ELSEVIER

Statistics & Probability Letters 34 (1997) 53–58

STATISTICS & PROBABILITY LETTERS

On the asymptotic mean integrated squared error of a kernel density estimator for dependent data

Jan Mielniczuk

Institute of Computer Science, Polish Academy of Sciences, Ordona 21, 01-237 Warsaw, Poland

Received October 1995; revised June 1996

Abstract

Hall and Hart (1990) proved that the mean integrated squared error (MISE) of a marginal kernel density estimator from an infinite moving average process $X_1, X_2, \ldots$ may be decomposed into the sum of MISE of the same kernel estimator for a *random* sample of the same size and a term proportional to the variance of the sample mean. Extending this, we show here that the phenomenon is rather general: the same result continues to hold if dependence is quantified in terms of the behaviour of a remainder term in a natural decomposition of the densities of (X_1, X_{1+i}), $i = 1, 2, \ldots$.

Keywords: Kernel estimator; Long-range dependence; Mean integrated square error

1. Introduction

Let $X_1, X_2, \ldots$ be real random variables having the same distribution, but not necessarily independent. The aim is to estimate the marginal density f of X_1, which is assumed to be differentiable. More specifically, we assume that $(X_i)_{i=1}^{\infty}$ is a strictly stationary process such that $EX_1^2 < \infty$ and the bivariate density f_i of (X_1, X_{i+1}) exists for $i = 1, 2, \ldots$. The density f is to be estimated by means of a kernel estimator

$$f_n(x) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b_n}\right), \quad x \in \mathbb{R}, \tag{1.1}$$

based on the first n observations $X_1, \ldots, X_n$, where K is a density function and b_n is a sequence of positive constants tending to 0. A popular choice of an appropriate bandwidth b_n minimizes some estimator of the mean integrated squared error $\mathrm{MISE}(f_n) := \int E(f_n - f)^2$. That is why much interest has been devoted to derive approximate forms of MISE. The asymptotic form of MISE for independent data is easily obtained (cf. Silverman, 1986). When dependence is present, natural questions to ask are: When is the behaviour of MISE, or of its pointwise version MSE, the same as under independence? What can be said in general? The first question is decided affirmatively when the data satisfy appropriate mixing conditions or a certain condition is imposed on f_i (cf. Rosenblatt, 1971 and Remark 1 below). An important message concerning the second question has been provided by Hall and Hart (1990). They proved that if the data are generated by an infinite

PII S0167-7152(96)00165-4

moving average process with mean μ and $\sim$ stands for asymptotic equality as $n \to \infty$, then

$$\int E(f_n - f)^2 \sim \int E(f_n^0 - f)^2 + E(\bar{X}_n - \mu)^2 \int f'^2, \tag{1.2}$$

where $\bar{X}_n$ is the sample mean and f_n^0 denotes the kernel estimator based on a *random* sample of size n from f, using the same K and b_n as in f_n. It follows that the answer to the first question depends on how the magnitudes of the two terms in (1.2) compare. Furthermore, since the second term does not depend on b_n, its rate of convergence provides a ceiling on the rate of convergence of MISE of f_n. It also follows from (1.2) that using a 'small' bandwidth results in a MISE of f_n asymptotically equivalent to MISE of f_n^0 (see Hall and Hart (1990) for further detailed comments on the consequences of (1.2)). An analogous phenomenon has been established concerning the asymptotic distributions of f_n and f_n^0 by Ho (1993) and, for regression estimation, by Csörgő and Mielniczuk (1995).

The purpose of this note is to show that the asymptotic representation (1.2) is true under quite general circumstances, covering a wide range of situations. Following Giraitis et al. (1994), we consider the following decomposition of the density f_i:

$$f_i(s,t) = f(s)f(t) + r(i)f'(s)f'(t) + h_i(s,t), \quad s,t \in \mathbb{R}, \tag{1.3}$$

