Ann. Inst. Statist. Math.
Vol. 57, No. 3, 443–453 (2005)

KERNEL ESTIMATION FOR STATIONARY DENSITY OF MARKOV CHAINS WITH GENERAL STATE SPACE

VIVIANE S. M. CAMPOS[1]* AND CHANG C. Y. DOREA[2]**

[1] *Depto. de Matematica, Universidade Federal do Rio Grande do Norte, 59072-970 Natal-RN, Brazil,*
e-mail: viviane@ccet.ufrn.br
[2] *Depto. de Matematica, Universidade de Brasilia, 70910-900 Brasilia-DF, Brazil,*
e-mail: cdorea@mat.unb.br

(Received July 22, 2003; revised March 11, 2004)

Abstract. Let $\{X_n\}_{n\geq 0}$ be a Markov chain with stationary distribution $f(x)\nu(dx)$, ν being a σ-finite measure on $E \subset R^d$. Under strict stationarity and mixing conditions we obtain the consistency and asymptotic normality for a general class of kernel estimates of $f(\cdot)$. When the assumption of stationarity is dropped these results are extended to geometrically ergodic chains.

Key words and phrases: Kernel estimator, general state space, mixing condition, geometric ergodicity.

1. Introduction

Kernel density type estimators for real-valued and strictly stationary Markov chains were considered by Roussas (1969, 1991), Rosenblatt (1970) and Athreya and Atuncar (1998). They extended to the Markov chain case the results of kernel density estimates of a sequence of independent and identically distributed (i.i.d.) random variables. For the stationary density $f(\cdot)$ kernel estimators of the type

$$(1.1) \qquad f_n(x) = \frac{1}{nh}\sum_{k=1}^{n} K\left(\frac{x - X_k}{h}\right), \qquad h = h_n \downarrow 0 \quad \text{for } n \to \infty,$$

were studied. Then under regularity conditions on $K(\cdot)$ and h the consistency and asymptotic normality were obtained. Roussas (1969) considered chains satisfying Doob's Condition D_0 with an unique ergodic set and no cyclically moving subsets. Rosenblatt (1970) replaced these assumptions by φ-mixing conditions. Athreya and Atuncar (1998) weakened these conditions by assuming Harris recurrency. Other results about kernel estimators in a Markov chain can be found in Roussas (1991).

We consider a more general setting: a Markov chain $\{X_n\}_{n\geq 0}$ with general state space $E \subset R^d$ and replace the stationary density by the existence of a density $f(\cdot)$ with respect to a σ-finite measure ν on E. Under this setting the estimators in (1.1) can be

*Partially supported by CAPES.
**Partially supported by CNPq, PROCAD/CAPES, PRONEX/FAPDF and FINATEC/UnB.

redefined by,

$$f_n(x) = \frac{1}{n}\sum_{k=1}^{n} W(h, x, X_k), \qquad h = h_n, \tag{1.2}$$

where $W(h, x, \cdot)$ is an appropriate weight function. Note that, in the usual case ν is the Lebesgue measure and we may think of W as K/h with h standing for $\nu((x - h/2, x + h/2))$. Further motivation for the use of (1.2) can be found in Campos and Dorea (2001), where the i.i.d. case was treated.

In Section 2, we gather the basic assumptions and some preliminary results. In Section 3, under strict stationarity and φ-mixing we show the strong consistency and asymptotic normality. And in Section 4, unlike the previous works on this matter, the assumption of strict stationarity is dropped.

2. Preliminaries

Let $\{X_n\}_{n\geq 0}$ be a Markov chain with transition kernel $\{P(x, A) : x \in E, A \in \mathcal{E}\}$ where $E \subset R^d$ and $\mathcal{E}$ is a σ-field of subsets of E. Assume that the chain possesses a stationary density $f(\cdot)$ with respect to the σ-finite measure ν on E, that is,

$$\int_A f(x)\nu(dx) = \int_E P^n(x, A) f(x)\nu(dx), \qquad \forall A \in \mathcal{E},\ \forall n$$

where for $n \geq 2$

$$P^n(x, A) = \int_E P^k(y, A) P^{n-k}(x, dy), \qquad 1 \leq k \leq n-1.$$

In the usual case the asymptotic properties of (1.1) were derived at continuity points of $f(\cdot)$. Here we consider ν-continuity points.

