Stochastic Processes and their Applications 21 (1986) 179-193
North-Holland

ON SMOOTHED PROBABILITY DENSITY ESTIMATION FOR STATIONARY PROCESSES

J.V. CASTELLANA and M.R. LEADBETTER*
Department of Statistics, University of North Carolina, Chapel Hill, NC 27514, USA

Received 12 June 1984
Revised 28 June 1985

Aspects of estimation of the (marginal) probability density for a stationary sequence or continuous parameter process, are considered in this paper. Consistency and asymptotic distributional results are obtained using a class of smoothed function estimators including those of kernel type, under various decay of dependence conditions for the process. Some of the consistency results contain convergence rates which appear to be more delicate than those previously available, even for i.i.d. sequences.

density estimation * stationary processes * stochastic processes

1. Introduction

For very many years there has been a great deal of interest in the problem of estimating the underlying probability density function (p.d.f.) from an i.i.d. sample $X_1, \ldots, X_n$ (cf. [3, 11] for reviews of this literature). More recently some attention has been given to estimating the marginal p.d.f. in dependent (and especially Markov) contexts (e.g. [8, 6]). The present paper is concerned with the case of *stationary* sequences and continuous parameter processes — a topic considered also in [5] but with somewhat different objectives and assumptions.

The estimators to be used are of 'smoothing function type', which may be of kernel form or may be based on the somewhat more general 'δ-sequences' and 'δ-families' satisfying appropriate axiom schemes. A perhaps more important point concerns the nature of the assumed dependence structure in the sequence or process. Here we use a 'dependence index' (developing a condition used in [8]), based simply on the differences between bivariate p.d.f. and product of univariate p.d.f.'s, in order to obtain consistency results. This is of course much easier to verify than the stronger mixing conditions (used, e.g., in [5]) and seems more likely to hold in practical cases (even though it is certainly possible for, e.g., a strongly mixing process to have no bivariate densities). An (array) form of the strong mixing assumptions will be used in connection with asymptotic distributional results where a stronger restriction is clearly needed.

* This research was supported by Contract N00014-75-C-0809 with the Office of Naval Research and Contract F49620-82-C-0009 with the Air Force Office of Scientific Research.

Preliminary results given in Section 2 are of a standard type and used in developing pointwise consistency for estimators in the case of sequences, in Section 3. These results mainly generalize familiar i.i.d. theorems in an obvious way but a sharpened form of variance convergence appears to be new even for i.i.d. cases. Section 4 concerns asymptotic normality in the sequence case. Finally, in Section 5, corresponding results are developed for continuous parameter processes observed in an interval $(0, T)$. Perhaps the most significant points of interest in this are the greater consistency rates arising from continuous sampling — a 'full rate $1/T$' of convergence of the variance to zero even being achieved for certain processes.

As noted, this paper concerns consistency and asymptotic distributional results for stationary sequences and continuous parameter processes. Further related results (involving global error measures and a.s. consistency) may be found in the thesis [1] or [2].

2. Notation and preliminary results

The following notation and assumptions will be used without comment in Sections 2-4. As noted $\{X_j; j=1,2,\ldots\}$ will be a stationary sequence with marginal probability density $f(x)$. It will be assumed when relevant that the bivariate distributions of the sequence are absolutely continuous, writing $f_j(x, y)$ for the joint density of X_1 and X_{1+j}, $j=1,2,\ldots$ (assumed finite for each j, x, y).

The primary measure of dependence of the sequence $\{X_j\}$ will be through the quantities (cf. [8])

$$\beta_n = \sup_{x,y} \sum_{i=1}^{n} |f_i(x,y) - f(x)f(y)|, \tag{2.1}$$

which is finite for each n. We refer to $\{\beta_n;\ n \geq 1\}$ as the *dependence index sequence* for the process $\{X_j;\ j=1,\ldots\}$. Clearly for i.i.d. sequences $\beta_n \equiv 0$ for all n, for sequences with high long range dependence β_n may tend to infinity, and in between β_n may converge to a finite limit at various rates. The following example illustrates the behavior of β_n in the particular case of certain stationary normal sequences.

Example. Let $\{X_i;\ i=1,2,\ldots\}$ be a stationary normal sequence with zero means, unit variances, and correlations $r_j = E(X_1 X_{1+j})$. If $|r_j| < \delta$ for all $j \neq 0$, and some $\delta < 1$, then

$$\beta_n \leq K \sum_{i=1}^{n} |r_i| \tag{2.2}$$

for some constant K. This may be readily checked by writing $\phi_r(x, y)$ for the standard bivariate normal density with correlation r and, regarding ϕ_r as a function r, using the mean value theorem to show that $|\phi_r(x,y) - \phi_0(x,y)| \leq K|r|$.

As noted, the marginal density $f(x)$ is to be estimated from the first n values $X_1, X_2, \ldots, X_n$ by the smoothed estimator

$$\hat{f}_n(x) \equiv \frac{1}{n} \sum_{i=1}^{n} \delta_n(x - X_i) \tag{2.3}$$

where the smoothing functions $\{\delta_n;\ n \geq 1\}$ are required to satisfy the following axiom scheme (cf. [4, 10]):

(i) $\int |\delta_n(x)|\, dx < A$, all n, some fixed A.
(ii) $\int \delta_n(x)\, dx = 1$, all $n \geq 1$.
(iii) $\delta_n(x) \to 0$ uniformly in $|x| > \lambda$, for any fixed $\lambda > 0$.
(iv) $\int_{|x| \geq \lambda} |\delta_n(x)|\, dx \to 0$, as $n \to \infty$, for any fixed $\lambda > 0$.

A sequence of functions $\{\delta_n(x);\ n \geq 1\}$ satisfying these axioms will be called simply a *δ-sequence.* An example of such a sequence is provided by the commonly used 'smoothing kernels' for which

$$\delta_n(x) = A_n^{-1} k(A_n^{-1} x) \tag{2.4}$$

where k is a bounded probability density on the real line such that $xk(x) \to 0$ as $x \to \infty$ and $A_n > 0$, $A_n \to 0$ as $n \to \infty$. The basic use of δ-sequences is through commonly used results of the following type.

