COMMUN. STATIST.—THEORY METH., 24(3), 581-591 (1995)

ON L_1-CONSISTENCY OF KERNEL - TYPE DENSITY ESTIMATOR FOR STATIONARY MARKOV PROCESSES

M.N. Mishra and B.L.S. Prakasa Rao

Department of Statistics
Sambalpur University
Sambalpur - 768 019
Orissa, India

Indian Statistical Institute
203, B.T. Road
Calcutta - 700 035
India

Key words and phrases : Recursive estimation; Stationary Markov process; L_1 - consistency; G_2 - condition.

Abstract

A convergence result for kernel type density estimators, proved by Devroye and Gyrofi (1985), is extended to stationary Markov processess satisfying G_2-condition introduced by Rosenblatt (1970).

1. INTRODUCTION

Nonparametric density estimation for independent and identically distributed observations is extensively studied and a comprehensive survey of various methods of density estimation and the properties of estimators is given in Prakasa Rao (1983). One of the main methods that has been extensively used in practice is the kernel type density estimation. Silverman (1985) gives several examples. Since the observations obtained over time are dependent in general, it is of interest to study density estimation in the stochastic processes frame work. Prakasa Rao (1977, 78, 79, 83) discussed generalization of kernel type methods and orthogonal series methods etc. for density estimation to stationary stochastic processes which are Markov or mixing in some sense. More recent work in the area is due to Bradley (1983), Hart (1984), Ioannides and Roussas (1987) and Tran (1989 a,b, 1990). For earlier work and more references, see Prakasa Rao (1983). Yakowitz (1989) discussed nonparametric density and regression estimation for Markov seqauences without mixing assumptions.

It has been observed that the standard kernel type density estimator is not recursive in nature. Acquisition of additional observations necessitate computation of the estimator all over again. In order to avoid this problem, recursive kernel type density estimators were studied for the case of dependent and identically distributed observations. For a detailed survey, see Prakasa Rao (1983), Chapter 5. It turns out that these type of estimators are amenable to analysis in the dependent case and have been found applicability in the recent literature on nonparametric inference for time series analysis. See Prakasa Rao (1994). More work in the area of recursive type density estimation for stationary processes is due to Nguyen (1979, 1981), Bosq (1987), Abdul-Al (1988), Isogai (1989), Tran (1989, 1990), Gyorfi and Masry (1990), and Hernandez-Lerma (1991) among others, Gillert and Wartenburg (1984) studied density estimation for non-stationary Markov processes.

In his study of density estimation for stationary Markov processes, Rosenblatt (1970) introduced the G_2-condition. Density estimation for continuous time stationary Markov processes was discussed in Nguyen (1979) and Prakasa Rao (1979). Chapter IV in Gyorfi et al. (1988) discusses recursive estimation when the stationary stochastic process satisfies a mixing condition.

Our aim in this paper is to extend a result on strong L_1-consistency of recursive kernel type density estimators, obtained by Devroye and Gyorfi (1985) in the i.i.d. case, to the case of stationary Markov process when the Markov process satisfies the Rosenblatt's G_2-condition. Proofs are analogous to those in Devroye and Gyorfi (1985), p. 194.

2. PRELIMINARIES

Suppose $Y_i, 1 \leq i \leq n$ are independent and identically distributed observations with a common density function f. One type of recursive estimator of the density f based on $Y_i, 1 \leq i \leq n$ is of the form

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{x - Y_i}{h_i}\right)$$

where $K(\cdot)$ is a suitable kernel and h_n is a suitable bandwidth sequence with $0 < h_n \to 0$ as $n \to \infty$ (cf. Prakasa Rao (1983)). Deheuvels (1974) proposed a variation of this estimator of the type

$$f_n(x) = \sum_{i=1}^{n} K\left(\frac{x - Y_i}{h_i}\right) / \sum_{i=1}^{n} h_i$$

and studied its properties (cf. Prakasa Rao (1983), p. 314).

Devroye and Gyorfi (1985), p. 194 investigated certain equivalence relations on L_1-convergence of the estimator f_n.

Here we propose to obtain a similar result for stationary Markov processes satisfying the G_2-condition.

Let $\{X_n, n \geq 1\}$ be a stationary process and define the transition operator H_n by

$$(H_n g)(x) = E[g(X_{n+1})|X_1 = x]$$

where g is any bounded measurable function defined on the real line. Define

$$|H_n|_2 = \sup_{[g: Eg(X_1)=0]} E^{1/2}(H_n g)^2 / E^{1/2}(g^2)$$

(cf. Prakasa Rao (1983), p. 322).