DEFINITION 2.1. For a function g defined on E we say that x is a ν-continuity point of g, in short $x \in C_\nu(g)$, if given $\epsilon > 0$ there exists $\delta > 0$ such that

$$\nu\{y : |y - x| < \delta, |g(x) - g(y)| > \epsilon\} = 0.$$

DEFINITION 2.2. We say that $\{X_n\}_{n\geq 0}$ satisfies φ-mixing condition if for all $A \in \mathcal{F}_0^k$, for all $B \in \mathcal{F}_{k+n}^\infty$, $k \geq 0$ and $n \geq 1$,

$$|P(A\cap B) - P(A)P(B)| \leq \varphi(n)P(A) \quad \text{and} \quad \varphi(n) \downarrow 0 \quad \text{as} \quad n \to \infty, \tag{2.1}$$

where $\mathcal{F}_l^{l+m} = \sigma(X_l, X_{l+1}, \ldots, X_{l+m})$.

DEFINITION 2.3. We say that the chain is geometrically ergodic if there exists a probability π on E and constants $\alpha > 0$ and $0 < \rho < 1$ such that

$$|P^n(x, A) - \pi(A)| \leq \alpha\rho^n, \qquad \forall x \in E,\ \forall A \in \mathcal{E},\ \forall n. \tag{2.2}$$

Note that, in this case, we necessarily have

$$\lim_{n\to\infty} P^n(x, A) = \pi(A) \quad \text{and} \quad \pi(A) = \int_A f(y)\nu(dy)\ \forall x \in E,\ \forall A \in \mathcal{E}.$$

CONDITION 1. For $x \in C_\nu(f)$ and $h = h_n \downarrow 0$ assume that $W(h, x, \cdot)$ are density functions with respect to ν. Moreover, given $\delta > 0$

$$|W_\delta(h,x,y)| \le K_\delta(x) < \infty, \qquad \text{and} \qquad \lim_{h\to 0} W_\delta(h,x,y) = 0. \tag{2.3}$$

where

$$W_\delta(h,x,y) = W(h,x,y)1_{\{z:|z-x|>\delta\}}(y). \tag{2.4}$$

CONDITION 2. Assume that Condition 1 holds and that for $\gamma_n(x) = \nu\{y : |y-x| \le h\}$ we have

$$\lim_{n\to\infty} \gamma_n(x) = \gamma(x) < \infty, \qquad \lim_{n\to\infty} n\gamma_n(x) = \infty \tag{2.5}$$

and

$$\gamma_n(x)W(h,x,y) \le K_1(x) < \infty \qquad \text{for} \quad h \text{ small}. \tag{2.6}$$

The following preliminary results will be needed in our proofs.

LEMMA 2.1. (Campos and Dorea (2001)) *Let $(S, \mathcal{S}, \lambda)$ be a σ-finite measurable space. For $x \in S$ fixed and $h > 0$ assume that $V(h, x, \cdot)$ are real-valued functions satisfying* (2.3) *and such that for h small*

$$\int_S |V(h,x,y)|\lambda(dy) \le K_0(x) < \infty.$$

Then if ψ is an integrable function and $x \in C_\lambda(\psi)$ we have

$$\lim_{h\to 0}\left|\int_S V(h,x,y)\psi(y)\lambda(dy) - \psi(x)\int_S V(h,x,y)\lambda(dy)\right| = 0.$$

LEMMA 2.2. (Roussas and Ioannides (1987)) *Let $\{X_n\}$ be φ-mixing and assume that ξ and η are, respectively, $\mathcal{F}_0^k$-measurable and $\mathcal{F}_{k+n}^\infty$-mensurable random variables such that*

$$E|\xi|^p < \infty, \qquad E|\eta|^q < \infty \qquad \textit{for} \quad p > 1,\ q > 1 \quad \textit{with}\ \frac{1}{p} + \frac{1}{q} = 1.$$

Then the covariance satisfies

$$|\operatorname{cov}(\xi,\eta)| \le 2\varphi^{1/p}(n)(E|\xi|^p)^{1/p}(E|\eta|^q)^{1/q}. \tag{2.7}$$

