\left((\phi_n'(x) - \phi'(x))^2 \right) = O(n^{-4/5}).$$

The proof of theorem 1 makes use of the following Lemma, which states a mixing type inequality for intervals, and is of independent interest.

Lemma 1 *Let ϕ be a diffeomorphism from $[0,1]$ to $[a,b]$ and let T be ϕ-conjugated to S_r. For any interval $B \subset [a,b]$, and any integer k,*

$$|\lambda_\phi(B \cap T^{-k}B) - \lambda_\phi(B)^2| \le 4\lambda_\phi(B) r^{-k}.$$

Proof. Let us first establish the result when T is a transform S_r, the general case following by conjugation.
Define the collection of sets

$$\alpha_k = \left\{ (ir^{-k}, (i+1)r^{-k}) : 0 \le i \le r^k - 1 \right\}.$$

By induction, we first see that $S_r^k(x) = r^k x \bmod 1$ for every x in $[0,1]$ and every positive integer k. Hence, for any set $A \in \alpha_k$ and any subset C of $[0,1]$, the equality $A \cap S^{-k}C = \{inf(A) + r^{-k}x : x \in C\} = inf(A) + r^{-k}C$ holds, as well as

$$\lambda(A \cap S^{-k}C) = r^{-k}\lambda(C). \tag{1}$$

Now, consider an interval B in $[0,1]$. We write

$$\lambda(B \cap S^{-k}B) = \sum_{\substack{A\in\alpha_k \\ A\subset B}} \lambda(A \cap B \cap S^{-k}B) + \sum_{\substack{A\in\alpha_k \\ A\not\subset B, A\cap B\neq\emptyset}} \lambda(A \cap B \cap S^{-k}B).$$

Since B is an interval, the second sum is over at most two sets in α_k, those whose closure overlaps with the endpoints of B. Using (1) to evaluate the first sum as well as the (at most two) terms in the second sum, we obtain

$$|\lambda(B \cap S^{-k}B) - \sum_{\substack{A\in\alpha_k \\ A\subset B}} r^{-k}\lambda(B)| \le \sum_{\substack{A\in\alpha_k \\ A\not\subset B, A\cap B\neq\emptyset}} \lambda(A \cap B \cap S^{-k}B) \le 2\lambda(B) r^{-k}.$$

The same endpoint argument yields

$$\lambda(B) - \sum_{\substack{A\in\alpha_k \\ A\subset B}} r^{-k} \le 2r^{-k}$$

and consequently, $\lambda(B)^2$ and $\sum_{\substack{A\in\alpha_k \\ A\subset B}} r^{-k}\lambda(B)$ differ by at most $2r^{-k}\lambda(B)$. This proves Lemma 1 for the transformations S_r.

To obtain the statement of the Lemma, observe that

$$\lambda_\phi(B) = \lambda(\phi B)$$

and

$$\lambda_\phi(B \cap T^{-k}B) = \lambda(\phi B \cap \phi T^{-k}\phi^{-1}\phi B) = \lambda(\phi B \cap S^{-k}\phi B).$$

The result follows by substituting ϕB for B in the result proved for S_r. ■

Proof of Theorem 1. In order to prove Theorem 1, we first derive a bound on the rate at which a transformation T mixes the intervals.

Lemma 2 *For every closed interval B and any positive integer n,*

$$\binom{n}{2}^{-1} \sum_{\substack{1\le i,j\le n-1 \\ j\neq k}} \left|\mathbb{E}\chi_{B\times B}(X_i, X_j) - (\mathbb{E}\chi_B(X_0))^2\right| \le \frac{16\ \lambda_\phi(B)}{n}.$$

Proof. By definition of the process and stationarity, if $i < j$,

$$\mathbb{E}\chi_{B\times B}(X_i, X_j) = \lambda_\phi(T^{-i}B \cap T^{-j}B) = \lambda_\phi(B \cap T^{i-j}B),$$

while

$$\mathbb{E}\chi_B(X_0) = \lambda_\phi(B).$$

We then write

$$\begin{aligned}\binom{n}{2}^{-1} & \sum_{\substack{1\le i,j\le n-1 \\ j\neq k}} |\lambda_\phi(B \cap T^{j-i}B) - \lambda_\phi(B)^2| \\ &= \frac{4}{n-1} \sum_{1\le k\le n-1} \left(1 - \frac{k}{n}\right) |\lambda_\phi(B \cap T^{-k}B) - \lambda_\phi(B)^2| \\ &\le \frac{16\ \lambda_\phi(B)}{(n-1)} \sum_{1\le k\le n-1} \left(1 - \frac{k}{n}\right) r^{-k}.\end{aligned}$$

The result follows by elementary algebra. ■

It is convenient to introduce the notation $K_n(\cdot) = h_n^{-1}\chi_{[-1/2,1/2]}(\cdot/h_n)$.
To prove Theorem 1, we decompose the quadratic error as the square of the bias and a variance term. The process being stationary, the bias term is the same as in the i.i.d. case.

Define

$$V_n = n^{-1}\mathrm{Var}K_n\,(x - X_0)$$

and

$$C_n = \binom{n}{2}^{-1} \sum_{\substack{1\le i,j\le n-1 \\ j\neq k}} \mathrm{Cov}\left(K_n(x - X_i), K_n(x - X_j)\right).$$

Still using stationarity, the variance term is simply $V_n + C_n$.

Setting $B = [x - h_n/2, x + h_n/2]$, we see that

$$\mathrm{Cov}\left(K_n(x - X_i), K_n(x - X_j)\right) = h_n^{-2}\left(\mathbb{E}\chi_{B\times B}(X_i, X_j) - (\mathbb{E}\chi_B(X_0))^2\right).$$

Thus, Lemma 2 and continuous differentiability of ϕ at x yield $C_n = O(1/nh_n)$. It follows that $\mathrm{Var}\phi_n'(x) = O(1/nh_n)$ and, when ϕ' is twice continuously differentiable at x, that

$$\mathbb{E}\left(\phi_n'(x) - \phi'(x)\right)^2 = O(h_n^4) + O(1/nh_n).$$

The result is then plain. ■

Let us emphasize here the specificity of Lemma 2. By restricting B to be an *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 (see Boyarsky and Góra [2], p.148, and Chernov [3]). The proof of Theorem 1 shows that these bounds are not sufficiently sharp to derive the result.

Note that the bound obtained in Lemma 2 obviously holds for ψ, ϕ^* and ϕ - mixing stochastic processes. However, (X_n) as above is purely deterministic when conditioned on X_0, and hence, is not even α - mixing.

Acknowledgement. The author would like to thank the referee for detailed and valuable comments which have substantially improved the presentation of the ideas in this paper.

References

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

[2] A. Boyarsky and P. Góra. *Laws of Chaos : Invariant Measures and Dynamical Systems in One Dimension.* Probability and its applications. Birkhaüser, 1997.

[3] N. Chernov. Limit theorems and markov approximations for chaotic dynamical systems. *Probab. Theory Relat. Fields*, 101:321–362, 1995.

[4] A. Lasota and M. Mackey. *Chaos, Fractals, and Noise : Stochastic Aspects of Dynamics.* Number 97 in Applied Mathematical Sciences. Springer-Verlag, 2nd edition, 1994.