Definition 2.1. The transition operator H_n is said to satisfy $G_2(m, \alpha)$ condition of Rosenblatt if there exists a positive integer m such that $|H_m|_2 \leq \alpha$ with $0 < \alpha < 1$.

If $\{X_n\}$ is a stationary Markov process satisfying the condition $G_2(m, \alpha)$, then it can be checked that

$$H_{m+n} = H_m H_n = H_n H_m$$

and for every $n > m \geq 1$,

$$H_n|_2 \leq \beta^n / \alpha \quad \text{where} \quad \beta = \alpha^{1/m} \in (0, 1).$$

It is well known that if a process satisfies G_2-condition, then it is exponentially strong mixing (cf. Rosenblatt (1971)). Moment bounds for strong mixing sequences have been discussed recently in Kim (1993).

3. MAIN RESULT

Let $\{X_n\}$ be a strictly stationary Markov process. Let $f(\cdot)$ be the one-dimensional marginal density of X_1 assuming that it exists. Suppose the process is observed up to "time" n. Then $f(x)$ can be estimated by a recursive estimator of the type

$$f_n(x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_i}\right) / \sum_{i=1}^{n} h_i$$

where $K(\cdot)$ is a bounded density and $\{h_n\}$ is a bandwidth sequence decreasing to zero.

Theorem 3.1 : Suppose the process X_n is a strictly stationary Markov process satisfying the condition $G_2(m, \alpha)$. Further assume that $K(\cdot)$ is a bounded density function satisfying the condition

$$\int_0^\infty \gamma(u)du < \infty \text{ where } \gamma(u) = \sup_{|x| \geq u} K(x), u \geq 0 \tag{3.0}$$

and the sequence $\{h_n\}$ satisfies the condition

$$h_n \downarrow 0 \text{ and } \sum_{i=1}^{n} h_i \simeq n^r \text{ where } 3/4 < r < 1$$

as $n \to \infty$. Then the following statements are equivalent :

(A) $f_n(x) \to f(x)$ almost surely, almost all x, all f ;

(B) $f_n(x) \to f(x)$ in probability, almost all x, some f ;

(C) $\lim_{n \to \infty} \sum_{i=1}^{n} h_i I(h_i > \varepsilon) / \sum_{i=1}^{n} h_i = 0$ for all $\varepsilon > 0$;

(D) $\int_{-\infty}^{\infty} |f_n(x) - f(x)| dx \to 0$ almost surely, all f ;

(E) $\int_{-\infty}^{\infty} |f_n(x) - f(x)| dx \to 0$ in probability, some f.

We first state and prove some lemmas which will be used in the sequel.

Lemma 3.1 : If K is a bounded density function satisfying (3.0) and $h_n \downarrow 0$. then

$$\frac{1}{h_n} E\left[K^p\left(\frac{x - X_1}{h_n}\right)\right] \to f(x) \int_{-\infty}^{\infty} K^p(y)dy \text{ as } n \to \infty$$

for almost all x and all $p > 0$.

Proof : See Devroye and Gyorfi (1985), p. 195.

Lemma 3.2 : Let $V_x(i) = K\left(\frac{x-X_i}{h_n}\right) - E\left[K\left(\frac{x-X_i}{h_n}\right)\right]$, $1 \leq i \leq n$ and

$g(n) = \sum_{i=1}^{n} h_i$. Suppose $g(n) \simeq n^r$ as $n \to \infty$ where $r > 3/4$. Then

$$\frac{1}{g(n)} \sum_{i=1}^{n} V_x(i) \to 0 \text{ as } n \to \infty \text{ almost surely,} \tag{3.1}$$

for almost all x.

Proof : We follow the technique employed by Loeve (1960), p. 487. Let $d^2 \le n \le (d+1)^2$ and

$$W(n) = \frac{1}{g(n)} \sum_{i=1}^{n} V_x(i).$$

Then

$$\begin{aligned} \frac{g(n)}{g(d^2)} W(n) - W(d^2) &= \frac{1}{g(d^2)} \sum_{i=d^2+1}^{n} V_x(i) \\ &= Y(d^2, n) \quad \text{(say)}. \end{aligned}$$

Let

$$\begin{aligned} U(d^2) &= \sup_{d^2 \le n \le (d+1)^2} |Y(d^2, n)| \\ &\le \frac{1}{g(d^2)} \sum_{i=d^2}^{(d+1)^2} |V_x(i)|. \end{aligned}$$