Moreover, if $|\eta| < M$ a.s. (almost surely) then

$$|E(\eta \mid \mathcal{F}_0^k) - E\eta| \le 2\varphi(n)M \qquad \textit{a.s.} \tag{2.8}$$

LEMMA 2.3. (Devroye (1991)) *Let $\mathcal{G}_0 = \{\emptyset, \Omega\} \subset \mathcal{G}_1 \subset \cdots \subset \mathcal{G}_n$ be a sequence of nested σ-algebras. Let U be a $\mathcal{G}_n$-measurable and integrable random variable, and define the Doob martingale $U_k = E(U \mid \mathcal{G}_k)$. Assume that there exist a $\mathcal{G}_{k-1}$-measurable random variables V_k and constants a_k such that $V_k \le U_k \le V_k + a_k$. Then given $\epsilon > 0$*

$$P(|U - EU| \ge \epsilon) \le 4 \exp\left\{ \frac{-2\epsilon^2}{\sum_{k=1}^n a_k^2} \right\}.$$

3. Estimation under φ-mixing

In this section, we assume that the chain is strictly stationary, that is,

$$P(X_n \in A) = \int_A f(y)\nu(dy), \qquad \forall A \in \mathcal{E},\ n = 0, 1, 2, \ldots$$

and is φ-mixing. Note that from (1.2) we have

$$E(f_n(x)) = \frac{1}{n} \sum_{k=1}^n E(W(h, x, X_k)) = \int_E W(h, x, y) f(y)\nu(dy)$$

and a direct application of Lemma 2.1 gives us the asymptotic unbiasedness.

PROPOSITION 3.1. *If Condition 1 holds then*

$$\lim_{n\to\infty} E(f_n(x)) = \lim_{h\to 0} EW(h, x, X_1) = f(x), \qquad \forall x \in C_\nu(f). \tag{3.1}$$

THEOREM 3.1. (Quadratic mean consistency) *If Condition 2 holds and $\sum_{n=1}^\infty \varphi^{1/2}(n) < \infty$, then*

$$\lim_{n\to\infty} E\{[f_n(x) - f(x)]^2\} = 0, \qquad \forall x \in C_\nu(f). \tag{3.2}$$

PROOF. By (3.1) it is enough to show that $\mathrm{var}(f_n(x)) \to 0$. Using (2.5) this can be accomplished by showing that $n\gamma_n(x)\,\mathrm{var}(f_n(x))$ is bounded. By stationarity we can write,

$$\begin{aligned} n\,\mathrm{var}(f_n(x)) &= \mathrm{var}(W(h, x, X_1)) \\ &\quad + \frac{2}{n} \sum_{k=1}^{n-1} (n-k)\,\mathrm{cov}[W(h, x, X_1), W(h, x, X_{1+k})]. \end{aligned}$$

Next, we show that for some $M < \infty$

$$\limsup_{n\to\infty} \gamma_n(x)\,\mathrm{var}(W(h, x, X_1)) < M \tag{3.3}$$

and

$$\gamma_n(x)\,\mathrm{cov}[W(h, x, X_1), W(h, x, X_{1+k})] \le 2M\varphi^{1/2}(k). \tag{3.4}$$

Then it will follow that,

$$n\gamma_n(x)\operatorname{var}(f_n(x)) \le M + 2M\sum_{k\ge 1}\varphi^{1/2}(k) < \infty.$$

To prove (3.3) note that from (2.5) and (3.1) we have

$$\lim_{n\to\infty}\gamma_n(x)[EW(h,x,X_1)]^2 = \gamma(x)f^2(x). \tag{3.5}$$

Define $V(h,x,y) = \gamma_n(x)W^2(h,x,y)$ then by (2.3), (2.4) and (2.6) we have

$$V_\delta(h,x,y) \le K_1(x)W_\delta(h,x,y) \le K_1(x)K_\delta(x).$$

Also,

$$\lim_{h\to 0} V_\delta(h,x,y) = 0 \quad \text{and} \quad \int_E V(h,x,y)\nu(dy) \le K_1(x) < \infty. \tag{3.6}$$

And from Lemma 2.1,

$$\lim_{n\to\infty}\gamma_n(x)E(W^2(h,x,X_1)) = f(x)\lim_{n\to\infty}\int_E V(h,x,y)\nu(dy) \le f(x)K_1(x). \tag{3.7}$$

