PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 12, Fasc. 1 (1991), pp. 25–33

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 $\boldsymbol{R}^d$-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 $\boldsymbol{R}^d$ and $\boldsymbol{R}^{2d}$, 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)

$$\mathrm{E}(I_t) = \mathrm{E}\int [\hat{\gamma}_t(x)-\gamma(x)]^2 dx = \int M_t(x) dx, \tag{1.1}$$

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

$$I_t := \int [\hat{\gamma}_t(x)-\gamma(x)]^2 dx, \tag{1.2}$$

and $M_t(x)$ is the *Mean Square Error*:

$$M_t(x) := \mathrm{E}[\hat{\gamma}_t(x)-\gamma(x)]^2. \tag{1.3}$$

(Unqualified integrals, as in (1.1) and (1.2), denote integration over all of $\boldsymbol{R}^d$.) 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_1-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 σ-algebra of R^d is denoted by $\mathscr{B}^d$. For a finite signed measure μ, $\|\mu\|$ denotes the variation norm, whereas for a function f on R^d, $\|f\|$ stands for the supremum norm. By convention, $0/0 = 0$.

2. Preliminaries. Let $\{x_t, t = 0, 1, \ldots\}$ be an R^d-valued homogeneous Markov process with transition density $q(y|x)$. Thus, given an arbitrary initial distribution μ_0, the distribution μ_t of x_t, for $t \geqslant 1$, is given by

$$\mu_t(B) = \int Q(B|x)\mu_{t-1}(dx), \qquad B \in \mathscr{B}^d, \tag{2.1}$$

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

ASSUMPTION 2.1. (a) The initial distribution μ_0 is absolutely continuous with a bounded density γ_0.

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

$$\|Q(\cdot|x) - Q(\cdot|y)\| \leqslant 2\alpha \quad \text{for all } x \text{ and } y \text{ in } R^d.$$

(c) For some constant $\bar{q}$, $q(y|x) \leqslant \bar{q}$ for all x and y in 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 μ such that

$$\|\mu_t - \mu\| \leqslant 2\alpha^t, \quad t = 0, 1, \ldots, \tag{2.2}$$

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

$$\gamma_t(y) = \int q(y|x)\gamma_{t-1}(x)dx \quad \text{and} \quad \gamma(y) = \int q(y|x)\gamma(x)dx \tag{2.3}$$

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

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

Finally, the latter inequality and (2.3) yield

$$|\gamma_t(y) - \gamma(y)| < \int q(y|x)|\gamma_{t-1}(x) - \gamma(x)|dx \leqslant 2\bar{q}\alpha^{t-1} \tag{2.4}$$

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

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

Notice also that both γ and γ_t are bounded: from (2.3) we obtain

$$\|\gamma\| \leqslant \bar{q} \quad \text{and} \quad \|\gamma_t\| \leqslant c_0 \text{ for all } t \geqslant 0, \tag{2.5}$$

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

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

LEMMA 2.3. $\sup_{x,y} |f_t(x, y) - f(x, y)| \leqslant \bar{q}\,\|\gamma_t - \gamma\| \to 0$ *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 $\mathbf{R}^d$ such that, for all x, y, x', and y' in $\mathbf{R}^d$,

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

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

It is easily verified that Assumption 2.4 (a) implies that γ and γ_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 $\boldsymbol{R}^d$ and let $\{b_t\}$ be a sequence of positive numbers. The WW estimate $\hat{\gamma}_t$ of γ is defined for $x \in \boldsymbol{R}^d$ and $t \geqslant 1$ by

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

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

$$\hat{f}_t(x, y) := t^{-1} \sum_{n=0}^{t-1} \bar{u}_n(x_n - x, x_{n+1} - y) \tag{3.1}$$

and

$$\hat{q}_t(y|x) := \hat{f}_t(x, y)/\hat{\gamma}_t(x),$$

respectively, where

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

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, $b_0 \leqslant 1$, and it satisfies some of the following conditions as $t \to 0$:

(B1) $b_t \to 0$;

(B2) $tb_t^d \to \infty$; (B2′) $tb_t^{2d} \to \infty$;

(B3) $\sum_t t^{-3/2} b_t^{-d} < \infty$; (B3′) $\sum_t t^{-3/2} b_t^{-2d} < \infty$.

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

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

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

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

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

$$\mathrm{E}(I_t) = \int \mathrm{Var}[\hat{\gamma}_t(x)] dx + \int B_t^2(x) dx. \tag{3.2}$$

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* $\int \mathrm{Var}[\hat{\gamma}_t(x)] dx \to 0$.

