ON INTEGRATED SQUARE ERRORS
OF RECURSIVE NONPARAMETRIC ESTIMATES
OF NONSTATIONARY MARKOV PROCESSES*

BY
ONÉSIMO HERNÁNDEZ-LERMA (MEXICO CITY)

Abstract. The integrated square error (ISE) and the mean integrated square error (MISE) for a class of recursive estimators of the transition density function of a vector-valued nonstationary Markov process are considered. Conditions are given under which the MISE converges, and the ISE converges in probability and almost surely.

1. Introduction. Let \( \{ x_t, t = 0, 1, \ldots \} \) be an \( \mathbb{R}^d \)-valued nonstationary Markov process with a transition density function \( q(y|x) = f(x, y)/\gamma(x) \), where \( \gamma(x) \) and \( f(x, y) \) are probability densities on \( \mathbb{R}^d \) and \( \mathbb{R}^{2d} \), respectively. Given a density estimator, say, \( \hat{\gamma}_t(x) \) of \( \gamma(x) \), a widely used measure of the global performance of \( \hat{\gamma}_t \) is the Mean Integrated Square Error (MISE)

\[
E(I_t) = E \int [\hat{\gamma}_t(x) - \gamma(x)]^2 dx = \int M_t(x) dx,
\]

where \( I_t \) is the Integrated Square Error (ISE):

\[
I_t := \int [\hat{\gamma}_t(x) - \gamma(x)]^2 dx,
\]

and \( M_t(x) \) is the Mean Square Error:

\[
M_t(x) := E[\hat{\gamma}_t(x) - \gamma(x)]^2.
\]

(Unqualified integrals, as in (1.1) and (1.2), denote integration over all of \( \mathbb{R}^d \).) In this paper, we let \( \hat{\gamma}_t \) be the recursive estimators introduced by Wolverton and Wagner [23], and further studied by Yamato [25] and other authors (cf. Chapters 5 and 6 in [15]), and give conditions under which both the MISE and the ISE converge to zero as \( t \to \infty \), the convergence of \( I_t \) being almost surely (a.s.). These results are also extended to estimators \( \hat{f}_t(x, y) \) and \( \hat{q}_t(y|x) \) of \( f(x, y) \) and \( q(y|x) \), respectively.

* This research was supported in part by the Consejo Nacional de Ciencia y Tecnología (CONACYT) under Grant PCEXCNA-050156.
The MISE in (1.1) was introduced by Rosenblatt [16] and it has also been studied by many authors, mainly for independent and identically distributed (i.i.d.) sequences (see [1], [7], [15], [19]) as well as for stationary mixing processes (e.g., [4], [13]). On the other hand, the study of nonparametric estimation problems for Markov processes was initiated in the late 1960's with the pioneering works of Roussas [20] and Rosenblatt [17], [18], and by now there is a large number of papers on the subject; see, e.g., [3], [11], [24] and the extensive list of references in [15], Chapter 6. All of these papers, however, have in common that they only consider stationary Markov processes. Two exceptions are [6] and [10], which are closely related to our present work. In [6], Gillert and Wartenberg study, among other things, the mean square error (1.3) and the MISE (but not the ISE) for scalar nonstationary Markov processes, using the well-known nonrecursive Parzen–Rosenblatt [14], [16] density estimators. In contrast, [10] studies the recursive Wolverton–Wagner (WW) density estimates (cf. Section 3 below), which are shown to be uniformly consistent in mean square as well as strongly pointwise consistent and strongly consistent in the \(L_2\)-norm. As far as the Markov process \(\{x_t\}\) is concerned, the context in the present paper is essentially the same as in [6] and [10].

We begin in Section 2 by introducing the assumptions on the Markov process \(\{x_t\}\) and we also summarize some of its properties. The WW estimates are introduced in Section 3 together with our main results, whose proofs are all collected in Section 4.