Hence

$$\begin{aligned} E|U(d^2)|^2 &\le \frac{1}{g^2(d^2)} E\left[\sum_{i=d^2}^{(d+1)^2} |V_x(i)|\right]^2 \\ &= \frac{1}{g^2(d^2)} \left\{ \sum_{i=d^2}^{(d+1)^2} E|V_x(i)|^2 \right. \\ &\quad \left. + \sum_{i=d^2}^{(d+1)^2} \sum_{j=d^2}^{(d+1)^2} E|V_x(i) V_x(j)| \right\}. \\ &\le \frac{1}{g^2(d^2)} \left\{ \sum_{i=d^2}^{(d+1)^2} E|V_x(i)|^2 \right. \\ &\quad \left. + \sum_{i=d^2}^{(d+1)^2} \sum_{j=d^2}^{(d+1)^2} (E|V_x(i)|^2 E|V_x(j)|^2)^{1/2} \right\} \\ &= \frac{1}{g^2(d^2)} \left[\sum_{i=d^2}^{(d+1)^2} \{\text{var}(V_x(i))\}^{1/2} \right]^2 \end{aligned} \tag{3.2}$$

Note that

$$\frac{1}{h_i}\operatorname{var}(V_x(i)) \le \frac{1}{h_i}E\left[K^2\left(\frac{x-X_i}{h_i}\right)\right] \tag{3.3}$$

and the term on the right side of (3.3) has a limit as $i \to \infty$ for almost all by Lemma 3.1. Hence there exists a function $L^2(x) < \infty$ a.e. such that

$$\frac{1}{h_i}\operatorname{var}(V_x(i)) \le L^2(x) < \infty \text{ a.e. for all } i \ge 1. \tag{3.4}$$

Here a.e. refers to that the statement might not hold in a set of Lebesgue measure zero. Therefore, it follows from (3.2) and (3.4) that

$$\begin{aligned}
E|U(d^2)|^2 &\le \frac{1}{g^2(d^2)}L^2(x)\left[\sum_{k=d^2}^{(d+1)^2} h_i^{1/2}\right]^2 \\
&\le \frac{1}{g^2(d^2)}L^2(x)\left[\sum_{k=d^2}^{(d+1)^2} h_i\right][(d+1)^2-d^2] \\
&\quad \text{(By Cauchy-Schwartz inequality)} \\
&= \frac{L^2(x)}{g^2(d^2)}[g((d+1)^2)-g(d^2)][(d+1)^2-d^2] \\
&\le c_1(x)\frac{L^2(x)(2d+1)d^{2r-2}(2d+1)}{d^{4r}} \\
&\le c_2(x)\,\frac{L^2(x)}{d^{2r}}
\end{aligned}$$

for some functions $c_1(x)$ and $c_2(x)$ depending on x for almost all x and hence

$$\sum_{d=1}^{\infty} E|U(d^2)|^2 < \infty$$

since $r > \frac{3}{4}$ by hypothesis.

Therefore by Tchebysheff's inequality and Borel-Cantelli lemma, it follows that

$$U(d^2) \to 0 \quad \text{a.s. as} \quad d \to \infty$$

for almost all x. In particular it follows that

$$\frac{g(n)}{g(d^2)}W(n) - W(d^2) \to 0 \text{ a.s. as } d \to \infty. \tag{3.5}$$

Now,

$$\sum_{d=1}^{\infty} E|W(d^2)|^2 = \sum_{d=1}^{\infty}\frac{1}{g^2(d^2)}E|\sum_{i=1}^{d^2}V_x(i)|^2$$

$$= \sum_{d=1}^{\infty} \frac{1}{g^2(d^2)} \left\{ \sum_{i=1}^{d^2} \sum_{j=1}^{d^2} \mathrm{cov}(V_x(i),\ V_x(j)) \right\}.$$

But there exists a function $L_0(x) < \infty$ a.e. such that

$$\mathrm{cov}(V_x(i),\ V_x(j)) \leq \beta^{|i-j|} L_0(x) \text{ for all } i \text{ and } j$$

by computations similar to those described in Prakasa Rao (1983), p. 323-324. since the process $\{X_n\}$ satisfies the $G_2(m, \alpha)$ condition and $K(\cdot)$ is a bounded kernel. Hence there exists a function $L_1(x) < \infty$ a.e. such that

$$\begin{aligned} \sum_{d=1}^{\infty} E|W(d^2)|^2 &\leq \left[\sum_{d=1}^{\infty} \frac{1}{d^{4r}} \left\{ \sum_{i=1}^{d^2} \sum_{j=1}^{d^2} \beta^{|i-j|} \right\} \right] L_1(x) \\ &\leq \left\{ \sum_{d=1}^{\infty} \frac{1}{d^{4r-2}} \right\} L_2(x) \end{aligned}$$

which is finite, provided $4r - 2 > 1$ or $r > 3/4$, for some function $L_2(x) < \infty$ a.e.

