

ON BERRY-ESSEEN THEOREM FOR NONPARAMETRIC DENSITY ESTIMATION IN MARKOV SEQUENCES

Basu, A. K.
Calcutta University

Sahoo, D. K.
Calcutta University

<https://doi.org/10.5109/13467>

出版情報 : Bulletin of informatics and cybernetics. 30 (1), pp.25-39, 1998-03. Research
Association of Statistical Sciences

バージョン :

権利関係 :

ON BERRY-ESSÉEN THEOREM FOR NONPARAMETRIC DENSITY ESTIMATION IN MARKOV SEQUENCES

By

A.K. BASU * and D.K. SAHOO †

Abstract

For a stationary sequence $\{X_i\}$ the Markov assumption G_2 , which is weaker than the Doeblin's condition D_0 , is discussed and is used to estimate nonparametric density and transition density. Under the G_2 -assumptions, the rate of convergence to normality of the estimated density is derived. Similar type of results are also derived for estimating the joint density and the estimated transition density.

Key words : Nonparametric density estimation, Berry-Esséen theorem, G_2 -Markov sequence.

AMS Subject Classification No. : Primary - 62G07, Secondary - 60F05.

1. Introduction

Non parametric estimators of the density function of a population based on a sample of independent observations have been considered by several authors. An excellent survey of the results in this area is given by Rosenblatt (1971). Some authors also consider the estimates of the density function when the observations are dependent in the sense that they are sampled from a stationary Markov sequence. Roussas (1967, 1969) considers density estimation from Markov processes and obtained consistency and asymptotic normality of Kernel type density estimators, under the Doeblin's condition D_0 as given in Doob (1953).

In this paper we consider Markov assumption G_2 which is discussed and employed by Rosenblatt (1970) and Yakowitz (1985, 1989) and others.

A stationary Markov sequence $\{X_i\}$ is a G_2 -sequence if there is a positive number $\rho < 1$ and a positive integer n such that for any bounded Borel function $h(\cdot)$ satisfying $E[h(X_i)] = 0$, we have

$$E_{X_1}[E(h^2(X_n)|X_1)] \leq \rho^2 E[h^2(X_1)]. \quad (1.1)$$

This assumption is weaker than D_0 . In fact Rosenblatt (1970) considered an example where D_0 is not satisfied but G_2 holds. He also considered Kernel type density estimators

* Calcutta University

† R.K. Mission College, Narendrapur

when the observations are sampled from a stationary Markov process and observed that these estimators have the same character as in the case of independence. He proved the asymptotic normality of the Kernel type estimators but does not find out the rate of convergence. In this paper our aim is to obtain Berry-Esséen type bound of the Kernel type density estimator under the Markov assumption G_2 . Yakowitz and his colleagues (1985) while analysing river flow data found that non-parametric model suits more than the popular autoregressive-moving-average (ARMA) model. He also found that under some restrictions ARMA model can be transformed into two-dimensional G_2 -Markov sequence. In fact it is clear from the discussion of Rosenblatt (1970) and Yakowitz (1985) that Markov assumption does not play significant role but stationarity and G_2 -condition make the proofs of our results valid and moreover the conclusion of our results is true if the first component of a two-dimensional Markov sequence satisfies G_2 -condition. Under G_2 -assumption we also obtain rate of convergence to normality for estimated joint density and transition density. In this connection it is to be noted that Prakasa Rao (1977) derived the Berry-Esséen bound for estimating only the density under stronger Doeblin's condition. Berry-Esséen bounds are also obtained by Basu and Sahoo (1988, 1989) for density estimates under different conditions and set up. In Section 2 we state the Markov assumption and consider the method for estimating the unknown stationary density. Using the methodology of Section 3 asymptotic normality together with rate of convergence of the Section 3 and Section 4 are used to prove the corresponding results for the estimate of the transition density in Section 5. Last of all some concluding remarks are added to Section 6. Sometimes we use the Vinogradov symbol \ll to indicate an inequality containing some unspecified positive constant factor.

2. The Markov Assumption and the Estimation of the density

Let $\{X_i\}$ be a stationary Markov sequence with a continuous stationary density function $f(x) > 0$ and a continuous transition p.d.f. $f(y/x)$. Let $f_{j-i}(y, x)$ and $f_{j-i}(y|x)$ denote the joint probability density of X_j and X_i and the conditional probability density of X_j given X_i respectively, for $i \neq j$.

Obviously then,

$$f_{j-i}(y, x) = f_{j-i}(y|x).f(x). \quad (2.1)$$

Define

$$\hat{f}_n(x) = (nh)^{-1} \sum_{j=1}^n K((x - X_j)/h) \quad (2.2)$$

as the estimate of $f(x)$ where

$$i) h = h(n) > 0 \text{ are such that } h(n) \rightarrow 0 \text{ and } nh(n) \rightarrow \infty \text{ as } n \rightarrow \infty \quad (2.3)$$

ii) $K(\cdot)$ is a given bounded continuous density function such that

$$K(-u) = K(u), \int uK(u)du = 0, \int u^i k^j(u)du < \infty \quad (2.4)$$

for $i = 0, 1, 2; j = 1, 2, 3,$

and iii) $f(x)$ is continuously differentiable upto 2nd order. (2.5)

We note that $\hat{f}_n(x)$ is the same sort of estimate of $f(x)$ as in the case of independent observations with

$$V(\hat{f}_n(x)) = (nh)^{-1}f(x) \int K^2(x)dx + O\left(\frac{1}{n}\right). \quad (\text{by computation and also see Rosenblatt (1970)}). \quad (2.6)$$

