JOURNAL OF MULTIVARIATE ANALYSIS 41, 89–101 (1992)

Quadratic Deviation of Penalized Mean Squares Regression Estimates

PAUL DOUKHAN AND ELISABETH GASSIAT

Université Paris-Sud, U.R.A. 743, bat. 425, 91405 Orsay Cédex, France

Communicated by C. R. Rao

Our aim in this work is to give a global test of hypothesis concerning the smoothing spline estimate of a regression function. For this, we prove a central limit theorem for integrated squares of such estimates. That leads to a test whose confidence sets are either continuous or discrete L^2-balls. We consider the case of nonperiodic splines and the periodic splines for which we get explicit expressions of the constants involved in such a test.

I. INTRODUCTION

The aim of this work is to investigate further probabilistic properties of spline regression estimates in the model

$$y_i = g(t_1) + \varepsilon_i. \tag{I.1}$$

In the following, $t_1, ..., t_n$ denote the observation points belonging to the compact interval $[0, 1]$, $\varepsilon_1 \cdots \varepsilon_n$ is an i.i.d. centered sequence, and g is the unknown function. Define the functional

$$J_n(f) = \frac{1}{n} \cdot \sum_{i=1}^{n} (y_1 - f(t_1))^2 + \lambda \cdot \int (f^{(m)}(t))^2 \, dt, \tag{I.2}$$

where λ is the smoothing parameter and depends on n; however, the subscript will be often omitted.

We shall consider the Sobolev space W_2^m of real valued functions on $[0, 1]$ with an absolutely continuous $(m-1)$th derivative and square

Received December 6, 1990; revised September 24, 1991.

AMS subject classifications: 60F05, 62G10, 62G99, 62J02, 62J99.

Key words and phrases: limit theorem; nonparametric interference: hypothesis testing, penalized mean squares estimates; regression and correlation: general nonlinear regression, smoothing splines regression estimates.

0047-259X/92 $3.00

integrable m-th derivative and the subspace K_2^m of W_2^m satisfying the periodic boundary conditions

$$K_2^m = \{f \in W_2^m, f^{(p)}(0) = f^{(p)}(1), p = 0, ..., m-1\}.$$

Knowing the observations $y_1 \cdots y_n$ and the fact that g belongs to W_2^m, we consider the penalized mean squares estimate $g_{n,\lambda}$ minimizing $J_n(f)$ on W_2^m (resp. on K_2^m if g is known to be periodic).

This method of estimation was introduced by Wahba (see [6, 13, 18], for instance), and the estimate is known to be a smoothing spline of degree $2m-1$.

We are interested here in giving asymptotic statistical tests on g for the whole range of the variable t.

Several authors give local asymptotic statistical tests on the value $g(t)$ for a fixed value of t (see [9] or [15]). For this, they use the Bayesian interpretation of the spline estimate (see [13]); the test follows from classical Bayesian computation of the posterior variance. As usual, for non parametric estimation, no functional result can hold because of a defect of tightness; indeed, it may be shown (see [11]) that $(g_{n,\lambda}(t) - Eg_{n,\lambda}(t))$ $(\mathrm{Var}\, g_{n,\lambda}(t))^{-1/2}$ is asymptotically a white noise. We give here a central limit theorem for the centered mean squares and integrated squares of errors, with the normalisation $n\lambda^{1/(4m)}$. The optimal regularization parameter λ is covered by the result. This result leads to a global test of hypothesis, provided that the constants appearing in the development are attainable. We only know one earlier result giving such a global test proved in Eubank and Spiegelman [10] for cubic splines and Gaussian errors. This kind of result is similar to those obtained by Hall [11] concerning kernel estimates and by Doukhan and Leon [8] concerning projection estimates.

The paper is organised as follows: in the first section, we recall the linear form of the estimate $g_{n,\lambda}$ and formulate the main results, the central limit theorem and the asymptotic form of the constants in the periodic case, leading to the construction of asymptotic global tests for g. The proofs are developed in the last section.

II. The Central Limit Theorem

We now define the sum of squares

$$R_{n,\lambda} = \frac{1}{n} \sum_{i=1}^{n} (g_{n,\lambda}(t_i) - g(t_i))^2. \tag{II.1}$$

The asymptotic behaviour of its expectation is known to be $ER_{n,\lambda} = O(\lambda + (n\lambda^{1/(2m)})^{-1})$ in the nonperiodic case, and $ER_{n,\lambda} = O(\lambda^2 + (n\lambda^{1/(2m)})^{-1})$ in the periodic case.