To prove (3.4) we make use of Lemma 2.2 by taking $p = q = 2$, $\xi = \gamma_n^{1/2}(x)W(h,x,X_1)$ and $\eta = \gamma_n^{1/2}(x)W(h,x,X_{1+k})$. By (3.6) and stationarity we get (3.4) with $M = K_1(x)$. $\square$

THEOREM 3.2. (Strong consistency) *If Conditon 2 holds, $\sum_{n\ge 1}\varphi(n) < \infty$ and if for all $\beta > 0$*

$$\sum_{n\ge 1}\exp\{-n\beta\gamma_n^2(x)\} < \infty. \tag{3.8}$$

Then

$$\lim_{n\to\infty} f_n(x) = f(x), \qquad \nu\text{-a.s.}, \quad \forall x\in C_\nu(f).$$

PROOF. The proof makes use of some of the ideas from Dorea and Zhao (2002).

(i) From Proposition 3.1 we have $\lim_{n\to\infty}E(f_n(x)) = f(x)$ thus enough to show

$$\lim_{n\to\infty}[f_n(x) - E_f(f_n(x))] = 0, \qquad \nu\text{-a.s.} \tag{3.9}$$

Define the auxiliary functions

$$\psi_k(X_k) = \sum_{j\ge 0}[E(\gamma_n(x)W(h,x,X_{k+j}) \mid \mathcal{F}_1^k) - E(\gamma_n(x)W(h,x,X_{k+j}))]$$

and

$$\psi_{k+1}(X_k) = \sum_{j\ge 0}[E(\gamma_n(x)W(h,x,X_{k+j+1}) \mid \mathcal{F}_1^k) - E(\gamma_n(x)W(h,x,X_{k+j+1}))]$$

where $\mathcal{F}_1^k = \sigma(X_1, \ldots, X_k)$. That they are well-defined follows from (2.8) by taking $\eta = \gamma_n(x)W(h, x, X_{k+l+j})$ and observing that by (2.6) we have $|\eta| \leq K_1(x)$. It follows that

$$|E(\eta \mid \mathcal{F}_1^k) - E\eta| \leq \varphi(l+j)K_1(x)$$

and for $L(x) = 2K_1(x)\sum_{n\geq 1}\varphi(n)$,

$$\psi_k(X_k) \leq L(x) \quad \text{and} \quad \psi_{k+1}(X_k) \leq L(x). \tag{3.10}$$

On the other hand, we can write

$$\psi_k(X_k) - \psi_{k+1}(X_k) = \gamma_n(x)[W(h, x, X_k) - EW(h, x, X_k)]$$

and

$$\begin{aligned} n\gamma_n(x)[f_n(x) - E(f_n(x))] &= \sum_{k=1}^{n} \gamma_n(x)[\psi_k(X_k) - \psi_{k+1}(X_k)] \\ &= \psi_1(X_1) - \psi_{n+1}(X_n) + \sum_{k=2}^{n}[\psi_k(X_k) - \psi_k(X_{k-1})]. \end{aligned}$$

Moreover,

$$\begin{aligned} P(|f_n(x) - E(f_n(x))| \geq \epsilon) \leq\ & P\left(|\psi_1(X_1) - \psi_{n+1}(X_n)| \geq \frac{n\gamma_n(x)\epsilon}{2}\right) \\ &+ P\left(\left|\sum_{k=2}^{n}[\psi_k(X_k) - \psi_k(X_{k-1})]\right| \geq \frac{n\gamma_n(x)\epsilon}{2}\right). \end{aligned} \tag{3.11}$$

(ii) We will show that given $\epsilon > 0$ there exists $\beta(\epsilon) > 0$ such that

$$P(|f_n(x) - Ef_n(x)| \geq \epsilon) \leq 4\exp\{-n\gamma_n^2(x)\beta(\epsilon)\}. \tag{3.12}$$

By (3.8) and application of the Borel-Cantelli Lemma we have the desired convergence (3.9). By (3.10) we have $|\psi_1(X_1) - \psi_{n+1}(X_n)| \leq 2L(x)$ and by (2.5) we have $n\gamma_n(x) \to \infty$. Thus, for n large

$$P\left(|\psi_1(X_1) - \psi_{n+1}(X_n)| \geq \frac{n\gamma_n(x)\epsilon}{2}\right) = 0.$$

