BULLETIN BRAZILIAN MATHEMATICAL SOCIETY

Bull Braz Math Soc, New Series 33(3), 409-418

Strong consistency of kernel estimators for Markov transition densities*

C. C. Y. Dorea

Abstract. Let $P(x, dy) = t(x, y)\nu(dy)$ be the transition kernel of a Markov chain, where $t(x, y)$ is a density with respect to a σ-finite measure ν on $(E, \mathcal{E})$, with $E \subset R^d$. In this note, we propose a general class of estimates for $t(x, y)$ that are strongly consistent and that extend the classical results for continuous densities on R^d.

Keywords: kernel estimates; transition density.

Mathematical subject classification: 62G05, 62M99.

1 Introduction

For estimating a density function $p(x)$, Rosenblatt (1956) and Parzen (1962) studied a general class of consistent estimators $p_n(x)$ as a kernel weighted average over the empirical distribution $F_n(\cdot)$,

$$p_n(x) = \int_{-\infty}^{+\infty} \frac{1}{h} K(\frac{x-y}{h}) dF_n(y) = \frac{1}{nh} \sum_{k=1}^{n} K(\frac{x - X_k}{h}), \qquad (1)$$

where the kernel $K(\cdot)$ and the window $h = h_n > 0$ are suitably chosen, and $X_1, X_2, \dots, X_n$ are independent random variables with a common density $p(x)$. The d-variate case can be obtained by replacing $1/nh$ by $1/nh^d$, see, for example, Prakasa Rao (1983).

Roussas (1969) extended the use of kernel estimators for real-valued and strictly stationary Markov chains $\{X_n\}_{n\geq 1}$ that possess a continuous transition density $t(x, y)$ and a stationary density $p(\cdot)$,

$$P(x, A) = \int_A t(x, y)dy \text{ and } \int_A p(z)dz = \int_{-\infty}^{+\infty} P(x, A)p(x)dx.$$

Received 2 June 2002.

*Research partially supported by CNPq, CAPES/PROCAD, FINEP/PRONEX-Brazil.

Under above setting there are in the literature several papers. The usual assumptions are: strict stationarity, that is, X_1 has density $p(\cdot)$; and ergodicity conditions or some mixing conditions with decay requirements (see, for example, Roussas (1991)). Under somehow weaker conditions, namely, Harris recurrence, Athreya and Atuncar (1998) also studied this problem.

In this note, we consider the problem of kernel estimates for Markov chains that possess transition densities with respect to a σ-finite measure ν on $(E, \mathcal{E})$ where $E \subset R^d$ and $\mathcal{E}$ is a σ-field of subsets of E. That is, the transition kernel is given by

$$P(X_{k+1} \in A | X_k = x) = P(x, A) = \int_A t(x, y)\nu(dy), \quad \forall\, x \in E, \quad \forall\, A \in \mathcal{E} \tag{2}$$

and

$$P^{m+n}(x, A) = \int_A P^n(y, A) P^m(x, dy).$$

The kernel $K(\cdot)$ of the estimator (1) can be replaced by a family of weight functions $W(h, x, \cdot)$

$$p_n(x) = \frac{1}{n} \sum_{k=1}^{n} W(h, x, X_k), \quad h = h_n \tag{3}$$

and for the transition density $t(x, y)$ we can define the estimators

$$t_n(x, y) = \frac{\displaystyle\sum_{k=1}^{n} W(h, x, X_k) W(h, y, X_{k+1})}{\displaystyle\sum_{j=1}^{n} W(h, x, X_j)}. \tag{4}$$

Our main result, Theorem 1, gives sufficient conditions for the strong consistency of $t_n(x, y)$. Also, it extends the classical results for densities on R^d and for discrete probability transitions with finite state space (cf. Remark 4 (b)).

2 Preliminaries and Main Result

Let $\{X_n\}_{n \geq 1}$ be a Markov chain with transition kernel (2) and let $p(\cdot)$ be a stationary density, that is,

$$\int_A p(x)\nu(dx) = \int_E P^n(y, A) p(y)\nu(dy), \quad \forall\, A \in \mathcal{E}, \quad n = 1, 2, \ldots . \tag{5}$$

Note that if the chain is ergodic, then for

$$P_\infty(A) = \int_A p(y)\nu(dy) \tag{6}$$

we have $P^n(x,\cdot) \to P_\infty(\cdot)$ under some appropriate norm. We shall use the total variation norm $\|\cdot\|$, that is, for a signed measure μ

$$\|\mu\| = \sup_{A\in\mathcal{E}} \mu(A) - \inf_{B\in\mathcal{E}} \mu(B).$$

Definiton 1. We say that the chain is *uniformly ergodic* if there exists a probability P_∞ on E such that

$$\sup_{x\in E} \|P^n(x,\cdot) - P_\infty(\cdot)\| \underset{n\to\infty}{\longrightarrow} 0.$$

Condition 1. (a) The chain is uniformly ergodic with stationary density $p(\cdot)$.

