Ann. Inst. Statist. Math.
30 (1978), Part A, 321-328

DENSITY ESTIMATION FOR MARKOV PROCESSES USING DELTA-SEQUENCES

B. L. S. PRAKASA RAO

(Received Nov. 25, 1977; revised July 31, 1978)

1. Introduction

Estimators of the density function of a population based on a sample of independent observations have been proposed by several authors. An excellent survey of the results in this area is given in Rosenblatt [7]. Recently Walter and Blum [9] proposed a method for density estimation using delta-sequences i.e., sequences of functions that converge to the generalized function δ in a suitable sense (cf. Korevaar [4]). An advantage of their approach is that all the earlier methods like kernel method, orthogonal series method, interpolation method and the characteristic function approach for density estimation are included in their method via delta-sequences.

Rosenblatt [6] and Roussas [8] considered kernel type of density estimators when the observations are assumed to be sampled from a stationary Markov process. Rosenblatt [6] has shown that these estimators have the same behavior as those of a density estimator in the independent and identically distributed case. Prakasa Rao [5] obtained a Berry-Esseen type bound for the distribution function of a density estimator in the Markov case generalizing a similar result of Wertz [10] in the independent case.

In this paper, we use the method of delta-sequences of Walter and Blum [9] to obtain estimators of density for stationary Markov processes. We shall obtain bounds for the mean square error of the proposed estimators.

We mention that the method of generalized functions was used by Borwanker [2] to show the non-existence of unbiased estimators in the ordinary sense for density when the observations are from a stationary process and he has also studied asymptotic properties of estimators via delta-sequences but has not pursued them in detail. Some of the details of these results can be found in Basawa and Prakasa Rao [1].

2. Preliminaries

Consider a probability space $(R, \mathcal{B}, P)$ where R is the real line, $\mathcal{B}$ the σ-field of Borel sets of R and P a probability measure. Let $\{X_n, n\geqq 1\}$ be a Markov process taking values in $(R, \mathcal{B}, P)$ with stationary transition measure $p(\xi, A)=\mathrm{P}(X_{n+1}\in A|X_n=\xi)$. Assume that $p(\xi, A)$ is a measurable function of ξ for fixed A and a probability measure on $\mathcal{B}$ for fixed ξ. Such a transition measure together with an initial probability measure gives rise to a Markov process by Doob [3]. Assume that the process $\{X_n: n\geqq 1\}$ satisfies Doeblin's condition (D_0) as given in Doob [3], p. 221 viz. there is a finite-valued measure τ on $\mathcal{B}$ with $\tau(R)>0$, an integer $\nu\geqq 1$ and $\varepsilon>0$ such that

$$p^{(\nu)}(\xi, A)\leqq 1-\varepsilon \qquad \text{if } \tau(A)\leqq\varepsilon$$

and there is only one ergodic set $E\subset R$ with $\tau(E)>0$ and this set contains no cyclically moving subsets. (Here $p^{(n)}(\cdot, \cdot)$ is the n-step transition measure.) Under (D_0), it can be shown that there exist positive constants $r\geqq 1$, $0<\rho<1$ and a unique stationary probability distribution $\pi(\cdot)$ such that

$$|p^{(n)}(\xi, A)-\pi(A)|\leqq r\rho^n$$

for $n\geqq 1$. The distribution $\pi(\cdot)$ taken as the initial distribution together with the stochastic transition function $p(\cdot, \cdot)$ determines a stationary Markov process. We shall assume that the initial distribution is always the stationary distribution.

Suppose that $p(\xi, \cdot)$ and $\pi(\cdot)$ are absolutely continuous with respect to Lebesgue measure on $(R, \mathcal{B})$ and let $f(\xi, \cdot)$ and $f(\cdot)$ be the corresponding densities. The problem which we consider is to obtain estimators of $f(\cdot)$. Let P_f be the probability measure on $(R^\infty, \mathcal{B}^\infty)$ corresponding to $f(\cdot, \cdot)$ and $f(\cdot)$. We shall assume that $f(\cdot, \cdot)$ and $f(\cdot)$ are continuous.

Before we proceed further, we shall state a few definitions from generalized functions.