The same result holds for the integrated square errors. The optimization of the order of convergence leads to the classical minimax bound $O(n^{-2m/(2m+1)})$, according to Bretagnolle and Huber [2] and Speckman [16] in the nonperiodic case, and $O(n^{-4m/(4m+1)})$ in the periodic case. An optimal choice of the sequence of λ is here, respectively, $\lambda = c_0 \cdot n^{-2m/(2m+1)}$ and $\lambda = c_1 \cdot n^{-2m/(4m+1)}$. Let

$$S_{n,\lambda} = n\lambda^{1/(4m)}(R_{n,\lambda} - ER_{n,\lambda}), \tag{II.2}$$

and $Z_{n,\lambda}$ is the analogous centered expression for the integrated squares:

$$Z_{n,\lambda} = n\lambda^{1/(4m)}(\|g_{n,\lambda} - g\|_2^2 - E\|g_{n,\lambda} - g\|_2^2). \tag{II.3}$$

We now give the main results in the different following situations.

II.1. *Nonperiodic Case*

We assume classically that the empirical distribution function, F_n, of the observation points t_i converges to the distribution F, that F is absolutely continuous with respect to the Lebesgue measure, and that $f = dF/dt$ is bounded above and below:

$$0 < a \leqslant f(t) \leqslant b < \infty.$$

We shall assume

$$\lambda_n \to 0, \qquad n\lambda_n^{1/(2m)} \to \infty. \tag{H}$$

Denoting by d_n, $d_n = \sup_t |F_n(t) - F(t)|$, we assume that

$$d_n \cdot \lambda_n^{-1/m} = O(1), \tag{K}$$

$$d_n \cdot \lambda_n^{3/2m} \to \infty. \tag{L}$$

We shall moreover assume that $E\varepsilon_j = 0$, $E\varepsilon_j^2 = \sigma^2$, and $E\varepsilon_j^8 < \infty$ in the following.

THEOREM 1. *Under the assumptions* (H), (K), (L) *and for* $\lambda = O(n^{-2m/(2m+1)})$,

> $S_{n,\lambda}/\sigma_n$ *converges in law to the standard Gaussian distribution, where* $\sigma_n^2 = \mathrm{var}(S_{n,\lambda})$.

THEOREM 2. *Under the assumptions* (H), (K), (L) *and for* $\lambda = O(n^{-2m/(2m+1)})$,

$Z_{n,\lambda}/\sigma_n$ *converges in law to the standard Gaussian distribution.*

This result leads to a tractable test as soon as the constants σ_n and $ER_{n,\lambda}$ are evaluated. In the periodic case, a precise evaluation is possible, thanks to the circulant form of the corresponding influence matrix; see Section II.2 below. However, here, general arguments given in Silverman [14, Sect. 5] show that these expansions do not generalize in the nonperiodic case due to boundary effects.

The expression of asymptotic variance satisfies (see Section III for the calculations)

$$\liminf_n \sigma_n^2 \geqslant S = 2\sigma^4 \cdot \int f^{1/(2m)}(x)\, dx \cdot \int (1+s^{2m})^{-4}\, ds;$$

thus replacing σ_n^2 by S leads to a conservative test at a fixed level.

The following suboptimal result is straightforward; if $\lambda = o(n^{-2m/(2m+1)})$ then $n\lambda^{1/(4m)} S_{n\lambda} - \lambda^{1/(4m)}\sigma^2 \operatorname{Tr} A^2$ converges to the centered Gaussian law with variance $2\sigma^4 \cdot \int f^{1/(2m)}(x)\, dx \cdot \int (1+s^{2m})^{-4} ds$. This result involves no functional of the unknown function g, but of course is not available for the optimal speed of convergence.

Note that σ^2 can be estimated as recalled by Eubank and Spiegelman [10, Sect. 3.2].

Using a part of the available data, it is possible to estimate non-linear parameters as Sobolev norms with procedures given, for intance, by Bickel and Ritov [1]; they give a central limit theorem for such estimates (under the assumption $g \in W_2^p$, for some $p > 2m + \frac{1}{4}$). Thus, combining both tests gives rise to a valid explicit test.

