Statistics & Probability Letters 18 (1993) 179–182
North-Holland
15 October 1993

The asymptotic mean squared error of L-smoothing splines

Felix P. Abramovich
Tel Aviv University, Ramat Aviv, Israel

Received March 1992
Revised February 1993

Abstract: We establish the asymptotical equivalence between L-spline smoothing and kernel estimation. The equivalent kernel is used to derive the asymptotic mean squared error of the L-smoothing spline estimator. The paper extends the corresponding results for polynomial spline smoothing.

Keywords: Nonparametric regression; equivalent kernel; differential operation; Green function.

1. Introduction

Consider the model

$$y_i = g(t_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $0 \leqslant t_1 \leqslant \cdots \leqslant t_n \leqslant 1$, $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. normal random variables with zero mean and variance σ^2, $g(\cdot)$ is a fixed but unknown function.

Define a nonparametric estimate $\hat{g}(t)$ as a solution to the following minimization problem:

$$\hat{g}(t) = \arg \min_{f \in \mathscr{D}(L)} \left\{ \frac{1}{n} \sum_{i=1}^{n} (y_i - f(t_i))^2 + \int_0^1 (Lf)^2 \right\} \tag{1}$$

where $L : \mathscr{D}(L) \subset L_2 \to L_2$ is a general differential operator of order m. This method, known as L-spline smoothing, which is considered in Kimeldorf and Wahba (1971) and Wahba (1990), is a natural generalization of polynomial spline smoothing (where $L = kD^m$) which has become one of the most popular approaches in nonparametric regression (a wide list of references may be found in Eubank, 1988; Wahba, 1990). It is well-known that the solution of (1) $\hat{g}(t)$ is a natural L-spline for the differential operator L with the knots $\{t_i\}$, that is $\hat{g}(\cdot)$ satisfies $L^* L \hat{g} = 0$ everywhere except, maybe, the knots $\{t_i\}$, where L^* is the adjoint operator to L, and conditions $L[\hat{g}(t)] \equiv 0$ on $[0, t_1]$ and $[t_n, 1]$, which imply $2m$ natural boundary conditions for endpoints 0 and 1. An explicit formula for $\hat{g}$ is given in Kimeldorf and Wahba (1971) and Wahba (1990).

From the quadratic nature of (1), $\hat{g}(\cdot)$ is linear in the observations $\{y_i\}$ and, hence, may be expressed as

$$\hat{g}(t) = \frac{1}{n} \sum_{i=1}^{n} W(t, t_i) y_i \tag{2}$$

for a certain weight function $W(t, t_i)$. Thus, the L-smoothing spline may be viewed as a general type of a kernel estimator with the kernel $W(\cdot, \cdot)$ called the *equivalent kernel.*

Correspondence to: Felix Abramovich, Department of Statistics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv 69978, Israel.

Rather suprisingly, the connection between spline smoothing and kernel estimation, originally based on different ideas, has much deeper foundation. Remarkably, the function W in (2) is closely related to the Green functions for the differential operator L^*L+I. It was shown for *polynomial* spline smoothing that the equivalent kernel $W(\cdot,\cdot)$ asymptotically may be well approximated by the Green function $G(\cdot,\cdot)$ for the differential operator $(-1)^m k^2 D^{2m}+I$ acting on the subspace of functions satisfying the natural boundary conditions (see Speckman, 1981; Cox, 1983; Messer, 1991). The explicit asymptotic expressions for $G(\cdot,\cdot)$ for the particular cases $m = 1, 2, 3$ were obtained in Silverman (1984).

In the following Section we shall generalize these results for L-smoothing splines and formulate the theorem that will show the asymptotic equivalence between L-spline smoothing and kernel estimation. This result will allow us to derive in Section 3 the asymptotic mean squared error for the L-smoothing spline estimator.

2. The equivalent kernel

In order to derive the equivalent kernel that corresponds to L-spline smoothing we shall first prove several lemmas given below.

Lemma 1. *The equivalent kernel $W(t, t_i)$ in (2) is the minimizer of*

$$\frac{1}{n}\sum_{j=1}^{n}\left(f(t_j)-n\delta_{ij}\right)^2+\int_0^1 (Lf)^2\,\mathrm{d}t \tag{3}$$

over all $f\in\mathscr{D}(L)$.

