DENSITY ESTIMATES AND MARKOV SEQUENCES†

M. ROSENBLATT

1 INTRODUCTION

Estimates of the density function of a population based on a sample of independent observations have been considered in a number of papers [1, 6–7]. Questions of bias, variance and asymptotic distribution of the estimates have been dealt with at greatest length. Our object is to look at such estimates of the density function when the observations are dependent. The results will not be dealt with in the most general context or under very general conditions. To obtain results in a simple and readily understandable form, the observations are assumed to be sampled from a stationary Markov sequence with a fairly strong condition on the Markov transition operator. However, the extent to which some of the conditions can be obviously relaxed will be indicated.

It should be noted that the asymptotic results we obtain in the case of dependent observations are essentially the same as those in the case of independent observations. This initially is surprising because it is certainly not true when estimating the distribution function by means of the sample distribution function. The more complicated nature of asymptotic results for this problem in the case of dependence can be seen in Billingsley [2]. However, the happy fact that the results we obtain in estimating the density have the same character as in the case of independence is due to the local character of the estimates.

2 REMARKS ON THE CASE OF INDEPENDENT SAMPLING

We just briefly make a few remarks on estimation of the density function of a population with a sample of independent observations. More detailed and precise results can be found [1, 6–7]. Consider a population with continuous positive density function $f(x)$. Let the sample of independent observations be $X_1, X_2, ..., X_n$. Let the estimate of $f(x)$ be

$$\hat{f}_n(x) = n^{-1}b(n)^{-1} \sum_{j=1}^{n} h(b(n)^{-1}(x - X_j)), \qquad (1)$$

[199]

where $h(x)$ is a given bounded continuous density function and $b(n)$ is a bandwidth parameter such that $b(n) \to 0$ and $nb(n) \to \infty$ as $n \to \infty$. The bias of the estimate is

$$b_n(x) = E\hat{f}_n(x) - f(x) = \int h(\alpha)\{f(x - b(n)\alpha) - f(x)\}\, d\alpha. \tag{2}$$

If f is continuously differentiable up to second order and the moments $m_i = \int \alpha^i h(\alpha)\, d\alpha, i = 1, 2$, exist then

$$b_n(x) = f'(x)\, m_1 b(n) + f''(x)\, m_2 b(n)^2/2 + o(b(n)^2). \tag{3}$$

Thus $b_n(x) = 0(b(n)^2)$ if $m_1 = 0$. The variance

$$\sigma^2(\hat{f}_n(x)) \cong n^{-1}b(n)^{-1}f(x)\int h^2(\alpha)\, d\alpha \quad \text{as} \quad n \to \infty. \tag{4}$$

It is clear that the estimates are asymptotically normally distributed. Specifically $\sqrt{(nb(n))}[\hat{f}_n(x) - E\hat{f}_n(x)]$ is asymptotically normally distributed with mean zero and variance $f(x) \int h^2(\alpha)\, d\alpha$. It is also interesting to look at the covariance of estimates at different points x, y.

The covariance

$$\text{cov}\,(\hat{f}_n(x), \hat{f}_n(y)) = n^{-1}b(n)^{-2}\,\text{cov}\,(h(b(n)^{-1}(x - X)), h(b(n)^{-1}(y - X))$$

$$= n^{-1}b(n)^{-2}\left\{\int h(b(n^{-1}(x - \alpha))\, h(b(n)^{-1}(y - \alpha))f(\alpha)\, d\alpha\right.$$

$$\left. -\int h(b(n)^{-1}(x - \alpha))\, f(\alpha)\, d\alpha \int h(b(n)^{-1}(y - \alpha))\, f(\alpha)\, d\alpha\right\}$$

$$= n^{-1}b(n)^{-1}\left\{\int h(\beta)\, h(b(n)^{-1}(y - x) + \beta)f(x - \beta b(n))\, d\beta\right.$$

$$\left. - b(n)\int h(\beta)f(x - b(n)\beta)\, d\beta \int h(\beta)f(y - b(n)\beta)\, d\beta\right\}. \tag{5}$$

This certainly implies that for fixed x, y with $x \neq y$

$$\sqrt{(nb(n))}[\hat{f}_n(x) - E\hat{f}_n(x)] \quad \text{and} \quad \sqrt{(nb(n))}[\hat{f}_n(y) - E\hat{f}_n(y)]$$

are asymptotically independent.