Lemma 2.1. *If $g(x)$ is continuous at $x = 0$, $\int |g(x)|\, dx < \infty$, and if $\{\delta_n(x);\ n \geq 1\}$ is a δ-sequence, then $\int g(x)\delta_n(x)\, dx \to g(0)$ as $n \to \infty$.*

The result holds without δ-axiom (iii) if g is bounded (but not necessarily in L_1).

Lemma 2.2. *If $\{\delta_n(x);\ n \geq 1\}$ is a δ-sequence and if $\alpha_n = \int \delta_n^2(x)\, dx < \infty$, then $\alpha_n \to \infty$, as $n \to \infty$. Further, the functions $\delta_n^*(x) \equiv \delta_n^2(x)/\alpha_n$ define a δ-sequence.*

Lemma 2.3. *Let $g(u, v)$ be a bounded measurable function which is continuous at the point (x, y), and let $\{\delta_n(x);\ n \geq 1\}$ be a δ-sequence. Then*

$$\iint \delta_n(u - x)\delta_n(v - y) g(u, v)\, du\, dv \to g(x, y) \qquad \text{as } n \to \infty.$$

Many variants of the axiom scheme and of these results are possible but those stated will be convenient for our purposes here. The proofs of Lemmas 2.1–2.3 are straightforward (cf. [4, 10, 1]) and will therefore be omitted.

3. Pointwise consistency

It is trivial to show that the estimator $\hat{f}_n(x)$ is asymptotically unbiased, precisely as for i.i.d. sequences. Specifically the following elementary result holds.

Theorem 3.1. *If the density f is continuous at the point x, then $E[\hat{f}_n(x)] \to f(x)$, as $n \to \infty$.*

Proof. This follows immediately since clearly

$$E[\hat{f}_n(x)] = \int \delta_n(x-u)f(u)\,\mathrm{d}u = \int \delta_n(u)f(x-u)\,\mathrm{d}u \tag{3.1}$$

which converges to $f(x)$ by Lemma 2.1 (with $g(u)=f(x-u)$). □

A discussion of the variance of $\hat{f}_n(x)$ relies on the following simple lemma.

Lemma 3.2. *If the stationary sequence $\{X_i;\ i=1,2,\ldots\}$ has dependence index sequence $\{\beta_n;\ n\geq 1\}$ then, for any fixed real x, y,*

$$\sum_{i=1}^{n} |\mathrm{Cov}(\delta_n(x-X_1),\ \delta_n(y-X_{1+i}))| = \mathrm{O}(\beta_n). \tag{3.2}$$

Proof. The left hand side of (3.2) clearly does not exceed

$$\sum_{i=1}^{n} \iint |\delta_n(x-u)\delta_n(y-v)| \cdot |f_i(u,v) - f(u)f(v)|\,\mathrm{d}u\,\mathrm{d}v$$

$$\leq \beta_n \iint |\delta_n(x-u)\delta_n(y-v)|\,\mathrm{d}u\,\mathrm{d}v \leq A^2\beta_n,$$

where A is as in δ-axiom (i). □

The first result for the asymptotic form of the variance now follows.

Theorem 3.3. *Let the δ-sequence $\{\delta_n;\ n\geq 1\}$ be such that $\alpha_n = \int \delta_n^2(x)\,\mathrm{d}x < \infty$ for each n. If the stationary sequence $\{X_j;\ j\geq 1\}$ has dependence index sequence $\{\beta_n;\ n\geq 1\}$ and if $\beta_n = \mathrm{o}(\alpha_n)$ as $n\to\infty$, then*

$$n\alpha_n^{-1}\,\mathrm{Var}[\hat{f}_n(x)] \to f(x) \quad \text{as } n\to\infty$$

for any continuity point x of f.

Proof. Clearly

$$n\alpha_n^{-1}\,\mathrm{Var}[\hat{f}_n(x)] = n^{-1}\alpha_n^{-1}\,\mathrm{Var}\left[\sum_{i=1}^{n} \delta_n(x-X_i)\right] = \alpha_n^{-1}\,\mathrm{Var}[\delta_n(x-X_i)]$$

$$+2\alpha_n^{-1}\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)\mathrm{Cov}[\delta_n(x-X_i),\ \delta_n(x-X_{1+i})]. \tag{3.3}$$

The first term on the right may be written as

$$\alpha_n^{-1}\int \delta_n^2(x-u)f(u)\,\mathrm{d}u - \alpha_n^{-1}\left(\int \delta_n(x-u)f(u)\,\mathrm{d}u\right)^2$$

$$= \int \delta_n^*(x-u)f(u)\,\mathrm{d}u - \alpha_n^{-1}(f(x)+\mathrm{o}(1))^2$$

using Lemma 2.1 and writing $\delta_n^*(u)=\delta_n^2(u)/\alpha_n$. By Lemmas 2.2 and 2.1, this converges to $f(x)$, as $n\to\infty$. On the other hand the modulus of the second term on the right of (3.3) is no greater than

$$2\alpha_n^{-1}\sum_{i=1}^{n}|\mathrm{Cov}(\delta_n(x-X_1),\,\delta_n(x-X_{1+i}))|=\mathrm{O}(\beta_n/\alpha_n) \tag{3.4}$$

by Lemma 3.2 and this tends to zero since $\beta_n=\mathrm{o}(\alpha_n)$. □

Theorem 3.3 shows that the variance of the estimator behaves much the same way in the dependent case as when the variables are i.i.d.

It may also be noted that for kernel estimators the condition $\beta_n=\mathrm{o}(\alpha_n)$ may be interpreted as requiring that the product (window size) × (dependence index) tends to zero.