where $r(i) = E((X_1 - EX_1)(X_{1+i} - EX_{i+1}))$, $i = 1,2,\ldots$. The bivariate function $g_i(s,t) := f(s)f(t) + r(i)f'(s)f'(t)$ integrating to 1 can be thought of as an approximation to f_i having the same marginals and covariance as f_i. When f_i is a bivariate normal density, h_i is the remainder term in the Taylor expansion of f_i with respect to $r(i)$. The decomposition (1.3) is the main tool to study the behaviour of MISE of f_n in the paper. As proof of Theorem 1 indicates, the two terms in (1.2) stem from the first two terms in (1.3), respectively. The term originating from the third summand in (1.3) is shown to be of negligible order provided assumption A1 or A2 below holds with $\varepsilon > 1$. The main result in Section 2 is split into two cases. The first corresponds to the situation of long-range dependence when the covariance function of $(X_i)_{i=1}^{\infty}$ is not absolutely summable, whereas the second pertains to the opposite situation. We refer to a recent book of Beran (1994) for a review of statistical problems with long-range dependence. We state two corollaries to Theorem 1. One generalizes Hall and Hart's result, the other deals with the case when the data are generated by a Gaussian subordinate model.

2. Main results

For any $\ell \in \mathscr{L}^1(\mathbb{R})$, let $\hat{\ell}$ denote the Fourier transform of ℓ and let $\phi = \hat{f}$, ω and κ_i stand for Fourier transforms of f, K and h_i of (1.3), respectively. Consider the following two conditions on κ_i and h_i:

A1. $\int_{\mathbb{R}} |\kappa_i(t,-t)|\,\mathrm{d}t = \mathcal{O}(|r(i)|^{\varepsilon})$ for some $\varepsilon > 0$ as $i \to \infty$;

A2. $|h_i(s,t)| \leqslant |r(i)|^{\varepsilon} g(s)g(t)$, $s,t \in \mathbb{R}$ for some $\varepsilon > 0$,

where $g \in \mathscr{L}^2(\mathbb{R})$. We assume throughout that K is a symmetric, bounded density with a compact support and write $b = b_n$ if it does not cause notational confusion.

Theorem 1. *Assume that $r(i) = L(i)i^{-\theta}$ for $0 < \theta < 1$ and $L(\cdot)$ is a function slowly varying at infinity. If either* A1 *or* A2 *holds with $\varepsilon > 1$ and $f' \in \mathscr{L}^1(\mathbb{R}) \cap \mathscr{L}^2(\mathbb{R})$, then*

$$\operatorname{MISE}(f_n) = \operatorname{MISE}(f_n^0) + \operatorname{Var}(\bar{X}_n)\int f'^2 + o(\operatorname{Var}(\bar{X}_n)), \tag{2.1}$$

where $\bar{X}_n = n^{-1}(X_1 + X_2 + \cdots + X_n)$ and f_n^0 is the kernel estimator based on a random sample of size n from f.

It follows from the proof of Theorem 1 and Karamata's theorem that $\operatorname{Var}(\bar{X}_n) \sim 2[(1-\theta)(2-\theta)]^{-1}L(n)n^{-\theta}$. If f has two continuous square integrable derivatives, then $\inf_{b_n} \operatorname{MISE}(f_n^0) \sim Cn^{-4/5}$ for some $C > 0$. Thus, $\operatorname{MISE}(f_n)$ is dominated by the term $\operatorname{Var}(\bar{X}_n)\int f'^2$ if $\theta < 4/5$. On the other hand, if $\sum |r(i)| < \infty$ then $\operatorname{Var}(\bar{X}_n) = \mathcal{O}(n^{-1})$, and in this case we have

Theorem 2. *Assume that* $\sum_{i=1}^{\infty} |r(i)| < \infty, f' \in \mathscr{L}^1(\mathbb{R}) \cap \mathscr{L}^2(\mathbb{R})$ *and either A1 or A2 is satisfied with* $\varepsilon \geq 1$. *Then*

$$\operatorname{MISE}(f_n) = \operatorname{MISE}(f_n^0) + \mathcal{O}(n^{-1}).$$

Since $\operatorname{MISE}(f_n^0) \geq \int \operatorname{Var}(f_n^0) \sim \int K^2/(nb_n)$, here we have $\operatorname{MISE}(f_n) \sim \operatorname{MISE}(f_n^0)$.

Remark 2.1. Rosenblatt (1971) proved that if $\sum |f_i(x,y) - f(x)f(y)| < \infty$ for all x, y then $E(f_n(x) - f(x))^2 \sim E(f_n^0(x) - f(x))^2$ for all continuity points x of f (see also Theorem 3.3 in Castellana and Leadbetter (1986) for a generalization of this result). If $(X_i)_{i=1}^{\infty}$ is a Gaussian sequence with absolutely summable covariances then the mean value theorem implies that Rosenblatt's condition holds.