3   THE MARKOV ASSUMPTION

Let us now assume that $X_1, X_2, ..., X_n$ is a sequence of observations on a stationary Markov process with stationary continuous density function $f(x) > 0$ and continuous transition probability density function $f(y|x)$. Let $f_k(y|x)$ and $f_k(y, x)$ denote the conditional probability density of

X_{k+1} given X_1 and the joint probability density of X_{k+1} and X_1 respectively, $k = 1, 2, \ldots$. Clearly

$$f_k(y, x) = f_k(y|x) f(x) \tag{6}$$

and

$$f_{k+1}(y|x) = \int f_k(y|z) f(z|x) \, dz \quad (k = 1, 2, \ldots). \tag{7}$$

We use the same sort of estimate $\hat{f}_n(x)$ of the density function $f(x)$ as in the case of independent observations. The bias of the estimate in the case of a Markov sequence is the same as in the case of independent observations. Let us now look at the covariance of the estimate when one has dependence. Clearly

$$\operatorname{cov}[\hat{f}_n(x), \hat{f}_n(y)] = n^{-2}b(n)^{-2} \sum_{j=-n+1}^{n-1} (n - |j|) \operatorname{cov}[h(b(n)^{-1}(x - X_0)), h(b(n)^{-1}$$
$$\times (y - X_j))]. \tag{8}$$

In §5 we will determine the asymptotic behavior of the covariance and of the distribution of the estimates $\hat{f}_n(x)$ themselves, given certain asymptotic conditions on the character of the dependence of the Markov sequence. Some of these conditions will be introduced and discussed in §4.

4   REMARKS ON THE TRANSITION PROBABILITY OPERATOR

In order to determine the asymptotic behaviour of covariance of density estimates at different points as well as the asymptotic distribution of estimates we shall have to impose certain conditions on the transition probability function of the Markov sequence. The conditions discussed here are undoubtedly too strong. Certainly weaker conditions that ensure the desired results can be determined. However, these conditions have a certain interest even outside the problem of density estimation and the discussion will be given in a broader context. Let T be the transition probability operator of a stationary Markov sequence. In our problem

$$(Th)(y) = \int h(x) f(x|y) \, dx, \tag{9}$$

where h can be thought of as a bounded function. The condition to be imposed on T corresponds roughly to what has been called 'geometric ergodicity' (see Kendall [5]) in some of the probability literature. Let $\|h\|_p$, $1 \leq p \leq \infty$, be the L^p norm of the function h with respect to

the invariant measure μ $(d\mu = f(x)\,dx$ in our case of density estimation) so that

$$\|h\|_p = \begin{cases} \left[\int |h(x)|^p \mu(dx)\right]^{1/p} & \text{if} \quad 1 \leqslant p < \infty, \\ \operatorname*{ess\,sup}_x |h(x)| & \text{if} \quad p = \infty, \end{cases} \tag{10}$$

where ess sup is understood as referring to the measure μ. The L^p norm of the operator T^n, $n = 1, 2, \ldots$,

$$\|T^n\|_p = \sup_{h \neq 0} \frac{\|T^n h\|_p}{\|h\|_p} = 1 \tag{11}$$

for $1 \leqslant p \leqslant \infty$ since $T^n 1 = 1$. However, since we are interested in the rate at which $T^n h \to Eh = \int h(x)\,\mu(dx)$ as $n \to \infty$, the modified norm

$$\begin{aligned} |T^n|_p &= \sup_h \frac{\|T^n h - Eh\|_p}{\|h - Eh\|_p} \\ &= \sup_{h \perp 1} \frac{\|T^n h\|_p}{\|h\|_p} \end{aligned} \tag{12}$$

is introduced where by $h \perp 1$ one means that $\int h(x)\,\mu(dx) = 0$.

The transition probability operator T is said to satisfy the condition G_p $(1 \leqslant p \leqslant \infty)$ if there is some positive integer n such that

$$|T^n|_p \leqslant \alpha < 1. \tag{13}$$

The reason for 'geometric ergodicity' is apparent since if (13) holds, it follows that

$$\lim_{m \to \infty} |T^m|_p / \alpha^{m/n} \leqslant 1. \tag{14}$$

Lemma. If T is a transition probability operator of a stationary Markov sequence satisfying condition G_p for some $p, 1 < p < \infty$, then T satisfies the condition G_q for all $q, 1 < q < \infty$.