Theorems 3.1 and 3.3 demonstrate that the bias and variance of $\hat{f}_n(x)$ tend to zero provided $\alpha_n=\mathrm{o}(n)$ and hence that the estimator is then mean square consistent. However, Theorem 3.3 gives a rate of convergence of the variance whereas Theorem 3.1 merely shows convergence of the bias to zero. It is possible by making further general assumptions about the density and δ-sequence to obtain a rate of convergence for the bias (and hence for the mean square error). This was shown first by Parzen [7] for i.i.d. sequences and as we have noted, the bias is in no way altered by the introduction of dependence. Further, an extension of this method gives a convergence result for the variance which is yet more precise than that of Theorem 3.3, and which appears to be new even for i.i.d sequences. We show this, since it is very simple, in the important case when the density f has a bounded second derivative. More general cases (e.g. involving the existence of a 'characteristic coefficient' as used in [7] or similar assumptions of subsequent literature, e.g. [5]) may be similarly considered. The bias result of Parzen involved smoothing functions of kernel form, but may be stated in this present context as follows.

Theorem 3.4. *Let the density f have a continuous, bounded second derivative f''. Let $\{\delta_n;\ n\geq 1\}$ be a non-negative δ-sequence such that for each n, $\delta_n(x)$ is even, $\theta_n\equiv\int x^2\delta_n(x)\,\mathrm{d}x<\infty$ and $\theta_n^{-1}\int_{\{|x|>\lambda\}}x^2\delta_n(x)\,\mathrm{d}x\to 0$ for each $\lambda>0$. Then the bias $b[\hat{f}_n(x)]=E[\hat{f}_n(x)]-f(x)$ satisfies*

$$b[\hat{f}_n(x)]=\tfrac{1}{2}\theta_n f''(x)+\mathrm{o}(\theta_n)\quad \text{as } n\to\infty.$$

Proof. This follows simply from (3.1) and Lemma 2.1 by writing

$$f(x-u)=f(x)-uf'(x)+\tfrac{1}{2}u^2f''(x-\eta_u),\quad |\eta_u|<u,$$

and noting that $\theta_n^{-1}u^2\delta_n(u)$ satisfies (i), (ii), and (iv) of the δ-axioms. □

The sharper form of Theorem 3.3 may now be obtained. In this the error rate is more precisely defined (even for i.i.d. sequences) and the effect of dependence is made clear.

Theorem 3.5. *Let the density f have a bounded second derivative f''. Let $\{\delta_n(x);\ n\geq 1\}$ be a non-negative δ-sequence with each δ_n even, and such that $\alpha_n=\int\delta_n^2(x)\,dx<\infty$, $\theta_n^*=\int x^2\delta_n^*(x)\,dx<\infty$ (where $\delta_n^*(x)=\delta_n^2(x)/\alpha_n$) and $\theta_n^{*-1}\int_{\{|x|>\lambda\}}x^2\delta_n^*(x)\,dx\to 0$ for each $\lambda>0$. Then*

$$\frac{n}{\alpha_n}\operatorname{Var}[\hat{f}_n(x)]=f(x)+\tfrac{1}{2}\theta_n^*f''(x)(1+o(1))-\alpha_n^{-1}f^2(x)(1+o(1))+O(\beta_n/\alpha_n). \tag{3.5}$$

where $\{\beta_n;\ n\geq 1\}$ is the dependence index sequence for the process $\{X_j;\ j\geq 1\}$.

Proof. As in Theorem 3.3 we obtain

$$n\alpha_n^{-1}\operatorname{Var}[\hat{f}_n(x)]=\int\delta_n^*(x-u)f(u)\,du-\alpha_n^{-1}f^2(x)(1+o(1))+O(\beta_n/\alpha_n). \tag{3.6}$$

The first term in (3.6) is just the expected value of an estimator $\hat{f}_n^*(x)$ of $f(x)$ based on δ_n^* rather than on δ and hence by Theorem 3.4 is just

$$f(x)+\tfrac{1}{2}\theta_n^*f''(x)(1+o(1))$$

giving (3.5). □

Note that the relative magnitude of θ_n^*, α_n^{-1} and β_n/α_n determine which terms should be kept in (3.5). In the i.i.d. case (when $\beta_n=0$) the final term drops out giving a sharper result than is usually stated in that case. For dependent cases where $\beta_n\neq 0$ the term $\alpha_n^{-1}f^2(x)(1+o(1))$ should be omitted since it is no larger than the final term. Also the conditions of the theorem are readily checked for (non-negative, even) kernel functions k satisfying $\int x^2k(x)\,dx<\infty$.

The final result of this section gives a rate of convergence to zero for the covariance between $\hat{f}_n(x)$ and $\hat{f}_n(y)$ when $x\neq y$. Since the methods involve similar calculations to those above, the proof will be sketched only.

Theorem 3.6. *Let $\{X_j;\ j\geq 1\}$ be a stationary sequence, with dependence index sequence $\{\beta_n;\ n\geq 1\}$. Then if x and y are distinct continuity points of the probability density function f,*

$$n\operatorname{Cov}[\hat{f}_n(x),\hat{f}_n(y)]=-f(x)f(y)(1+o(1))+O(\beta_n)$$

where the term $O(\beta_n)$ does not exceed $2A^2\beta_n$, A being the constant in δ-axiom (i).