Proof of Theorem 1. Since $\operatorname{MISE}(f_n) = \int \operatorname{Var}(f_n) + \int (Ef_n - f)^2$ and the integrated squared bias is the same for f_n and f_n^0, it is sufficient to prove the theorem with $\operatorname{MISE}(f_n)$ and $\operatorname{MISE}(f_n^0)$ replaced by $\int \operatorname{Var}(f_n)$ and $\int \operatorname{Var}(f_n^0)$, respectively. Hall and Hart (1990) proved in their Lemma 4.1 that $\int \operatorname{Var}(f_n)$ is equal

$$\int \operatorname{Var}(f_n^0) + \frac{1}{n\pi} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \int |\omega(bt)|^2 \left\{ \operatorname{Re} E\mathrm{e}^{\mathrm{i}t(X_1 - X_{j+1})} - |\phi(t)|^2 \right\} \mathrm{d}t. \tag{2.2}$$

In view of the representation in (1.3), for all $t \in \mathbb{R}$ we have

$$E\mathrm{e}^{\mathrm{i}t(X_1 - X_{j+1})} = \int \mathrm{e}^{\mathrm{i}t(z_1 - z_2)} f(z_1) f(z_2)\, \mathrm{d}z_1\, \mathrm{d}z_2 + r(j) \int \mathrm{e}^{\mathrm{i}t(z_1 - z_2)} f'(z_1) f'(z_2)\, \mathrm{d}z_1\, \mathrm{d}z_2$$
$$+ \int \mathrm{e}^{\mathrm{i}t(z_1 - z_2)} h_j(z_1, z_2)\, \mathrm{d}z_1\, \mathrm{d}z_2 .$$

Notice that the first two terms here are $|\phi(t)|^2$ and $r(j)|\widehat{f'}(t)|^2$, respectively. By Parseval's theorem and the fact that $\widehat{f'}(t)\omega(bt)$ is the characteristic function of $f' * K_b(t)$, where $*$ denotes convolution and $K_b(\cdot) := b^{-1}K(\cdot\, b^{-1})$, we see that the second term in (2.2) is

$$\frac{2}{n} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \left[r(j) \int [K_b * f']^2 + \frac{1}{2\pi} \int |\omega(bt)|^2 \operatorname{Re} \int \mathrm{e}^{\mathrm{i}t(z_1 - z_2)} h_j(z_1, z_2)\, \mathrm{d}z_1\, \mathrm{d}z_2\, \mathrm{d}t \right]$$
$$= \frac{2}{n} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) r(j) \int [K_b * f']^2 + \frac{1}{n\pi} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \int |\omega(bt)|^2 \operatorname{Re} \kappa_j(t, -t)\, \mathrm{d}t$$
$$=: S_n + T_n .$$

Observe now that the conditions on f and K imply $\int [K_b * f']^2 = \int f'^2 + \mathrm{o}(1)$. Indeed, on the one hand, $K_{b_n} * f'(t) \to f'(t)$ for almost all t in view of assumptions on K and $f' \in \mathscr{L}^1(\mathbb{R})$ (cf. Devroye, 1986, Theorem 2.8), and thus by Fatou's lemma it follows that $\int f'^2 \leq \liminf_n \int [K_{b_n} * f']^2$. On the other hand, writing

$$\int [K_b * f']^2 = \int K(x)K(y) f'(t - xb) f'(t - yb)\, \mathrm{d}t\, \mathrm{d}x\, \mathrm{d}y$$

and using the inequality $f'(t-xb)f'(t-yb) \leqslant [f'^2(t-xb)+f'^2(t-yb)]/2$ we get that $\int [K_b * f']^2 \leqslant \int f'$. The observation and the equation

$$\mathrm{Var}(\bar{X}_n) = \frac{2}{n}\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)r(j) + \frac{r(0)}{n},$$

in which the first term dominates, imply that $S_n = \mathrm{Var}(\bar{X}_n)\int f'^2 + \mathrm{o}(\mathrm{Var}(\bar{X}_n))$. Noticing finally that in view of Karamata's theorem

$$\frac{1}{n}\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)|r(j)|^\eta = \mathcal{O}\left(|L(n)|^\eta n^{-\eta\theta}\right) \tag{2.3}$$

for any $\eta > 0$ such that $\eta\theta < 1$, the inequality $\int |\omega(bt)|^2 |\mathrm{Re}\,\kappa_j(t,-t)|\,\mathrm{d}t \leqslant \int |\kappa_j(t,-t)|\,\mathrm{d}t$ and assumption A1, along with simple properties of slowly varying functions, yield $T_n = \mathrm{o}(\mathrm{Var}(\bar{X}_n))$ and hence (2.1) under A1.