Let U be such that $Uh = Th - Eh$. Notice that

$$\|h - Eh\|_q \leqslant \|h\|_q + |Eh| \leqslant 2\|h\|_q \tag{15}$$

and that

$$U^n h = U^{n-1}(Th - Eh) = T^{n-1}(Th - Eh). \tag{16}$$

This implies that

$$\|U^n\|_q \leqslant 2|T^n|_q \leqslant 2, \quad 1 \leqslant q \leqslant \infty. \tag{17}$$

However

$$|T^n|_q \leqslant \|U^n\|_q. \tag{18}$$

Since $\|U^n\|_1, \|U^n\|_\infty \leqslant 2$ the Riesz convexity theorem [4] can be applied to this situation. Suppose that for some p, $1 < p < \infty$ (for example $p = 2$), there is a positive integer n such that

$$|T^n|_p = \alpha < 1. \tag{19}$$

It then follows that $\left|T^{nk}\right|_p \leqslant \alpha^k$. Consider any other q with $1 < q < \infty$. For $q > p$ we have $1/q = (p/q)(1/p) + (1 - p/q)$, while if $q < p$ then

$$1/q = \left(\frac{1 - 1/q}{1 - 1/p}\right) 1/p + \left(1 - \frac{1 - 1/q}{1 - 1/p}\right).$$

The Riesz convexity theorem tells us that if $q > p$

$$\log \|U^m\|_q \leqslant \frac{p}{q} \log \|U^m\|_p + \left(1 - \frac{p}{q}\right) \log \|U^m\|_\infty \tag{20}$$

and if $q < p$

$$\log \|U^m\|_q \leqslant \left(\frac{1 - 1/q}{1 - 1/p}\right) \log \|U^m\|_p + \left(1 - \frac{1 - 1/q}{1 - 1/p}\right) \log \|U^m\|_1. \tag{21}$$

Since $\left\|U^{nk}\right\|_p \leqslant 2\alpha^k$ and $\|U^m\|_\infty$, $\|U^m\|_1 \leqslant 2$ it is clear that for large enough k

$$\left\|U^{nk}\right\|_q \leqslant \beta < 1 \tag{22}$$

and therefore

$$\left|T^{nk}\right|_q \leqslant \beta < 1. \tag{23}$$

It is of some interest to consider the well known Doeblin condition and see how it relates to the conditions G_p. Let $P(x, A)$ be the transition probability function of the stationary Markov sequence defined for each real x and each Borel set A. Then

$$(Th)(x) = \int P(x, dy)\, h(y) \tag{24}$$

for bounded h and the n step transition probability function $P_n(x, A)$ is defined recursively via

$$P_{n+1}(x, A) = \int P(x, dy)\, P_n(y, A) \quad (n = 1, 2, \ldots). \tag{25}$$

The Doeblin condition D is said to be satisfied if there is a finite measure ϕ on the Borel sets with positive mass, an integer $n \geqslant 1$, and $\epsilon > 0$ such that

$$P_n(x, A) \leqslant 1 - \epsilon \tag{26}$$

for all x if $\phi(A) \leqslant \epsilon$. Doob (see [3]) considers the more stringent condition D_0 where one assumes that D is satisfied, there is only one ergodic set, and there are no cyclically moving sets. Under condition D_0, in Lemma 7.2 on p. 224 of Doob's book on stochastic processes, it is shown that there are constants γ and ρ, with $0 < \rho < 1$, such that if h is bounded

$$(|h| \leqslant M < \infty)$$

then

$$\left|(T^k h)(x) - Eh\right| \leqslant 2\gamma \rho^k. \tag{27}$$

However, it is easily seen that (27) is just the condition G_∞. By the Riesz convexity theorem, it follows that if G_∞ is satisfied then G_p holds for $1 < p \leqslant \infty$. Similarly, if G_1 is satisfied then G_p holds for $1 \leqslant p < \infty$.

Notice that the condition D_0 is satisfied if we have an ergodic aperiodic stationary Markov sequence with continuous stationary probability density $f(x)$ and transition probability density $p(y|x)$ satisfying

$$p(y|x) \leqslant K(1+y^2)^{-\alpha} \tag{28}$$

for some constant K and $\alpha > 1/2$.