To estimate the second part of (3.11) we make use of Lemma 2.3. Let $\mathcal{G}_l = \mathcal{F}_1^l$, $U = \sum_{k=2}^{n}[\psi_k(X_k) - \psi_k(X_{k-1})]$, $U_l = E(U \mid \mathcal{G}_l)$, $V_l = U_{l-1} - 2L(x)$ and $a_l = 4L(x)$. Note that we have for $k > l$,

$$\begin{aligned} E(\psi_k(X_k) \mid \mathcal{G}_l) &= \sum_{j\geq 0}\{E[E(\gamma_n(x)W(h, x, X_{k+j}) \mid \mathcal{G}_k) \mid \mathcal{G}_l] \\ &\qquad - E[E(\gamma_n(x)W(h, x, X_{k+j})) \mid \mathcal{G}_l]\} \\ &= \sum_{j\geq 0}[E(\gamma_n(x)W(h, x, X_{k+j}) \mid \mathcal{G}_l) - E(\gamma_n(x)W(h, x, X_{k+j}))] \\ &= \psi_k(X_l). \end{aligned}$$

Similarly $E(\psi_k(X_{k-1}) \mid \mathcal{G}_l) = \psi_k(X_l)$. It follows that

$$\begin{aligned} U_l &= \sum_{k=2}^{l} [E(\psi_k(X_k) \mid \mathcal{G}_l) - E(\psi_k(X_{k-1}) \mid \mathcal{G}_l)] \\ &\quad + \sum_{k=l+1}^{n} [E(\psi_k(X_k) \mid \mathcal{G}_l) - E(\psi_k(X_{k-1}) \mid \mathcal{G}_l)] \\ &= \sum_{k=2}^{l} [\psi_k(X_k) - \psi_k(X_{k-1})] \end{aligned}$$

and

$$|U_l - U_{l-1}| \leq |\psi_l(X_l)| + |\psi_l(X_{l-1})| \leq 2L(x).$$

Since $EU = 0$ we have

$$P\left(\left| \sum_{k=2}^{n} [\psi_k(X_k) - \psi_k(X_{k-1})] \right| \geq \frac{n\gamma_n(x)\epsilon}{2} \right) \leq 4\exp\{-n\gamma_n^2(x)\beta(\epsilon)\}$$

where $\beta(\epsilon) = \frac{\epsilon^2}{32L^2(x)}$, and (3.12) follows. □

Remark 1. (a) Note that, in the above proof, the assumption of strict stationarity is not used, except where the asymptotic unbiasedness from Proposition 3.1 was needed.

(b) The difficulty in replacing the φ-mixing condition (2.1) by a weaker α-mixing condition

$$|P(A \cap B) - P(A)P(B)| \leq \alpha(n) \quad \text{and} \quad \alpha(n) \downarrow 0 \quad \text{as } n \to \infty,$$

relies on the fact that we can no longer guarantee the inequality (2.8) used in the proof of the strong consistency (cf. Theorem 3.2).

Next, we provide conditions under which the asymptotic normality can be obtained. Assume that the transition kernel possesses a density function $p(x, y)$ with respect to $\nu(dy)$, that is,

$$P(x, A) = \int_A p(x, y)\nu(dy), \qquad \forall x \in E, \ \forall A \in \mathcal{E}. \tag{3.13}$$

It follows that for $n \geq 1$ and $p^{(0)}(x, y) = 1$,

$$P^{(n)}(x, A) = \int_A p^{(n)}(x, y)\nu(dy) \tag{3.14}$$

where

$$p^{(n)}(x, y) = \int_E p^{(n-1)}(x, z)p(z, y)\nu(dz).$$

PROPOSITION 3.2. *Assume that the chain possesses the transition density* (3.13) *and that*

$$\lim_{n\to\infty} \gamma_n(x) \int W^2(h, x, y)\nu(dy) = \tau^2(x) \tag{3.15}$$

exists. Moreover, assume that Condition 2 holds, $\sum_{n\geq 1} n\varphi^{1/2}(n) < \infty$ *and* $(x,x) \in C_{\nu^2}(p^{(n)})$ *for* $n = 1, 2, \ldots$ *Then we have*

$$\lim_{n\to\infty} n\gamma_n(x)\operatorname{var}(f_n(x)) = \sigma^2(x) \tag{3.16}$$

where

$$\sigma^2(x) = f(x)\tau^2(x) - \gamma(x)f^2(x) + 2\gamma(x)\sum_{k=1}^{\infty} p_{1,k+1}(x,x) < \infty. \tag{3.17}$$