Proof. It is readily seen that

$$n\operatorname{Cov}[\hat{f}_n(x),\hat{f}_n(y)]=n^{-1}\sum_{i=1}^{n}\operatorname{Cov}[\delta_n(x-X_i),\ \delta_n(y-X_i)]$$
$$+n^{-1}\sum_{i\neq j}\operatorname{Cov}[\delta_n(x-X_i),\ \delta_n(y-X_j)] \tag{3.7}$$

The first term on the right of (3.7) is

$$\int \delta_n(x-u)\delta_n(y-u)f(u)\,\mathrm{d}u - \int \delta_n(x-u)f(u)\,\mathrm{d}u \int \delta_n(y-v)f(v)\,\mathrm{d}v. \tag{3.8}$$

By splitting the range of integration of the first term in (3.8) into the parts $\{|x-u|\leq\lambda\}$, $\{|x-u|>\lambda, |y-u|\leq\lambda\}$, and $\{|x-u|>\lambda, |y-u|>\lambda\}$ and applying the δ-axioms, it is seen that this term tends to zero and hence (3.8) converges to the same limit as its second term, viz. $-f(x)f(y)$. That is the first term on the right of (3.7) converges to $-f(x)f(y)$.

The last term in (3.7) is the sum of

$$\sum_{i=1}^{n-1}\left(1-\frac{1}{n}\right)\mathrm{Cov}[\delta_n(x-X_1), \delta_n(y-X_{1+i})] \tag{3.9}$$

and the corresponding sum formed by interchanging x and y. The sum in (3.9) does not exceed, in absolute value,

$$\sum_{i=1}^{n}\iint |\delta_n(x-u)\delta_n(y-v)|\cdot|f_i(u,v)-f(u)f(v)|\,\mathrm{d}u\,\mathrm{d}v$$

$$\leq \beta_n\iint|\delta_n(x-u)\delta_n(y-v)|\,\mathrm{d}u\,\mathrm{d}v\leq A^2\beta_n.$$

Since the same is true when x and y are interchanged the second term in (3.7) does not exceed $2A^2\beta_n$ in absolute value, from which the desired result follows. □

Note that for i.i.d. sequences $\beta_n=0$ and this result reduces to a standard result in the i.i.d. case, with 'full' convergence rate n^{-1}. The introduction of dependence does not change the rate unless $\beta_n\to\infty$ and then the covariance converges to zero at least as fast as β_n/n.

4. Asymptotic normality

In order to derive the asymptotic normality of the estimator, $\hat{f}_n$, a slight modification to an 'array form' of the strong mixing condition due to Rosenblatt will be used, requiring the following definition.

Definition 4.1. Let $\{X_k; k\geq 1\}$ be a stationary sequence and let $\mathcal{M}_i^j$ denote the σ-algebra of events generated by $\{X_k; i<k\leq j\}$. Then

$$\alpha_{n,l}\equiv \max_{1\leq i\leq n-l}\ \sup_{A\in\mathcal{M}_o^i, B\in\mathcal{M}_{i+l}^n}|P(A\cap B)-P(A)P(B)|$$

for $1\leq l\leq n-1$. The array of positive constants $\alpha_{n,l}$ (defined for $1\leq l\leq n-1$) will be called the strong mixing coefficients.

The following lemma due to Volkonskii and Rozanov [9] will be used in what follows.

Lemma 4.2. *Let* $\eta_1, \eta_2, \ldots, \eta_m$ *be random variables measurable with respect to* $\mathcal{M}_{i_1}^{j_1}, \mathcal{M}_{i_2}^{j_2}, \ldots, \mathcal{M}_{i_m}^{j_m}$, *respectively, where* $0 \leq i_1 < j_1 < j_2 < \cdots < i_m < j_m \leq n$, $i_{k+1} - j_k \geq l \geq 1$ *and* $|\eta_k| \leq 1$, $k = 1, 2, \ldots, m$. *Then*

$$\left| E\left[\prod_{k=1}^{m} \eta_k\right] - \prod_{k=1}^{m} E[\eta_k]\right| \leq 16(m-1)\alpha_{n,l}$$

where $\alpha_{n,l}$ *is the strong mixing coefficient.*

Theorem 4.3. *Let* $\{\delta_n(x); n \geq 1\}$ *be a* δ*-sequence such that* $\alpha_n = \int \delta_n^2(x)\,dx = o(n)$ *and such that for some constant* K_0, $|\delta_n(x)| \leq K_0 \alpha_n$ *for all* $x \in \mathbb{R}$ *and all* $n \geq 1$. *Assume that the stationary sequence* $\{X_k; k \geq 1\}$ *has a dependence index sequence* $\{\beta_n; n \geq 1\}$ *which satisfies* $\beta_n = o(\alpha_n)$ *as* $n \to \infty$. *Suppose there exists a sequence of integers* $\{k_n; n \geq 1\}$ *for which* $(n\alpha_n)^{1/2}\alpha_{n,k_n} \to 0$ *and* $k_n = o(n/\alpha_n)^{1/2}$ *as* $n \to \infty$. *If* u *is a continuity point of the probability density* f, *with* $f(u) \neq 0$, *then*

$$n^{1/2}(f_n(u) - E[f_n(u)])/(\alpha_n f(u))^{1/2}$$

has the standard normal limiting distribution.

Proof. Clearly there exist constants $\lambda_n \to \infty$ such that $\lambda_n k_n = o(n/\alpha_n)^{1/2}$ and $\lambda_n (n\alpha_n)^{1/2}\alpha_{n,k_n} \to 0$. Define integers τ, τ' (depending on n) by $\tau = [\lambda_n^{-1}(n/\alpha_n)^{1/2}]$($[\cdot]$ denoting integer part), $\tau' = k_n$. It follows at once that

$$\text{(i)}\ \ \tau = o(n/\alpha_n)^{1/2}, \quad \text{(ii)}\ \ \tau' = o(\tau), \quad \text{(iii)}\ \ \frac{n}{\tau}\alpha_{n,\tau'} \to 0. \tag{4.1}$$