3. Speed of Convergence to Asymptotic normality of the estimated density

Consider $W_n = (nh)^{1/2}[\hat{f}_n(x) - Ef_n(x)]/[f(x) \int K^2(x)dx]^{1/2}$

$$= \sum_{j=1}^n (nh)^{-1/2} [K((x - X_j)/h) - EK((x - X_j)/h)] / (f(x) \int K^2(x)dx)^{1/2}. \quad (3.1)$$

The asymptotic normality and its speed of convergence is obtained by writing the sum as a sum of big blocks separated by small blocks where the big blocks are approximately independent and then applying a well known lemma for the sum of two sequences of random variables which is stated below :

LEMMA 3.1. *Let $W_n = Z_n + Y_n, n = 1, 2, \dots$ where $\{W_n\}_{n \geq 1}$ and $\{Z_n\}_{n \geq 1}$ are sequences of random variables. Let ψ be a distribution function with a bounded derivative. Suppose $\{b_n\}_{n \geq 1}$ is a sequence of positive constants satisfying*

$$\sup_t |P[Z_n \leq t] - \psi(t)| = O(b_n), n \rightarrow \infty. \quad (3.2)$$

Then for any sequence of positive constants $\{a_n\}_{n \geq 1}$ we have

$$\sup_t |P[W_n \leq t] - \psi(t)| = O(b_n) + O(a_n) + P[|Y_n| > a_n], n \rightarrow \infty. \quad (3.3)$$

Write

$$W_n = \sum_1^k (A_l + B_l) + H \quad (3.4)$$

where

$$\begin{aligned} A_l &= \sum_{j=(l-1)(m+r)+1}^{lm+(l-1)r} T_{nj} / (f(x) \int K^2(x)dx)^{1/2} \\ B_l &= \sum_{j=lm+(l-1)r+1}^{l(m+r)} T_{nj} / (f(x) \int K^2(x)dx)^{1/2} \\ H &= \sum_{j=k(m+r)+1}^n T_{nj} / (f(x) \int K^2(x)dx)^{1/2} \end{aligned} \quad (3.5)$$

and $T_{nj} = (nh)^{-1/2} [K((x - X_j)/h) - E\{K((x - X_j)/h)\}]$.

Here A_l and B_l are big and small blocks respectively with $m = m(n) \rightarrow \infty$ and $r = r(n) \rightarrow \infty$ as $n \rightarrow \infty$ but $m(n) = o(n)$ and $r(n) = o(m(n))$.

Further $k = k(n) = [n/(m+r)]$, the greatest integer containing $n/(m+r)$, tends to infinity as $n \rightarrow \infty$, since $m, r = o(n)$. Here H accounts for the additional few terms at the end not included in big blocks or small blocks.

Assumption : We select m, h, k and r such that

$$\begin{aligned} m(n) &= m = O(n^{7/10}) \\ h(n) &= h = O(n^{-1/5}) \\ k(n) &= k = O(n^{3/10}) \\ r(n) &= r < O(n^{1/5}) \text{ such that } r = o(m). \end{aligned} \quad (3.6)$$

We write W_n as $W_n = Z_n + Y_n$ where $Z_n = \sum_1^k A_l$ and $Y_n = \sum_1^k B_l + H$.

Here A_l are treated as independent variables with the same marginal distributions. Also $E(A_l) = 0$ and $S_n^2 = V(\sum_1^k A_l) = o(nh)$ (see Rosenblatt (1970)). Now we have the following Theorem.

THEOREM 3.1. *Let Φ be the distribution function of the univariate standard normal distribution then under the assumption (3.6)*

$$\sup_t |P[Z_n \leq t] - \Phi(t)| = O(n^{-1/5}). \quad (3.7)$$

PROOF. Under (3.6)

$$S_n^2 = V(\sum_1^k A_l) \ll n^{4/5}. \quad (3.8)$$

Now by repeated application of C_r -inequality we have,

$$\begin{aligned} E|A_l|^3 &= (nh)^{-3/2} \{f(x) \int K^2(x) dx\}^{-3/2} E \left| \sum_{j=(l-1)(m+r)+1}^{lm+(l-1)r} [K((x-X_j)/h) \right. \\ &\quad \left. - EK((x-X_j)/h)]^3 \right| \ll m^2 (nh)^{-3/2} \sum_j [EK^3((x-X_j)/h) \\ &\quad + E^3 K((x-X_j)/h)] \\ &= m^2 (nh)^{-3/2} \sum_j [h \int_{-\infty}^{\infty} K^3(u) f(x-hu) du + \{h \int_{-\infty}^{\infty} K(u) f(x-hu) du\}^3] \\ &= m^2 (nh)^{-3/2} \sum_j [O(h^3) + hf(x) \int_{-\infty}^{\infty} K^3(u) du + \{hf(x) + O(h^3)\}^3] \\ &\quad \text{by Taylor's expansion and using (2.3), (2.4) and (2.5)} \\ &= O(m^3 (n^3 h)^{-1/2}). \end{aligned}$$

This implies

$$\begin{aligned} \sum_{l=1}^k E|A_l|^3 &= O(km^3 n^{-3/2} h^{-1/2}) \\ &= O(n) \text{ under (3.6)}. \end{aligned} \quad (3.9)$$

Hence by Berry-Esséen theorem for independent random variables we get

$$\sup_t |P[\sum_1^k A_l \leq t] - \Phi(t)| \leq \text{const.} \sum_{l=1}^k E|A_l|^3 / s_n^3$$

Therefore $\sup_t |P[\sum_1^k A_l \leq t] - \Phi(t)| = O(n^{-1/5})$ using (3.8) and (3.9) i.e.

$$\sup_t |P[Z_n \leq t] - \Phi(t)| = O(n^{-1/5}).$$

Hence the Theorem 3.1 is proved.