(b) For $h = h_n > 0$ and $\gamma_h(x) = \nu\{y : |y-x| \le h\}$ we have

$$\lim_{n\to\infty} h_n = 0 \quad \text{and} \quad \lim_{n\to\infty} \gamma_h(x) = \gamma(x) < \infty. \tag{7}$$

(c) $W(h,x,\cdot)$ is a density with respect to ν and satisfies : given $\delta > 0$ for $W_\delta(h,x,y) = W(h,x,y)1_{\{z:|z-x|>\delta\}}(y)$ we have

$$\lim_{n\to\infty} W_\delta(h,x,y) = 0 \ \text{ and } \ W_\delta(h,x,y) \le K_\delta(x) < \infty. \tag{8}$$

Moreover, for n large

$$\gamma_h(x)W(h,x,y) \le L(x) < \infty. \tag{9}$$

Remark 1. In the classical case when ν is the Lebesgue measure we have $\gamma_h(x) = h$ and $\gamma(x) = 0$. The weight function is taken to be

$$W(h,x,y) = \frac{1}{h}K(\frac{x-y}{h})$$

where the kernel $K(\cdot)$ is a density function satisfying regularity conditions that includes $\lim_{|z|\to\infty} |z|K(z) = 0$. And this justifies assumptions (8) and (9). Also, to assure consistency of $p_n(x)$ it is further required that x is a continuity point of $p(\cdot)$. This leads us to the following definition.

Definition 2. For a real-valued function g on E we say that x is a ν-*continuity point* of g, or $x \in C_\nu(g)$, if, given $\epsilon > 0$ there exists $\delta > 0$ such that

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

Lemma 1. (Campos and Dorea (2001)). *Let g be an integrable function and $x \in C_\nu(g)$. Assume that Condition 1(b) and 1(c) hold then*

$$\lim_{h \to 0} \int_E W(h, x, y) g(y) \nu(dy) = g(x). \tag{10}$$

Remark 2. (a) If $W(h, x, \cdot)$ is just an integrable function, not necessarily a density, then (10) becomes

$$\lim_{h \to 0} \left| \int_E W(h, x, y) g(y) \nu(dy) - g(x) \int_E W(h, x, y) \nu(dy) \right| = 0. \tag{11}$$

(b) If g is an integrable function on E^d and W satisfies the corresponding hypotheses then we also have (11) with $\nu^d = \nu \times \ldots \times \nu$ in place of ν. In particular, if $W'(h, (x, y), (u, v)) = W(h, x, u) W(h, y, v)$ and $\gamma'_h(x, y) = \nu^2\{(u, v) : |(y, v) - (x, y)| \leq h\}$ then W' is a density with respect to ν^2 and $\lim_{h \to 0} \gamma'_h(x, y) = \gamma(x, y) < \infty$ since

$$\gamma_{\frac{h}{\sqrt{2}}}(x) \gamma_{\frac{h}{\sqrt{2}}}(y) \leq \gamma'_h(x, y) \leq 2 \gamma_h(x) \gamma_h(y).$$

Moreover, for $W'_\delta(h, (x, y), \cdot) = W'(h, (x, y), \cdot) 1_{\{|(u,v)-(x,y)|>\delta\}}(\cdot)$ we have

$$W'_\delta(h, (x, y), (u, v)) \leq W_{\frac{\delta}{\sqrt{2}}}(h, x, u) + W_{\frac{\delta}{\sqrt{2}}}(h, y, v)$$

so that (8) holds. Also, (9) is satisfied since

$$\gamma'_h(x, y) W'(h, (x, y), (u, v)) \leq 2 L(x) L(y).$$

Thus, if $(x, y) \in C_{\nu^2}(g)$ we have

$$\lim_{h \to 0} \int_{E^2} W(h, x, u) W(h, y, v) g(u, v) \nu^2(du dv) = g(x, y). \tag{12}$$