Write also $k = \tau + \tau'$, $m = [n/k]$, and divide the integer set Δ_n $(1, \ldots, n)$ into subsets of alternate 'length' τ, τ', writing

$$\Delta_{i,n} = ((i-1)k+1, (i-1)k+2, \ldots (i-1)k+\tau),$$

$$\Delta'_{i,n} = ((i-1)k+\tau+1, (i-1)k+\tau+2, \ldots ik)$$

for $1 \leq i \leq m$, and $\Delta_{m+1,n} = (mk+1, mk+2, \ldots n)$. Then clearly

$$(\hat{f}_n(u) - E\hat{f}_n(u))/(\operatorname{Var}\hat{f}_n(u))^{1/2} = \sum_{i=1}^{n} X_{n,i} \tag{4.2}$$

where

$$X_{n,i} \equiv (\delta_n(u - X_i) - \mu_n)/n^{1/2}\sigma_n,$$

$$\mu_n \equiv E[\delta_n(u - X_1)],$$

$$\sigma_n^2 \equiv \operatorname{Var}[\delta_n(u - X_1)] + 2\sum_{j=1}^{n-1}\left(1 - \frac{j}{n}\right)\operatorname{Cov}[\delta_n(u - X_1), \delta_n(u - X_{1+j})].$$

For any set of integers Δ, let $\eta_n(\Delta)=\sum_{j\in\Delta} X_{n,j}$ so that (4.2) may be written as

$$\eta_n(\Delta_n)=\sum_{k=1}^{m}\eta_n(\Delta_{k,n})+\sum_{k=1}^{m}\eta_n(\Delta'_{k,n})+\eta_n(\Delta_{m+1,n}). \tag{4.3}$$

The remainder of the proof will be accomplished in two lemmas.

Lemma 4.4. *The variances of the second and third terms on the right of* (4.3) *tend to zero, and* $\sum_{k=1}^{m}\operatorname{Var}\eta_n(\Delta_{k,n})\to 1$ *as* $n\to\infty$.

Proof. The variance of the second term on the right of (4.3) is

$$\operatorname{Var}\left(\sum_{k=1}^{m}\sum_{j\in\Delta'_{k,n}} X_{n,j}\right)\le\frac{m\tau'}{n\sigma_n^2}\operatorname{Var}[\delta_n(u-X_1)]$$
$$+\frac{2m\tau'}{n\sigma_n^2}\sum_{k=1}^{n}|\operatorname{Cov}[\delta_n(u-X_1),\delta_n(u-X_{1+k})]|. \tag{4.4}$$

The first term on the right of (4.4) tends to zero since $m\tau'/n\sim\tau'/\tau\to 0$ and $\sigma_n^2=n\operatorname{Var}\hat{f}_n(u)\sim\alpha_n f(u)$ (Theorem 3.3), whereas $\operatorname{Var}\delta_n(u-X_1)\sim\alpha_n f(u)$ as in the proof of Theorem 3.3. By (3.4) the second term on the right of (4.4) may be written as $(m\tau'/n\sigma_n^2)\mathrm{O}(\beta_n)=\mathrm{O}(\tau'/\tau)(\beta_n/\alpha_n)\to 0$ as $n\to\infty$. Hence both terms on the right of (4.4) tend to zero so that the variance of the second term on the right of (4.3) tends to zero as asserted. A similar (and even simpler) calculation gives the same result for the third term.

The final assertion also follows since again similar calculations give (using (3.4) again)

$$\sum_{k=1}^{m}\operatorname{Var}\{\eta_n(\Delta_{k,n})\}=\frac{m\tau}{n\sigma_n^2}\operatorname{Var}\delta_n(u-X_1)+\frac{m\tau}{n\sigma_n^2}\mathrm{O}(\beta_n)$$
$$=1+\mathrm{o}(1)+\mathrm{O}(\beta_n/\alpha_n)\to 1\quad\text{as } n\to\infty$$

as required. □

Since the variances of the second and third terms in (4.3) tend to zero, as $n\to\infty$, the asymptotic distribution of $\eta_n(\Delta_n)$ is the same as that for $\zeta_n(\Delta_n)\equiv\sum_{k=1}^{m}\eta_n(\Delta_{k,n})$, if it exists.

Lemma 4.5. *The asymptotic distribution of the sum* $\zeta_n(\Delta_n)$, *if it exists, is the same as if the summands* $\eta_n(\Delta_{k,n})$ *were independent. Further, the Lindeberg condition holds, viz.*

$$\sum_{k=1}^{m}E\{\eta_n^2(\Delta_{k,n})I(|\eta_n(\Delta_{k,n})|\ge\varepsilon)\}\to 0\quad\textit{for each } \varepsilon>0. \tag{4.5}$$

Proof. For the first assertion note that, from Lemma 4.2,

$$\left| E[\exp(\mathrm{i}t\zeta_n(\Delta_n))] - \prod_{k=1}^{m} E(\exp(\mathrm{i}t\eta_n(\Delta_{k,n}))] \right|$$

is bounded by

$$16(m-1)\alpha_{n,\tau'} \sim 16\left(\frac{n}{\tau}\right)\alpha_{n,\tau'} \to 0$$

by (4.1), (iii). Hence the first statement follows.

To verify the Lindeberg condition (4.5) note that $|X_{n,i}| \leq K(\alpha_n/n)^{1/2}$ for some constant K since $\mu_n \to f(u)$, $\sigma_n^2 \sim \alpha_n f(u)$ and $|\delta_n(x)| \leq K_0\alpha_n$ for all x. Hence

$$|\eta_n(\Delta_{1,n})| \leq \sum_{k=1}^{\tau} |X_{n,k}| \leq \tau K(\alpha_n/n)^{1/2} < \varepsilon$$

for all sufficiently large n by (4.1)(i), so that $P\{|\eta_n(\Delta_{1,n})| > \varepsilon\} = 0$ for all sufficiently large n, from which (4.5) follows trivially. □

It follows from Lemma 4.5 and the last statement of Lemma 4.4 that $\sum_{k=1}^{m} \eta_n(\Delta_{k,n})$ has the standard normal limiting distribution and hence so does $\eta_n(\Delta_n)$ by the remark prior to Lemma 4.5, concluding the proof of the theorem. □

5. Continuous parameter processes

In this section we consider a (strictly) stationary (measurable) stochastic process $\{X_t;\ t \geq 0\}$ with absolutely continuous marginal distribution function F, and corresponding density f. The density f is to be estimated from knowledge of the process X_t up to time T, by an estimator of the form

$$\hat{f}_T(x) \equiv T^{-1} \int_0^T \delta_T(x - X_s)\,\mathrm{d}s \tag{5.1}$$

where $\{\delta_T(x);\ T > 0\}$ is a family of smoothing functions defined for each $T > 0$ (here called a δ-family) satisfying the same axioms as in Section 2 for δ-sequences with 'n' and '$n \geq 1$' replaced by 'T' and '$T > 0$' respectively.