However, it is easy to give a simple example which doesn't satisfy D or G_∞ but for which G_p, $1 < p < \infty$, holds. This example is provided by the aperiodic Gaussian stationary Markov sequences. For convenience, assume that the random variables of the sequence have mean zero and variance one. The invariant probability density is then

$$f(x) = \frac{1}{\sqrt{(2\pi)}} \exp\left(-x^2/2\right)$$

and the transition probability density is

$$p(y|x) = \frac{1}{\sqrt{(2\pi(1-\rho^2))}} \exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right) \tag{29}$$

with ρ a constant such that $|\rho| < 1$. Mehler's formula indicates that (29) can be written

$$p(y|x) = \sum_{\nu=0}^{\infty} \rho^\nu h_\nu(y) h_\nu(x) \frac{e^{-\nu^2/2}}{\sqrt{(2\pi)}}, \tag{30}$$

where the h_ν are the Hermite polynomials orthonormal with respect to the standard Gaussian density $f(x)$. By using (3) one can see that

$$|T^n|_2 = |\rho|^{2n} \tag{31}$$

so that condition G_2 is certainly satisfied. By the Lemma of this section it follows that G_p is satisfied for $1 < p < \infty$.

5 ASYMPTOTIC BEHAVIOR OF THE COVARIANCE AND DISTRIBUTION OF ESTIMATES

The asymptotic behavior of the covariance of a density estimate at two points x, y will first be determined before the asymptotic distribution is considered. Assume that h is a continuous density function with

$$h(u) = o(|u|^{-1}) \quad \text{as} \quad |u| \to \infty.$$

The density f is taken to be bounded and continuous. Consider the term corresponding to $j = 0$ in the sum (8).

Then $\qquad \operatorname{cov}\{h(b(n)^{-1}(x-X)), h(b(n)^{-1}(y-X))\}$

$$= b(n) \int h(u)\, h(b(n)^{-1}(y-x)+u) f(x-b(n)\, u)\, du$$

$$- b(n)^2 \int h(u) f(x-b(n)\, u)\, du \int h(u) f(y-b(n)\, u)\, du$$

$$\cong \begin{cases} b(n) f(x) \int h^2(u)\, du & \text{if} \quad x = y \\ 0(b(n)^2) & \text{if} \quad x \neq y \end{cases} \tag{32}$$

when $f(x) > 0$ as $n \to \infty$. The terms with $j \neq 0$ have the following asymptotic behavior

$$\operatorname{cov}\{h(b(n)^{-1}(x-X_0)), h(b(n)^{-1}(y-X_j))\}$$

$$= b(n^2) \int h(u)\, h(v) f_j(x-b(n)\, u, y-b(n)\, v)\, du\, dv$$

$$- b(n)^2 \int h(u) f(x-b(n)\, u)\, du \int h(u) f(y-b(n)\, u)\, du$$

$$\cong b(n)^2 \{f_j(x,y) - f(x) f(y)\} \tag{33}$$

as $n \to \infty$ if the joint density functions $f_j(\cdot, \cdot)$ are bounded continuous functions. We shall now get a bound on (33) under the assumption that the Markov sequence satisfies condition G_2. Let

$$g_x(\alpha) = h(b(n)^{-1}(x-\alpha)) - \int h(b(n)^{-1}(x-\alpha)) f(\alpha)\, d\alpha. \tag{34}$$

Then $\qquad |\operatorname{cov}\{h(b(n)^{-1}(x-X_j)), h(b(n)^{-1}(y-X_0))\}|$

$$= |E(g_y(X)\,(T^j g_x)\,(X))|$$

$$\leqslant \{E|g_y(X)|^2\, E|(T^j g_x)\,(X)|^2\}^{\frac{1}{2}}$$

$$\leqslant M\rho^j \{E|g_y(X)|^2\, E|g_x(X)|^2\}^{\frac{1}{2}}$$

$$\leqslant b(n)\, M'\rho^j \int h^2(u)\, du\, \sqrt{(f(x) f(y))} \tag{35}$$

for sufficiently large n with M, M' and ρ constants, $0 < \rho < 1$, if $f(x)$ and $f(y)$ are positive. The estimates (32), (33) and (35) indicate that

$$\lim_{n \to \infty} nb(n) \operatorname{cov}[\hat{f}_n(x), \hat{f}_n(y)]$$

$$= \begin{cases} f(x) \int h^2(u)\, du & \text{if} \quad x = y \\ 0 & \text{if} \quad x \neq y \end{cases} \tag{36}$$

under the conditions assumed.