Let S be the space of infinitely differentiable rapidly decreasing functions on R and S' be the dual space of continuous linear functionals on S. Members of S' are called generalized functions. The generalized function δ corresponds to the measure with unit mass at zero and it is defined by $\langle\delta, \varphi\rangle=\varphi(0)$ for $\varphi\in S$.

If X is a random variable, then $\delta(X-x)$ is the map which takes an element φ of S into the random variable $\varphi(X)$ i.e., $\langle\delta(X-x), \varphi(x)\rangle =\varphi(X)$. If X has the continuous density function $f(x)$, then the expected value of $\delta(X-x)$ may be calculated by

(2.1) $$\mathrm{E}\,\langle \delta(X-x),\,\varphi(x)\rangle=\int \varphi(x)f(x)dx=\langle f,\varphi\rangle$$

i.e. $\mathrm{E}\,\delta(X-x)=f(x)$

in the sense of S'. Convergence in the space S' is the weak convergence i.e. $f_n\to f$ in S' if $\langle f_n,\varphi\rangle\to\langle f,\varphi\rangle$ for all $\varphi\in S$.

3. Generalized function estimators

Suppose the process $\{X_n,\ n\geqq 1\}$ is observed up to time n. We introduce the function

$$\hat{f}_n(x)=\frac{1}{n}\sum_{i=1}^{n}\delta(X_i-x)$$

as a generalized function estimator for $f(x)$. Clearly $\hat{f}_n(x)$ is an unbiased estimator for $f(x)$ in the generalized sense. It can be seen by (2.1) since the process is stationary. Note that

(3.1) $$\begin{aligned}&\mathrm{E}\,[\langle \hat{f}_n,\varphi\rangle-\langle f,\varphi\rangle]^2\\&=\mathrm{E}\left[\left\langle\frac{1}{n}\sum\delta(X_i-x),\,\varphi(x)\right\rangle-\langle f,\varphi\rangle\right]^2\\&=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\,[\langle\delta(X_i-x),\,\varphi(x)\rangle]\\&\quad+\frac{1}{n^2}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}\mathrm{Cov}\,[\langle\delta(X_i-x),\,\varphi(x)\rangle,\,\langle\delta(X_j-x),\,\varphi(x)\rangle]\\&=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\,[\varphi(X_i)]+\frac{1}{n^2}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}\mathrm{Cov}\,[\varphi(X_i),\,\varphi(X_j)]\\&=\frac{1}{n}\mathrm{Var}\,[\varphi(X_1)]+\frac{1}{n^2}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}\mathrm{Cov}\,[\varphi(X_i),\,\varphi(X_j)]\,.\end{aligned}$$

Suppose that $\mathrm{Var}\,[\varphi(X_1)]<\infty$. It is clear that $\mathrm{Var}\,[\varphi(X_i)]<\infty$ and $\mathrm{Cov}\,[\varphi(X_i),\,\varphi(X_j)]<\infty$ for all i and j in view of stationarity. The first term on the R.H.S. of (3.1) tends to zero as $n\to\infty$. The second term is bounded in absolute value by

$$\begin{aligned}&\frac{2}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{j}2r^{1/2}\rho^{(j-i)/2}\,\mathrm{E}\,[\varphi(X_i)]^2\\&=\frac{4r^{1/2}}{n^2}\mathrm{E}\,[\varphi(X_1)]^2\left\{\sum_{j=1}^{n}\sum_{i=1}^{j}\rho^{(j-i)/2}\right\}\\&=\frac{4}{n^2}r^{1/2}\,\mathrm{E}\,[\varphi(X_1)]^2\left\{\frac{n\rho^{1/2}}{1-\rho^{1/2}}-\frac{\rho(1-\rho^{(n-1)/2})}{(1-\rho^{1/2})^2}\right\}\end{aligned}$$

by Doob [3], p. 222 (cf. Lemma 2.2, Prakasa Rao [5]) and the last term tends to zero as $n \to \infty$. Hence

$$\mathrm{E}\,[\langle \hat{f}_n, \varphi\rangle - \langle f, \varphi\rangle]^2 \to 0$$

which might be called weak mean square convergence. Note that the usual mean square convergence does not make sense here since δ^2 cannot be interpreted as an element of S'.

Observe that $\hat{f}_n$ cannot be considered as an ordinary estimator since δ is not an ordinary function but a generalized function. In order to get an estimator consisting of ordinary random variables, we approximate δ by delta-sequences. This is what we do in the next section.