In order to get an estimate of $\sup_x |P[\sum_1^k A_l \leq x] - \Phi(x)|$ we have to consider the characteristic function of $\sum_1^k A_l$ and to compare this characteristic function with the product of the characteristic functions of the individual A_l 's. Following Rosenblatt (1970, p. 207) we find that

$$|E\{\exp(it \sum_1^k A_l)\} - \prod_{l=1}^k E\{\exp(it A_l)\}| \leq (k-1)M\rho^{r(n)} \rightarrow 0 \quad (3.10)$$

where M is a constant and $0 < \rho < 1$ and also $k(n)\rho^{r(n)} \rightarrow 0$.

Now from basic inequality (Loeve) and Esséen's lemma we have,

$$|P[\sum_1^k A_l \leq x] - \Phi(x)| \leq |P[\sum_1^k A_l \leq x] - G(x)| + |G(x) - \Phi(x)| \quad (3.11)$$

$$\leq \frac{2}{\pi} \int_0^u t^{-1} |E[e^{it \sum_1^k A_l}] - \prod_{l=1}^k E[e^{it A_l}]| dt + \frac{24}{\pi u} \sup_x G'(x) + O(n^{-1/5})$$

where $u > 0$ and $G(x)$ is the distribution function of $\sum_1^k A_l$ when A_l are treated as independent but with same marginal distributions as defined in (3.5).

$$\leq \frac{2}{\pi} (k-1)M\rho^{r(n)} \log |u| + \frac{24}{\pi u} \sup_x G'(x) + O(n^{-1/5}).$$

For a given $0 < \rho < 1$, selecting a suitable constant satisfying $r = 0(\log n)$ and selecting $u = O(n^{1/5})$ and assuming $\sum_1^k A_l$ has finite density at x ,

$$\begin{aligned} |P[\sum_1^k A_l \leq x] - \Phi(x)| &= O(n^{-1/5} \log n) + O(n^{-1/5}) \\ &= O(n^{-1/5} \log n). \end{aligned} \quad (3.12)$$

We can represent Y_n in the following form -

$$\begin{aligned}
 Y_n &= \sum_{l=1}^k B_l + H \\
 &= \frac{\sum_{j=m+1}^{m+r} T_{nj} + \sum_{j=2m+r+1}^{2(m+r)} T_{nj} + \dots + \sum_{j=km+(k-1)r+1}^{k(m+r)} T_{nj} + \sum_{j=k(m+r)+1}^n T_{nj}}{(f(x) \int K^2(x) dx)^{1/2}} \\
 &= \left[\sum_{j \leq k+1} \chi_{nj} \right] / (f(x) \int K^2(x) dx)^{1/2} \quad \text{where } \chi_{n1} = T_{nm+1} + \dots + T_{nm+r} \\
 \chi_{nk} &= T_{n \ k m + k - 1 r + 1} + \dots + T_{n \ k(m+r)} \\
 \chi_{nk+1} &= T_{n \ k(m+r)+1} + \dots + T_{n \ n}.
 \end{aligned} \tag{3.13}$$

under this representation and also under assumption (3.6) we have the following theorem:

THEOREM 3.2.

$$P[|Y_n| \geq \varepsilon] = O(n^{-1}) \tag{3.14}$$

where $\varepsilon = O(n^{-1/4}(\log n)^3)$.

PROOF. Since $\|\chi_{nj}\|_\infty = O(r(n)n^{-2/5}) = c_n$, we have for $i \leq j \leq k+1$ and $O < t < c_n^{-1}$, following Philipp (1969)

$$E(e^{t\chi_{nj}}) < e^{t^2 c_n^2}. \tag{3.15}$$

Also following (4.1) of Rosenblatt (1970)

$$E|e^{tY_n} - \prod_j E(e^{t\chi_{nj}})| \leq kM\rho^r(n) \tag{3.16}$$

where M is a constant and $O < \rho < 1$.

Hence it follows from (3.6) and (3.15) that

$$E(e^{tY_n}) << \prod_{j \leq k+1} E(e^{t\chi_{nj}}) << e^{t^2 c_n^2 (k+1)}. \tag{3.17}$$

To conclude the proof we state the following well known result (without proof).

RESULT 3.1. *If $\{V_n\}$ is a sequence of random variables such that $E(e^{tV_n})$ exists for some $t > 0$ then*

$$P[|V_n| \geq \varepsilon] \leq E(e^{tV_n})/e^{t\varepsilon} \quad \text{for all } \varepsilon > 0. \tag{3.18}$$

Now setting $t = \text{const. } n^{1/4} \log n$, $\varepsilon = \text{const. } n^{-1/4}(\log n)^3$ and using (3.6) and (3.17) on the Result 3.1 we get the R.H.S. of (3.14) by choosing $r(n) = O(\log n)$ so that $r(n) < O(n^{1/5})$ is satisfied.

Hence the Theorem 3.2.