(c) *If both* (B1) *and* (B2) *hold, then the MISE* $\mathrm{E}(I_t) \to 0$, *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

$$\bar{I}_t := \int\int [\hat{f}_t(x, y) - f(x, y)]^2 dxdy,$$

and the MISE can be written as

$$\mathrm{E}(\bar{I}_t) := \int\int \bar{M}_t(x, y) dxdy,$$

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

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

with $\bar{B}_t(x, y) := \mathrm{E}\hat{f}_t(x, y) - f(x, y)$, 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* $\int\int \bar{B}_t^2(x, y) dxdy \to 0$.

(b) *If* (B2′) *holds, then* $\int\int \mathrm{Var}[\hat{f}_t(x, y)] dxdy \to 0$.

(c) *If both* (B1) *and* (B2′) *hold, then* $\mathrm{E}(\bar{I}_t) \to 0$, *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

$$f_t^*(x, y) := t^{-1} \sum_{n=0}^{t-1} b_n^{-d} u[(x_n - x)/b_n^{1/2}] u[(x_{n+1} - y)/b_n^{1/2}].$$

Then Theorem 3.2 holds when $\hat{f}_t$ is replaced by f_t^*, and conditions (B2′) and (B3′) are replaced by (B2) and (B3), respectively. However, we decided to use $\hat{f}_t$ (and not 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 $f_t^*(x, y)$, 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 $\hat{q}_t(y|x)$:

$$J_t(x) := \int [\hat{q}_t(y|x) - q(y|x)]^2 dy, \qquad x \in \mathbf{R}^d.$$

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

(a) *If* (B2′) *holds, then* $\mathrm{E}J_t(x) \to 0$, *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.,* $\hat{\gamma}_t(\cdot)$ *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* $x \in \mathbf{R}^d$, *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* $\sup_{x,y}|\bar{B}_t(x, y)| \to 0$.

(b′) *If* (B1) *and* (B2′) *hold, then* $\sup_{x,y} \bar{M}_t(x, y) \to 0$.

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

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

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

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

$$\Gamma_{nm}(x) := \mathrm{Cov}[u_n(x_n - x),\, u_m(x_m - x)]. \tag{4.1}$$

By Ueno's [21] Lemma 3, we have

$$|\Gamma_{nm}(x)| \leqslant \|u\|\, b_m^{-d}\alpha^{m-n}\mathrm{E}u_n(x_n - x) \quad \text{for all } 0 \leqslant n \leqslant m,$$

where

$$\mathrm{E}u_n(x_n - x) = \int u_n(y - x)\gamma_n(y)dy = \int \gamma_n(b_n y + x)u(y)dy,$$

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

$$\int |\Gamma_{nm}(x)|\,dx \leqslant \|u\|\, b_t^{-d}\alpha^{m-n} \quad \text{for } 0 \leqslant n \leqslant m \leqslant t, \tag{4.2}$$

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

$$\begin{aligned} \int \mathrm{Var}[\hat{\gamma}_t(x)]\,dx &\leqslant t^{-2}\sum_{n,m}\int |\Gamma_{nm}(x)|\,dx \\ &\leqslant \|u\|\, t^{-2} b_t^{-d}\sum_{n,m}\alpha^{|m-n|} \leqslant C t^{-1} b_t^{-d} \\ &\qquad\qquad \text{with } C := 2\|u\|/(1-\alpha), \end{aligned} \tag{4.3}$$

which results from the inequality $\sum_{n,m}\alpha^{|m-n|} \leqslant 2t/(1-\alpha)$. Thus condition (B2) implies part (b).

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

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

$$I_t \leqslant 2\int [\hat{\gamma}_t(x) - \mathrm{E}\hat{\gamma}_t(x)]^2 dx + 2\int B_t^2(x)dx.$$

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 $t \to \infty$,

$$Y_t := \int [\hat{\gamma}_t(x) - \mathrm{E}\hat{\gamma}_t(x)]^2 dx \to 0 \text{ a.s.} \tag{4.4}$$

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

$$Y_t \to L \text{ a.s.} \quad \text{for some finite limit } L, \tag{4.5}$$

since, by part (b), $\mathrm{E}(Y_t) = \int \mathrm{Var}[\hat{\gamma}_t(x)]\,dx \to 0$, 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 $\hat{\gamma}_t(x)$,

$$\hat{\gamma}_{t+1}(x) = (t+1)^{-1}[t\hat{\gamma}_t(x) + u_t(x_t - x)]. \tag{4.6}$$