We now consider the asymptotic distribution of

$$\sqrt{(nb(n))}\,[\hat{f}_n(x_j) - E\hat{f}_n(x_j)] \quad (j = 1, \ldots, m),$$

for any finite m-tuple of points x_j. These random variables are asymptotically jointly normal and independent with means zero and variances given by (36) as $n \to \infty$ under the assumptions made. So as to limit the notation the argument on asymptotic normality will be given for a single point, that is, $m = 1$. However, exactly the same argument and estimates can be used to obtain the multivariate result by applying them to any given linear combination

$$\sqrt{(nb(n))}\sum_{j=1}^{m}\alpha_j(\hat{f}_n(x_j) - E\hat{f}_n(x_j))$$

of the random variables. This is the standard reduction of a multivariate limit theorem to a univariate limit theorem. Consider

$$\sqrt{(nb(n))}\,[\hat{f}_n(x) - E\hat{f}_n(x)].$$

The argument for asymptotic normality will proceed by the classic device of proving such a limit theorem for dependent random variables, that is, by writing the sum as a sum of big blocks separated by small blocks. The contribution due to the small blocks is shown to be negligible and the big blocks approximately independent. A standard central limit theorem for independent random variables (the Liapounov theorem) is then used to get asymptotic normality. Now,

$$\sqrt{(nb(n))}\,[\hat{f}_n(x) - E\hat{f}_n(x)]$$

$$= \sum_{j=1}^{n}\frac{1}{\sqrt{(nb(n))}}\,[h(b(n)^{-1}(x - X_j)) - Eh(b(n)^{-1}(x - X_j))]$$

$$= \sum_{l=1}^{k}(A_l + B_l) + H \tag{37}$$

with

$$A_l = \sum_{j=(l-1)(m+r)+1}^{lm+(l-1)r}, \quad B_l = \sum_{j=lm+(l-1)r+1}^{l(m+r)} \tag{38}$$

and

$$H = \sum_{j=k(m+r)+1}^{n}. \tag{39}$$

Here $m = m(n)$ and $r = r(n)$ are the summation ranges of the big and small blocks A_l and B_l respectively. They will be taken so that

$$m(n), r(n) \to \infty \quad \text{as} \quad n \to \infty \quad \text{but} \quad m(n) = o(n) \quad \text{and} \quad r(n) = o(m(n)).$$

Further $k = k(n) = [n/(m+r)]$ is the greatest integer less than or equal to $n/(m+r)$. The term H accounts for the additional few terms at the

end not included in the big blocks or small blocks. Notice that $k(n) \to \infty$ as $n \to \infty$ since $m, r = o(n)$. Estimates like those used in (32), (33) and (35) imply that

$$E \left| \sum_{l=1}^{k} B_l + H \right|^2 \leqslant 2 \left(\sum_{l=1}^{k-1} E |B_l|^2 + E |B_k + H|^2 \right)$$

$$\leqslant C \left(\frac{kr(n)}{n} + \frac{m}{n} \right) = o(1), \tag{40}$$

with C a constant.

Thus $\sum_{l=1}^{k} B_l + H$ can be neglected. Consider the characteristic function of $\sum_{l=1}^{k} A_l$. We use the G_2 property of the Markov sequence to compare this characteristic function with the product of the characteristic functions of the individual A_l's and find that

$$\left| E \left\{ \exp \left(it \sum_{l=1}^{k} A_l \right) \right\} - \prod_{l=1}^{k} E \{ \exp (it A_l) \} \right|$$

$$\leqslant \sum_{j=2}^{k} \left| E \left\{ \exp \left(it \sum_{l=1}^{j} A_l \right) \right\} - E \{ \exp (it A_j) \} E \left\{ \exp \left(it \sum_{l=1}^{j-1} A_l \right) \right\} \right|$$