II.2. *Periodic Case*

For the periodic estimation, we assume that the knots t_i are equally spaced in the interval $[0, 1]$: $t_i = 1/n$. In this case, all the matrices that appear in the calculations are circulant; thus the constants involved in the central limit theorem can be precisely evaluated.

THEOREM 3. *If g possesses $3m$ derivatives, $n\lambda^{1/(4m)} \cdot (\|g_{n,\lambda} - g\|_2^2 - E\|g_{n,\lambda} - g\|_2^2)/\sigma_n$ converges in law to the standard Gaussian distribution and*

$$\sigma_n^2 = \frac{\sigma^4}{\pi} \cdot \int_0^\infty \frac{dy}{(1+y^{2m})^4}(1 + O(\lambda^{1/2m})) + \sigma^2 n \lambda^{2+1/2m} \frac{1}{2} \|g^{(2m)}\|^2 (1 + O(\lambda + \lambda^{-2} n^{1-2m})) \quad \text{(II.4)}$$

$$n\lambda^{1/4m}E\|g_{n,\lambda}-g\|^2=\lambda^{-1/4m}\frac{\sigma^2}{\pi}\int_0^\infty\frac{dy}{(1+y^{2m})^2}(1+O(\lambda^{1/4m}))$$
$$+n\lambda^{2+1/4m}\frac{1}{2}\cdot\|g^{(2m)}\|^2(1+O(n^{-1}+(\lambda n^{2m})^{-1})). \quad \text{(II.5)}$$

III. *Proofs of the Theorems*

The existence and the linear form of $g_{n,\lambda}$ are widely known (see, for example, [4] or [6]); as soon as $n \geq m$,

$$g_{n,\lambda}(t)=\frac{1}{n}\sum_{i=1}^{n}G_n(t,t_i)\cdot y_i.$$

Let us denote by A the influence matrix (for the sake of simplicity A is not indexed by n though it varies with n) $A_{i,j}=(1/n)\,G_n(t_i,t_j)$. Here $A_{i,j}$ denotes the (i,j)th element of the matrix A. The matrix A has a different expression in the periodic and non periodic case, so that we study its properties and their consequences in two subsections.

III.1. *The Non Periodic Case*

We follow a result of Silverman to investigate further asymptotic properties of the influence matrix.

Silverman [14, Proposition 2 and Remark 4] shows that there exists $C>0$ only depending on f with

$$\forall s\in[0,1],\ \forall t\in[0,1],$$
$$\left|h(t)\,f(t)\,G(s,t)-K\left(\frac{s-t}{h(t)}\right)\right|\leq C\left[h(t)+\frac{d_n}{h(t)}+e(t)\right], \quad \text{(III.1)}$$

where $h(t)=(\lambda/f(t))^{1/(2m)}$, $e(t)=\exp[-(t\wedge(1-t)/h(t)\sqrt{2})]$, and K is the function with Fourier transform $1/(1+\omega^{2m})$.

LEMMA 1. *Under the assumptions* (H) *and* (K),

$$T_n=\frac{1}{n}\cdot\sum_{i=1}^{n}|G_n(s,t_i)|=O(1), \qquad \text{uniformly over } s \text{ in } [0,1]. \quad \text{(III.2)}$$

Use Silverman's result to write $T_n=T_{n,1}+T_{n,2}+O(1+d_n\cdot\lambda^{-1/m})$:

$$T_{n,1}=\frac{1}{n}\cdot\sum\frac{1}{h(t_i)}\left|K\left(\frac{s-t_i}{h(t_i)}\right)\right|\leq d_n\cdot\lambda^{-1/(2m)}+\int_0^1\frac{f(t)}{h(t)}\left|K\left(\frac{s-t}{h(t)}\right)\right|dt.$$

Set $u(s-t)\cdot\lambda^{-1/(2m)}$:

$$T_{n,1}\leqslant d_n\lambda^{-1/(2m)}+\frac{1}{n}\cdot\int|K(u\cdot f(s-u\lambda^{-1/(2m)}))|\,du=O(1),$$

$$T_{n,2}=\frac{1}{n}\cdot\sum\frac{1}{h(t_i)}\cdot e(t_i)\leqslant d_n\lambda^{-1/(2m)}+\int_0^1 e(t)\frac{f(t)}{h(t)}\,dt.$$

Now, decompose the integral for t in $[0,\frac{1}{2}]$ and t in $[\frac{1}{2},1]$, in the second one change $1-t$ in t, and use the same variable change as for $T_{n,1}$; then $T_{n,2}$ appears as a $O(1)$ in the same way.