Hence, defining

$$Z_t(x) := \hat{\gamma}_t(x) - \mathrm{E}\hat{\gamma}_t(x) \quad \text{and} \quad U_t(x) := u_t(x_t - x) - \mathrm{E}u_t(x_t - x),$$

we obtain

$$\begin{aligned} Z_{t+1}^2(x) &= Z_t^2(x) + (t+1)^{-2}[U_t^2(x) + 2tU_t(x)Z_t(x) - (2t+1)Z_t^2(x)] \\ &\leqslant Z_t^2(x) + (t+1)^{-2}[U_t^2(x) + 2tU_t(x)Z_t(x)], \end{aligned}$$

whence

$$Y_{t+1} = \int Z_{t+1}^2(x)dx \leqslant Y_t + (t+1)^{-2}\int [U_t^2(x) + 2tU_t(x)Z_t(x)]\,dx.$$

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

$$Y_t' := (t+1)^{-2}\mathrm{E}\{\int [U_t^2(x) + 2tU_t(x)Z_t(x)]\,dx\,|\,\mathscr{F}_t\}.$$

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

$$\sum_t \mathrm{E}|Y_t'| < \infty. \tag{4.7}$$

To prove this, note that $\mathrm{E}\int U_t^2(x)dx = \int \Gamma_{tt}(x)dx < \|u\|\,b_t^{-d}$ by (4.2), whereas by (4.3) and repeated applications of the Schwartz inequality we obtain

$$\begin{aligned} \mathrm{E}\int |U_t(x)|\,|Z_t(x)|\,dx &\leqslant (\int \mathrm{Var}[\hat{\gamma}_t(x)]\,dx)^{1/2} (\int \Gamma_{tt}(x)dx)^{1/2} \\ &\leqslant (Ct^{-1}b_t^{-d})^{1/2}(\|u\|\,b_t^{-d})^{1/2} \leqslant C_1 t^{-1/2} b_t^{-d} \end{aligned}$$

for some constant C_1. Thus, for some constant C_2, $\mathrm{E}|Y_t'| \leqslant C_2 t^{-3/2} b_t^{-d}$, so that (4.7) follows from condition (B3). This completes the proof of (4.5), which, as noted earlier, yields part (d). ∎

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

$$\bar{\Gamma}_{nm}(x, y) := \mathrm{Cov}[\bar{u}_n(x_n - x,\, x_{n+1} - y),\, \bar{u}_m(x_m - x,\, x_{m+1} - y)]$$

can be estimated by using Ueno's [21] Lemma 3 again, to obtain

$$|\bar{\Gamma}_{nm}(x, y)| \leqslant \begin{cases} b_m^{-2d}\|u\|^2\alpha^{m-n-1}\mathrm{E}\bar{u}_n(x_n-x, x_{n+1}-y) & \text{for } n+1 \leqslant m, \\ b_n^{-2d}\|u\|^2\mathrm{E}\bar{u}_n(x_n-x, x_{n+1}-y) & \text{for } n = m. \end{cases}$$

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

$$\int\int|\bar{\Gamma}_{nm}(x, y)|\,dxdy \leqslant \begin{cases} b_t^{-2d}\|u\|^2\alpha^{m-n-1} & \text{for } n+1 \leqslant m \leqslant t, \\ b_t^{-2d}\|u\|^2 & \text{for } n = m \leqslant t. \end{cases}$$

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* $x \in \mathbf{R}^d$, *as* $t \to \infty$,
(a) $\int \bar{M}_t(x, y)dy = \int \mathrm{E}[\hat{f}_t(x, y) - f(x, y)]dy \to 0$, *and*
(b) $\int|\hat{f}_t(x, y) - f(x, y)|^2 dy \to 0$ *a.s.*

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

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

we obtain

$$|\hat{q}_t(y|x) - q(y|x)| \leqslant 2\varepsilon^{-2}\{\gamma(x)|\hat{f}_t(x, y) - f(x, y)| + f(x, y)|\hat{\gamma}_t(x) - \gamma(x)|\}.$$

Hence, by the inequality $(a+b)^2 \leqslant 2(a^2+b^2)$, the ISE $J_t(x)$ satisfies

$$\begin{aligned} J_t(x) &\leqslant 8\varepsilon^{-4}\{\gamma^2(x)\int|\hat{f}_t(x, y) - f(x, y)|^2 dy + |\hat{\gamma}_t(x) - \gamma(x)|^2 \int f^2(x, y)dy\} \\ &\leqslant C\{\int|\hat{f}_t(x, y) - f(x, y)|^2 dy + |\hat{\gamma}_t(x) - \gamma(x)|^2\} \end{aligned}$$

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

$$\mathrm{E}J_t(x) \leqslant C\{\int \bar{M}_t(x, y)dy + M_t(x)\},$$

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

REFERENCES

[1] P. J. Bickel and M. Rosenblatt, *On some global measures of the deviations on density function estimates*, Ann. Statist. 1 (1973), pp. 1071–1095.