Define

$$\mathcal{F}_k = \sigma(X_1, \ldots, X_k) \quad \text{and} \quad \mathcal{F}_k^\infty = \sigma(X_k, X_{k+1}, \ldots). \tag{13}$$

Lemma 2. *Assume that Condition 1(a) holds and let* η *be a bounded and* $\mathcal{F}_k^\infty$*-measurable function. Then there exist constants* β *and* $0 \le \rho < 1$ *such that*

$$\left| E(\eta|\mathcal{F}_j) - \int \eta \, dP_\infty \right| \le \beta\rho^{k-j}, \quad j = 1, 2, \ldots, k \tag{14}$$

where P_∞ *is defined by (6).*

Remark 3. Take $\eta = 1_A$ with $A \in \mathcal{F}_k^\infty$ then $E(\eta|\mathcal{F}_j) = P^{k-j}(X_j, A)$ and from (14) we have $|P^{k-j}(X_j, A) - P_\infty(A)| \le \beta\rho^{k-j}$. It follows that

$$\|P^n(x, \cdot) - P_\infty(\cdot)\| \le \beta\rho^n. \tag{15}$$

In fact, Theorem 16.0.2 from Meyn and Tweedie (1994) shows that a uniformly ergodic chain converges at a geometric rate (15). Thus Lemma 2 can be proved by applying standard convergence arguments.

Lemma 3. **(Devroye (1991)).** *Let* $\mathcal{G}_0 = \{\emptyset, E\} \subset \mathcal{G}_1 \subset \ldots \subset \mathcal{G}_n$ *be a sequence of nested* σ*-algebras. Let* U *be a* $\mathcal{G}_n$*-measurable and integrable random variable and define the Doob martingale* $U_k = E(U|\mathcal{G}_k)$*. 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* $\epsilon > 0$

$$P\left(|U - EU| \ge \epsilon\right) \le 4\exp\left\{ -\frac{2\epsilon^2}{\sum_{k=1}^n a_k^2} \right\}. \tag{16}$$

Theorem 1. *Let* $x \in C_\nu(p)$ *with* $p(x) > 0$ *and let* $(x, y) \in C_{\nu^2}(t)$*. Assume that*

$$\sum_{n\ge1} \exp\{-n\gamma_h^2(x)\gamma_h^2(y)\alpha\} < \infty, \quad \forall\alpha > 0 \tag{17}$$

and that $W(h, x, \cdot)$ *and* $W(h, y, \cdot)$ *satisfy Condition 1. Then*

$$P(\lim_{n\to\infty} t_n(x, y) = t(x, y)) = 1. \tag{18}$$

Remark 4. (a) In the independent case it is assumed that x is a continuity point of $p(\cdot)$ and that $\sum exp\{-n\, h_n^2\alpha\} < \infty$, and this justifies assumption (16).

(b) In Roussas (1991), for transition densities with respect to the Lebesgue measure, the strong consistency (18) is proved assuming continuity of $p(\cdot)$ and existence of bounded second order derivatives of $t(\cdot, \cdot)$.

3 Proof of the Results

Proof of Theorem 1. (i) Define

$$p_n(x) = \frac{1}{n}\sum_{k=1}^{n} W(h, x, X_k) \text{ and } g_n(x, y) = \frac{1}{n}\sum_{k=1}^{n} W(h, x, X_k)W(h, y, X_{k+1}).$$

To prove (18) enough to show

$$P(\lim_{n\to\infty} p_n(x) = p(x)) = 1 \quad \text{and} \quad P(\lim_{n\to\infty} q_n(x, y) = p(x)t(x, y)) = 1. \tag{19}$$

so that (18) follows.

(ii) First, we show the asymptotic unbiasedness of p_n and q_n. Since $p(\cdot)$ is a stationary density we have $X_1, X_2, \ldots$ identically distributed and by Lemma 1,

$$E(p_n(x)) = \int_E W(h, x, y)p(y)\nu(dy) \underset{n\to\infty}{\longrightarrow} p(x). \tag{20}$$

Similarly, by (2)

$$\begin{aligned} E(q_n(x, y)) &= E(W(h, x, X_1)W(h, y, X_2)) \\ &= \int_{E^2} W(h, x, u)W(h, y, v)p(u)t(u, v)\nu^2(dudv). \end{aligned}$$