To see that (2.1) holds when A1 is replaced by A2, note that $\omega(\cdot)$ is a real function by the symmetry of K, and therefore $|\omega(bt)|^2 = \omega^2(bt)$ is the value at t of the characteristic function of $K_b * K_b = (K*K)_b =: L_b$. Thus, changing the order of integration and using the definition of an inverse transform, we have

$$\begin{aligned}\mathrm{Re}\int |\omega(bt)|^2 \int \mathrm{e}^{\mathrm{i}t(z_1-z_2)} h_j(z_1,z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2\,\mathrm{d}t &= \mathrm{Re}\int\int \omega^2(bt)\mathrm{e}^{-\mathrm{i}t(z_2-z_1)}\,\mathrm{d}t\, h_j(z_1,z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2 \\ &= 2\pi \int L_b(z_2-z_1)h_j(z_1,z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2.\end{aligned}$$

By A2 the absolute value of the last integral is not greater than

$$\begin{aligned}2\pi\left|\int\int L(u)h_j(z_2-ub,z_2)\,\mathrm{d}u\,\mathrm{d}z_2\right| &\leqslant 2\pi|r(j)|^\varepsilon \int\int L(u)g(z_2-ub)g(z_2)\,\mathrm{d}z_2\,\mathrm{d}u \\ &\leqslant 2\pi|r(j)|^\varepsilon \int\int \frac{L(u)}{2}\left[g^2(z_2-ub)+g^2(z_2)\right]\mathrm{d}z_2\,\mathrm{d}u = 2\pi|r(j)|^\varepsilon \int g^2(z_2)\,\mathrm{d}z_2.\end{aligned}$$

Therefore, reasoning as in the first part of the proof, (2.1) follows. □

Proof of Theorem 2. Notice that

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n} + 2\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)r(i) \leqslant \frac{1}{n}\left[1+\frac{2}{n}\sum_{i=1}^{n-1}|r(i)|\right] = \mathcal{O}\left(\frac{1}{n}\right).$$

Thus the proof follows the lines of that of Theorem 1 with one obvious change. □

Remark 2.2. (a) The proof of Theorem 1 shows that (2.1) may in fact be written as

$$\mathrm{MISE}(f_n) = \mathrm{MISE}(f_n^0) + \mathrm{Var}(\bar{X}_n)\int [K_b * f']^2 + \mathcal{O}(\mathrm{Var}(\bar{X}_n)n^{-\alpha}),$$

where $\alpha = \min(1-\theta,\ \theta(\varepsilon-1)) - \delta$ for an arbitrary $\delta > 0$. This follows since the sum in (2.3) is $\mathcal{O}(n^{-1})$ for $\eta\theta > 1$ and $\mathcal{O}(L_0(n)n^{-1})$ for $\eta\theta = 1$, where $L_0(\cdot)$ is some function slowly varying at infinity.

(b) It is also easy to see that the representation (2.1) still holds when the assumption $r(i) = L(i)i^{-\theta}$, $0 < \theta < 1$, is replaced by a weaker set of two conditions: $n\,\mathrm{Var}(\bar{X}_n) \to \infty$ and $n^{-1}\sum_{i=1}^n (1-i/n)|r(i)|^\varepsilon = \mathrm{o}(\mathrm{Var}(\bar{X}_n))$ for some $\varepsilon > 1$.

3. Corollaries

Now we consider two special cases when $(X_i)_{i=1}^{\infty}$ is either an infinite moving average process or an instantaneous transform of a Gaussian sequence. Assume first that

$$X_i = \mu + \sum_{j\leqslant i} b_{i-j}\zeta_j, \tag{3.1}$$

where $(\zeta_j)_{j=-\infty}^{\infty}$ is an i.i.d. sequence such that $E\zeta_0 = 0$ and $E\zeta_0^2 = 1$ and $(b_j)_{j\geqslant 0}$ is a sequence such that $\sum_j b_j^2 < \infty$. It is known that if $b_j = L_1(j)j^{-(1+\theta)/2}$, where $L_1(\cdot)$ is a slowly varying function and $0 < \theta < 1$, then $r(j) = \sum_{k\geqslant 0} b_k b_{k+j} \sim L_1^2(j)j^{-\theta}$. The following corollary is a generalization of Theorem 2.1 in Hall and Hart (1990) in the case when the coefficients b_j determining a more general infinite moving average process considered in their paper vanish on one side of the origin. Note that the conditions on ζ_0 listed below are satisfied by the exponential, gamma and normal distributions.