PROOF. (i) First, we show that

$$\lim_{n\to\infty} \int_{E^2} W(h,x,u)W(h,x,v)f(u)p^{(k)}(u,v)\nu(du)\nu(dv) = f(x)p^{(k)}(x,x). \tag{3.18}$$

Let $g(u,v) = f(u)p^{(k)}(u,v)$ and $V(h,x,u,v) = W(h,x,u)W(h,x,v)$. Then both g and V are density functions with respect to ν^2. Moreover $(x,x) \in C_{\nu^2}(g)$ and for $V_\delta(h,x,u,v) = V(h,x,u,v)1_{(|(u,v)-(x,x)|>\delta)}(u,v)$ we have

$$V_\delta(h,x,u,v) \leq W_{\delta/\sqrt{2}}(h,x,u) + W_{\delta/\sqrt{2}}(h,x,v) \leq 2K_{\delta/\sqrt{2}}(x) < \infty.$$

Thus the hypotheses of Lemma 2.1 are satisfied and (3.18) follows.

(ii) By stationarity we have

$$\begin{aligned} n\gamma_n(x)\operatorname{var}(f_n(x)) &= \gamma_n(x)\operatorname{var}(W(h,x,X_1)) \\ &\quad + \frac{2}{n}\sum_{k=1}^{n-1}(n-k)\operatorname{cov}\{W(h,x,X_1), W(h,x,X_{1+k})\}. \end{aligned}$$

From (3.5), (3.15) and Lemma 2.1,

$$\lim_{n\to\infty} \gamma_n(x)E(W^2(h,x,X_1)) = f(x)\tau^2(x)$$

and

$$\lim_{n\to\infty} \gamma_n(x)\operatorname{var}(W(h,x,X_1)) = f(x)\tau^2(x) - f^2(x)\gamma(x).$$

From (3.18) we have

$$\lim_{n\to\infty} \sum_{k=1}^{n-1} \gamma_n(x)\operatorname{cov}\{W(h,x,X_1), W(h,x,X_{1+k})\} = \sum_{k\geq 1}\gamma(x)f(x)p^{(k)}(x,x)$$

and from (3.4),

$$\sum_{k=1}^{n-1} k\gamma_n(x)\operatorname{cov}\{W(h,x,X_1), W(h,x,X_{1+k})\} \leq 2M\sum_{k=1}^{n-1} k\varphi^{1/2}(k).$$

Since $\sum_{k=1}^{n-1} k\varphi^{1/2}(k) < \infty$ we obtain (3.17). □

The proof of asymptotic normality uses the same techniques as the proof of the Central Limit Theorem for Markov chains from Doob (1953). The basic difference is the replacement of Doob's condition D_0 by φ-mixing condition. The major difficulty was to compute the limiting variance, which was done in Proposition 3.2. A detailed proof can be found in Campos (2001).

THEOREM 3.3. (Asymptotic normality) *Assume that the hypotheses of Proposition 3.2 are satisfied. Let $f(x) > 0$ and $E\{|\sqrt{\gamma_n(x)}W(h,x,X_1)|^3\} < \infty$. Then for $\sigma^2(x)$ given by* (3.17) *we have*

$$\sqrt{n\gamma_n(x)}[f_n(x) - E(f_n(x))] \xrightarrow{\mathcal{D}} N(0,\sigma^2(x)). \tag{3.19}$$

4. Estimation under geometric ergodicity

In this section, we replace the assumptions of φ-mixing and strict stationarity by the geometric ergodicity (2.2). Let $\mu_0(dy)$ be an arbitrary initial distribution and let $\mu_n(dy)$ denote the distribution of X_n, that is,

$$P_{\mu_0}(X_n \in A) = \int_A \mu_n(dy) \qquad \forall A \in \mathcal{E}$$

where P_{μ_0} indicates that the initial distribution is μ_0. Similarly we will use the notation E_{μ_0} instead of E for the expectation.