It is now easy to see as before that

$$W(d^2) \to 0 \text{ a.s. for almost all } x \text{ as } d \to \infty. \tag{3.6}$$

Relations (3.5) and (3.6) imply that

$$\frac{g(n)}{g(d^2)} W(n) \to 0 \text{ a.s. as } d \to \infty.$$

Since

$$\frac{g(n)}{g(d^2)} \to 1 \text{ as } d \to \infty,$$

it follows that

$$W(n) \to 0 \text{ a.s. for almost all } x \text{ as } n \to \infty.$$

This proves the relation (3.1).

Lemma 3.3 : Let f_n be a density estimator and f be a density on R. If $f_n(x) \to f(x)$ in probability (almost surely) as $n \to \infty$ for almost all x, then $\int_{-\infty}^{\infty} |f_n(x) - f(x)| dx \to 0$ in probability (almost surely) as $n \to \infty$.

Proof : See Glick (1974).

Proof of Theorem 3.1 : Obviously (A) $\Rightarrow$ (B). Lemma 3.3 shows that (B) $\Rightarrow$ (E). Hence (A) $\Rightarrow$ (B) $\Rightarrow$ (E). It follows again that (A) $\Rightarrow$ (D) $\Rightarrow$ (E). It is sufficient to prove that

$$(C) \Rightarrow (A) \text{ and } (E) \Rightarrow (C).$$

Assume that (C) holds. Let $\varepsilon > 0$. Define

$$f_n^*(x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_i}\right) I_{(h_i \le \varepsilon)} / \sum_{i=1}^{n} h_i.$$

Then, it follows by (C) that

$$|f_n(x) - f_n^*(x)| \le \frac{M \sum_{i=1}^{n} I_{(h_i > \varepsilon)}}{\sum_{i=1}^{n} h_i} = o(1)$$

where M is a bound on the kernal $K(\cdot)$. Note that

$$\begin{aligned}
&|f_n^*(x) - f(x)| \\
\le\ & \frac{\left|\sum_{i=1}^{n} h_i \left[h_i^{-1} K\left(\frac{x - X_i}{h_i}\right) - E\left(h_i^{-1} K\left(\frac{x - X_i}{h_i}\right)\right)\right] I_{(h_i \le \varepsilon)}\right|}{\sum_{i=1}^{n} h_i} \\
&+ \frac{\left|\sum_{i=1}^{n} h_i \left|\left[E(h_i^{-1} K\left(\frac{x - X_i}{h_i}\right)\right) - f(x)\right]\right| I_{(h_i \le \varepsilon)}}{\sum_{i=1}^{n} h_i} \\
&+ \sum_{i=1}^{n} h_i f(x) I_{(h_i > \varepsilon)} / \sum_{i=1}^{n} h_i. \\
=\ & T_1 + T_2 + T_3 \quad \text{(say)}.
\end{aligned}$$

Note that

$$T_2 \le \sup_{|h_i| < \varepsilon} \left| E\left(h_i^{-1} K\left(\frac{x - X_1}{h_i}\right)\right) - f(x)\right|$$

and Lemma 3.1 implies that $T_2 \to 0$ as $n \to \infty$ for sufficiently small ε for almost all x. Condition (C) implies that $T_3 \to 0$ as $n \to \infty$. It is sufficient to prove taht $T_1 \to 0$ a.s. for almost all x to conclude that (C) $\Rightarrow$ (A). Note that

$$T_1 = \frac{\left|\sum_{i=1}^{n} V_x(i) I(h_i \le \varepsilon)\right|}{\sum_{i=1}^{n} h_i}$$

and the last term tends to zero a.s. by Lemma 3.2. This proves (C) $\Rightarrow$ (A).

We complete the proof by showing that (E) $\Rightarrow$ (C). The condition (C) is a consequence of the arguments similar to those in Devroye and Gyorfi (1985). p. 198 by noting that the characteristic function of $E(f_n)$ is

$$\varphi_n(t) = \frac{\sum_{i=1}^{n} h_i \varphi(t)\beta(h_i t)}{\sum_{i=1}^{n} h_i}, \quad t \in R$$

where $\varphi(t)$ is the characteristic function of the marginal density f and $\beta(t)$ is the characteristic function of the kernel $K(\cdot)$. We omit the details.

ACKNOWLEDGEMENT

The authors thank the referee for his comments and suggestions.