**NOTATION.** The Borel \(\sigma\)-algebra of \(\mathbb{R}^d\) is denoted by \(\mathcal{B}^d\). For a finite signed measure \(\mu\), \(\|\mu\|\) denotes the variation norm, whereas for a function \(f\) on \(\mathbb{R}^d\), \(\|f\|\) stands for the supremum norm. By convention, \(0/0 = 0\).

2. Preliminaries. Let \(\{x_t, t = 0, 1, \ldots\}\) be an \(\mathbb{R}^d\)-valued homogeneous Markov process with transition density \(q(y|x)\). Thus, given an arbitrary initial distribution \(\mu_0\), the distribution \(\mu_t\) of \(x_t\), for \(t \geq 1\), is given by

\[
\mu_t(B) = \int Q(B|x)\mu_{t-1}(dx), \quad B \in \mathcal{B}^d,
\]

where \(Q(B|x) = \int q(y|x)dy\) is the transition probability measure. Throughout we suppose the following

**Assumption 2.1.** (a) The initial distribution \(\mu_0\) is absolutely continuous with a bounded density \(\gamma_0\).

(b) There is a positive number \(\alpha < 1\) such that

\[
\|Q(\cdot|x) - Q(\cdot|y)\| \leq 2\alpha \quad \text{for all } x \text{ and } y \text{ in } \mathbb{R}^d.
\]

(c) For some constant \(\bar{q}\), \(q(y|x) \leq \bar{q}\) for all \(x\) and \(y\) in \(\mathbb{R}^d\).

Assumption 2.1 (b) is a well-known ergodicity condition for Markov processes ([2], [12], [21]), and also for Markov control (or decision) processes ([5], [8], [9]) in which the transition measure \(Q(\cdot|x, a)\) depends also on an "action" (or control) variable \(a \in A\) for some Borel space \(A\). Sufficient
conditions for Assumption 2.1 (b) are given, e.g., in [5], [8], [9] — taking $A$ as a one-point set — and in [3], Section 2.1, for Markov processes of the form $x_{t+1} = F(x_t) + \xi_t$, where $\{\xi_t\}$ is a sequence of i.i.d. random vectors.

Assumption 2.1 (b) guarantees the existence of a probability measure $\mu$ such that

$$\|\mu_t - \mu\| \leq 2x^{t}, \quad t = 0, 1, \ldots,$$

for an arbitrary initial distribution. Clearly, $\mu$ is the unique invariant distribution of $\{x_t\}$. On the other hand, (2.1) and Assumption 2.1 (a) yield that $\mu_t$ is absolutely continuous for all $t \geq 0$, and this, in turn, combined with (2.2), implies that also $\mu$ is absolutely continuous. The corresponding densities of $\mu_t$ and $\mu$, denoted by $\gamma_t$ and $\gamma$, respectively, satisfy

$$\gamma_t(y) = \int q(y|x)\gamma_{t-1}(x)dx \quad \text{and} \quad \gamma(y) = \int q(y|x)\gamma(x)dx$$

for all $y \in \mathbb{R}^d$ and $t \geq 1$, and since $\|\mu_t - \mu\| = \int |\gamma_t(x) - \gamma(x)|dx$, we can rewrite (2.2) as

$$\int |\gamma_t(x) - \gamma(x)|dx \leq 2x^{t}, \quad t = 0, 1, \ldots$$

Finally, the latter inequality and (2.3) yield

$$|\gamma_t(y) - \gamma(y)| < \int q(y|x)|\gamma_{t-1}(x) - \gamma(x)|dx \leq 2\bar{q}x^{t-1}$$

for all $y \in \mathbb{R}^d$ and $t \geq 1$, where $\bar{q}$ is the constant in Assumption 2.1 (c). Therefore, we conclude

**Lemma 2.2.** $\|\gamma_t - \gamma\| = \sup_x|\gamma_t(x) - \gamma(x)| \leq 2\bar{q}x^{t-1} \to 0$ as $t \to \infty$.

Notice also that both $\gamma$ and $\gamma_t$ are bounded: from (2.3) we obtain

$$\|\gamma\| \leq \bar{q} \quad \text{and} \quad \|\gamma_t\| \leq c_0 \text{ for all } t \geq 0,$$

where $c_0 := \max\left\{\bar{q}, \|\gamma_0\|\right\}$.