Since $x \in C_\nu(p)$ and $(x, y) \in C_{\nu^2}(t)$ we have from (12)

$$\lim_{n\to\infty} E(q_n(x, y)) = p(x)t(x, y). \tag{21}$$

(iii) To prove the first part of (19) enough to show that given $\epsilon > 0$ there exits $\alpha_\epsilon > 0$, independent of n such that

$$P(|p_n(x) - E(p_n(x))|) \le 4\exp\{-n\gamma_h^2(x)\alpha_\epsilon\}. \tag{22}$$

By (20) we have $P(|E(p_n(x)) - p(x)| \ge \epsilon) = 0$ for n large. Since

$$\sum_{n\ge1} \exp(-n\gamma_h^2(x)\alpha_\epsilon) < \infty$$

we have by Borel-Cantelli lemma the desired convergence.

To prove (22), let $\mu_n = \gamma_h(x)E(W(h, x, X_k)$ and for $s \geq 0$ define

$$A_s(X_j) = \sum_{r \geq s} \left[E\left(\gamma_h(x)W(h, x, X_{j+r}) \,|\, \mathcal{F}_j\right) - \mu_n\right].$$

That $A_s(X_j)$ is well-defined follows from (9) and (14)

$$|A_s(X_j)| \leq \sum_{r \geq s} \beta\rho^r = A < \infty. \tag{23}$$

Note that $A_0(X_j) - A_1(X_j) = \gamma_h(x)W(h, x, X_j) - \mu_n$ and that

$$\begin{aligned} U &= \sum_{j=2}^{n} \left[A_0(X_j) - A_1(X_{j-1})\right] \\ &= n\gamma_h(x)\left[p_n(x) - E(p_n(x))\right] - \left[A_0(X_1) - A_1(X_n)\right]. \end{aligned}$$

We will show that U satisfies the hypotheses of Lemma 3 with $\mathcal{G}_k = \mathcal{F}_k$, $V_k = E(U|\mathcal{G}_{k-1}) - 2A$ and $a_k = 4A$. Since $EU = 0$ we have

$$\begin{aligned} &P\big(|p_n(x) - E(p_n(x))| \geq \epsilon\big) = P\big(|U + [A_0(X_1) - A_1(X_n)| \geq n\gamma_h(x)\epsilon\big) \\ &\leq P\left(|A_0(X_1) - A_1(X_n)| \geq \frac{n\gamma_h(x)\epsilon}{2}\right) + P\left(|U - EU| \geq \frac{n\gamma_h(x)\epsilon}{2}\right). \end{aligned}$$

Since $|A_0(\cdot) - A_1(\cdot)|$ is bounded, the first term is 0 for n large. From (16) we have

$$P\left(|U| \geq \frac{n\gamma_h(x)\epsilon}{2}\right) \leq 4\exp\left\{-\frac{n\gamma_h^2(x)\epsilon^2}{8A^2}\right\}$$

and (22) follows. It remains to verify the hypotheses of Lemma 3. Clearly, U is $\mathcal{G}_n$-measurable and V_k is $\mathcal{G}_{k-1}$-measurable. Now, for $k > j$ we have $E(A_s(X_j)|\mathcal{G}_k) = A_s(X_j)$ and for $k \leq j$

$$\begin{aligned} E(A_s(X_j)|\mathcal{G}_k) &= \sum_{r \geq s} E\left\{E(\gamma_h(x)W(h, x, X_{j+r})|\mathcal{F}_j) - \mu_n \,|\, \mathcal{G}_k\right\} \\ &= \sum_{r \geq s} \left[E(\gamma_h(x)W(h, x, X_{j+r} \,|\, \mathcal{F}_k) - \mu_n\right] = A_{j+k+s}(X_k). \end{aligned}$$

Thus, for $2 \le k \le n$

$$\begin{aligned} U_k = E(U|\mathcal{G}_k) &= \sum_{j=2}^{k-1}\big[A_0(X_j) - A_1(X_{j-1})\big] + A_0(X_k) - A_1(X_{k-1}) \\ &\quad + \sum_{j=k+1}^{n}\big[A_{j-k}(X_k) - A_{j-1+k+1(X_k)}\big] \\ &= \sum_{j=2}^{k}\big[A_0(X_j) - A_1(X_{j-1})\big]. \end{aligned}$$

