Reverse chaos may not be a curse

Examples of stationary reverse chaotic sequences whose density can be estimated with optimal i.i.d. rate

Salim Lardjane

Université de Bretagne Sud & CREST
Laboratoire de Statistique & Modélisation
ENSAI, Campus de Ker Lann, rue Blaise Pascal
35170 Bruz, France

e-mail: lardjane@ensai.fr

Abstract

The author introduces a class of reverse chaotic stochastic processes for which nonparametric estimation of the invariant density can be done with optimal i.i.d. rate. This class includes standard examples of nonmixing linear and nonlinear autoregressive stochastic processes.

KEYWORDS: Nonparametric density estimation, Reverse chaos, $r$-adic transformations, Mixing.

Published in the Journal of Nonparametric Statistics Volume 17, Number 8/December 2005, pp. 885-889.
The purpose of this paper is to give an example of a class of stationary nonmixing stochastic processes which satisfy a deterministic recurrence equation in reverse time and for which the estimation of the marginal density can be done at the same rate as in the situation where the data are independent and identically distributed.

To motivate this work, let us introduce some standard examples of reverse chaotic stochastic processes. Let us first consider the linear autoregressive stochastic process \((X_n)\) given by

\[
X_{n+1} = \frac{1}{r} (X_n + \epsilon_n), \quad n \geq 0
\]

where \(r\) is an integer greater or equal to 2, \(X_0\) is uniformly distributed on \([0, 1]\), and \((\epsilon_n)\) is a sequence of discrete independent random variables with uniform distribution on \([0, \ldots, r-1]\), independent from \(X_0\). Although the process \((X_n)\) is strictly stationary, it is not strongly mixing. To see this, note that, for any \(n\) positive, \(X_n = S_r (X_{n+1})\) where \(S_r\) is the \(r\)-adic transform

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

Consequently, \(X_0, X_1, \ldots, X_n\) is a truncated trajectory of the \(r\)-adic dynamical system in reverse time, and \(X_0, X_1, \ldots, X_n\) is perfectly deterministic when conditioned on \(X_n\).

The process \((X_n)\) is quoted as a standard example of a nonmixing autoregressive stochastic process since Andrews [1], and has been considered by Bartlett [3], Chernick [6, 7], Lawrance [11] and Tong [16]. It is mentioned by Bosq [4] (p. 18), Doukhan [8] (p. 19) and Rosenblatt [15] (p. 51), among others. In the discussion of this article by Bartlett, the origin of this ‘curious example’ is attributed to both Rosenblatt and Whittle while Rosenblatt refers in the same discussion to ‘ancestral folklore’. Note that, since the process \((X_n)\) is linear, the results of Carbon and Tran [5] and Hallin and Tran [9] on nonparametric density estimation for linear stochastic processes are relevant for \((X_n)\).

It has been pointed by Tong [16] that another interesting example of reverse chaos is the non-linear autoregressive process given by

\[
Y_{n+1} = \frac{1}{2} \left( 1 + (2\epsilon_{n+1} - 1)(1 - Y_n)^{1/2} \right), \quad n \geq 0
\]

where \(Y_0\) has a \(\beta(1/2, 1/2)\) distribution on \([0, 1]\) and \((\epsilon_n)\) is a sequence of i.i.d. \(\mathcal{B}(1/2)\) random variables, independent from \(Y_0\).

\((Y_n)\) is strictly stationary and for every \(n \geq 0\),

\[
Y_n = L(Y_{n+1})
\]

where \(L\) is the logistic transform

\[
L : x \in [0, 1] \mapsto 4x(1-x).
\]
The purpose of this paper is to generalize these examples to a large family of strictly stationary reverse chaotic stochastic processes, and prove, that in spite of their nonmixingness, the estimation of their marginal densities can be done with optimal i.i.d. rate.

To specify this family, let us first introduce some definitions. A transformation $T$ is called conjugated with $S_r$ if there exists an interval $[a, b]$ and a diffeomorphism $\phi$ from $[a, b]$ to $[0, 1]$ such that $S_r \circ \phi = \phi \circ T$. Sometimes, we will say that $T$ is $\phi$-conjugated with $S_r$. We will say that $T$ and $S_r$ are positively conjugated if $\phi$ is increasing.

Let us now consider a map $T$, positively $\phi$-conjugated to some $S_r$. Then, $T$ identifies $\phi, a, b$ and $S_r$ [10]. Moreover, since $\phi$ is assumed to be increasing, it can be viewed as the cumulative distribution of a probability measure $\lambda_\phi$ on $[a, b]$.

The $r$-adic transform $S_r$ has $r$ bijective branches

$$S_{r,i} : \quad x \in \alpha_i \mapsto r x - i, \quad i = 0, \ldots, r - 1,$$

where $\alpha_i = [ir^{-1}, (i + 1)r^{-1})$. Since $\phi$ is bijective, $T$ also has $r$ bijective branches

$$T_i : \quad x \in \phi^{-1}\alpha_i \mapsto \phi^{-1} \circ S_{r,i} \circ \phi(x), \quad i = 0, \ldots, r - 1,$$

Now, let $X_0$ be a point drawn randomly from the distribution with c.d.f. $\phi$. Let us define the sequence $(X_n)$ almost surely by the random recurrence equation

$$X_{n+1} = T_{\theta_{n+1}}^{-1}(X_n), \quad n \geq 0, \quad (1)$$

where $(\theta_n)_{n \geq 1}$ is a sequence of independent random variables uniformly distributed on $\{0, \ldots, r - 1\}$ and independent from $X_0$.