LEMMA 2. *Under the assumptions* (H), (K), *and* (L), $d_n^{-2}\lambda^{3/(2m)}\to\infty$,

$$\lambda^{1/(2m)}\cdot\sum_{i=1}^{n}(A^2)_{i,i}^2=o(1). \tag{III.3}$$

Proof. $(A^2)_{ii}=\sum_{j=1}^n A_{ij}^2$. For simplicity, note $\mu=\lambda^{1/(2m)}$ and $\phi(t)=f(t)^{-1/(2m)}$:

$$(A^2)_{ii}=\frac{1}{n^2}\cdot\sum_j G_1^2(t_i,t_j)+\frac{1}{n^2}\cdot\sum_j R^2(t_i,t_j)+\frac{2}{n^2}\cdot\sum_j G_1(t_i,t_j)\cdot R(t_i,t_j)$$

with

$$G_1(s,t)=\frac{1}{\mu f(t)\,\phi(t)}\cdot K\left(\frac{s-t}{\mu\phi(t)}\right),\qquad |R(s,t)|\leqslant C\left(1+\frac{d_n}{\mu^2}+e(t)/\mu\right).$$

First

$$\frac{1}{n^2}\cdot\sum_j G_1^2(s,t_j)=\frac{1}{n}\cdot\int G_1^2(s,t)f(t)\,dt+O\left(\frac{d_n}{n\mu^2}\right);$$

set

$$u=\frac{s-t}{\mu\phi(s)}=\frac{\phi(s)}{n\cdot\mu}\int\frac{K^2(u\phi(s)/\phi(s-u\phi(s)))}{(f\phi^2)(s-u\phi(s))}\,du$$
$$=f(s)^{1/(2m)-1}\|K\|_2^2/n\lambda^{1/(2m)}(1+o(1)).$$

Now $(1/n^2)\cdot\sum_j R^2(s,t_j)=O(1/n+d_n^2/n\mu^4)=o(1/n\lambda^{1/(2m)})$ under the assumption on λ. We then have

$$(A^2)_{ii}=f(t_i)^{1/(2m)-1}\|K\|_2^2/n\lambda^{1/(2m)}(1+o(1))$$

$$\sum_i(A^2)_{ii}^2=\int f(t)^{1/m-1}\,dt\cdot\|K\|_2^4/n\lambda^{1/m}(1+o(1))$$

and the lemma follows.

LEMMA 3. *Under the assumptions* (H), (K), *and* (L),

$$\mathrm{Tr}(A^2)=\lambda^{-1/2m}\|K\|_2^2\int_0^1 (f(t))^{1/(2m)}dt\cdot(1+o(1)). \tag{III.4}$$

Proof.

$$\mathrm{Tr}(A^2)=\sum_{i=1}^n (A^2)_{ii}=\lambda^{-1/(2m)}\|K\|_2^2\cdot\frac{1}{n}\sum_{i=1}^n f(t_i)^{1/(2m)-1}(1+o(1))$$

and $(1/n)\sum f(t_i)^{1/(2m)-1}=\int (f(t))^{1/(2m)}dt\,(1+o(1))$.

LEMMA 4. *Under the assumptions* (H), (K), *and* (L),

$$\mathrm{Tr}(A^4)=\lambda^{-1/(2m)}\cdot\int(1+s^{2m})^{-4}ds\cdot\int f(v)^{1/(2m)}dv(1+o(1)), \tag{III.5}$$

$$\mathrm{Tr}(A^8)=O(\lambda^{-1/(2m)}). \tag{III.6}$$

Proof. We shall give the proof concerning the result for $\mathrm{Tr}(A^4)$; the result for $\mathrm{Tr}(A^8)$ using similar arguments will be omitted:

$$\begin{aligned}\mathrm{Tr}(A^4)&=\sum_{i,j}(A^2)_{ij}^2\\&=\frac{1}{n^4}\cdot\sum_{ij}\left(\sum_k\frac{1}{\mu\phi(t_i)\,\phi(t_j)\,f(t_i)\,f(t_j)}K\left(\frac{t_i-t_k}{\mu\phi(t_i)}\right)K\left(\frac{t_j-t_k}{\mu\phi(t_j)}\right)\right)^2\\&\quad+\text{remainder term } Q.\end{aligned}$$