PROPOSITION 4.1. *There exists constants $\alpha > 0$ and $0 < \rho < 1$ such that*

$$\sup_{A\in\mathcal{E}} \left| P_{\mu_0}(X_n \in A) - \int_A f(y)\nu(dy) \right| \leq \alpha\rho^n \tag{4.1}$$

and for $A \in \mathcal{F}_0^k$, $B \in \mathcal{F}_{k+n}^\infty$, $\varphi(n) = 2\alpha\rho^n$ we have

$$|P_{\mu_0}(A\cap B) - P_{\mu_0}(A)P_{\mu_0}(B)| \leq \varphi(n)P_{\mu_0}(A). \tag{4.2}$$

PROOF. Using (2.2) we have

$$\begin{aligned}\left| P_{\mu_0}(X_n \in A) - \int_A f(y)\nu(dy) \right| b &= \left| \int_E P^n(x,A)\mu_0(dx) - \int_E\int_A f(y)\nu(dy)\mu_0(dx) \right| \\ &\leq \int_E |P^n(x,A) - \pi(A)|\mu_0(dx) \leq \alpha\rho^n.\end{aligned}$$

As for (4.2) enough to use the fact that

$$\begin{aligned}|E_{\mu_0}\{1_{(X_{k+n}\in B)} \mid \mathcal{F}_0^k\} - P_{\mu_0}(X_{k+n} \in B)| &\leq |P^n(X_k,B) - \pi(B)| \\ &\quad + |P_{\mu_0}(X_{k+n} \in B) - \pi(B)| \\ &\leq \alpha\rho^n + \alpha\rho^{k+n} \leq 2\alpha\rho^n = \varphi(n). \qquad \square\end{aligned}$$

THEOREM 4.1. *Let $\{X_n\}$ satisfying the geometric ergodicity condition* (2.2) *and let μ_0 be any initial distribution. Assume that Condition* 2 *holds then*

$$\lim_{n\to\infty} E_{\mu_0}\{(f_n(x)-f(x))^2\}=0, \qquad \forall x\in C_\nu(f). \tag{4.3}$$

If, in addition, (3.8) *holds then*

$$\lim_{n\to\infty} f_n(x)=f(x), \qquad \nu\text{-a.s.}, \quad \forall x\in C_\nu(f).$$

Moreover, if the hypotheses of Theorem 3.3 *are satisfied we have asymptotic normality* (3.19).

PROOF. (i) First, we show the asymptotic unbiasedness. From Lemma 2.1 we have for the initial stationary distribution $\pi(dx)=f(x)\nu(dx)$

$$\lim_{n\to\infty}\left|\int_E W(h,x,y)f(y)\nu(dy)-f(x)\right| = \lim_{n\to\infty}|E_\pi W(h,x,X_k)-f(x)|=0. \tag{4.4}$$

Since $E_{\mu_0}W(h,x,X_k)=\int_E W(h,x,y)\mu_k(dy)$ we have from (2.6) and (4.1)

$$\left|\int_E W(h,x,y)[\mu_k(dy)-f(y)]\nu(dy)\right| \le \frac{K_1(x)}{\gamma_n(x)}2\alpha\rho^k. \tag{4.5}$$

It follows that,

$$\left|\frac{1}{n}\sum_{k=1}^n[E_{\mu_0}(W(h,x,X_k))-E_\pi(W(h,x,X_k))]\right| \le 2\frac{\alpha K_1(x)}{n\gamma_n(x)}\sum_{k=1}^n\rho^k.$$

By (2.5) and (4.4) we have,

$$\lim_{n\to\infty}E_{\mu_0}(f_n(x))=f(x). \tag{4.6}$$

(ii) To prove the consistency in quadratic mean we proceed as in Theorem 3.1 by showing (3.3) and (3.4) without the assumption of stationarity. From (2.6) and (4.6) we have,

$$\begin{aligned}
\frac{\gamma_n(x)}{n}\sum_{k=1}^n \operatorname{var}_{\mu_0}(W(h,x,X_k)) &= \frac{\gamma_n(x)}{n}\sum_{k=1}^n[E_{\mu_0}(W^2(h,x,X_k))-(E_{\mu_0}(W(h,x,X_k)))^2]\\
&\le \frac{\gamma_n(x)}{n}\sum_{k=1}^n E_{\mu_0}(W^2(h,x,X_k))\\
&\le \frac{K_1(x)}{n}\sum_{k=1}^n E_{\mu_0}(W(h,x,X_k))<\infty.
\end{aligned}$$