4. Estimation via delta-sequences

A sequence $\{\delta_m\}$ of L^∞ functions on R is called a *delta-sequence* of positive type $\alpha > 0$, if $\delta_m(x) \geqq 0$, $m \geqq 1$ and

(i) there exist $A > 0$, $B > 0$ such that

$$\left|1 - \int_{-A}^{B} \delta_m(x)dx\right| = O(m^{-\alpha})\ ; \tag{4.1}$$

(ii) $\sup \{|\delta_m(x)| : |x| \geqq m^{-\alpha}\} = O(m^{-\alpha})$;

(iii) $\|\delta_m\|_\infty = \text{ess sup}\, \{|\delta_m(x)| : x \in R\} \sim m$.

An example of a delta-sequence of the above type is

$$\delta_m = m\chi_{[0, m^{-1}]}\ , \qquad m = 1, 2, \cdots$$

where $\chi_{[0,m^{-1}]}$ is the indicator function of the interval $[0, m^{-1}]$. Note that this sequence is a delta-sequence of positive type 1. Other examples are discussed in Walter and Blum [9].

Suppose now that we have observed the process $\{X_n, n \geqq 1\}$ up to time n. Let $\{\delta_m\}$ be a delta-sequence of positive type α. Define

$$\hat{f}_{nm}(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_m(x - X_i)\ , \qquad n, m = 1, 2, \cdots .$$

We choose $\hat{f}_{nm}(x)$ as an estimator for $f(x)$. Since $\hat{f}_{nm}$ are random variables with finite variance, we can use mean square error as indicator of rate of convergence. In fact

$$\begin{aligned} &\mathrm{E}\,[\hat{f}_{nm}(x) - f(x)]^2 \\ &= \mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n} \delta_m(x - X_i) - f(x)\right\}^2 \\ &= \frac{1}{n^2}\,\mathrm{E}\left\{\sum_{i=1}^{n} \delta_m(x - X_i)\right\}^2 + f^2(x) - \frac{2}{n} f(x)\,\mathrm{E}\left\{\sum_{i=1}^{n} \delta_m(x - X_i)\right\} \end{aligned} \tag{4.2}$$

$$=\frac{1}{n^2}\{n\,\mathrm{E}\,[\delta_m^2(x-X_1)]+\sum_i\sum_{\substack{j\\ i\neq j}}\mathrm{E}\,[\delta_m(x-X_i)\delta_m(x-X_j)]\}$$
$$+f^2(x)-2f(x)\,\mathrm{E}\,(\delta_m(x-X_1))$$

by stationarity of the process. This can also be written in the form

$$\frac{1}{n}\mathrm{E}\,(\delta_m^2(x-X_1))+\frac{2}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{j-1}\{\mathrm{Cov}\,[\delta_m(x-X_i),\,\delta_m(x-X_j)]$$
$$+(\mathrm{E}\,[\delta_m(x-X_i)])^2\}+f^2(x)-2f(x)\,\mathrm{E}\,(\delta_m(x-X_1))\,. \tag{4.3}$$

Since

$$|\mathrm{Cov}\,[\delta_m(x-X_i),\,\delta_m(x-X_j)]|\leqq 2r^{1/2}\rho^{(j-i)/2}\,\mathrm{E}\,[\delta_m^2(x-X_i)]$$