Similar results can be obtained for the joint Markov process $z_t := (x_t, x_{t+1})$, $t = 0, 1, \ldots$. For each $t \geq 0$, $z_t$ has a density $f_t(x, y) = q(y|x)\gamma_t(x)$ for $(x, y)$ in $\mathbb{R}^{2d}$, which is uniformly bounded, since $|f_t(x, y)| \leq \bar{q}c_0 \leq c_0^t$ by (2.5) and Assumption 2.1 (c). Furthermore, Lemma 2.2 yields

**Lemma 2.3.** $\sup_{x, y}|f_t(x, y) - f(x, y)| \leq \bar{q} \cdot \|\gamma_t - \gamma\| \to 0$ as $t \to \infty$, where $f(x, y) = q(y|x)\gamma(x)$.

For the results in the following section we need additional assumptions:

**Assumption 2.4.** There exist bounded measurable functions $g$ and $h$ on $\mathbb{R}^d$ such that, for all $x, y, x', y'$ in $\mathbb{R}^d$,

(a) $|q(y|x) - q(y'|x)| \leq |g(x)| |y - y'|$,

(b) $|q(y|x) - q(y'|x')| \leq |h(y)| |x - x'|$.

It is easily verified that Assumption 2.4 (a) implies that $\gamma$ and $\gamma_t$ are uniformly Lipschitz-continuous, and similarly for $f$ and $f_t$ when both Assumptions 2.4 (a) and (b) hold.
3. Recursive estimation. In this section we consider the recursive Wolverton-Wagner (WW) nonparametric estimates of \( \gamma(x) \), \( f(x, y) \), and \( q(y|x) \), and state our main results (Theorems 3.1, 3.2, 3.3).

Let \( u(x) \) be a given probability density on \( \mathbb{R}^d \) and let \( \{b_t\} \) be a sequence of positive numbers. The WW estimate \( \hat{\gamma}_t \) of \( \gamma \) is defined for \( x \in \mathbb{R}^d \) and \( t \geq 1 \) by

\[
\hat{\gamma}_t(x) := t^{-1} \sum_{n=0}^{t-1} u_n(x_n - x) \quad \text{with} \quad u_n(x) := b_n^{-d} u(x/b_n).
\]

Similarly, the WW estimates of \( f(x, y) \) and \( q(y|x) \) are defined for \( (x, y) \in \mathbb{R}^{2d} \) by

\[
\hat{f}_t(x, y) := t^{-1} \sum_{n=0}^{t-1} \hat{u}_n(x_n - x, x_{n+1} - y)
\]

and

\[
\hat{q}_t(y|x) := \frac{\hat{f}_t(x, y)}{\hat{\gamma}_t(x)},
\]

respectively, where

\[
\hat{u}_n(x, y) := u_n(x)u_n(y) = b_n^{-2d} u(x/b_n)u(y/b_n).
\]

We assume throughout that \( u \) is bounded and \( \int |x| u(x) dx < \infty \). Concerning the sequence \( \{b_t\} \), we assume that it is nonincreasing, \( b_0 \leq 1 \), and it satisfies some of the following conditions as \( t \to 0 \):

(B1) \( b_t \to 0 \);

(B2) \( t b_t^d \to \infty \); \quad (B2') \( t b_t^{2d} \to \infty \);

(B3) \( \sum t^{-3/2} b_t^{-d} < \infty \); \quad (B3') \( \sum t^{-3/2} b_t^{-2d} < \infty \).