It is easily seen that the sequence $(X_n)$ is not strongly mixing since it satisfies the reverse recurrence equation

$$X_n = T(X_{n+1}), \quad n \geq 0. \quad (2)$$

We claim that $(X_n)$ is strictly stationary with marginal distribution $\lambda_\phi$.

Since $(X_n)$ is a Markov chain, it is sufficient to prove that $X_1$ has the same distribution as $X_0$. Let $A$ be a borel set of $[a, b]$. Let us denote by $\lambda$ the Lebesgue measure on $[0, 1]$. Then,

$$P(X_1 \in A) = E(P(T_{\theta_1}^{-1}X_0 \in A|\theta_1)) = \sum_{i=0}^{r-1} r^{-1} P(T_i^{-1}X_0 \in A)$$

$$= r^{-1} \sum_{i=0}^{r-1} \lambda_\phi(T_i(A \cap \phi^{-1}\alpha_i)) = r^{-1} \sum_{i=0}^{r-1} \lambda(S_r \phi(A \cap \phi^{-1}\alpha_i))$$

$$= r^{-1} \sum_{i=0}^{r-1} r \lambda(\phi A \cap \alpha_i) = \lambda_\phi(A).$$
The marginal density of \((X_n)\) is the derivative \(\phi'\). Based on \(X_0, \ldots, X_{n-1}\), the Parzen-Rosenblatt estimator of the density \(\phi'\) at point \(x\) with a uniform kernel is
\[
\phi_n'(x) = \frac{1}{nh_n} \sum_{i=0}^{n-1} \chi_{[-1/2,1/2]}((x - X_i)/h_n)
\]
where \((h_n)\) is a sequence of positive numbers.

The following result shows that in spite of the strong dependence induced by (2), the estimation of the density \(\phi'\) 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 both \(\lim_{n \to \infty} h_n = 0\) and \(\lim_{n \to \infty} nh_n = \infty\), and if \(\phi'\) is continuous at point \(x\), then \(\phi_n'(x)\) is a pointwise consistent estimator of \(\phi'(x)\) in quadratic mean. Moreover, if \(\phi'\) is twice continuously differentiable at point \(x\), and \(h_n = c n^{-1/5}\), then
\[
\mathbb{E} \left( (\phi_n'(x) - \phi'(x))^2 \right) = O(n^{-4/5}).
\]

**Proof of Theorem 1.** In order to prove Theorem 1, we first derive a covariance inequality for \((X_n)\).

**Lemma 1** For every closed interval \(B\) and any positive integer \(n\),
\[
\left( \frac{n}{2} \right)^{-1} \sum_{0 \leq i, j \leq n-1 \atop i \neq j} | \mathbb{E}_{B \times B} (X_i, X_j) - (\mathbb{E}_{B}(X_0))^2 | \leq \frac{16 \lambda(B)}{n}.
\]

**Proof.** By definition of the process and stationarity, if \(i < j\),
\[
\mathbb{E}_{B \times B} (X_i, X_j) = \mathbb{E}_{B \times B} (T^{-j}X_j, X_j) = \lambda(B \cap T^{-j}B),
\]
while \(\mathbb{E}_{B}(X_0) = \lambda(B)\). But, according to lemma 1 in Lardjane [10], for any \(k\) positive\(^1\),
\[
|\lambda(B \cap T^{-k}B) - \lambda(B)\|^2 \leq 4\lambda(B)r^{-k}.
\]

We then write
\[
\left( \frac{n}{2} \right)^{-1} \sum_{0 \leq i, j \leq n-1 \atop i \neq j} | \lambda(B \cap T^{-j}B) - \lambda(B)\|^2 |
\leq \frac{16 \lambda(B)}{(n-1)} \sum_{1 \leq k \leq n-1} \left( 1 - \frac{k}{n} \right) r^{-k}.
\]

\(^1\)Inequalities analogous to (3) can actually be derived from general covariance inequalities obtained in Ergodic Theory (Baladi [2], Liverani [12, 13]). Building on these inequalities, some consistency results on nonparametric density estimation for one-dimensional dynamical systems have been obtained by Prieur [14]. In the context of this paper and of Lardjane [10], a direct use of these inequalities would have led us to an upper bound of the form \(C \cdot \lambda(B) \cdot r^{-k}\), where \(C\) is a unspecified positive constant. Inequality (3) is still of interest since it provides an explicit value for \(C\) in the special case considered here, and given the comparatively elementary and self-contained nature of its proof [10].
The result follows by elementary algebra. ■

Let us 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, and the variance term can be decomposed as \( V_n + C_n \), where

\[
V_n = n^{-1} \operatorname{Var} K_n(x - X_0) \quad \text{and} \quad C_n = \left( \begin{array}{c} n \\ 2 \end{array} \right)^{-1} \sum_{0 \leq i, j \leq n-1 \atop i \neq j} \operatorname{Cov} (K_n(x - X_i), K_n(x - X_j)) .
\]

Setting \( B = [x - h_n/2, x + h_n/2] \), we see that

\[
\operatorname{Cov} (K_n(x - X_i), K_n(x - X_j)) = h_n^{-2} \left( \mathbb{E} X_B \times B (X_i, X_j) - (\mathbb{E} X_B (X_0))^2 \right).
\]

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

\[
\mathbb{E} (\phi_n'(x) - \phi'(x))^2 = O(h_n^4) + O(1/nh_n),
\]
hence the result. ■

References


