Smoothing splines as locally weighted averages

Douglas Nychka*

Department of Statistics

North Carolina State University

August 30, 1989

Abstract

A smoothing spline is a nonparametric estimate that is defined as the solution to a minimization problem. Because of the form of this definition it is most convenient to represent a spline estimate relative to an orthonormal set of basis functions. One problem with this representation is that it obscures the fact that a smoothing spline, like most other nonparametric estimates, is a local, weighted average of the observed data. This property has been used extensively to study the limiting properties of kernel estimates and it is advantageous to apply similar techniques to spline estimates. Although equivalent kernels have been identified for a smoothing spline these functions are either not accurate enough for asymptotic approximations or are restricted to equally spaced points. Rather than improve these approximations, this work concentrates on bounding the size of the kernel function. It is shown that the absolute value of the equivalent kernel decreases exponentially away from its center.

This bound is used to derive the asymptotic form for the pointwise bias of a first order smoothing spline estimate. The pointwise bias has the usual form that would be expected for a second order kernel estimate with a variable bandwidth.

Another potential application of this bound is in establishing the consistency of a spline estimate uniformly with respect to the smoothing parameter. Such results are important for studying data-based methods for selecting the smoothing parameter.

AMS Subject Classifications: Primary 62G05 Secondary 62G15

Key Words and Phrases: Nonparametric regression, Symbolic computation, Kernel estimates, Cross-validation.

*Research supported by NSF grant DMS-8715756.

Section 1 Introduction

The basic model considered in nonparametric regression is

$$(1.1) \qquad Z_i = \theta(t_i) + \epsilon_i, \quad 1 \leq i \leq n,$$

where $\underset{\sim}{Z}$ is an observation vector depending on a smooth function θ, observation points, $0 = t_1 \leq t_2 \, \ldots \, \leq t_n = 1$ and errors, ϵ_i, that are assumed to be independent, have mean zero and have common variance σ^2. A statistical problem that is posed by this model is to estimate θ without having to assume a parametric form for this function.

One class of estimators that has been studied extensively consists of weighted local averages of the observations where the weights are specified by a kernel function. For example the kernel estimate attributed to Nadaryra (1964) and Watson (1964) has the form:

$$(1.2) \qquad \hat{\theta}(t) = \frac{1}{n} \sum_{j=1}^{n} w(t,t_j) \, Z_j \, ,$$

where

$$(1.3) \qquad w(t,t_j) = \frac{1}{h} \, K\left[\frac{t-t_j}{h}\right] \Big/ \sum_{k=1}^{n} \frac{1}{nh} \, K\left[\frac{t-t_k}{h}\right].$$

The function K is assumed to be continuous, to be symmetric about zero and to integrate to one. The bandwidth parameter, h, controls the relative weight given to observations as a function of their distance from t. Determining a suitable value for this free parameter objectively from the data is an important practical issue. The reader is referred to (Eubank 1988, Härdle, Hall and Marron 1988) for background on this important topic.

It is easy to study the theoretical properties of kernel estimators because the estimate is an explicit function of the observations and the weight function, w(.,.), has a simple form. The formulation of the kernel estimate, however, constrains it to local operations on the data. This restriction may make it difficult to apply kernel methods to more complicated observational models than (1.1). An alternative to kernel estimators are those based on maximizing a penalized likelihood. For example, under the assumption that the distribution of the errors may be approximated by a normal distribution, one has the log likelihood

$$L(\underset{\sim}{Z}|\theta,\sigma) = -\sum_{k=1}^{n} (Z_k - \theta(t_k))^2 / 2\sigma^2 \; - n \, ln(\sqrt{2\pi}\sigma).$$

Furthermore we will assume that

$$\theta \in W_2^m[0,1] = \{\theta: \theta, \dots, \theta^{(m-1)} \text{ absolutely continuous}, \theta^{(m)} \in L_2[0,1]\}$$

and take as a roughness penalty

$$J(\theta) = \int_{[0,1]} (\theta^{(m)})^2 dt.$$

For $\delta > 0$ and we have the penalized log likelihood

(1.4) $$L(\underset{\sim}{Z} | \theta, \sigma) - \delta J(\theta).$$

A maximum penalized likelihood estimate (MPLE) is the function that maximizes (1.4) for all $\theta \in W_2^m[0,1]$. This approach has the advantage of being very flexible in adapting to more complicated models but one problem is that the estimate is defined implicitly. See Nychka et al. (1983), O'Sullivan, Yandell and Raynor (1985) and Silverman (1985) for some examples of applications of splines to observational models that differ from (1.1).