THEOREM 3.3 RATE OF CONVERGENCE TO NORMALITY. *If Φ be the distribution function of the univariate standard normal distribution and x be a point of continuity of f then under assumption (3.6) with $r(n) = O(\log n)$*

$$\sup_t |P\{(\hat{f}_n(x) - E\hat{f}_n(x))/(\sqrt{V}[\hat{f}_n(x)]) \leq t\} - \Phi(t)| = O(n^{-1/5}). \quad (3.19)$$

PROOF. Note that, by (2.6) for $0 < u \leq O(n^{3/5})$, $P(|\sqrt{nh}V(\hat{f}_n(x))/f(x) \int K^2(x)dx - 1| > u) = 0$ as $n \rightarrow \infty$. Selecting $b_n = n^{-1/5}$, $a_n = \varepsilon = O(n^{-1/4}(\log n)^3)$, taking ψ to be Φ and lastly applying the result of Theorem 3.1 and Theorem 3.2 on lemma 3.1 and on the well known result of Michel and Pfanzagl (1971) we have,

$$\sup_t |P[W_n \leq t] - \Phi(t)| = O(n^{-1/5}) + O(n^{-1/4}(\log n)^3) + O(n^{-1})$$

Therefore, $\sup_t |P\{(\hat{f}_n - E(\hat{f}_n))/\sqrt{V}[\hat{f}_n] \leq t\} - \Phi(t)| = O(n^{-1/5})$ [proved]

4. Speed of Convergence to normality of the estimated stationary density

Let $f(x, y)$ be the stationary probability density function of (X_j, X_{j+1}) . We assume that all joint distributions with a finite number of distinct X_j 's are absolutely continuous with uniformly bounded continuous density functions. A natural estimator for the stationary probability is therefore,

$$\hat{f}_n(x, y) = (nh^2)^{-1} \sum_{j=1}^{n-1} K((x - X_j)/h, (y - X_{j+1})/h) \quad (4.1)$$

where $K(x, y)$ is a bounded continuous density function satisfying $\int u^2 K(u, v) du dv < \infty$ and $\int v^2 K(u, v) du dv < \infty$. Note that $V(\hat{f}_n(x, y)) = (nh^2)^{-1} f(x, y) \int K^2(x, y) dx dy + O(1/nh)$. Choosing $m(n) = O(n^{18/35})$, $k(n) = O(n^{17/35})$, $r(n) = O(\log n)$ and following exactly similar procedure as in Section 3 we have the following theorem.

THEOREM 4.1. *If Φ be the distribution function of standard normal distribution and (x, y) be a point of continuity of 'f' then*

$$\sup_t |P\{(\hat{f}_n(x, y) - E\hat{f}_n(x, y))/(V[\hat{f}_n(x, y)])^{1/2} \leq t\} - \Phi(t)| = O(n^{-1/7}(\log n)^3) \quad (4.2)$$

where $h = O(n^{-1/5})$.

To reduce the bias term in Theorem 4.2 and Theorem 5.2 we assume well known bias reduction criterion $\int \int u^i v^j K(u, v) du dv = 0$ for $i, j = 0, 1$ and $i + j \neq 0$.

THEOREM 4.2. *Under the conditions of Theorem 4.1 and $h_n = n^{-\alpha}$, $\alpha = \frac{1}{5}$,*

$$\sup_t |P\left\{\frac{\hat{f}_n(x, y) - f(x, y)}{\sqrt{\text{Var}\{\hat{f}_n(x, y)\}}} \leq t\right\} - \Phi(t)| = O(n^{-\frac{1}{10} + \tau}); \quad 0 < \tau < \frac{1}{10} \quad (4.3)$$

provided the partial derivatives of $f(x, y)$ upto 2nd order are continuous and 3rd order derivatives are finite.

PROOF. We know that

$$\begin{aligned} E\hat{f}_n(x, y) &= \frac{1}{nh^2} \sum_{j=1}^{n-1} \int \int K[(x - x_j)/h, (y - x_{j+1})/h] f(x_j, x_{j+1}) dx_j dx_{j+1}, \\ &= \frac{n-1}{nh^2} \int \int K[(x - x_1)/h, (y - y_1)/h] f(x_1, y_1) dx_1 dy_1. \end{aligned}$$

Using a change of variable and Taylor's expansion of $f(x, y)$ yields

$$\begin{aligned} E(\hat{f}_n(x, y)) &= \frac{n-1}{n} \int \int K(u, v) f(x - uh, y - vh) dudv \\ &= \frac{n-1}{n} \int \int K(u, v) [f(x, y) - huf_x(x, y) - hvf_y(x, y) \\ &\quad + \frac{h^2u^2}{2} f_{XX}(x, y) + \frac{h^2v^2}{2} f_{YY}(x, y) + h^2uvf_{XY}(x, y) + O(h^3)] dudv \\ &= f(x, y) + O(h^2) \end{aligned}$$

Therefore,

$$|E\hat{f}_n(x, y) - f(x, y)| = O(h^2) = O(n^{-2\alpha}). \quad (4.4)$$

Also,

$$Var[\hat{f}_n(x, y)] = O\left(\frac{1}{nh^2}\right) = O(n^{-1(1-2\alpha)}). \quad (4.5)$$

Define

$$\frac{\hat{f}_n(x, y) - f(x, y)}{\sqrt{Var\{\hat{f}_n(x, y)\}}} = Z_n^* + Y_n^* \quad (4.6)$$

where

$$Z_n^* = \frac{\hat{f}_n(x, y) - E\hat{f}_n(x, y)}{\sqrt{Var\{\hat{f}_n(x, y)\}}} \quad \text{and} \quad Y_n^* = \frac{E\hat{f}_n(x, y) - f(x, y)}{\sqrt{Var\{f_n(x, y)\}}}.$$