Lemmas 2.1–2.3 then have immediate analogues obtained by simply replacing n by T in all cases. Further Theorem 3.1 also has an obvious analogue — showing asymptotic unbiasedness of $\hat{f}_T(x)$ exactly as for sequences, i.e. giving

Theorem 5.1. *If $\{\delta_T(x);\ T > 0\}$ is a δ-family, and the density f is continuous at the point x, then $E[\hat{f}_T(x)] \to f(x)$ as $T \to \infty$.*

The bias of the estimator clearly does not involve the dependence structure of the process in any way. Hence the obvious continuous parameter analogue of Theorem 3.4 holds, giving the rate of bias convergence. Moreover, the bias of $\hat{f}_T$ (which is based on all observations X_t for $0 \le t \le T$) is just the same as would be obtained by using the sequence estimator (with $\delta_n = \delta_T$ for $T = n$) for the sequence obtained by sampling the process X_t at $t = 1, 2, 3, \ldots$. Hence, in a sense, the continuous measurement of X_t brings no improvement in the rate of convergence of the bias. The situation for the variance and covariances can be radically different, however, leading to significantly faster convergence for continuous measurement, as one would expect, and as we see next.

Our first result shows that if the local dependence of X_s and X_t is sufficiently restricted when $s \neq t$, then it is possible to obtain a 'full rate' $1/T$ of convergence of the variance to zero. This contrasts sharply with the sequence case where $(n/\alpha_n)\operatorname{Var}[\hat{f}_n(x)]$ typically converges to a non-zero limit, and may be explained intuitively by the fact that the sampling collects a whole continuum of 'somewhat independent' random variables. Further comments will be given in the specific example following the theorem. In the following specific result (and subsequently) $f_s(x, y)$ will denote the joint density of X_0 and X_s, assumed to exist for all $s \neq 0$.

Theorem 5.2. *Let* $|f_s(u, v) - f(u)f(v)| \le \Psi(s) \in L_1((0, \infty))$, *for all* u, v. *If* f_s *is continuous at* (x, x) *and* f *is continuous at* x, *then*

$$T\operatorname{Var}[\hat{f}_T(x)] \to 2\int_0^\infty [f_s(x, x) - f^2(x)]\,\mathrm{d}s \tag{5.2}$$

as $n \to \infty$.

Proof. It is readily checked that

$$\begin{aligned} T\operatorname{Var}[\hat{f}_T(x)] &= 2\int_0^T \left(1 - \frac{s}{T}\right)\operatorname{Cov}[\delta_T(x - X_0), \delta_T(x - X_s)]\,\mathrm{d}s \\ &= 2\iint \delta_T(x-u)\delta_T(x-v)\int_0^T\left(1 - \frac{s}{T}\right) \\ &\qquad \times [f_s(u, v) - f(u)f(v)]\,\mathrm{d}s\,\mathrm{d}u\,\mathrm{d}v \end{aligned}$$

The inner integral is bounded above in absolute value by $\int_0^\infty \Psi(s)\,\mathrm{d}s$ for all u, v and differs from $\int_0^\infty [f_s(u, v) - f(u)f(v)]\,\mathrm{d}s$ by no more than $\int_0^T (s/T)\Psi(s)\,\mathrm{d}s + \int_T^\infty \Psi(s)\,\mathrm{d}s$ which converges to zero by dominated convergence and the fact that $\Psi \in L_1$. This convergence is trivially uniform in (u, v) and hence it follows simply that

$$T\operatorname{Var}[\hat{f}_T(x)] = 2\iint \delta_T(x-u)\delta_T(x-v)\int_0^\infty [f_s(u, v) - f(u)f(v)]\,\mathrm{d}s\,\mathrm{d}u\,\mathrm{d}v + \mathrm{o}(1)$$

as $T \to \infty$. But the function $g(u, v) = \int_0^\infty [f_s(u, v) - f(u)f(v)]\,\mathrm{d}s$ is continuous at (x, x) since if (u_n, v_n) is any sequence converging to (x, x), $g(u_n, v_n) \to g(x, y)$ by dominated convergence, using continuity of f_s and f. Hence (5.2) follows at once from the continuous version of Lemma 2.3. □

Note that a discrete analogue of the condition $\Psi(s) \in L_1$ would be that β_n converge to a finite limit — a stronger assumption than made in Theorem 3.3. However, the main restriction in the assumption $\Psi(s) \in L_1$ is a strong dependence limitation between X_t and X_{t+s} when $s \to 0$, a feature that does not have a discrete time analogue.

Some insight into the applicability of the above result may be obtained by looking at a class of stationary normal processes for which it holds. Specifically let $\{X_t;\ t \geq 0\}$ be a stationary normal process with zero mean and covariance function

$$r(\tau) = 1 - C|\tau|^\alpha + o(|\tau|^\alpha) \quad \text{as } \tau \to 0 \tag{5.3}$$

with $0 < \alpha < 2$. Let $f_s(x, y)$ denote the bivariate normal density with correlation $r(s)$. It then follows that $f_s(x, y) \leq Ks^{-\alpha}$ in any neighborhood of $s = 0$, for some constant K (depending on the neighborhood). Hence $|f_s(x, y) - f(x)f(y)| \leq 1 + Ks^{-\alpha/2}$ on such a neighborhood of $s = 0$. On the other hand, if the covariance $r(\tau)$ is bounded away from 1 outside some neighborhood of $\tau = 0$ and integrable, then $|f_s(x, y) - f(x)f(y)| \leq K'|r(s)|$ (as noted in Section 2) for some constant K', and the function $\Psi(s)$ which is $1 + Ks^{-\alpha/2}$ in a neighborhood of $s = 0$ and $K'|r(s)|$ outside that neighborhood, satisfies the condition of the theorem.