by Doob [3], p. 222, it follows that

$$\mathrm{E}\,[\hat{f}_{nm}(x)-f(x)]^2 \tag{4.4}$$
$$\leqq \mathrm{E}\,(\delta_m^2(x-X_1))\left\{\frac{1}{n}+\frac{4r^{1/2}}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{j-1}\rho^{(j-i)/2}\right\}$$
$$+\frac{2}{n^2}\left(\sum_{j=1}^{n-1}j\right)[\mathrm{E}\,(\delta_m(x-X_1))]^2+f^2(x)-2f(x)\,\mathrm{E}\,(\delta_m(x-X_1))$$
$$=\mathrm{E}\,(\delta_m^2(x-X_1))\left\{\frac{1}{n}+\frac{4r^{1/2}}{n^2}\left(\frac{n\rho^{1/2}}{1-\rho^{1/2}}-\frac{\rho(1-\rho^{(n-1)/2})}{(1-\rho^{1/2})^2}\right)\right\}$$
$$+\frac{(n-1)}{n}(\mathrm{E}\,(\delta_m(x-X_1)))^2+f^2(x)-2f(x)\,\mathrm{E}\,(\delta_m(x-X_1))$$
$$\leqq\frac{\mathrm{E}\,(\delta_m^2(x-X_1))}{n}\left\{1+\frac{4r^{1/2}\rho^{1/2}}{1-\rho^{1/2}}\right\}+\frac{n-1}{n}[\mathrm{E}\,(\delta_m(x-X_1))]^2$$
$$+f^2(x)-2f(x)\,\mathrm{E}\,(\delta_m(x-X_1))$$
$$=\frac{\mathrm{E}\,(\delta_m^2(x-X_1))}{n}\left\{1+\frac{4r^{1/2}\rho^{1/2}}{1-\rho^{1/2}}\right\}-\frac{1}{n}[\mathrm{E}\,(\delta_m(x-X_1))]^2$$
$$+[\mathrm{E}\,(\delta_m(x-X_1))-f(x)]^2$$
$$=\frac{1}{n}\{\mathrm{E}\,(\delta_m^2(x-X_1))-[\mathrm{E}\,(\delta_m(x-X_1))]^2\}+[\mathrm{E}\,(\delta_m(x-X_1))-f(x)]^2$$
$$+\frac{4r^{1/2}\rho^{1/2}}{n(1-\rho^{1/2})}\,\mathrm{E}\,(\delta_m^2(x-X_1))\,.$$

We shall now prove the following theorem.

THEOREM 4.1. *Let $\{\delta_m\}$ be a sequence of positive type α; let $\{X_n, n\geqq 1\}$ be a stationary Markov process with marginal density function $f(x)$. If $f\in\mathrm{Lip}\,\lambda$ for some $0<\lambda\leqq 1$, then*

$$\sup_{x\in R}\mathrm{E}\,(\hat{f}_{mn}(x)-f(x))^2\leqq C_0mn^{-1}+C_1m^{-2\alpha\lambda}$$

where C_0 and C_1 are constants.

PROOF. It is clear from (iii) that

$$\text{(4.5)} \qquad E(\delta_m^2(x-X_1)) \leqq \|\delta_m\|_\infty E(\delta_m(x-X_1)) \sim m\, E(\delta_m(x-X_1)) .$$

Arguments in Walter and Blum [9], p. 6 show that

$$\text{(4.6)} \qquad |E(\delta_m(x-X_1)) - f(x)| \leqq C_2 m^{-\alpha\lambda} + C_3 m^{-\alpha}$$

uniformly in x which implies in particular that

$$\text{(4.7)} \qquad |E(\delta_m(x-X_1))| \leqq C_4$$

uniformly in x since f will be uniformly bounded in x by Lipschitz condition on f. (4.5) and (4.7) together show that

$$\text{(4.8)} \qquad E(\delta_m^2(x-X_1)) - [E(\delta_m(x-X_1))]^2 \leqq C_5 m$$

for some constant C_5 independent of x. (4.5)–(4.8) combined together give the relation

$$\begin{aligned} \sup_{x \in R} E[\hat{f}_{nm}(x) - f(x)]^2 &\leqq \frac{1}{n}(C_5 m) + (C_2 m^{-\alpha\lambda} + C_3 m^{-\alpha})^2 + \frac{C_6 m}{n} \\ &\leqq \frac{C_0 m}{n} + C_1 m^{-2\alpha\lambda} \end{aligned}$$

in view of inequality (4.4). This proves the theorem

Remark. Note that the bound on the mean square error obtained above is the same as in the independent case as far as the rate of convergence is concerned. In particular if $m_n = [n^{1/(1+2\alpha\lambda)}]$ and $\hat{f}_n = \hat{f}_{m_n, n}$, then

$$E[\hat{f}_n(x) - f(x)]^2 = O(n^{-1+1/(1+2\alpha\lambda)})$$

uniformly in $x \in R$.