Using (4.4) and (4.5) we see that

$$|Y_n^*| = O\left(\frac{n^{-2\alpha}}{n^{-(1-2\alpha)/2}}\right) = O(n^{\frac{1-6\alpha}{2}}). \quad (4.7)$$

We select 'a' such that $n^{-a} > n^{\frac{1-6\alpha}{2}}$ (i.e. 'a' is selected such that $6\alpha - 2a - 1 > 0$) and then

$$P[|Y_n^*| > n^{-a}] = 0. \quad (4.8)$$

Also Theorem 4.1 gives

$$\sup_t |P[Z_n^* \leq t] - \Phi(t)| = O(n^{-1/7}(\log n)^3).$$

Now applying the Lemma 3.1 we get

$$\begin{aligned} \sup_t |P\left[\frac{\hat{f}_n(x, y) - f(x, y)}{\sqrt{Var\{\hat{f}_n(x, y)\}}} \leq t\right] - \phi(t)| \\ = O(n^{-1/7}(\log n)^3 + n^{-a}). \end{aligned} \quad (4.9)$$

In particular for $a = \frac{1}{10} - \tau$, $0 < \tau < \frac{1}{10}$ and $\alpha = \frac{1}{5}$ we get the right hand side of (4.9) as $O(n^{-(\frac{1}{10} - \tau)})$ and hence the theorem.

5. Speed of Convergence to Asymptotic Normality of the Estimated Transition Density

If $f(x, y)$ be the stationary p.d.f. for (X_j, X_{j+1}) then $f(y/x) = f(x, y)/f(x)$ is the transition p.d.f. of X_{j+1} given $X_j = x$. Let I be an interval on R and F_0 be the class of all distributions having the density 'f' such that, $0 < \varepsilon_1 \leq \inf_{x \in I} f(x) \leq \sup_{x \in I} f(x) \leq \varepsilon_2$.

Further we assume, in this section, K is differentiable.

A logical estimator for the transition density is, therefore,

$$\hat{f}_n(y/x) = \hat{f}_n(x, y)/\hat{f}_n(x) \quad (5.1)$$

where $\hat{f}_n(x, y)$ and $\hat{f}_n(x)$ are as in (4.1) and (2.2) respectively. Under some assumptions Yakowitz (1985) proved that,

$$(nh^2)[\hat{f}_n(y/x) - f(y/x)] \overset{d}{\sim} N(0, f(x, y) \left[\int K^2(\alpha, \beta) d\alpha d\beta \right] / f^2(x)). \quad (5.2)$$

Now, $(nh^2)^{1/2}[\hat{f}_n(y/x) - f(y/x)]$

$$\begin{aligned} &= (nh^2)^{1/2} \left[\frac{\hat{f}_n(x, y)/f(x)}{\hat{f}_n(x)/f(x)} - \frac{f(x, y)}{f(x)} \right] \\ &= (nh^2)^{1/2} [\hat{f}_n(x, y)/f(x) - f(y/x) \cdot \hat{f}_n(x)/f(x)] / [\hat{f}_n(x)/f(x)] \\ &= W_n/T_n, \text{ say} \end{aligned} \quad (5.3)$$

where

$$\begin{aligned} W_n &= (nh^2)^{1/2} [\hat{f}_n(x, y)/f(x) - f(y/x) \hat{f}_n(x)/f(x)] \\ &= Z_n + Y_n, \end{aligned} \quad (5.4)$$

$$\begin{aligned} Z_n &= (nh^2)^{1/2} [\hat{f}_n(x, y) - f(x, y)] / f(x), \\ Y_n &= -(nh^2)^{1/2} f(y/x) [\hat{f}_n(x) - f(x)] / f(x), \end{aligned} \quad (5.5)$$

and

$$T_n = \hat{f}_n(x)/f(x). \quad (5.6)$$

Also we have seen in section 4 that

$$\begin{aligned} \sup_t |P[Z_n \leq t] - \Phi(t)| &= O(n^{-(\frac{1}{10}-\tau)}) \\ &= O(b_n), \text{ say.} \end{aligned} \quad (5.7)$$

Under the above representation we have the following results :

RESULT 5.1. For any $a_n > 0, \delta > 0$ and $C_1 > 0$,

$$P[|Y_n| > a_n] \leq 2 \cdot \exp[-\delta(Aa_n - C_1 h^2 \sqrt{nh^2})^2 / (A_1^2 h^2)] \quad (5.8)$$

if $A = f(x)/f(y/x) < \infty$ and $O < A_1 = K(x) < \infty$.

RESULT 5.2. For any $\eta > 0, \delta > 0$ and $C_1 > 0$,

$$P[|T_n - 1| > \eta] \leq 2 \cdot \exp[-n\delta(\eta f(x) - C_1 h^2)^2 / A_1^2]$$

where $0 < A_1 = K(x) < \infty$.

Before proving the above two results we state and prove the following theorem :

THEOREM 5.1. *For a stationary Markov G_2 -sequence $\{X_i\}$ satisfying (1.1), if $F_n(y)$ and $F(y)$ are the empirical and true distribution functions respectively then for any $d > 0$ and $\delta > 0$,*

$$\sup_y P[|F_n(y) - F(y)| \geq d] \ll 2 \exp(-nd^2 \delta).$$

PROOF. We have $F_n(y) - F(y)$

$$\begin{aligned} &= \left[\sum_{i=1}^n I_y(X_i) - E \sum_{i=1}^n I_y(X_i) \right] / n \quad \text{where } I_y(X_i) = 1 \text{ if } y \geq X_i \\ &= 0 \quad \text{otherwise} \\ &= \frac{1}{n} \sum_{i=1}^n Z_i \quad \text{where } Z_i = I_y(X_i) - EI_y(X_i). \end{aligned}$$

Here Z_1, Z_2, \dots, Z_n are uniformly bounded random variables with $E(Z_k) = 0$, $V(Z_k) = E(Z_k^2) \leq \rho^2 \sigma^2 \leq F(y)(1 - F(y)) \leq 1/4$ and $|Z_k| \leq 1$.