The MPLE just described actually takes the form of an m^{th} order smoothing spline. Multiplying (1.4) by $-1/n$ and setting $\lambda = \delta/n$ one may define the MPLE, $\hat{\theta}_\lambda$, as the minimizer of

(1.5) $$\mathcal{L}(\theta) = \frac{1}{n} \sum_{k=1}^{n} (Z_k - \xi(t_k))^2 + \lambda \int_{[0,1]} (\xi^{(m)})^2 dt.$$

over all $\xi \in W_2^m[0,1]$.

Although the MPLE is a linear function of the observations, and, thus can be written in the same form as (1.2), the weight functions do not have a closed form. The work of Silverman (1984) and Messer (1989) identify a kernel that will approximate weight functions but these results are limited. Although Silverman's approximation provides excellent intuition about how a spline estimate weights the data relative to the distribution of the observation points, it is not accurate enough for establishing asymptotic properties. Messer's Fourier analysis gives a high order approximation to w(.,.) but depends on $\{t_k\}$ being equally spaced. This article takes a slightly different approach than these authors. Besides characterizing the form of w(s,t), bounds are given on how these weights decrease as $|t-\tau|$ increases. Such bounds are useful for analyzing the asymptotic properties of the estimate and more will be said about this below.

The main result of this article will be stated. Let F_n denote the empirical distribution function for $\{t_j\}_{1 \le j \le n}$, let F be a distribution function with a continuous and strictly positive density function f on [0,1] and let

$$D_n = \sup_{t\in[0,1]} |F_n - F|.$$

Following the work of Cox (1983,1984b), the weight function for an m^{th} order spline estimate can be approximated by a Green's function to a particular $2m^{th}$ order differential equation. This integral kernel will be denoted by $G_\lambda(t,\tau)$ (see (2.5) for a complete definition of this Green's function). The following assumption is the key to our approximation.

Assumption A *Let* $\rho=\lambda^{1/2m}$. *There are positive constants* α, ϵ, $K<\infty$ *such that for all* $t,\tau \in [0,1]$

(1.6) $$|G_\lambda(t,\tau)| \le (K/\rho^{1+\epsilon})\, e^{-\alpha|t-\tau|/\rho}$$

(1.7) $$\left|\frac{\partial}{\partial t}G_\lambda(t,\tau)\right| \le (K/\rho^{2+\epsilon})\, e^{-\alpha|t-\tau|/\rho}$$

If $s\neq t$ then

(1.8) $$\left|\frac{\partial^2}{\partial t\partial\tau}G_\lambda(t,\tau)\right| \le (K/\rho^{3+\epsilon})\, e^{-\alpha|t-\tau|/\rho}$$

Either $\frac{\partial^2}{\partial t\partial\tau}G_\lambda(t,\tau)$ *exits for* $s=t$ *in which case* (1.8) *holds*

or for all $h\in C_0[0,1]$

(1.9) $$\left|\frac{d}{dt}\int_{[0,1]}\left(\frac{\partial}{\partial\tau}G_\lambda(t,\tau)\right)h(\tau)d\tau\right| \le (K/\rho^{3+\epsilon})\left[\int_{[0,1]}\tfrac{1}{2}\, e^{-\alpha|t-\tau|/\rho}|h(\tau)|d\tau + |h(\tau)|\right]$$

Theorem 1.1 *Let* $\rho=\lambda^{1/2m}$ *and* $\delta_n = 4KD_n/\rho^{2+\epsilon}$. *Under Assumption A, if* $\delta_n<1$ *and* $\rho<\alpha$

$$|w(t,t_j)| < \frac{K}{(1-\delta_n)\rho^{1+\epsilon}}\, e^{-\alpha|t-t_j|/\rho},$$

and

$$|w(t,t_j)-G_\lambda(t,t_j)| < \frac{\delta_n K}{(1-\delta_n)\rho^{1+\epsilon}}\, e^{-\alpha|t-t_j|/\rho},$$

uniformly over $t, t_j \in [0,1]$.