Normal processes with $\alpha < 2$ in (5.3) have 'irregular' sample paths in contrast to the more regular case $\alpha = 2$. The irregular nature of the paths corresponds to less correlation and hence 'more information' in the measurement of X_t leading to the maximal rate of convergence of the variance to zero. It should be noted that the class with $\alpha < 2$ does contain interesting cases — such as the Ornstein-Uhlenbeck process ($\alpha = 1$).

One might expect in more regular cases that the variance of $\hat{f}_T$ would converge to zero at the rate α_T/T ($\alpha_T = \int \delta_T^2(x)\,dx$) by analogy with the sequence rate. However, faster convergence is possible, and even typical up to, of course, the 'full rate' $1/T$ which applies to the irregular case above.

The previous result provided an exact rate of convergence of the variance to zero for the class of processes considered. It seems likely that a convenient result giving exact (but slower) rates could be obtained for more general classes. However, here we give a result which provides lower bounds for the convergence rate. To obtain this it is convenient to define a continuous parameter analogue of the dependence index sequence. Specifically if again $f_s(x, y)$ denotes the joint density of X_0 and X_s, the dependence index function will be defined to be the function of $T > \gamma > 0$ given by

$$\beta_T(\gamma) = \sup_{(x,y)\in\mathbb{R}^2} \int_\gamma^T |f_s(x, y) - f(x)f(y)|\,ds \tag{5.4}$$

assumed finite for all $T > \gamma > 0$. The following lemma provides the basic calculations needed.

Lemma 5.3. *Let* $\{\delta_T(x);\ T>0\}$ *be a* δ*-family and* $\alpha_T=\int \delta_T^2(u)\,du<\infty$ *for each* T. *Then* $\alpha_T\to\infty$ *as* $T\to\infty$, *and*

$$\alpha_T^{-1}\operatorname{Var}[\delta_T(x-X_0)]\to f(x), \tag{5.5}$$

as $T\to\infty$, *at each continuity point* x *of* f. *Further, for* $0<\gamma<T$,

$$\int_\gamma^T |\operatorname{Cov}[\delta_T(x-X_0),\ \delta_T(x-X_s)]|\,ds\le A^2\beta_T(\gamma) \tag{5.6}$$

where A *is the constant in the first* δ*-axiom.*

Proof. The fact that $\alpha_T\to\infty$, as $T\to\infty$, and that (5.5) holds may be shown as in the discrete case (Lemma 2.2 and the proof of Theorem 3.3), writing $\operatorname{Var}[\delta_T(x-X_0)]=\alpha_T\int\delta_T^*(x-u)f(u)\,du-(\int\delta_T(x-u)f(u)\,du)^2$ with $\delta_T^*(u)=\alpha_T^{-1}\delta_T^2(u)$.

To prove (5.6), note that the left hand side does not exceed

$$\int_\gamma^T\!\!\iint |\delta_T(x-u)\delta_T(x-v)|\cdot|f_s(u,v)-f(u)f(v)|\,du\,dv\,ds$$

from which the result follows by the definition of $\beta_T(\gamma)$ and the first δ-axiom. □

The following result gives an 'intermediate' convergence rate between T^{-1} and α_T/T in cases where local dependence is greater than that required in Theorem 5.2.

Theorem 5.4. *Let* $\{\delta_T(x);\ T\ge 0\}$ *be a* δ*-family with* $\alpha_T=\int\delta_T^2(u)\,du<\infty$. *Let* $\{\gamma_T;\ T\ge 0\}$ *be positive constants with* $\gamma_T\to 0$, *as* $T\to\infty$, *and suppose that the dependence index function* $\beta_T(\gamma)$ *for the stationary process* $\{X_t;\ t\ge 0\}$ *satisfies*

$$\gamma_T^{-1}\beta_T(\gamma_T)=o(\alpha_T)\quad \text{as } T\to\infty. \tag{5.7}$$

Then if x *is a continuity point of* f,

$$\limsup_{T\to\infty} T\gamma_T^{-1}\alpha_T^{-1}\operatorname{Var}[\hat f_T(x)]\le 2f(x). \tag{5.8}$$

Proof. It follows in an obvious way from the first two lines of the proof of Theorem 5.2 (splitting the integration range $(0,T)$ into $(0,\gamma_T)$ and (γ_T,T), that

$$T\operatorname{Var}[\hat f_T(x)]\le 2\gamma_T\operatorname{Var}[\delta_T(x-X_0)]+2A^2\beta_T(\gamma_T)$$

from which the result follows simply by Lemma 5.3. □

It may be seen also from this proof that the requirement $o(\alpha_T)$ in (5.7) may be replaced by $O(\alpha_T)$, and then the right hand side of (5.8) is replaced by some finite constant K.