$$\leqslant (k-1) M \rho^{r(n)}, \tag{41}$$

where M is a constant and $0 < \rho < 1$. The sequences $m(n)$, $r(n)$ will be chosen so that $k(n) \rho^{r(n)} \to 0$. However this means that we have asymptotic normality if we can show that $\sum_{l=1}^{k} A_l$ is asymptotically normal when the A_l's are treated as independent random variables with the same marginal distributions. This will be shown to be true by using the Liapounov theorem. For this we require an estimate of the fourth moment of the A_l's. Given that the joint density functions up to fourth order are continuous and bounded, estimates like those obtained in (32) and (33) show that

$$E |A_l|^4 \leqslant \sum_{j=1}^{k} M_j (mb(n))^j, \tag{42}$$

where the M_j's are constants. But then

$$\frac{k E |A_l|^4}{(E |\sum A_l|^2)^2} \leqslant \frac{M_1}{nb(n)} + M_2 \frac{m}{n} + M_3 \frac{m^2}{n} b(n) + M_4 \frac{m^3}{n} b(n)^2. \tag{43}$$

If $nb(n) \to \infty$, $b(n) \to 0$, but $m(n) = o(n^{\frac{1}{3}})$ the expression (43) tends to zero as $n \to \infty$ and the Liapounov theorem is applicable. We can still choose $r(n) = o(m(n))$ so that (41) tends to zero as $n \to \infty$.

Theorem 1. *Let* $X_n, n = \dots, -1, 0, 1, \dots$, *be a stationary Markov sequence with bounded continuous density function* $f(x)$ *and bounded transition probability density* $p(y|x)$ *continuous in the variables* (y, x). *Consider the estimate* $\hat{f}_n(x)$ *of the density* $f(x)$ *given by*

$$\hat{f}_n(x) = n^{-1} b(n)^{-1} \sum_{j=1}^{n} h(b(n)^{-1}(x - X_j)),$$

where h *is a given uniformly continuous density function. If* f *is continuously differentiable up to second order and the moments* $m_i = \int \alpha^i h(\alpha) \, d\alpha$, $i = 1, 2$, *of* h *exist then the bias*

$$E\hat{f}_n(x) - f(x) = m_1 f'(x) b(n) + m_2 f''(x) b(n)^2 + o(b(n))^2,$$

where the bandwidth $b(n) \to 0$ *as* $n \to \infty$. *If the sequence* $\{X_n\}$ *satisfies condition* G_2 *and* $nb(n) \to \infty$ *as* $n \to \infty$ $(b(n) \to 0)$ *any* m-*tuple* (m *finite*) *of the normalized deviations*

$$\sqrt{(nb(n))} \, [\hat{f}_n(x_i) - E\hat{f}_n(x_i)] \quad (i = 1, \dots, m)$$

(x_i *distinct*) *are asymptotically normal and independent with mean zero and variances*

$$f(x_i) \int h^2(\alpha) \, d\alpha \quad (i = 1, \dots, m).$$

One could equally well consider estimation of the joint density function of a fixed number of random variables. Under conditions corresponding to those assumed in Theorem 1, one again finds that the asymptotic results are the same as those in the case of independent observations. Actually we shall explicitly note one such result under somewhat broader conditions of dependence. Now $X_n, n = \dots, -1, 0, 1, \dots,$ is assumed to be a stationary sequence of random variables. *Assume that all joint distributions of a finite number of distinct* X_j's *are absolutely continuous with uniformly bounded continuous density functions.* Let $f(x, x')$ be the joint density function of X_j, X_{j+1}. A natural estimate of $f(x, x')$ is given by

$$f_n(x, x') = n^{-1} b(n)^{-2} \sum_{j=1}^{n-1} h(b(n)^{-1}(x - X_j), b(n)^{-1}(x' - X_{j+1})), \quad (44)$$

where $h(x, x')$ is a bounded continuous density function. Let $\mathscr{B}_n$ be the Borel field generated by $X_k, k \leqslant n$, and $\mathscr{F}_n$ be the Borel generated by $X_k, k \geqslant n$. We shall say that $\{X_m\}$ *satisfies the condition* S *if for every random variable* g *with* $Eg^2 < \infty, Eg = 0$, *that is* $\mathscr{F}_{n+k}$ *measurable it follows that*

$$E|E(g|\mathscr{B}_n)|^2 \leqslant a(k) E|g|^2, \quad (45)$$

where $a(k) = O(k^{-4-\epsilon})$ for some $\epsilon > 0$ as $k \to \infty$. By making estimates quite similar to those in the proof of the Theorem one can prove the following result.