Proof. Let $\hat{g}_{(i)}(t)$ be a minimizer of (3). Note that (3) is a particular case of (1) when the data-vector y is $y_{(i)}=(0,\dots,0,n,0,\dots,0)'$.

Then, according to (2),

$$\hat{g}_{(i)}(t)=\frac{1}{n}\sum_{j=1}^{n}W(t,t_j)y_{(i)j}=W(t,t_i). \qquad \square$$

Define in L_2 with the usual inner product $\langle f_1,f_2\rangle=\int f_1(t)f_2(t)\,\mathrm{d}t$ the differential operator $S=L^*L+I$ and consider S acting on the subspace of functions from $\mathscr{D}(L^*L)$ satisfying the same $2m$ natural boundary conditions at the endpoints 0 and 1 as $\hat{g}(t)$.

Lemma 2. *S is a self-adjoint operator.*

Proof.

$$\begin{aligned}\langle Sf_1,f_2\rangle &= \langle L^*Lf_1,f_2\rangle+\langle f_1,f_2\rangle\\ &= \langle Lf_1,Lf_2\rangle+\langle f_1,f_2\rangle\\ &= \langle f_1,L^*Lf_2\rangle+\langle f_1,f_2\rangle\\ &= \langle f_1,Sf_2\rangle, \qquad \square\end{aligned}$$

Lemma 3. *S is a positive definite operator; that is there exists a strictly positive constant q, such that for every $f\in\mathscr{D}(S)$, $\langle Sf,f\rangle\geqslant q\|f\|^2$.*

Proof.

$$\begin{aligned}\langle Sf,f\rangle &= \langle L^*Lf,f\rangle+\langle f,f\rangle\\ &= \langle Lf,Lf\rangle+\langle f,f\rangle\geqslant\|f\|^2. \qquad \square\end{aligned}$$

Consider the equation $Sf=u$, $u\in L_2$. From the theory of functional analysis it is known (e.g. Vulikh, 1967, p. 337) that for a positive definite, self-adjoint operator S,

(i) there exists a unique solution of the above equation;
(ii) this solution is the minimizer of the function

$$\langle Sf,f\rangle-2\langle f,u\rangle$$

over all $f\in\mathscr{D}(S)$.

From (i) it follows that there exists an inverse operator G, called the *Green operator*, such that $Gu=f$. It is well-known that the operator G is an integral type operator with the kernel function $G(\cdot,\cdot)$ being the Green function for the operator S acting on $\mathscr{D}(S)$,

$$G[u(t)]=\int_0^1 G(t,s)u(s)\,\mathrm{d}s.$$

Note that being a positive definite operator, S has only positive eigenvalues. It is easy to show that the minimal eigenvalue is 1 of multiplicity m and the corresponding subspace of eigenfunctions is the m-dimensional kernel space of operator L,

$N(L)$. Thus, G also has positive eigenvalues λ's with $\lambda_{\max} = 1$ and, therefore, $\|G\| = \lambda_{\max} = 1$.

We have now finished all the preliminary results and are ready to prove the main theorem of this section. Note that $S[G(t, s)] = \delta(t - s)$. From (ii), $G(t, t_i)$ has to be a minimizer of

$$\langle Sf, f\rangle - 2\langle f, \delta(t - t_i)\rangle = \|Lf\|^2 + \|f\|^2 - 2f(t_i). \tag{4}$$

Require that the knots placement converges to a uniform distribution as $n \to \infty$ in the sense of Cox (1983), that is

$$\int_0^1 |t - F_n(t)| \, dt \to 0$$

where $F_n(t)$ is the c.d.f. of the probability measure which assigns the mass $1/n$ to each knot t_j (some equivalent convergence conditions are mentioned in Cox's paper).