Moreover, by (23)

$$U_{k-1} - 2A \le U_k \le U_{k-1} + 2A$$

and

$$V_k \le U_k \le V_k + 4A.$$

(iv) The proof of the second part of (19) uses the same type of arguments as in (iii). It is enough to show that given $\epsilon > 0$ there exists $\beta_\epsilon > 0$ such that

$$P(|q_n(x,y) - E(q_n(x,y))| \ge \epsilon) \le 4\exp\{-n\gamma_h^2(x)\gamma_h^2(y)\beta_\epsilon\}. \tag{24}$$

Let $\rho_n = \gamma_h(x)\gamma_h(y)E\{W(h,x,X_k)W(h,y,X_{k+1})\}$ and for $s \ge 0$ and $j \ge 1$ define

$$B_s(\mathcal{F}_{j+1}) = \sum_{r\ge s}\big[E(\gamma_h(x)\gamma_h(y)W(h,x,X_{j+r})W(h,y,X_{j+1+r})\,|\,\mathcal{F}_{j+1}) - \rho_n\big].$$

To verify that $B_s(\cdot)$ is well-defined we have for $s < 2$

$$|B_s(\mathcal{F}_{j+1})| \le 2L(x)L(y) + |B_{s+1}(\mathcal{F}_{j+1})|.$$

And using (15) for $s \ge 2$

$$\begin{aligned} B_s(\mathcal{F}_{j+1}) &= \sum_{r\ge s}\int_{E^2}\gamma_h(x)\gamma_h(y)W(h,x,u)W(h,y,v) \\ &\qquad\qquad [P^{r-1}(X_{j+1},du) - P_\infty(du)]P(u,dv) \\ |B_s(\mathcal{F}_{j+1})| &\le L(y)\sum_{r\ge s}\int_E \gamma_h(x)W(h,x,u)|P^{r-1}(X_{j+1},du) - P_\infty(du)| \\ &\le L(x)L(y)\sum_r \beta\rho^r. \end{aligned}$$

Let B such that $|B_s(\mathcal{F}_{j+1})| \leq B < \infty$. Write

$$\begin{aligned} U &= \sum_{j=2}^{n} [B_0(\mathcal{F}_{j+1}) - B_1(\mathcal{F}_j)] \\ &= n\gamma_h(x)\gamma_h(y)[q_n(x, y) - E(q_n(x, y))] - [B_0(\mathcal{F}_2) - B_1(\mathcal{F}_{n+1})]. \end{aligned}$$

For $\mathcal{G}_k = \mathcal{F}_{k+1}$ we have

$$U_k = E(U|\mathcal{G}_k) = \sum_{j=2}^{k} \big[B_0(\mathcal{F}_{j+1}) - B_1(\mathcal{F}_j)\big].$$

And the hypotheses of Lemma 3 are verified by taking $V_k = U_{k-1} - 2B$ and $a_k = 4B$. □

References

[1] Athreya, K.B. and Atuncar, G.S. - *Kernel estimations for real-valued Markov chains*, Sankhya, **60** (1998), 1–17.

[2] Campos, V.S.M. and Dorea, C.C.Y. - *Kernel density estimation: the general case*, Statistics & Probability Letters, **55** (2001), 173–180.

[3] Devroye, L. - *Exponential inequalities in nonparametric estimation*, In: G. G. Roussas (ed), Nonparametric Functional Estimation and Related Topics, Kluwer Ac. Publ., 31–44, (1991).

[4] Meyn, S.P. and Tweedie, R.L. - *Markov Chains and Stochastic Stability*, Springer Verlag, N.Y., (1994).

[5] Parzen, E. - *On estimation of a probability function and its mode*, Annals of Math. Statistics, **33** (1962), 1065–1076.

[6] Prakasa Rao, B.L.S. - *Nonparametric Functional Estimation*, Academic Press, N.Y., (1983).

[7] Rosenblatt, M. - *Remarks on some nonparametric estimates of a density function*, Annals of Math. Statistics, **27** (1956), 832–837.

[8] Roussas, G.G. - *Nonparametric estimation in Markov processes*, Annals of the Inst. of Statistical Math., **21** (1969), 73–87.

[9] Roussas, G.G. - *Estimation of transition distribution function and its quantiles in Markov processes: strong consistency and asymptotic normality*, In: G.G. Roussas (ed), Nonparametric Functional Estimation and Related Topics, Kluwer Ac. Publ., 443–462, (1991).

C. C. Y. Dorea
Departamento de Matemática
Universidade de Brasília
Caixa Postal 04322, 70919-970 Brasília-DF
BRAZIL
E-mail: cdorea@mat.unb.br