Theorem 2. Let $\{X_m, m = \ldots, -1, 0, 1, \ldots\}$ be stationary with uniformly bounded continuous joint density functions. Consider the estimate $\hat{f}_n(x, x')$ of the joint density $f(x, x')$ of X_j, X_{j+1} given by (44). If f is continuously differentiable up to second order and the moments,

$$m_{i,j} = \int \alpha^i \beta^j h(\alpha, \beta) \, d\alpha \, d\beta \quad (i, j = 0, 1, 2),$$

when $i + j \leqslant 2$ exist then the bias

$$Ef_n(x, x') - f(x, x') = [m_{1,0}f_x(x, x') + m_{0,1}f_{x'}(x, x')] b(n)$$
$$+ \tfrac{1}{2}[m_{2,0}f_{x,x} + 2m_{1,1}f_{x,x'} + m_{0,2}f_{x',x'}] b(n)^2 + o(b(n)^2) \quad (46)$$

as $b(n) \to 0$. If $\{X_m\}$ satisfies the condition S and $nb(n)^2 \to \infty$ as

$$n \to \infty \quad (b(n) \to 0)$$

then any finite m-tuple of the normalized deviations

$$\sqrt{(nb(n)^2)}[\hat{f}_n(x, x') - E\hat{f}_n(x, x')] \tag{47}$$

corresponding to distinct points (x, x') are independent with mean zero and variances

$$f(x, x') \int h^2(\alpha, \beta) \, d\alpha \, d\beta. \tag{48}$$

Regression estimates have been considered in [8] and [10]. Without going into details, one can show that under conditions similar to those specified in [8], the asymptotic behavior of these in the case of dependent observations is the same as the asymptotic behavior in the case of independent observations. In [9] estimation of the transition probability density for Markov sequences is discussed under condition D_0.

6 ACKNOWLEDGEMENT

REFERENCES

[1] Bartlett, M. S. (1963). Statistical estimation of density functions. *Sankhyā*, Ser. A **25**, 245–54.

[2] Billingsley, P. (1968). *Convergence of probability measures.* New York: John Wiley.

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

[4] Dunford, N. and Schwartz, J. L. (1958). *Linear Operators*, Part I. Interscience.

[5] Kendall, D. G. (1959). Unitary dilations of Markov transition operators, and corresponding integral representations for transition probability matrices. In *Probability and statistics* (ed. U. Grenander). New York: John Wiley.

[6] Parzen, E. (1962). On estimation of a probability density and mode, *Ann. Math. Statist.* **33**, 1065–76.

[7] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.* **27**, 832–5.

[8] Rosenblatt, M. (1969). Conditional probability density and regression estimators. In *Multivariate Analysis*, vol. II. Academic Press.

[9] Roussas, G. (1967). Nonparametric estimation in Markov processes, *Technical Report No.* 110, Department of Statistics, University of Wisconsin, Madison, Wisconsin.

[10] Watson, G. S. (1964). Smooth regression analysis, *Sankhyā*, Ser. A **26**, 359–72.

DISCUSSION ON ROSENBLATT'S PAPER

MICHAEL WOODROOFE

I believe that Professor Rosenblatt's paper constitutes a significant contribution to the nonparametric literature and therefore feel quite honored to be a discussant of it. The paper not only contributes a major result by giving sufficient conditions for the asymptotic normality of density estimates from stationary Markov processes, but should also stimulate more research on the problem of density estimation from such processes. Indeed, it is to be hoped that Professor Rosenblatt's present paper on density estimation from dependent observations will stimulate as much research as did his first paper on density estimation from independent observations [4]. Among other things, I will indicate some directions in which I feel this research might profitably proceed.