The proof for this theorem is given in Section 3. One attractive feature of this result is that this bound is not asymptotic and holds exactly for finite sample sizes. The only requirement is that the sizes of the equivalent bandwidth, ρ, and D_n must be balanced such that $\delta_n < 1$ and $\rho < \alpha$.

Assumption A of this theorem places certain restrictions on how G_λ and its derivatives must decrease as $|t-\tau|$ increases. In this article we show that this assumption holds for m=1, and any $\epsilon > 0$ when f has a continuous derivative. For the more interesting case of cubic smoothing splines (m=2), when F is a uniform distribution Assumption A holds with $\epsilon = 0$ and $\alpha = 1$. Our same proof strategy can also be used to extend this result to more general densities (see the remarks at the end of Section 5) but a rigorous proof is beyond the scope of this article. In general we conjecture that Assumption A will hold for all m and a wide class of design densities. It is for this reason that Theorem 1.1 has been phrased in terms of this more general assumption rather our specific results for the cases when m=1,2.

One application of Theorem 1.1 is to derive an asymptotic form for the bias of a spline estimate. To our knowledge this is the first characterization of the pointwise bias of a smoothing spline estimate.

Theorem 1.2 *Assume that* $\hat{\theta}_{n\lambda}$ *is a first order* (m=1) *smoothing spline estimate and assume that* $D_n = O(n^{-1})$. *Suppose that* $\theta \in C_2[0,1]$ *and satisfies the Holder condition:*

$$\left|\theta^{(2)}(t) - \theta^{(2)}(\tau)\right| \le M|t-\tau|^\beta \quad \textit{for some } \beta > 0 \textit{ and } M < \infty.$$

Choose $\Delta > 0$,

(1.10) $\qquad \lambda_n \sim n^{-1/2}\log(n)$ *and* $\Lambda_n \to 0$ *as* $n \to \infty$.

then

(1.11) $\qquad E\,\hat{\theta}_{n\lambda}(t) - \theta(t) = (\lambda/f(t))\theta^{(2)}(t) + o(\lambda)$

uniformly for $\lambda \in [\lambda_n, \Lambda_n]$ *as* $n \to \infty$ *and* $t \in [\Delta, 1-\Delta]$ such that $\theta^{(2)}(t) \neq 0$.

This form is similar to what one would expect for a positive kernel estimate (Silverman 1984, Section 3). By identifying an equivalent bandwidth with $\sqrt{\lambda/f(t)}$ one may see that the spline estimate actually varies the amount of smoothing according to the local density of the observations.

A potential application of Theorem 1.1 is in analyzing the convergence of this estimator uniformly with respect to the smoothing parameter. For example, suppose $\hat{\theta}_\lambda$ is an m^{th} order smoothing spline estimate and let

$$\mathrm{WASE}(\lambda)= \frac{1}{n} \sum_{j=1}^{n} \omega(t_j)(\theta(t_j)-\hat{\theta}_\lambda(t_j))^2$$

denote the weighted average squared error (WASE) for $\omega \geq 0$. One basic problem in nonparametric regression is to prove that under suitable conditions on λ_n and Λ_n,

$$\sup_{\lambda\in[\lambda_n,\Lambda_n]} \left|\frac{\mathrm{WASE}(\lambda)-\ \mathrm{EWASE}(\lambda)}{\mathrm{EWASE}(\lambda)}\right| \xrightarrow{p} 0 \quad \text{as } n\to\infty. \tag{1.12}$$

The work in Speckman (1982), Li (1986), and Cox (1984b) consider this problem in the context of demonstrating that the smoothing parameter chosen by generalized cross-validation yields an asymptotically optimal procedure for choosing λ. Speckman and Li require the stringent assumption that the errors follow a normal distribution. Cox's work has weakened this assumption by first developing a Gaussian approximation to the spline estimate and then applying the theory developed for cross-validation under the assumption of normality. This work is limited, however, to the situation where $\omega \equiv 1$.