To state the consistency results in a compact form, let us write the mean square error \( M_t(x) \) in (1.3) as

\[
M_t(x) = \text{Var}[^\hat{\gamma}_t(x)] + B_t^2(x),
\]

where \( \text{Var} \) denotes the variance, and \( B_t(x) \) is the bias function of \( \hat{\gamma}_t(x) \), i.e.,

\[
\text{Var}[\hat{\gamma}_t(x)] := E[\hat{\gamma}_t(x) - E\hat{\gamma}_t(x)]^2, \quad B_t(x) := E\hat{\gamma}_t(x) - \gamma(x).
\]

Thus, we can write the MISE in (1.1) as

\[
E(I_t) = \int \text{Var}[\hat{\gamma}_t(x)] dx + \int B_t^2(x) dx.
\]

**Theorem 3.1.** Suppose that Assumptions 2.1 and 2.4 (a) hold and let \( t \to \infty \).

(a) If (B1) holds, then \( \int B_t^2(x) dx \to 0 \).

(b) If (B2) holds, then \( \int \text{Var}[\hat{\gamma}_t(x)] dx \to 0 \).

(c) If both (B1) and (B2) hold, then the MISE \( E(I_t) \to 0 \), and therefore the ISE \( I_t \to 0 \) in probability.

(d) If (B1)-(B3) hold, then \( I_t \to 0 \) almost surely (a.s.).
Theorem 3.1, as well as Theorems 3.2 and 3.3 below, are proved in Section 4.

The corresponding result for the 2d-dimensional estimator \( \hat{f}_t(x, y) \) is a natural extension of Theorem 3.1 (see the Remark following Theorem 3.2). The ISE for \( \hat{f}_t \) is

\[
\bar{I}_t := \iint [\hat{f}_t(x, y) - f(x, y)]^2 \, dx \, dy,
\]

and the MISE can be written as

\[
E(\bar{I}_t) := \iint \bar{M}_t(x, y) \, dx \, dy,
\]

where \( \bar{M}_t \) is the mean square error

\[
\bar{M}_t(x, y) := E[\hat{f}_t(x, y) - f(x, y)]^2 = \text{Var}[\hat{f}_t(x, y)] + B_t^2(x, y)
\]

with \( B_t(x, y) := B_t^f(x, y) - f(x, y) \), the bias function. With this notation, we have

**Theorem 3.2.** Suppose that Assumptions 2.1 and 2.4 (both (a) and (b)) hold and let \( t \to \infty \).

(a) If (B1) holds, then \( \iint B_t^2(x, y) \, dx \, dy \to 0 \).

(b) If (B2') holds, then \( \iint \text{Var}[\hat{f}_t(x, y)] \, dx \, dy \to 0 \).

(c) If both (B1) and (B2') hold, then \( E(\bar{I}_t) \to 0 \), and \( \bar{I}_t \to 0 \) in probability.

(d) If (B1), (B2') and (B3') hold, then \( \bar{I}_t \to 0 \) a.s.

**Remark.** Suppose that instead of the estimator \( \hat{f}_t \) in (3.1) we consider

\[
\hat{f}_t^*(x, y) := \frac{1}{i-1} \sum_{n=0}^{i-1} b_n^{-d} \left[ ((x_n - x)/b_n^{1/2}) u[(x_{n+1} - y)/b_n^{1/2}] \right].
\]

Then Theorem 3.2 holds when \( \hat{f}_t \) is replaced by \( \hat{f}_t^* \), and conditions (B2') and (B3') are replaced by (B2) and (B3), respectively. However, we decided to use \( \hat{f}_t \) (and not \( \hat{f}_t^* \)) as an estimator of \( f(x, y) \) because with \( \hat{f}_t \) the calculations in the proof as Theorem 3.1 are more directly extended to the 2d-dimensional situation of Theorem 3.2. Other authors, of course, use \( \hat{f}_t^*(x, y) \), or some variant, to estimate \( f(x, y) \); see, e.g., Prakasa Rao [15], p. 320.

Finally, let us consider the ISE for the estimator \( \hat{q}_t(y|x) \):

\[
J_t(x) := \iint [\hat{q}_t(y|x) - q(y|x)]^2 \, dy, \quad x \in \mathbb{R}^d.
\]

**Theorem 3.3.** Suppose that Assumptions 2.1 and 2.4 hold together with conditions (B1), (B2), and (B3). Let \( x \in \mathbb{R}^d \) be such that \( \gamma(x) > 0 \) and let \( t \to \infty \).