Corollary 1. *Assume that X_i has representation* (3.1), $b_j = L_1(j)j^{-(1+\theta)/2}$ *with $L_1(\cdot)$ a slowly varying function,* $0 < \theta < 1$, *and* $E|\zeta_0|^3 < \infty$, $|Ee^{iu\zeta_0}| < C(1+|u|)^{-\delta}$ *for some* $\delta > 0$. *Then the decomposition* (2.1) *holds true.*

Proof. The corollary follows from Theorem 1 and the proof of Lemma 2 of Giraitis et al. (1994) upon noting that their conclusions used here are valid for a general μ in (3.1). Namely, they prove that $|\hat{f}_j(u_1,u_2)| \leqslant C_1(k)(1+|u|)^{-k}$, $u = (u_1, u_2) \in \mathbb{R}^2$ and $|\hat{f}(u)| \leqslant C_2(k)(1+|u|)^{-k}$, $u \in \mathbb{R}$, for an arbitrary $k \in \mathbb{N}$, where the constant $C_1(k)$ does not depend on j. Consider the decomposition

$$\int_{\mathrm{R}} |\kappa_j(t,-t)|\,\mathrm{d}t = \int_{|t|>j^d} |\kappa_j(t,-t)|\,\mathrm{d}t + \int_{|t|\leqslant j^d} |\kappa_j(t,-t)|\,\mathrm{d}t, \tag{3.2}$$

where d is an arbitrary positive number less than $\min(\theta/7, (1-\theta)/12)$. Observe that $\kappa_j(z_1,z_2) = \hat{f}(z_1,z_2) + (r(j)z_1z_2 - 1)\hat{f}(z_1)\hat{f}(z_2)$. Thus, it follows from the above bounds on the characteristic functions of f_j and f that

$$\int_{|t|>j^d} |\kappa_j(t,-t)|\,\mathrm{d}t \leqslant j^{-dk}\int_{|t|>j^d} |t|^k|\kappa_j(t,-t)|\,\mathrm{d}t = \mathcal{O}(j^{-dk}) = \mathcal{O}(|r(j)|^{\varepsilon})$$

for $k > \varepsilon\theta/d$. At the same time Eq. (2.20) in Giraitis et al. (1994) implies that $\sup_{|t|\leqslant j^d} |\kappa_j(t,-t)| = \mathcal{O}(j^{-\theta-3d})$. Thus, the second integral in (3.2) is $\mathcal{O}(j^{-\theta-2d})$, which implies condition A1 in view of the form of $r(j)$ and the properties of $L_1^2(j)$. □

Consider now a mean-zero stationary Gaussian sequence $(Z_j)_{j=1}^{\infty}$ with covariance $E(Z_1Z_{1+j}) = L(j)j^{-\theta}$ for some $\theta \in (0,1)$ and a slowly varying function $L(\cdot)$. Let $G(\cdot)$ be a monotone function such that $EG^2(Z_1) < \infty$, and suppose that $X_j := G(Z_j)$, $j = 1, 2, \ldots$. It is known (cf. Taqqu, 1975) that $r(j) = E((X_1 - EX_1)(X_{1+j} - EX_{j+1})) \sim c_1^2 L(j)j^{-\theta}$, where c_1 is a constant different from zero. Denote by ϕ the univariate standard normal density.

Corollary 2. *Assume that* $H := G^{-1}$ *is differentiable and* $|H'(x)| \leqslant \exp(\eta H^2(x))$, $x \in \mathbb{R}$, *for some* $\eta < 1/2$. *Then* (2.1) *holds true with* $\int f'^2$ *replaced by* $\int f_H'^2$ *where* $f_H(s) := \phi(H(s))$ *for* $s \in \mathbb{R}$.