Then, asymptotically $\|f\|^2 = \int f^2(t) \, dt$ is approximated by $(1/n)\sum_{j=1}^n f^2(t_j)$ and (4) may be replaced by

$$\int_0^1 (Lf)^2 \, dt + \frac{1}{n}\sum_{j=1}^n \left(f(t_j) - n\delta_{ij}\right)^2 - n. \tag{5}$$

But according to the first lemma, the minimizer of (5) is $W(t, t_i)$. Thus, the following result has been proved:

Theorem 1. *If the density of knots converges to a uniform distribution, the L-smoothing spline asymptotically behaves like a kernel estimator. The corresponding equivalent kernel if the Green function of the differential operator $S = L^*L + I$ acting on $\mathscr{D}(S)$ defined above.* □

3. Asymptotic mean squared error for the L-smoothing spline

Using the asymptotic equivalence between L-spline smoothing and kernel estimation established by the previous theorem we derive now a formula for the asymptotic mean squared error (MSE) of the L-smoothing spline estimator.

We start with the bias term $b(t)$. Suppose that $g(\cdot) \in \mathscr{D}(L)$. According to the theorem it follows from (2) that

$$\hat{g}(t) = \frac{1}{n}\sum_{i=1}^n G(t, t_i) y_i.$$

Replacing asymptotically the sum by the integral we have

$$E\hat{g}(t) = \int G(t, s) g(s) \, ds.$$

Since the operator S is self-adjoint, $G(t, s)$ is a symmetric function, that is $G(t, s) = G(s, t)$ for all s and t. Thus, $G(t, s)$ satisfies

$$L_s^* L_s[G(t, s)] + G(t, s) = \delta(s - t) \tag{6}$$

where L_s shows that differentiation is performed with respect to s.

Multiplying both parts of (6) by $g(s)$ and integrating them one immediately has

$$\int_0^1 L_s^* L_s[G(t, s)] g(s) \, ds + E\hat{g}(t) = g(t).$$

We note that since $G(t, s)$ satisfies the natural boundary conditions,

$$\int L_s^* L_s[G(t, s)] g(s) \, ds = \int L_s[G(t, s)] L_s[g(s)] \, ds.$$

So, finally

$$b(t) = \int_0^1 L_s[G(t, s)] L_s[g(s)] \, ds. \tag{7}$$

For the particalar case $L = kD^m$, which yields polynomial smoothing splines of order $2m - 1$, the corresponding Green function $G_0(t, s)$ (7) gives

$$b(t) = k^2 \int_0^1 G_0^{(0,m)}(t, s) g^{(m)}(s) \, ds$$

which coincides with known results for polynomial spline smoothing obtained in Speckman (1981).

The asymptotic variance follows immediately from (2) by replacing the sum by the integral:

$$\operatorname{Var} \hat{g}(t) = \frac{\sigma^2}{n}\int_0^1 G^2(t, s) \, ds \tag{8}$$

Combining (7) and (8) yields the asymptotic MSE:

$$\mathrm{MSE} = \left[\int_0^1 L_s[G(t, s)]\, L_s[g(s)]\, \mathrm{d}s\right]^2 + \frac{\sigma^2}{n}\int_0^1 G^2(t, s)\, \mathrm{d}s.$$

Remark. We would like to make brief comments about the notation $\int f(s)\delta(s-t)\,\mathrm{d}s = f(t)$ used in the paper. Being formally not correct in the sense of Riemann–Lebesgue integration, it is, however, a customary and usual notation for generalized functions. More rigorous proofs of the main results can be developed by first introducing a smoothed version of the δ-function and then taking appropriate limits.

Acknowledgment

The author would like to thank David Steinberg for helpful discussions and valuable remarks.

References

Cox, D.D. (1983), Asymptotics for M-type smoothing spline, *Ann. Statist.* **11**, 530–551.

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

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

Messer, K. (1991), A comparison of a spline estimate to its equivalent kernel estimate, *Ann. Statist.* **19**, 817–829.

Silverman, B.W. (1984), Spline smoothing: the equivalent variable kernel method, *Ann. Statist.* **12**, 898–916.

Speckman, P. (1981), The asymptotic integrated mean square error for smoothing noisy data by splines, preprint, Univ. of Oregon (Eugene, OR).

Vulikh, B.Z. (1967), *Introduction to Functional Analysis* (Nauka, Moscow). [In Russian.]

Wahba, G. (1990), *Spline Models for Observational Data* (SIAM, Philadelphia, PA).