Using the fact that F_n converges to F and the fact that the kernel K is continuous and bounded it is easy to see that the first term can be expressed as

$$\int\frac{dF(u)\,dF(v)}{\mu^4\phi^2(u)\,\phi^2(v)\,f^2(u)\,f^2(v)}\cdot\left[\int K\left(\frac{u-t}{\mu\phi(u)}\right)K\left(\frac{v-t}{\mu\phi(v)}\right)dF(t)\right]^2(1+o(1)).$$

Using the variable changes $s=(u-t)/\mu\phi(u)$, and then $w=(v-u)/\mu\phi(v)$, this term appears to be equivalent to

$$\lambda^{-1/(2m)}\left[\int\left(\int K(s)\,K(s+w)\,ds\right)^2dw\right]\cdot\int f(v)^{1/(2m)}dv.$$

Using the fact that K is symmetric and using the Parseval identity, this last expression equals

$$\lambda^{-1/(2m)} \cdot \int \frac{1}{(1+s^{2m})^4}\, ds \cdot \int f(v)^{1/(2m)}\, dv \cdot (1+o(1)).$$

The lemma follows when it is noted that the remainder term Q is $o(\lambda^{-1/(2m)})$, which follows from the fact that

$$Q = O\left((1/n^4) \cdot \sum_{ij} 1/\mu^2\phi^2(t_j)\, f^2(t_j)\right) \cdot \left(\sum_k R_{ik} \cdot K((t_j - t_k)/\mu\phi(t_j))\right)^2 + O\left((1/n^4) \cdot \sum_{ij}\left(\sum_k R_{ik} R_{jk}\right)^2\right),$$

and calculations similar to those developed for Lemma 1–3.

We now have the elements to prove Theorems 1 and 2.

The following development is straightforward,

$$S_{n,\lambda} = \lambda^{1/(4m)}\left[\sum_{i=1}^{n} (A^2)_{ii}(\varepsilon_i^2 - \sigma^2) + 2\sum_{i=1}^{n}\sum_{j<l} A_{ij}A_{jl}\varepsilon_j\varepsilon_l + 2\sum_i (AB)_i \varepsilon_i\right]$$

with $B = (A - Id)X$, $X = (g(t_i))_{i=1\cdots n}$. Moreover the variance of $S_{n,\lambda}$ is

$$\sigma_n^2 = \lambda^{1/(2m)}\left[(E\varepsilon_1^4 - 2\sigma^4)\left(\sum (A^2)_{ii}^2\right) + 2\sigma^4 \operatorname{Tr}(A^4) + 4\sigma^2 \sum (AB)_j^2\right]. \qquad \text{(III.7)}$$

Using Lemma 2, Lemma 4, the fact that A has all its eigenvalues between 0 and 1, and usual asymptotic bounds for the bias (see, for instance, [6, Lemma 4]), we can straightfully remark that $0 < \liminf_n \sigma_n^2 \leqslant \limsup_n \sigma_n^2 = O(1)$ as soon as $\lambda = O(n^{-2m/(2m+1)})$, which is the case for the optimal choice of λ that leads to the optimal speed of convergence for $ER_{n,\lambda}$.

Proof of Theorem 2. Let $W_n = n\lambda^{1/(4m)} \int (g_{n,\lambda} - g)^2 (dF_n(t) - dF(t))$. Following Cox [5] there exists a constant C such that $|W_n| \leqslant C \cdot d_n \|g_{n\lambda} - g\|_1^2 \cdot n\lambda^{1/(4m)}$, where $\|\cdot\|_p$ denotes the usual norm in the Sobolev space of order p. Then $E|W_n| = d_n \lambda^{-1/m} O(\lambda + n^{-1}\lambda^{-1/(2m)}) n\lambda^{1/(4m)}$, W_n converges to 0 in probability under the assumptions of the theorem, and the asymptotic result of Theorem 2 is then a consequence of Theorem 1.