Proof. Let $f_\rho(x,y) = [2\pi(1-\rho^2)^{1/2}]^{-1}\exp(-(x^2+y^2-2\rho xy)/2(1-\rho^2))$ be the normal density with a correlation coefficient ρ. It is easy to check that

$$\left.\frac{\partial f_\rho}{\partial \rho}\right|_{\rho=0}(x,y) = (1/2\pi)xy\exp(-(x^2+y^2)/2) = \varphi'(x)\varphi'(y).$$

Moreover,

$$\left|\frac{\partial^2 f_\rho(x,y)}{\partial \rho^2}\right| \leqslant |Q_\rho(x,y)| \exp\left(-(x^2+y^2-2\rho xy)/2(1-\rho^2)\right), \tag{3.3}$$

where $Q_\rho(x,y)$ is a polynomial such that $\sup_{|\rho|\leqslant 1-\eta}|Q_\rho(x,y)| < Q(x)Q(y)$ for any $\eta > 0$ and a certain univariate polynomial $Q(x)$. Taking into account the inequality $2\rho xy \leqslant (x^2+y^2)\rho$, it follows that that the bound in (3.3) for ρ such that $\rho < 1-\eta$ may be replaced by $Q(x)Q(y)\exp(-(x^2+y^2)/4)$. By a Taylor expansion this means that for $G(x)=x$ condition A2 is satisfied with $\varepsilon = 2$ and $g(x) = Q(x)\exp(-x^2/4)$. For a general G assume, without loss of generality, that G is nondecreasing and observe that the density of $(G(Z_1), G(Z_{1+j}))$ is equal to $f_{r(j)}(H(x), H(y))H'(x)H'(y)$. Consider decomposition (1.3) for $f_{r(j)}(H(x),H(y))$ with both sides multiplied by $H'(x)H'(y)$ and apply to it the reasoning from the proof of Theorem 1. It follows that Corollary 2 holds upon noting that $g(x) := Q(H(x))H'(x)\exp(-H^2(x)/4)$ belongs to $\mathscr{L}^2(\mathbb{R})$, when $|H'(x)| \leqslant \exp(\eta H^2(x))$ with $\eta < \frac{1}{2}$.

Remark 3.1. Theorem 2.2 in Hall et al. (1994) provides a decomposition of MISE under conditions more stringent than those in Corollary 2. The term of order n^{-1} is of the form $n^{-1}\sum_{i=1}^{n-1}(1-i/n)\{\int f_i(s,s) - f^2(s)\}\,\mathrm{d}s$. Note that under the conditions of Corollary 2 this is equal to $n^{-1}\sum_{i=1}^{n-1}(1-i/n)r(i)\{\int f_H'^2 + \mathcal{O}(|r(i)^\varepsilon|)\}$ with some $\varepsilon > 0$ and can be written as $\mathrm{Var}(\bar{X}_n)\int f_H'^2 + \mathrm{o}(\mathrm{Var}(\bar{X}_n))$.

Acknowledgements

I wish to thank Sándor Csörgő for drawing my attention to Giraitis, Koul and Surgailis (1994) and for his insightful comments.

References

Beran, J. (1994), *Statistics for Long-Memory Processes* (Chapman & Hall, London).

Castellana, J.V. and M.R. Leadbetter (1986), On smoothed probability density estimation for stationary processes, *Stochastic Process. Appl.* **21**, 179–193.

Csörgő, S. and J. Mielniczuk (1995), Random-design regression under long-range dependent errors, Statistics Report No. 248, University of Michigan.

Devroye, L. (1986), *A Course in Density Estimation* (Birkhäuser, Boston).

Giraitis, L., H. Koul and D. Surgailis (1994), Asymptotic normality of regression estimators with long memory errors, Michigan State University, preprint.

Hall, P. and J.D. Hart (1990), Convergence rates in density estimation for data from infinite-order moving average process, *Probab. Theory Related Fields* **87**, 253–274.

Hall, P., S.N. Lahiri and Y.K. Truong (1994), On bandwidth choice for density estimation with dependent data, Statistics Report No. SR6/94, The Australian National University.

Ho, H.-C. (1993), Asymptotic properties in density estimation for sequences of long-range dependence, Academia Sinica, preprint.

Rosenblatt, M. (1971), Curve estimates, *Ann. Math. Statist.* **42**, 1815–1842.

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

Taqqu, M.S. (1975), Weak convergence to fractional Brownian motion and to the Rosenblatt process, *Z. Wahrsch. Verw. Gebiete* **31**, 287–302.