Let $Z = \sum_1^n Z_i$. Therefore $F_n(y) - F(y) = Z/n$.

Now

$$\begin{aligned} E(e^{tZ_k}) &= E\left[1 + tZ_k + \frac{t^2 Z_k^2}{2} + \frac{t^3 Z_k^3}{3!} + \frac{t^4 Z_k^4}{4!} + \dots\right] \\ &\leq 1 + \frac{t^2}{2} E Z_k^2 + E\left[\frac{t^3}{3!} Z_k^3 + \frac{t^4}{4!} Z_k^4 + \dots\right] \\ &\quad \text{(since } Z_i \text{'s are bounded by one)} \\ &\leq 1 + \frac{t^2 \rho^2 \sigma^2}{2} + \rho^2 \sigma^2 \left[\frac{t^3}{3!} + \frac{t^4}{4!} \dots\right] \quad \text{(because of } G_2 \text{ - property)} \\ &= 1 + \rho^2 \sigma^2 t^2 / 2 + \rho^2 \sigma^2 \sum_{n=3}^{\infty} t^n / n! \\ &= 1 + \rho^2 \sigma^2 \frac{t^2}{2} + \rho^2 \sigma^2 t^2 \sum_{n=3}^{\infty} \frac{t^{n-2}}{n!} \\ &= 1 + (\rho^2 \sigma^2 t^2 / 2) + \frac{\rho^2 \sigma^2 t^2}{6} \sum_{n=3}^{\infty} \frac{t^{n-2}}{(n-3)!} \quad \text{(since } \frac{1}{n!} \leq \frac{1}{6} \frac{1}{(n-3)!} \text{ for } n \geq 3) \\ &= 1 + (\rho^2 \sigma^2 t^2 / 2) \{1 + te^t / 3\} \\ &\leq \exp\left\{\frac{1}{2} \rho^2 \sigma^2 t^2 (1 + te^t / 3)\right\} \quad \text{(since } 1 + x \leq e^x \text{ for } x \geq 0) \\ &\leq \exp\left\{\frac{1}{8} \rho^2 t^2 (1 + te^t / 3)\right\}. \end{aligned}$$

Again from Rosenblatt (1970) we have,

$|E(\frac{n}{1} e^{tZ_i}) - \frac{n}{1} E(e^{tZ_i})| \leq (n-1)M\rho^n$ where M is a constant independent of y and $0 < \rho < 1$. This implies

$$-(n-1)M\rho^n + \frac{n}{i=1} E(e^{tZ_i}) \leq M(t) \leq (n-1)M\rho^n + \frac{n}{i=1} E(e^{tZ_i})$$

where $M(t) = E(e^{tZ}) = E(e^{t\Sigma Z_i}) = E(\frac{n}{1}e^{tZ_i})$.

$$\text{Therefore } M(t) \leq (n-1)M\rho^n + \frac{n}{1} \exp(\frac{1}{2}\rho^2\sigma^2t^2(1+te^t/3))$$

$$\ll 2 \cdot \exp\{\frac{n}{2}\rho^2\sigma^2t^2(1+te^t/3)\}$$

$$= 2 \cdot e^{nc}, \text{ where } c = \frac{1}{2}\rho^2\sigma^2t^2(1+te^t/3)$$

$$\leq e^{2nc}, \text{ because } 2 < e < 3$$

$$\leq \exp\{2 \cdot n \cdot \frac{1}{2}\rho^2\sigma^2t^2(1+te^t/3)\}$$

$$\leq \exp\{\frac{n}{4}t^2\rho^2(1+te^t/3)\}$$

$$= \exp\{\frac{1}{2}t^2D^2(1+te^t/3)\}, \text{ where } D^2 = \frac{n}{2}\rho^2.$$

Also we have the inequality (see Renyi p. 384)

$$P[Z \geq M + \frac{T + \log M(t)}{t}] \leq e^{-T}$$

where M is the expectation of Z .

Here $M = 0$ and therefore

$$P(Z \geq \{T + \frac{1}{2}t^2D^2(1+te^t/3)\}/t) \leq e^{-T}.$$

Put $t = \sqrt{2T}/D$ and therefore

$$P(Z \geq D\sqrt{(2T)}\{1 + \frac{\sqrt{(2T)}}{6D}e^{\sqrt{(2T)}/D}\}) \leq e^{-T}.$$

Now substitution of $\lambda = \sqrt{(2T)}$ gives

$$P(Z \geq \lambda D\{1 + (\lambda/6D)e^{\lambda/D}\}) \leq e^{-\lambda^2/2}.$$

Applying the obtained result to $-Z$, we find that

$$P(|Z| \geq \lambda D\{1 + (\lambda + 6D)e^{\lambda/D}\}) \leq 2e^{-\lambda^2/2}.$$

In order to write this in a more convenient form, we restrict ourselves to the case $\frac{\lambda}{D} \leq 1$.