Proof of Theorem 1. We shall make use of a theorem in Hall and Heyde [12, p. 58]: Let $X_n = \sum_{j=1}^n X_{j,n}$ be a martingale; $V_n^2 = \sum_{j=1}^n E_{j-1} X_{j,n}^2$, where E_k designates the expectation with respect to the σ-field generated by

$X_1, ..., X_k$. If $\sigma_n^2 = EV_n^2 \to \tau^2$, $V_n^2/\sigma_n^2 \xrightarrow{P} 1$, $\sum_{j=1}^n E_{j-1} X_{j,n}^4 \xrightarrow{P} 0$, then X_n converges in law to the centered Gaussian distribution with variance τ^2.

We shall verify the conditions of this theorem for the sequence

$$X_{j,n} = \lambda^{1/(4m)} \left[(A^2)_{jj}(\varepsilon_j^2 - \sigma^2) + (AB)_j \varepsilon_j + 2 \sum_{1<j} (A^2)_{jl} \varepsilon_j \varepsilon_l \right].$$

X_n is then obviously a martingale:

$$V_n^2 = \lambda^{1/(2m)} \left[\mu_4 \sum_j (A^2)_{jj}^2 + \sigma^2 \sum_j (AB)_j^2 + 4\sigma^2 \sum_{k<j, l<j} (A^2)_{jk} (A^2)_{jl} \varepsilon_l \varepsilon_k \right.$$
$$\left. + 4\sigma^2 \sum_j (AB)_j \sum_{l<j} (A^2)_{jl} \varepsilon_l \right],$$

with $\mu_4 = E(\varepsilon^2 - \sigma^2)^2$.

Now we also have $EV_n^2/\sigma_n^2 = 1$ and we calculate

$$V_n^2 - \sigma_n^2 = \lambda^{1/(2m)} \left[4\sigma^2 \sum_j \sum_{l<j} (A^2)_{jl}^2 (\varepsilon_l^2 - \sigma^2) + 4\sigma^2 \sum_{\substack{k<j, l<j \\ k \neq l}} (A^2)_{jk} (A^2)_{jl} \varepsilon_l \varepsilon_k \right.$$
$$\left. + 4\sigma^2 \sum_j (AB)_j \sum_{l<j} (A^2)_{jl} \varepsilon_l \right] = A_n + B_n + C_n.$$

Calculation shows that

$$EA_n^2 = \lambda^{1/m} O(\operatorname{Tr} A^8) = o(1)$$

and

$$EB_n^2 = \lambda^{1/m} O(\operatorname{Tr} A^8) = o(1),$$

using Lemma 4. $EC_n^2 = \lambda^{1/m} O(\|AB\|^2 (1 + O(\sum_l (A^2)_{ll}^2)) + \operatorname{Tr} A^8) = o(1)$ using Lemmas 4 and 2 and the fact that $\|AB\|^2 = O(n\lambda)$. Then $V_n^2 - \sigma_n^2 \to 0$ in probability, and V_n^2/σ_n^2 tends to 1 in probability because σ_n^2 is bounded and always larger than a positive number.

Now $\sum_j E_{j-1} X_{jn}^4$ is the sum of different terms,

$$D_n = \lambda^{1/m} \left[\sum_j (A^2)_{jj}^4 + \sum_j (AB)_j^4 + \sum_j (AB)_j^2 (A^2)_{jj}^2 \right],$$

$$H_n = \lambda^{1/m} \left[\sum_j \left(\sum_{l<j} (A^2)_{jl} \varepsilon_l \right)^4 \right],$$

$$I_n = \lambda^{1/m} \left[\sum_j (AB)_j \left(\sum_{l<j} (A^2)_{jl} \varepsilon_l \right)^3 \right],$$

$$J_n = \lambda^{1/m} \left[\sum_j (AB)_j^3 \sum_{l \leqslant j} (A^2)_{jl} \varepsilon_l \right],$$

$$K_n = \lambda^{1/m} \left[\sum_j ((AB)_j^2 + (A^2)_{jj}^2) \left(\sum_{l<j} (A^2)_{jl} \varepsilon_l \right)^2 \right],$$