(a) If (B2') holds, then \( E J_t(x) \to 0 \), and therefore \( J_t(x) \to 0 \) in probability.

(b) If (B2') and (B3') hold, then \( J_t(x) \to 0 \) a.s.

**4. Proofs.** For ease of reference, we restate here some results from [10].

**Lemma 4.1.** Suppose that Assumptions 2.1 and 2.4 (a) hold and let \( t \to \infty \).

(a) If (B1) holds, then \( \sup_x B_t(x) \to 0 \), i.e., \( \hat{g}_t(\cdot) \) is uniformly asymptotically unbiased.
(b) If (B1) and (B2) hold, then $\sup_x M_t(x) \to 0$, i.e., $\hat{\gamma}_t(\cdot)$ is uniformly consistent in mean square.

(c) If (B1), (B2), and (B3) hold, then $\hat{\gamma}_t(x) \to \gamma(x)$ a.s. for all $x \in \mathbb{R}^d$, i.e., $\hat{\gamma}_t(\cdot)$ is strongly pointwise consistent.

Suppose, in addition, that Assumption 2.4 (b) holds. Then the corresponding results for $\hat{f}_t(x, y)$ are the following:

(a') If (B1) holds, then $\sup_x |B_t(x, y)| \to 0$.

(b') If (B1) and (B2') hold, then $\sup_x M_t(x, y) \to 0$.

(c') If (B1), (B2') and (B3') hold, then $\hat{f}_t(x, y) \to f(x, y)$ a.s. for all $(x, y) \in \mathbb{R}^{2d}$.

Proof. See Theorems 3.1 and 4.1 in [10].

Proof of Theorem 3.1. (a) This part follows from Lemma 4.1 (a) since

$$\int B_t^2(x) dx < \sup_x |B_t(x)| \int |B_t(x)| dx \to 0 \quad \text{as} \quad t \to \infty.$$

(b) Let us write the variance of $\hat{\gamma}_t(x)$ as $\text{Var}[\hat{\gamma}_t(x)] = t^{-2} \sum_{n,m} \Gamma_{nm}(x)$, where the sum is over $n, m = 0, 1, \ldots, t-1$, and $\Gamma_{nm}(x)$ is the covariance function:

$$\Gamma_{nm}(x) := \text{Cov}[u_n(x_n-x), u_m(x_m-x)].$$

By Ueno’s [21] Lemma 3, we have

$$|\Gamma_{nm}(x)| \leq \|u\| b_m d^{m-n} \text{Eu}_n(x_n-x) \quad \text{for all} \quad 0 \leq n \leq m,$$

where

$$\text{Eu}_n(x_n-x) = \int u_n(y-x)\gamma_n(y)dy = \int \gamma_n(b_n y + x)u(y)dy,$$

and $\alpha \in (0, 1)$ is the coefficient of ergodicity in Assumption 2.1 (b). Therefore, since $\{b_n\}$ is nonincreasing and $\int \text{Eu}_n(x_n-x)dx = 1$, we obtain

$$\int |\Gamma_{nm}(x)| dx \leq \|u\| b^{-d} t^{n-m} \quad \text{for} \quad 0 \leq n \leq m \leq t,$$

so that the variance of $\hat{\gamma}_t$ satisfies

$$\int \text{Var}[\hat{\gamma}_t(x)] dx \leq t^{-2} \sum_{n,m} \int |\Gamma_{nm}(x)| dx,$$

$$\leq \|u\| t^{-2} b^{-d} \sum_{n,m} \alpha^{m-n} \leq Ct^{-1} b^{-d}$$

with $C := 2 \|u\|/(1-\alpha)$, which results from the inequality $\sum_{n,m} \alpha^{m-n} \leq 2\alpha/(1-\alpha)$. Thus condition (B2) implies part (b).