Readers familiar with the asymptotic theory for kernel estimators (Marron and Härdle, 1985, 1986) will note that it is not necessary to develop a Gaussian approximation to the estimate as an intermediate step in order to prove (1.12) . Instead one relies on the local averaging property of these estimators (see for example the condition stated at (3.2), Marron and Härdle, 1986) and the existence of higher moments of the error distribution. One reason for developing the bounds in Theorem 1.1 was to apply these simpler strategies from the work on kernel estimates to splines. Also, this approach would lend itself to consider general weight functions for the WASE. In doing do so one could then handle the important case of cross-validation restricted to an interior interval or study the properties of local cross-validation.

The remainder of this article will be outlined. The next section reviews Cox's representation of a spline estimate and contrasts this form with an expansion using an orthogonal basis. Section 3 gives a proof of Theorem 1.1 and Section 4 gives a proof of Theorem 1.2.

The most difficult part of this analysis is characterizing a Green's function that corresponds to a $2m^{th}$ order differential equation. Section 5 discusses these functions and derives an explicit formula for the case m=1 for uniform densities. The more important case when m=2 can be computed using symbol manipulation software and a program for computing the Green's function using MACSYMA is included in an appendix. Section 5 also contains a proof that verifies Assumption A for first order splines. One interesting side result from this proof is an approximation to the general Green's function using the uniform Green's function and a rescaling of the independent variable that locally is proportional to $1/\sqrt{f}$.

Section 2 Representations of a smoothing spline estimate

Although smoothing splines are often represented using an orthogonal decomposition the equivalent kernel representation is based on a slightly different viewpoint. This section contrasts these two representations and explains some problems working with the orthogonal expansion.

One way of solving the spline minimization problem is via an orthogonal decomposition that simultaneously diagonalizes the mean residual sum of squares and the roughness penalty. Using the theory for reproducing kernel Hilbert spaces there is a set of functions $\{\phi_{\nu n}\}_{1\leq\nu\leq n}$ $\subseteq W_2^m[0,1]$ with the property that the solution to (1.5) is contained in the linear subspace spanned by $\{\phi_{\nu n}\}_{1\leq\nu\leq n}$ (Eubank 1988). Furthermore, the Demmler-Reinsch basis may be chosen to satisfy the orthogonality relations:

(2.1) $$\frac{1}{n}\sum_{k=1}^{n} \phi_{\nu n}(t_k)\, \phi_{\mu n}(t_k) = \delta_{\nu\mu}$$

(2.2) $$\int_{[0,1]} \phi''_{\nu n}(t)\phi''_{\mu n}(t)\, dt = \gamma_{\nu n}\, \delta_{\nu\mu} \qquad 0\leq\gamma_{1n}\leq\gamma_{2n} \ldots \leq\gamma_{nn}$$

where $\delta_{\nu\mu}$ is Kronecker's delta.

Thus, the spline estimate will have the explicit form

$$\hat{\theta}_\lambda(t) = \sum_{\nu=1}^{n} \alpha_\nu\, \phi_{\nu n}(t), \qquad \text{for } \underset{\sim}{\alpha} \varepsilon \Re^n$$

and by (2.1) and (2.2)

(2.3) $$\mathcal{L}\left(\sum_{\nu=1}^{n} \alpha_\nu\, \phi_{\nu n}\right) = \sum_{\nu=1}^{n} (z_\nu - \alpha_\nu)^2 + \lambda \sum_{\nu=1}^{n} \gamma_{\nu n}\alpha_\nu^2$$

where $z_\nu = \frac{1}{n}\sum_{k=1}^{n} \phi_{\nu n}(t_k)\, Z_k$. Minimizing (2.2) over $\underset{\sim}{\alpha} \in \mathbb{R}^n$ gives

$$\alpha_\nu = \frac{z_\nu}{1+\lambda\gamma_{\nu n}}$$

and thus

$$\hat\theta_\lambda(t) = \sum_{\nu=1}^{n} \frac{z_\nu}{1+\lambda\gamma_{\nu n}}\phi_{\nu n}(t).$$