and now $\sum_j (A^2)_{jj}^4 = o(1)$, using Lemma 2. Note that $|(AB)_j| \leqslant \max_k |B_k| \cdot \sum_k |A_{jk}|$. But $A_{jk} = 1/n \cdot G(t_j, t_k)$, and thus Lemma 1 applied to $s = t_j$ shows that $\sum_k |A_{jk}| \leqslant$ const. On the other hand if g is in W_m^2 then g is bounded, and thus $|B_j| \leqslant$ const by the same trick and $\sum_j (AB)_j^4 \leqslant \text{const} \cdot \sum_j (AB)_j^2 = O(n\lambda)$. Now $\sum_j (AB)_j^2 (A^2)_{jj}^2 \leqslant \text{const} \cdot \sum_j (A^2)_{jj}^2 = o(1)$ by Lemma 2 and $D_n = o(1)$. $E(H_n^2)$ is the product of $\lambda^{2/m}$, terms in which appear moments of ε up to order 8 and sums with the form

$$\left(\sum_j \left(\sum_{l<j} (A^2)_{jl}^2 \right)^2 \right)^2 \leqslant (\operatorname{Tr} A^8)^2 = O(\lambda^{-1/m}).$$

Now, $(\sum_j \sum_{l<j} (A^2)_{jl}^4)^2 \leqslant (\sum_j (\sum_{l<j} (A^2)_{jl}^2)^2)^2 = O(\lambda^{-1/m})$ so that H_n tends to 0 in probability.

The same result holds for I_n, J_n, and K_n using similar calculations.

Remark. We could have used the central limit theorem for generalized quadratic forms proved by De Jong [7] as suggested in the conclusion of [10]; however, there is no real gain with respect to assumptions as well as calculations that are similar.

III.2. *Periodic Case*

In the periodic case, and with equally spaced knots t_i, the influence matrix A is circulant (see [18] for the details). $A = I - n\lambda M^{-1}(I-P)$, where P is the matrix with 1 everywhere, and $M = H + n\lambda I$; the eigenvalues of H are

$$v_{j,n} = n \cdot \sum_{k=-\infty}^{+\infty} \frac{1}{(2\pi nk - j)^{2m}} = \frac{n}{(2\pi j)^{2m}} + c_j,$$

$$0 < c_j \leqslant c, \quad \text{for } j = 1, ..., n - \tfrac{1}{2}, \qquad \text{(III.8)}$$

$$v_{j,n} = v_{n-j,n} \qquad \text{and} \qquad v_{0,n} = n.$$

The proof of the central limit theorem follows arguments similar to those for Theorem 1 and will be omitted. The important result is here to exact computation of the asymptotic development of the variance and of the expectation terms.

LEMMA 5.

$$\operatorname{Tr}(A^p) = 1 + \lambda^{-1/2m} \cdot \int_0^\infty \frac{dy}{\pi(1+y^{2m})^p} + O(1) + O((\lambda n^{2m})^{1-p} \cdot \lambda^{-1/2m}).$$

Proof.

$$\operatorname{Tr}(A^p) = 1 + \sum_{j=1}^{n-1} \left(\frac{v_{j,n}}{v_{j,n} + n\lambda}\right)^p = 1 + 2 \sum_{j=1}^{n-1/2} \left(\frac{1}{1+\lambda(2\pi j)^{2m}}\right)^p + R_1.$$

Using (15), $R_1 = O(1/\lambda n^{2m-1})$:

$$2 \sum_{j=1}^{n-1/2} \left(\frac{1}{1+\lambda(2\pi j)^{2m}}\right)^p \lambda^{-1/2m} \cdot \int_0^\infty \frac{dy}{\pi(1+y^{2m})^p} + O(1) + O((\lambda n^{2m})^{1-p} \cdot \lambda^{-1/2m}).$$

LEMMA 6.

$$\sum B_j^2 = n\lambda^2 \tfrac{1}{2} \cdot \|g^{(2m)}\|^2 (1 + O(n^{-1} + \lambda^{-1} n^{-2m}))$$

$$\sum (AB)_j^2 = n\lambda^2 \tfrac{1}{2} \cdot \|g^{(2m)}\|^2 (1 + O(\lambda + n^{1-2m}\lambda^{-2})).$$