We have then

$$e^{\lambda/D} \leq e < 3.$$

Now we put $\gamma = \lambda(1 + \lambda/2D)$ and then $\lambda \leq \gamma \leq \lambda(1 + \gamma/2D)$.

$$P(|Z| \geq \gamma D) \leq 2e^{-\frac{\gamma}{2}(1+\gamma/2D)^{-2}}$$

i.e.

$$P(|F_n(y) - F(y)| \geq \frac{\gamma D}{n}) \leq 2e^{-\{\gamma^2/2(1+\gamma/2D)^{-2}\}} \text{ for all } y.$$

i.e.

$$P(|F_n(y) - F(y)| \geq d) \leq 2e^{-\{n^2 d^2 / 2D^2(1 + \frac{nd}{2D^2})^{-2}\}} \text{ for all } y.$$

Now replacing D^2 by $\frac{n}{2}\rho^2$ we have

$$\begin{aligned} P(|F_n(y) - F(y)| \geq d) &\leq 2 \exp(-n\rho^2 d^2 / (d + \rho^2)^2) \\ &\leq 2 \exp(-nd^2 \delta). \end{aligned}$$

Since all the upper bounds in our calculations are independent of y ,

$$\sup_y P[|F_n(y) - F(y)| \geq d] \ll 2 \exp(-nd^2 \delta).$$

REMARK. The result in this theorem perhaps follows from P.K. Bhattacharyya (1972) but for completeness and independent interest we have given some details.

PROOF OF RESULT 5.1.

$$\begin{aligned} P[|Y_n| > a_n] &= P[(nh^2)^{1/2} / f(x) f(y/x) |\hat{f}_n(x) - f(x)| > a_n] \\ &= P[(nh^2)^{1/2} |\hat{f}_n(x) - f(x)| > \varepsilon], \text{ where } \varepsilon = A.a_n \\ &\leq P[(nh^2)^{1/2} \{|\hat{f}_n(x) - E\hat{f}_n(x)| + |E\hat{f}_n(x) - f(x)|\} > \varepsilon]. \end{aligned}$$

Now,

$$\begin{aligned} E(\hat{f}_n(x)) - f(x) &= O(h^2), \text{ as in (4.4)} \\ &\leq C_1 h^2. \end{aligned}$$

Therefore,

$$\begin{aligned} P[|Y_n| > a_n] &\leq P[|\hat{f}_n(x) - E\hat{f}_n(x)| > \{\varepsilon / (nh^2)^{1/2} - C_1 h^2\}] \\ &= P[|\hat{f}_n(x) - E\hat{f}_n(x)| > \{\varepsilon' / (nh^2)^{1/2}\}] \end{aligned}$$

where

$$\begin{aligned} \varepsilon' &= \varepsilon - C_1 h^2 (nh^2)^{1/2} \\ &= A.a_n - C_1 h^2 (nh^2)^{1/2} \\ &\leq \sup_y P[|F_n(y) - F(y)|] \left\| \frac{d}{dy} K((x-y)/h) \frac{1}{h} \right\|_1 > \varepsilon' / (nh^2)^{1/2} \} \end{aligned} \quad (5.10)$$

(integrating by parts).

Using (2.4), $\left\| \frac{d}{dy} K((x-y)/h) \frac{1}{h} \right\|_1 = \int_{-\infty}^{\infty} \frac{d}{dy} K((x-y)/h) \frac{1}{h} dy = A_1$.

From (5.10) we have,

$$\begin{aligned} P[|Y_n| > a_n] &\leq \sup_y P[|F_n(y) - F(y)| > \{\varepsilon' / A_1 (nh^2)^{1/2}\}] \\ &\leq 2 \exp[-n\delta \{\varepsilon'^2 / (A_1^2 nh^2)\}] \\ &= 2 \exp[-\delta \{Aa_n - C_1 h^2 (nh^2)^{1/2}\}^2 / (h^2 A_1^2)]. \end{aligned}$$

PROOF OF RESULT 5.2. Following the proof of Result 5.1 we have,

$$\begin{aligned}
P[|T_n - 1| > \eta] &\leq P[|\hat{f}_n(x) - f(x)| > \eta f(x)] \\
&= P[|\hat{f}_n(x) - f(x)| > \eta_1] \text{ where } \eta_1 = \eta \cdot f(x) \\
&\leq P[|\hat{f}_n(x) - E\hat{f}_n(x)| > \eta_1 - C_1 h^2] \\
&\leq \sup_y P[|\hat{F}_n(y) - F(y)| \geq (\eta'_1/A_1)], \\
&\quad \text{where } \eta'_1 = \eta_1 - C_1 h^2 \\
&= 2 \cdot \exp[-n\eta_1'^2 \delta/A_1^2] \\
&= 2 \cdot \exp[-n\delta(\eta_1 - C_1 h^2)^2/A_1^2] \\
&= 2 \cdot \exp[-n\delta(\eta f(x) - C_1 h^2)^2/A_1^2] \quad (\text{proved})
\end{aligned}$$

Now using (5.7), (5.8) and Lemma 3.1 we have the following result.

RESULT 5.3.

$$\begin{aligned}
\sup_t |P[W_n \leq t] - \Phi(t)| &= O(b_n) + O(a_n) + P[|Y_n| > a_n] \\
&<< n^{-(\frac{1}{10} - \tau)} + a_n + \exp[-\delta(Aa_n - C_1 h^2 \sqrt{(nh^2)})^2/(A_1^2 h^2)] \quad (5.11)
\end{aligned}$$

where $h = O(n^{-\frac{1}{5}})$.