(c) This part follows from (a), (b), and equation (3.2).

(d) Adding and subtracting $E\hat{\gamma}_t(x)$ inside the brackets in (1.2), and using the inequality $(a+b)^2 \leq 2(a^2+b^2)$, we see that the ISE $I_t$ satisfies

$$I_t \leq 2 \int [\hat{\gamma}_t(x) - E\hat{\gamma}_t(x)]^2 dx + 2 \int B_t^2(x) dx.$$
The second term on the right-hand side converges to zero (by part (a)), and therefore to prove (d) it suffices to show that, as \( t \to \infty \),

\[
Y_t := \int [\gamma_t(x) - E\gamma_t(x)]^2 \, dx \to 0 \text{ a.s.}
\]

(4.4)

In turn, to prove (4.4), it suffices to prove that

\[
Y_t \to L \text{ a.s. for some finite limit } L,
\]

(4.5)

since, by part (b), \( E(Y_t) = \int \text{Var}[\gamma_t(x)] \, dx \to 0 \), so that necessarily \( L = 0 \) a.s. To prove (4.5) we will use van Ryzin's Lemma of [22], pp. 1765 and 1766. First note that, by the definition of \( \gamma_t(x) \),

\[
\gamma_{t+1}(x) = (t+1)^{-1} [r\gamma_t(x) + u_t(x_t-x)].
\]

(4.6)

Hence, defining

\[
Z_t(x) := \gamma_t(x) - E\gamma_t(x) \quad \text{and} \quad U_t(x) := u_t(x_t-x) - Eu_t(x_t-x),
\]

we obtain

\[
Z^2_{t+1}(x) = Z^2_t(x) + (t+1)^{-2} [U^2_t(x) + 2tU_t(x)Z_t(x) - (2t+1)Z^2_t(x)]
\]

\[
\leq Z^2_t(x) + (t+1)^{-2} [U^2_t(x) + 2tU_t(x)Z_t(x)],
\]

whence

\[
Y_{t+1} = \int Z^2_{t+1}(x) \, dx \leq Y_t + (t+1)^{-2} \int [U^2_t(x) + 2tU_t(x)Z_t(x)] \, dx.
\]

Now for \( t = 1, 2, \ldots \) let \( \mathcal{F}_t \) be the \( \sigma \)-algebra generated by \( x_0, \ldots, x_{t-1} \), and notice that \( Y_t \) is \( \mathcal{F}_t \)-measurable, and also \( E(Y_{t+1} | \mathcal{F}_t) \leq Y_t + Y'_t \), where \( Y'_t \) is an \( \mathcal{F}_t \)-measurable random variable defined by

\[
Y'_t := (t+1)^{-2} E \left\{ \int [U^2_t(x) + 2tU_t(x)Z_t(x)] \, dx | \mathcal{F}_t \right\}.
\]

To conclude (4.5) from van Ryzin's Lemma [22], we have to show that

\[
\sum_t E|Y'_t| < \infty.
\]

(4.7)

To prove this, note that \( E \int U^2_t(x) \, dx = \int \Gamma_n(x) \, dx < \|u\| b_i^{-d} \) by (4.2), whereas by (4.3) and repeated applications of the Schwartz inequality we obtain

\[
E \int |U_t(x)||Z_t(x)| \, dx \leq \left( \int \text{Var}[\gamma_t(x)] \, dx \right)^{1/2} \left( \int \Gamma_n(x) \, dx \right)^{1/2}
\]

\[
\leq (C t^{-1} b_i^{-d})^{1/2} \left( \|u\| b_i^{-d} \right)^{1/2} \leq C_1 t^{-1/2} b_i^{-d}
\]

for some constant \( C_1 \). Thus, for some constant \( C_2 \), \( E|Y'_t| \leq C_2 t^{-3/2} b_i^{-d} \), so that (4.7) follows from condition (B3). This completes the proof of (4.5), which, as noted earlier, yields part (d). \( \blacksquare \)