Proof. $\sum (AB)_j = n \cdot \lambda^2 \cdot \sum_{j=1}^{n-1} \|g_{j,n}\|^2 \cdot (n^2 v_{j,n}^2/(n\lambda + v_{j,n}^4))$ and $\sum (B_j)^2 = n \cdot \lambda^2 \sum_{j=1}^{n-1} \|g_{j,n}\|^2 \cdot (n/(v_{j,n} + n\lambda^2)^2)$ with $g_{j,n} = (1/n) \cdot \sum_{k=1}^n g(k/n) \cdot \exp(2i\pi jk/n)$. We shall develop the computations for $\sum (AB)_j^2$ and omit those for $\sum B_j^2$, which are similar. It is easy to see that

$$\sum (AB)_j^2 = n\lambda^2 \cdot \sum_{j=1}^{n-1} \|g_{jn}\|^2 \cdot \frac{(2\pi j)^{4m}}{(1+\lambda(2\pi j)^{2m})^4} + O(n^{1-2m}\lambda^{-2}).$$

Denote by g_j the jth Fourier coefficient of g:

$$g_j = \int_0^1 g(t) \exp(-2i\pi jt)\, dt.$$

To evaluate $\|g_{jn}\|^2 - \|g_j\|^2$, use Taylor developments, the fact that g is periodic (so $\int g^{(p)} = 0$ for $p > 0$), and induction reasoning to show

$$\|g_{jn}\|^2 - \|g_j\|^2 = O((j/n)^m).$$

Then

$$\sum (AB)_j^2 = n\lambda^2 \cdot \sum_{j=1}^{n-1} \|g_j\|^2 \cdot \frac{(2\pi j)^{4m}}{(1+\lambda(2\pi j)^{2m})^4} + O(n^{1-2m}\lambda^{-2}) + O(n^{1-2m}\lambda^{-1})$$

and

$$\sum_{j=1}^{n-1} \|g_j\|^2 \cdot \frac{(2\pi j)^{4m}}{(1+\lambda(2\pi j)^{2m})^4} = \frac{1}{2} \|g^{(2m)}\|^2 + \|g^{(3m)}\|^2 \, O(\lambda + 1/n^{2m}).$$

The theorem is deduced now from (15) and the fact that

$$E(R_{n,\lambda}) = (1/n) \cdot \left(\sum B_j^2 + \sigma^2 \cdot \mathrm{Tr}(A^2) \right).$$

Acknowledgment

The authors aknowledge the referee for providing Refs. [7] and [10] and useful comments.

References

[1] Bickel, P. J., and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence. *Sankhā Ser. A* **50**, No. 3 381–393.

[2] Bretagnolle, J., and Huber, C. (1979). Estimation d'une densité: risques minimax. *Z. W.* **47** 119–137.

[3] Cogburn, R., and Davis, H. T. (1974). Periodic splines and spectral estimation. *Ann. Statist.* **2**, No. 6 1108–1126.

[4] Cox, D. D. (1983). Asymptotics for M-type smoothing function. *Ann. Statist.* **11** 530–555.

[5] Cox, D. D. (1984). Multivariate smoothing spline function *SIAM J. Numer. Anal.* **21**, No. 4 789–813.

[6] Craven, P., and Wahba, G. (1979). Smoothing noisy data with spline function. *Numer. Math.* **31** 377–403.

[7] De Jong, P. (1987). A central limit theorem for generalised quadratic forms. *Probab. Theory Related Fields* **25** 261–277.

[8] Doukhan, P., and Leon, J. R. (1990). *Déviation quadratique d'estimateurs d'une densité par projections orthogonales*; Note aux CRAS, Paris, Ser. 1, Vol. 310, pp. 425–430.

[9] Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*, Dekker, New York.

[10] Eubank, R. L., and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression technics. *J. Amer. Statist. Assoc.* **85**, No. 410 387–392.

[11] Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16.

[12] Hall, P., and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, San Diego.

[13] Kimeldorf, G., and Wahba, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* **33** 82–95.

[14] Silverman (1984). Spline smoothing: The equivalent variable kernel method. *Ann. Statist.* **12**, No. 3 898–916.

[15] SILVERMAN (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *J. Roy. Statist. Soc. Ser. B* **47**, No. 1 1–52.

[16] SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13**, No. 3 970–983.

[17] UTRERAS DIAZ, F. (1980). Sur le choix du paramètre d'ajustement dans le lissage par fonctions splines. *Numer. Math.* **34** 15–28.

[18] WAHBA, G. (1975). Smoothing noisy data with spline functions. *Numer. Math.* **24**, 383–393.