[2] R. L. Dobrushin, *Central limit theorem for nonstationary Markov chains, I*, Theory Probab. Appl. 1 (1956), pp. 65–80.

[3] P. Doukhan et M. Ghindès, *Estimation de la transition de probabilité d'une chaîne de Markov Döeblin-récurrente*, Stochastic Process. Appl. 15 (1983), pp. 271–293.

[4] P. Doukhan, J. León et F. Portal, *Une mesure de la déviation quadratique d'estimateurs non paramétriques*, Ann. Inst. H. Poincaré Probab. Statist. 22 (1986), pp. 37–66.
[5] J. P. Georgin, *Contrôle de chaînes de Markov sur des espaces arbitraires*, Ann. Inst. H. Poincaré, Sect. B, 14 (1978), pp. 255–277.
[6] H. Gillert and A. Wartenberg, *Density estimation for nonstationary Markov processes*, Math. Operationsforsch. u. Statist., Ser. Statist. 15 (1984), pp. 263–275.
[7] P. Hall, *Limit theorems for stochastic measures of the accuracy of density estimators*, Stochastic Process. Appl. 13 (1982), pp. 11–25.
[8] O. Hernández-Lerma, *Adaptive Markov Control Processes*, Springer-Verlag, New York 1989.
[9] — *Nonparametric density estimation and adaptive control*, Reporte Interno no. 12, Depto. de Matemáticas, CINVESTAV-IPN, México 1988.
[10] — S. O. Esparza and B. S. Duran, *Recursive nonparametric estimation of nonstationary Markov processes*, Bol. Soc. Mat. Mexicana 33 (1988), pp. 57–69.
[11] D. Ioannides and G. G. Roussas, *Note on the uniform convergence of density estimates for mixing random variables*, Statist. Probab. Lett. 5 (1987), pp. 279–285.
[12] M. Iosifescu, *On two recent papers on ergodicity in nonhomogeneous Markov chains*, Ann. Math. Statist. 43 (1972), pp. 1732–1736.
[13] A. Mokkadem, *Etude des risques des estimateurs à noyaux*, Note aux C.R.A.S., Série I, 301 (1985), pp. 447–450.
[14] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist. 33 (1962), pp. 1065–1076.
[15] B. L. S. Prakasa Rao, *Nonparametric Functional Estimation*, Academic Press, New York 1983.
[16] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist. 27 (1956), pp. 832–837.
[17] — *Conditional probability density and regression estimator*, in: *Multivariate Analysis II* (P. R. Krishnaiah, Ed.), Academic Press, New York 1969.
[18] — *Density estimates and Markov sequences*, in: *Nonparametric Techniques in Statistical Inference* (M. L. Puri, Ed.), Cambridge Univ. Press, Cambridge 1970.
[19] — *A quadratic measure of deviation of two-dimensional density estimates and a test of independence*, Ann. Statist. 3 (1975), pp. 1–14.
[20] G. G. Roussas, *Nonparametric estimation of the transition density function of a Markov process*, Ann. Math. Statist. 40 (1969), pp. 1386–1400.
[21] T. Ueno, *Some limit theorems for temporally discrete Markov processes*, J. Fac. Sci. Univ. Tokyo, Sect. 1, 7 (1957), pp. 449–462.
[22] J. van Ryzin, *On strong consistency of density estimates*, Ann. Math. Statist. 40 (1969), pp. 1765–1772.
[23] C. T. Wolverton and T. J. Wagner, *Recursive estimates of probability densities*, IEEE Trans. Syst. Sci. Cyber. 5 (1969), pp. 246–247.
[24] S. Yakowitz, *Nonparametric density estimation, prediction, and regression for Markov sequences*, J. Amer. Statist. Assoc. 80 (1985), pp. 215–221.
[25] H. Yamato, *Sequential estimation of a continuous probability density function and mode*, Bull. Math. Statist. 14 (1971), pp. 1–12.

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

Received on 6.12.1988