By way of example consider again a stationary normal process with zero mean and covariance $r(\tau)$ as in (5.3), but now with $\alpha=2$ (the 'regular' case). Let $\{\delta_T(x);\ T\geq 0\}$ be a δ-family with $\alpha_T=\int \delta_T^2(u)\,\mathrm{d}u$, such that $\int_0^T |r(s)|\,\mathrm{d}s=\mathrm{o}(\alpha_T)$ (which will hold trivially if r is integrable). Let $\{\gamma_T;\ T\geq 0\}$ be a family of constants tending to zero, as $T\to\infty$, satisfying

$$\begin{aligned} &\text{(i)} && \gamma_T^{-1}=\mathrm{o}(\alpha_T/\log(\alpha_T)) \\ &\text{(ii)} && \int_0^T |r(s)|\,\mathrm{d}s=\mathrm{o}(\alpha_T\gamma_T) \end{aligned} \tag{5.9}$$

as $T\to\infty$. (Note that if r is integrable (ii) is no restriction since it then follows from (i).) It is readily checked that (5.7) holds (provided $|r(s)|$ is bounded away from 1 in say $|s|\geq 1$) by writing

$$\beta_T(\gamma)\leq \int_\gamma^1 (1-r^2(s))^{-1/2}\,\mathrm{d}s+1+K\int_1^T |r(s)|\,\mathrm{d}s \tag{5.10}$$

for some constant K (again using the bound of the example in Section 2). The first term on the right of (5.10) does not exceed $-K'\log(\gamma)$ for some K' by (5.3) so that

$$\beta_T(\gamma_T)\leq -\log(\gamma_T)+1+\mathrm{o}(\alpha_T\gamma_T)$$

from which (5.7) follows at once by (5.9)(i).

It thus follows that (5.8) holds, so that the variance converges to zero at least as fast as $\gamma_T\alpha_T/T$. If $r(s)$ is integrable γ_T can be chosen from (5.9)(i) which requires that $\gamma_T\alpha_T$ tend to infinity faster than $\log(\alpha_T)$. Thus the convergence rate $\gamma_T\alpha_T/T$ may be chosen to be any rate slower than $(\log(\alpha_T))/T$. For example, if $\alpha_T=T^p$, for $0<p<1$, then a convergence rate slower than $(\log(T))/T$ is achieved. Thus, while the full rate $1/T$ of the irregular case is not attained, a rate close to it can be achieved.

Convergence results can also be obtained for covariances of the estimates $\hat{f}_T(x)$ and $\hat{f}_T(y)$ when $x\neq y$. For example, under the conditions of Theorem 5.2 it may be shown that

$$T\,\mathrm{Cov}[\hat{f}_T(x),\hat{f}_T(y)]\to 2\int_0^\infty [f_s(x,y)-f(x)f(y)]\,\mathrm{d}s$$

whereas under the conditions of Theorem 5.4 we have the obvious corollary

$$\limsup_{T\to\infty} T\gamma_T^{-1}\alpha_T^{-1}\,\mathrm{Cov}[\hat{f}_T(x),\hat{f}_T(y)]\leq 2(f(x)f(y))^{1/2}$$

though it seems likely that a sharper form of this latter inequality may be possible.

Finally we note that asymptotic distributional results may be obtained along similar lines to Theorem 4.3. For example, the following result may be shown. The notation developed above is used in this statement and σ_T^2 is written for $\mathrm{Var}\int_0^T \delta_T(u-X_s)\,\mathrm{d}s$, $(\alpha_T=\int \delta_T^2(x)\,\mathrm{d}x)$.

Theorem 5.5. *Let the stationary, measurable stochastic process* $\{X_t: t \geq 0\}$ *have strong mixing function* $\{\alpha_{T,s}: T > s > 0\}$ *Suppose that there exists a function* $\{k_T: T \geq 0\}$ *such that* $T\alpha_T^2\alpha_{T,k_T} \to 0$ *and* $\alpha_T^2 k_T / T^{1/2} \to 0$ *as* $T \to \infty$. *Also suppose that the dependence index function* $\beta_\gamma(T) = o(\alpha_T)$ *for each* $\gamma > 0$, *and let the* δ*-family be such that* $|\delta_T(x)| \leq K_0\alpha_T$ *for some constant* K_0, *all* $T \geq 0$, x. *Then*

$$T[\hat{f}_T(u) - E\hat{f}_T(u)]/\sigma_T$$

has an asymptotic standard normal distribution at all points u *such that* $\liminf_{T\to\infty} T^{-1}\sigma_T^2 > 0$. □

This result may be proved along similar lines to Theorem 4.3, using corresponding interval lengths τ, τ' given by $\tau' = k_T$, $\tau = \theta_T k_T$ where θ_T is chosen so that $\alpha_T = o(\theta_T)$, $T\alpha_T\theta_T\alpha_{T,k_T} \to 0$, and $\alpha_T\theta_T k_T / T^{1/2} \to 0$. Also previous calculations may be used to determine the asymptotic form of σ_T in some cases.

References

[1] J.V. Castellana, Nonparametric density estimates for stationary stochastic processes, Ph.D. dissertation (Statistics Department, University of NC at Chapel Hill, 1982).

[2] J.V. Castellana, Almost sure results in probability density estimation for stochastic processes, to appear.

[3] M.J. Fryer, A review of some non-parametric methods of density estimation, *J. Inst. Math. Applics.* 20, (1977) 335–354.

[4] M.R. Leadbetter, On the nonparameteic estimation of probability densities, Ph.D. dissertation (Statistics Department, University of NC at Chapel Hill, 1963).

[5] E. Masry, Probability density estimation from sampled data, IEEE Trans. Info. Theory, IT-29, (1983) 696–709.

[6] H.T. Nguyen, Density estimation in a continuous-time stationary Markov process, Ann. Statist. 7, (1978) 380–395.

[7] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Statist. 33, 1065–1076.

[8] M. Rosenblatt, Density estimates and Markov sequences, in: M. Puri, ed., Nonparametric Techniques in Statistical Inference (Cambridge University, London, 1970) pp. 199–210.

[9] V.A. Volkonskii and Y.A. Rozanov, Some limit theorems for random functions I, Theory Prob. Appl. 4, (1959) 178–197.

[10] G.S. Watson and M.R. Leadbetter, Hazard analysis II, Sankhya A 26, (1964) 101–116.

[11] E.J. Wegman, Nonparametric probability density estmation I, A summary of available methods, Technometrics 14, (1972) 533–546.