**Proof of Theorem 3.2.** This proof is, of course, the same as that of Theorem 3.1 with obvious changes. For instance, part (a) follows from Lemma 4.1 (a'), and, similarly, the covariance function

\[
\Gamma_{nm}(x, y) := \text{Cov}[\tilde{u}_n(x_n-x, x_{n+1}-y), \tilde{u}_m(x_m-x, x_{m+1}-y)]
\]
can be estimated by using Ueno's [21] Lemma 3 again, to obtain

\[ |F_{nn}(x, y)| \leq \begin{cases} \frac{b_{n-2d}}{u^2} \alpha^{m-n-1} E\{x_n - x, x_{n+1} - y\} & \text{for } n+1 \leq m, \\ \frac{b_{n-2d}}{u^2} E\{x_n - x, x_{n+1} - y\} & \text{for } n = m. \end{cases} \]

Thus, the 2d-dimensional analogue of (4.2) is

\[ \int \int |F_{nm}(x, y)| dx dy \leq \begin{cases} \frac{b_{n-2d}}{u^2} \alpha^{m-n-1} & \text{for } n+1 \leq m \leq t, \\ \frac{b_{n-2d}}{u^2} & \text{for } n = m \leq t. \end{cases} \]

The other changes in the proof of Theorem 3.1 are just as obvious. ■

Moreover, from Lemma 4.1 (a')--(c') and Theorem 3.2 the following can be seen:

**Lemma 4.2.** For each \( x \in \mathbb{R}^d \), as \( t \to \infty \),

(a) \( \int M_i(x, y) dy = \int E[\mathcal{F}(x, y) - f(x, y)] dy \to 0 \), and

(b) \( \int |\mathcal{F}(x, y) - f(x, y)|^2 dy \to 0 \) a.s.

**Proof of Theorem 3.3.** Let \( x \in \mathbb{R}^d \) be such that \( \gamma(x) > \epsilon > 0 \). Then, by Lemma 4.1 (c), \( \gamma(x) > \epsilon/2 \) a.s. for all \( t \) sufficiently large, and therefore, since

\[ \tilde{q}_t(y|x) - q(y|x) = [\gamma(x)\tilde{q}_t(x)]^{-1} \{ \gamma(x)[\tilde{f}_t(x, y) - f(x, y)] + f(x, y)[\gamma(x) - \tilde{q}_t(x)] \}, \]

we obtain

\[ |\tilde{q}_t(y|x) - q(y|x)| \leq 2e^{-2} \{ \gamma(x)[\tilde{f}_t(x, y) - f(x, y)] + f(x, y)[\gamma(x) - \tilde{q}_t(x)] \}. \]

Hence, by the inequality \((a + b)^2 \leq 2(a^2 + b^2)\), the ISE \( J_i(x) \) satisfies

\[ J_i(x) \leq 8e^{-4} \{ \gamma^2(x) \int |\tilde{f}_t(x, y) - f(x, y)|^2 dy + |\gamma_t(x) - \gamma(x)|^2 \int f^2(x, y) dy \} \]

\[ \leq C \{ \int |\tilde{f}_t(x, y) - f(x, y)|^2 dy + |\gamma_t(x) - \gamma(x)|^2 \} \]

for some constant \( C \). Thus, part (b) follows from Lemmas 4.1 (c) and 4.2 (b), and on the other hand, taking expectations, we obtain

\[ EJ_i(x) \leq C \{ \int M_i(x, y) dy + M_i(x) \}, \]

so that part (a) follows from Lemmas 4.1 (b) and 4.2 (a). ■

**REFERENCES**


Nonstationary Markov processes

[4] P. Doukhan, J. León et F. Portal, Une mesure de la déviation quadra-
tique d'estimateurs non paramétriques, Ann. Inst. H. Poincaré Probab. Statis-
[7] P. Hall, Limit theorems for stochastic measures of the accuracy of density estima-

Depto. de Matemáticas
CINVESTAV-IPN
Apdo. Postal 14-740
México, D.F. 07000, Mexico

Received on 6.12.1988