Similarly, writing

$$E_{\mu_0}W(h,x,X_{j+l})=\int_{E^2}W(h,x,v)\mu_j(dy)P^l(y,dv)$$

we have for $L_n(j,l) = |\operatorname{cov}_{\mu_0}\{W(h,x,X_j), W(h,x,X_{j+l})\}|$

$$\begin{aligned} L_n(j,l) &= \left| \int_{E^3} W(h,x,u)W(h,x,v)\mu_j(du)\mu_j(dy)[P^l(u,dv) - P^l(y,dv)] \right| \\ &\leq 2\alpha\rho^l \frac{K_1(x)}{\gamma_n(x)} \int_{E^2} W(h,x,u)\mu_j(du)\mu_j(dy) \\ &= 2\alpha\rho^l \frac{K_1(x)}{\gamma_n(x)} E_{\mu_0} W(h,x,X_j). \end{aligned}$$

It follows from (4.6) that

$$\frac{\gamma_n(x)}{n} \sum_{j=1}^{n-1} \sum_{l=1}^{n-j} L_n(j,l) \leq \frac{K_1(x)}{n} \sum_{j=1}^{n-1} \sum_{l=1}^{n-j} 2\alpha\rho^l E_{\mu_0} W(h,x,X_j) < \infty.$$

(iii) By Proposition 4.1 the chain is φ-mixing and $\varphi(n) = 2\alpha\rho^n$. Clearly $\sum_{n\geq 1} \varphi(n) < \infty$ and the strong consistency follows from (i) and Remark 1.

(iv) Since $\sum_{n\geq 1} n\varphi^{1/2}(n) < \infty$ using (4.1) we have for any measurable function φ

$$\lim_{n\to\infty} |P_{\mu_0}(\varphi(X_1,\dots,X_n) \in A) - P_\pi(\varphi(X_1,\dots,X_n) \in A)| = 0.$$

And the asymptotic normality (3.19) follows. □

References

Athreya, K. B. and Atuncar, G. S. (1998). Kernel estimations for real-valued Markov chains, *Sankhya: The Indian Journal of Statistics, Series A*, **60**, 1–17.

Campos, V. S. M. (2001). Asymptotic analysis of kernel type estimators for densities associated with general Markov chains, Ph.D. Thesis, Department of Mathematics, Universidade de Brasilia, Brasilia-DF, Brazil (in Portuguese).

Campos, V. S. M. and Dorea, C. C. Y. (2001). Kernel density estimation: The general case, *Statistics and Probability Letters*, **55**, 173–180.

Devroye, L. (1991). Exponential inequalities in nonparametric estimation, *Nonparametric Functional Estimations and Related Topics* (ed. G. G. Roussas), 31–44, Kluwer Academic Publishers, Dordrecht.

Doob, J. L. (1953). *Stochastic Processes*, John Wiley and Sons, New York.

Dorea, C. C. Y. and Zhao, L. C. (2002). Nonparametric density estimation in hidden Markov models, *Statistical Inference for Stochastic Processes*, **5**, 55–64.

Rosenblatt, M. (1970). Density estimates and Markov sequences, *Nonparametric Techniques in Statistical Inference* (ed. M. Puri), Cambridge University Press, Oxford.

Roussas, G. G. (1969). Nonparametric estimation in Markov processes, *Annals of the Institute of Statistical Mathematics*, **21**, 73–87.

Roussas, G. G. (1991). Estimation of transition distribution function and its quantiles in Markov processes: Strong consistency and asymptotic normality, *Nomparametric Functional Estimation and Related Topics* (ed. G. G. Roussas), 443–462, Kluwer Academic Publishers, Dordrecht.

Roussas, G. G. and Ioannides, D. (1987). Moment inequalities for mixing sequences of random variables, *Stochastic Analysis and Applications*, **5**(1), 61–120.