BIBLIOGRAPHY

Abdul-Al, K.I. (1988). "On recursive probability density estimation for stationary process", *Publications de ℓ' Institut de Statistique de ℓ'* Universite de Paris, **33**, 3-24.

Bosq, D. (1987), "La statistique nonparametrique des processus". *Rend. Sem. Mat. Univers. Politech. Torino* **45**, 1-24.

Bradley, R.C. (1983). "Asymptotic normality of some kernel-type estimators of probability density", *Statistics and Probability Letters* **1**, 295-300.

Deheuvels, P. (1974). "Estimation sequentielle de la densite", Ph.D. Thesis - Universite de Paris - VI.

Devroye, L. and Gyorfi, L. (1985). *Nonparametric Density Estimation, The L_1 view.* John Wiley, New York.

Gillert, H. and Wartenburg, A. (1984). "Density estimation for nonstationary Markov processess", *Math. Operation and Statistik, Ser. Statistics* **15**, 263-275.

Glick, N. (1974). "Consistency conditions for probability estimators and integrals of density estimators", *Utilitas Math.* **6**, 61-74.

Gyorfi, L., Hardle, W., Sarda, P. and Vieu, P. (1988). *Nonparametric Curve Estimation from Time Series*, Lecture Notes in Statistics # 60, Springer-Verlag, New York.

Gyorfi, L. and Masry, E. (1990). "The L_1 and L_2 strong consistency of recursive kernel density estimates from dependent samples", *IEEE Trans. Information Theory* Vol. IT-36, 531-539.

Hart, J.D. (1984). "Efficiency of a kernel density estimator under an autoregressive dependence model", *J. Amer. Stat. Assoc.* **79**, 110-117.

Hernandez-Lerma, O. (1991). "On integrated square errors of recursive nonparametric estiamtes of nonstationary Markov processes", *Probability and Mathematical Statistics*, **12**, 25-33.

Ioannides, D. and Roussas, G.G. (1987). "Note on the uniform convergence of density estimates for mixing random variables", *Statistics and Probability Letters*, **5**, 279-285.

Isogai, E. (1989). "Nonparametric recursive estimation in stationary Markov processes", *Communications in Statistics – Theory and Methods*, **18**, 1309-1323.

Kim, Tae Yoon (1993). "A note on moment bounds for strong mixing sequences", *Statistics and Probability Letters*, **16**, 163-168.

Loeve, M. (1960). *Probability Theory*, Van Nostrand, Boston.

Nguyen, H.T. (1979). "Density estimation in a continuous time stationary Markov process", *Ann. Statistics*, **7**, 341-348.

Nguyen, H.T. (1981). "Asymptotic normality of recursive density estimators in Markov processes", *Publications de l' Institut de Statistique de l' Universite de Paris*, **26**, 73-93.

Prakasa Rao, B.L.S. (1977). "Berry-Esseen bound for density estimators of stationary Markov processes", *Bull. Math. Statisti.* **17**, 15-21.

Prakasa Rao, B.L.S. (1978), "Density estimation for Markov processes using delta sequences", *Ann. Inst. Statist. Math.* **30**, 321-328.

Prakasa Rao, B.L.S. (1979). "Nonparametric estimation for continuous time Markov processes via delta families", *Publication de l' Institut de Statistique de l' Universite de Paris* **24**, 79-97.

Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation*, Academic Press, Orlando.

Prakasa Rao, B.L.S. (1994). "Nonparametric approach to time series analysis", *In Stochastic Processes and Statistical Inference* (Ed. B.L.S. Prakasa Rao and B.R. Bhat) Wiley Eastern, New Delhi (To appear).

Rosenblatt, M. (1970). "Density estimates and Markov sequences", In *Nonparametric Techniques in Statistical Inference* (Ed.M.L. Puri), Cambridge Univ. Press, 199-213.

Rosenblatt, M. (1971). *Markov Processes, Structure and Asymptotic Behaviour*, Springer-Verlag, Berlin.

Silverman, B.W. (1986). *Density Estimation and Data Analysis*, Chapman and Hall, London.

Tran. L.T. (1989a). "Recursive density estimation under dependence", *IEEE Transaction on Information Theory*, Vol. IT-35, 1103-1108.

Tran. L.T. (1989b). "The L_1 convergence of kernel density estimates under dependence", *Canadian Journal of Statistics*, **17**, 197-208.

Tran. L.T. (1990). "Kernel density estimation under dependence", *Statistics and Probability Letters*, **10**, 193-201.

Yakowitz, S. (1989). "Nonparametric density and regression estimation for Markov sequences without mixing assumptions", *J. Multivariate Analysis*, **30**, 124-136.

Received January, 1994; Revised October, 1994.