5. Examples of delta-sequences

Walter and Blum [9] have extensively studied the rates of convergence for different delta-sequences in the independent and identically distributed case. We shall consider only two cases. If $\delta_m = m\chi_{[0, m^{-1}]}$, then $\alpha = 1$ and

$$E(\hat{f}_n(x) - f(x))^2 = O(n^{-1+1/(1+2\lambda)}) .$$

Suppose $K(t)$ is any bounded measurable function such that $K(t) \geqq 0$, $K(t) = O(t^{-1-\beta})$ as $|t| \to \infty$ for some $\beta > 0$ and

$$\int_{-\infty}^{\infty} K(t)dt=1 .$$

Let

$$\delta_m(t)=mK(mt) .$$

It is not difficult to check that $\{\delta_m\}$ is a delta-sequence of type $\alpha=\beta/(\beta+2)$. Hence

$$\begin{aligned} \mathrm{E}\,(\hat{f}_n(x)-f(x))^2 &= O(n^{-1+1/(1+(2\beta\lambda/(\beta+2)))}) \\ &= O(n^{-1+(\beta+2)/(2+\beta+2\beta\lambda)}) . \end{aligned}$$

If $\beta=1$, then

$$\mathrm{E}\,(\hat{f}_n(x)-f(x))^2=O(n^{-1+3/(3+2\lambda)})$$

and further if $\lambda=1$, then

$$\mathrm{E}\,(\hat{f}_n(x)-f(x))^2=O(n^{-2/5})$$

uniformly for $x \in R$.

6. Delta-sequences of densities

A sequence of densities $\{\delta_m\}$ is a *delta-sequence of density-type* α, $\alpha>0$, if

(i) $\|\delta_m\|_\infty=O(m)$, and

(ii) $\int_{-\infty}^{-m^{-\alpha}} \delta_m(x)dx+\int_{m^{-\alpha}}^{\infty} \delta_m(x)dx=O(m^{-\alpha})$.

One can construct estimators $\hat{f}_{nm}(x)$ based on delta-sequences of density type α and the conclusions of Theorem 4.1 and the remark thereafter hold for this sequence also by a slight modification of the proof of Theorem 4.1 as pointed out by Walter and Blum [9] in the independent case. The sequences considered in the previous section are also of density type.

If Y is a random variable with mean zero, finite variance and bounded density function g and g_m is the density of $\bar{Y}_m=\sum Y_i/m$ where Y_i, $1\leqq i\leqq m$ are i.i.d., then Walter and Blum [9] have shown that $\{g_m\}$ is a delta-sequence of density type $\alpha=1/3$.

Remark. We have studied the asymptotic behavior of mean square error of the density estimators using delta-sequences. The problem remains to study of the asymptotic behavior of the distributions of these estimators. We shall pursue this problem in a future paper.

Acknowledgement

The author thanks Professor J. Blum for a sending a preprint of his paper.

INDIAN STATISTICAL INSTITUTE, NEW DELHI

REFERENCES

[1] Basawa, I. V. and Prakasa Rao, B. L. S. (1979). *Statistical Inference for Stochastic Processes: Theory and Methods*, Academic Press, London, (To appear).

[2] Borwanker, J. D. (1967). *Some Asymptotic Results for Stationary Processes*, Ph.D. Thesis, University of Minnesota, USA.

[3] Doob, J. L. (1953). *Stochastic Processes*, Wiley, New York.

[4] Korevaar, J. (1971). *Mathematical Methods*, Academic Press, New York.

[5] Prakasa Rao, B. L. S. (1977). Berry-Esseen type bound for density estimators of stationary Markov processes, *Bull. Math. Statist.*, **17**, 15-21.

[6] Rosenblatt, M. (1970). Density estimates and Markov sequences, In *Nonparametric Techniques in Statistical Inference*, Cambridge University Press, London.

[7] Rosenblatt, M. (1971). Curve estimates, *Ann. Math. Statist.*, **42**, 1815-1842.

[8] Roussas, G. G. (1969). Nonparametric estimation in Markov processes, *Ann. Inst. Statist. Math.*, **21**, 73-87.

[9] Walter, G. and Blum, J. (1976). Probability density estimation using delta-sequences, (Preprint) The University of Wisconsin, Milwaukee, USA.

[10] Wertz, W. (1971). Empirische betrachtungen und normal approximation bei dichteschatzungen, *Operat. Res. Verfahren*, **8**, 430-448.