In considering the asymptotic normality of the estimates, Professor Rosenblatt has solved a rather difficult problem and has left some less difficult results unstated. Since I feel these results to be important, I will take the liberty of stating some of them together with an indication of their proofs. Let X_n be a stationary process satisfying condition S with $a(k)$ summable. (This includes the case of a stationary Markov Process which satisfies G_2.) The argument of [3] then extends without difficulty to show that

$$M_n = \sup_x |\hat{f}_n(x) - f(x)| \to 0 \tag{1}$$

in probability as $n \to \infty$, provided that f is uniformly continuous, $\sqrt{[nb(n)]} \to \infty$, and the Fourier Transform of h is integrable. Thereafter, it follows, again as in [3], that the sample mode is a consistent estimate of the true mode, provided that the latter is unique. It would also be of interest to know the rate of convergence to zero of M_n and such related random variables as were studied in [6] and to determine conditions for and the exact rate of convergence to zero of the integrated mean square error.

$$I_n = \int_{-\infty}^{\infty} E[(f_n(x) - f(x))^2] \, dx.$$

To the best of my knowledge, the only other work which seriously considers density estimation from Markov processes is the unpublished report [5], in which consistency and asymptotic normality of kernel-type density estimates are obtained under the condition D_0. It is therefore of interest to consider Professor Rosenblatt's condition G_2 with some

[211]

care. It is quite intuitively appealing in that it requires the dependence between segments of the process to decrease as the distance between them increases. I was therefore intrigued by the G_2 condition and hoped that it might be satisfied by a large enough class of stationary Markov processes to provide a reasonable setting for asymptotic inference from such processes. I still have this hope, but in the course of attempting to verify G_2 for linear processes of the form

$$X_k = \sum_{j=0}^{\infty} \beta^j Y_{k-j} \quad (k = 0, 1, 2, \ldots), \tag{2}$$

where the Y_i are independent and identically distributed and $|\beta| < 1$, I discovered an example which has shaken that hope slightly. If the Y_i have the distribution function

$$F_1(y) = 1 - (\log y)^{-2} : x \geqslant e, \tag{3}$$

then (2) does not satisfy G_2. Indeed, by considering LaPlace transforms and applying the Tauberian theorems of [1], one may show that F, the stationary distribution function, has a slowly varying tail. Letting

$$g_n(x) = (1 - F(n))^{-\frac{1}{2}} \quad \text{for} \quad x > n \quad \text{and} \quad g_n(x) = 0 \quad \text{for} \quad x \leqslant n,$$

it is then easily verified that

$$|T^k g_n|_2^2 = \|T^k g_n\|_2^2 - o(1) \geqslant \left[\frac{1 - F(n\beta^{-k})}{1 - F(n)}\right] - o(1) \to 1 \quad \text{as} \quad n \to \infty,$$

for each fixed k. Since $|g_n|_2 \leqslant \|g_n\|_2^2 = 1$ for all n, G_2 cannot be satisfied.

The distribution function in (3) is, of course, sufficiently pathological to be of little practical interest, but I still find this example disturbing. Since $X_k = \beta^k X_0 + Z$, where Z is independent of X_0, the process defined by (2) has the property that the dependence of X_k on X_0 decreases geometrically as $k \to \infty$, which property led me to believe that any process of the form (2) should be geometrically ergodic, that it should satisfy G_2. I believe the generality of the G_2 condition to be a question on which research could profitably be done. If it should turn out to be reasonably general, it might well simplify subsequent work on asymptotic inference from Markov processes.

Another area in which I feel research could profitably be done is that of density estimation from general linear processes of the form

$$X_k = \sum_{j=0}^{\infty} c_j Y_{k-j} \quad (k = 0, 1, 2, \ldots),$$

where the Y_i are as in (2). What conditions on the coefficients c_j and/or the distribution of Y_i are sufficient to insure the consistency and asymptotic normality of density estimates from such processes? These questions could be attacked either by verifying an appropriate version of condition S or from scratch, perhaps by using the techniques of [2], Part II. The former approach would be more desirable since conditions similar to S appear throughout the probability literature (e.g. in [2]), but the latter approach would probably be the easier.

REFERENCES

[1] Feller, William (1966). *An Introduction to probability theory and its applications*, **2**. New York: John Wiley.

[2] Ibragimov, I. A. (1962). Some limit theorems for stationary processes, *Theory Probab. Applic.* **7**, 349–82.

[3] Parzen, E. (1962). On estimation of a probability density and mode, *Ann. Math. Statist.* **33**, 1065–76.

[4] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density, *Ann. Math. Statist.* **27**, 832–5.

[5] Roussas, G. (1967). Non-parametric estimation in Markov Processes, *Technical Report No.* 110, Department of Statistics, University of Wisconsin, Madison, Wisconsin.

[6] Woodroofe, M. (1967). On the maximum deviation of the sample density, *Ann. Math. Statist.* **38**, 475–81.