Let us state another result which is due to Michel and Pfanzagl (1971).

RESULT 5.4. *Let (Ω, \mathcal{F}, P) be a probability space and W_n, T_n be \mathcal{F} -measurable functions. Then, for any $\eta > 0$,*

$$\sup_t |P(W_n/T_n \leq t) - \Phi(t)| \leq \sup_y |P(W_n \leq y) - \Phi(y)| + P[|T_n - 1| > \eta] + \eta.$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Now we have the following theorem :

THEOREM 5.2. *If Φ be the distribution function of a standard normal distribution then under the assumption of Theorem 4.2*

$$\sup_{F_0} \sup_{x \in I} \sup_{n \geq n_0} \sup_t |P[\{(\hat{f}_n(y/x) - f(y/x))/(V(\hat{f}_n(y/x)))^{1/2}\} \geq t] - \Phi(t)| << n^{-\frac{1}{10} + \tau}.$$

PROOF. Applying Result 5.2 and Result 5.3 in Result 5.4 we have,

$$\begin{aligned}
&\sup_t |P[(W_n/T_n) \leq t] - \Phi(t)| \\
&<< n^{-\frac{1}{10} + \tau} + a_n + \exp[-\delta(Aa_n - C_1 h^2 \sqrt{(nh^2)})^2/h^2 A_1^2] \\
&\quad + \exp\{-n\delta(\eta f(x) - C_1 h^2)^2/A_1^2\} + \eta.
\end{aligned}$$

Putting $a_n = 0(n^{-\frac{1}{5}}(\log n)^{1/2})$ and $\eta = 0((\log n)/n)^{1/2}$, we get

$$\begin{aligned}
&\sup_t |P\{(\hat{f}_n(y/x) - f(y/x))/(V(\hat{f}_n(y/x)))^{1/2}\} \leq t] - \Phi(t)| \\
&<< n^{-\frac{1}{10} + \tau} + n^{-\frac{1}{5}}(\log n)^{1/2} + \exp(-\delta(Aa_n - C_1 h^2 \sqrt{(nh^2)})^2/h^2 A_1^2) \\
&\quad + \exp(-n\delta(\eta f(x) - C_1 h^2)^2 A_1^2) + ((\log n)/n)^{1/2} << n^{-\frac{1}{10} + \tau}.
\end{aligned}$$

6. Concluding Remarks

Prakasa Rao (1977) considered the problem of estimating the unknown density under much stronger condition D_0 and got the result $O(n^{-\frac{1}{3}+\tau})$, $0 < \tau < \frac{1}{3}$. Neither the problem of estimating the stationary density nor the transition density has been considered by him. Although our rate result is poorer than Prakasa Rao (1977), yet our results have some theoretical interest in the sense that there is no work in the available literature regarding the rate of convergence to normality for estimating the stationary density and transition density under much weaker assumption G_2 which has so much applications in time series data.

Also, in this context, one thing should be noted is that we have taken the window-width h_n as $O(n^{-\frac{1}{3}})$ which is optimal in the sense that it minimizes mean-integrated square error and this choice of window-width is different from that of Prakasa Rao (1977) where he took the window-width as $O(n^{-\frac{1}{3}})$.

In Theorem 4.2 and Theorem 5.2 we conjecture that the rates should be $O(n^{-\frac{1}{7}}(\log n)^3)$. Actually in our earlier version we obtained the above rate but unfortunately there was a technical flaw. Thanks to the referee for pointing out this.

References

- Basu, A.K. and Sahoo, D.K. (1989) : Berry-Esséen type of theorems for sequential density estimation. *Sequential Analysis*, 119-134.
- Basu, A.K. and Sahoo, D.K. (1989) : Rate of convergence for nonparametric density estimation in linear process. *Bull of Inform. and Cybernetics*, **23**, 209-224.
- Bhattacharya, P.K. (1972) : Probabilities of large deviations of sums of random variables. *Sankhyā*, Ser. A, **34**, 9-16.
- Doob, J.L. (1953) : Stochastic processes, Wiley, Newyork.
- Michel and Pfanzagl (1971) : The accuracy of the normal approximation for minimum contrast estimates, *Z. Wahrs Verw. Gebiete*, **18**, 73-84.
- Philipp (1969) : The remainder in the central limit theorem for mixing stochastic processes. *Ann. Math. Statist.*, **40**, 601-609.
- Prakasa Rao, B.L.S. (1977) : Berry-Esséen bound for density estimators of stationary-Markov processes. *Bull. Math. Statist.*, **17**, 15-21.
- Renyi (1970) : Probability theory, North Holland Publishing Co.
- Rosenblatt, M. (1970) : Density estimates and Markov sequences. Nonparametric techniques in statistical inference, ed. M. Puri, Oxford - Cambridge Univ. Press.
- Rosenblatt, M. (1971) : Curve estimates, *Ann. Math. Statist.*, **42**, 1815-1842.
- Roussas (1967) : Non-parametric estimation in Markov processes, Technical Report No. 110, Dept. of Statistics, Univ. of Wisconsin, Madison, Wisconsin.
- Roussas (1969) : Non-parametric estimation of the transition distribution function of a Markov process : *Ann. Math. Statist.*, **40**, 1386-1400.
- Yakowitz (1985) : Non-parametric estimation, prediction and regression for Markov sequences : *Journal of the Americal Statistical Association*, **80**, 215-221.

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

Received April 2, 1996

Revised December 13, 1996

Re-revised January 6, 1998