Note that with this representation there is also a simple expression for the average squared error

$$\mathrm{ASE}(\lambda) = \frac{1}{n}\sum_{j=1}^{n}(\theta(t_j)-\hat\theta_\lambda(t_j))^2 = \sum_{\nu=1}^{n}\left[\frac{z_\nu}{1+\lambda\gamma_{\nu n}} - \theta^*_{\nu n}\right]^2,$$

where $\theta^*_{\nu n} = \frac{1}{n}\sum_{k=1}^{n} \theta(t_k)\, \phi_{\nu n}(t_k)$.

Thus, we have an appealing decomposition of the expected ASE into the usual two parts: the average squared bias and the average variance.

$$\mathrm{EASE}(\lambda) = \lambda^2 \sum_{\nu=1}^{n}\left[\frac{\gamma_{\nu n}\theta^*_{\nu n}}{1+\lambda\gamma_{\nu n}}\right]^2 + \sigma^2/n \sum_{\nu=1}^{n}\frac{1}{1+\lambda\gamma_{\nu n}}$$

It is difficult, however to work with norms other than the two that are diagonalized by this basis. For example, consider weighted average squared error. For most interesting choices of ω it is not possible to get a simple expression for this error criterion in terms of λ and $\underset{\sim}{\theta}^*_n$. Also, it should be noted that except in the case of periodic splines the Demmler-Reinsch basis can not be written down explicitly. Therefore, even though one can make the identification

$$w(t,t_j) = \sum_{\nu=1}^{n} \frac{\phi_{\nu n}(t)\phi_{\nu n}(t_j)}{1+\lambda\gamma_{\nu n}}, \qquad (2.4)$$

this does not lead to useful methods for studying a spline's asymptotic properties.

Another way of characterizing the solution to (1.5) is to examine the functional equations obtained from setting the functional derivative equal to zero. A necessary and sufficient condition for $\hat\theta_\lambda$ to be a solution is that

$$\frac{1}{n}\sum_{k=1}^{n} -(Z_k-\hat\theta_\lambda(t_k))h(t_k) + \lambda\int_{[0,1]} \hat\theta_\lambda^{(m)} h^{(m)} dt = 0$$

for all h in a dense set of $W_2^m[0,1]$ (Cox 1983). One way of characterizing $w(\ \cdot\ ,t_j)$ is to note that this function is actually the smoothing spline estimate applied to the "data": $Z_k=n$ for $j=k$ and $Z_k=0$ otherwise. To simplify notation let $w_j(t)= w(t,t_j)$ thus we have

$$\frac{1}{n}\sum_{k=1}^{n} w_j(t_k))h(t_k) + \lambda \int_{[0,1]} w_j^{(m)} h^{(m)} dt - h(t_j) = 0$$

or

$$\int_{[0,1]} w_j\, h\, dF_n + \lambda \int_{[0,1]} w_j^{(m)} h^{(m)} dt - h(t_j) = 0$$

for $1 \leq j \leq n$.

The solution to this functional equation is given by Cox (1983,1984). Let $G_\lambda(t,\tau)$ be the Green's function for the differential operator $\lambda(-D^2)^m + f$ with domain

$$\mathfrak{D}_m = \{h \in C_{2m}[0,1]\colon h^{(\nu)}(0) = h^{(\nu)}(1) = 0 \ \text{ for } m \leq \nu \leq 2m-1\}. \tag{2.5}$$

(It should be noted that G_λ defined here differs from Cox's Green's function by the factor $f(t_j)$).

To simplify notation let $g_j(t) = G_\lambda(t,t_j)$. Thinking of G_λ as an integral kernel one can define the integral operator:

$$R_{n\lambda}(h)(t) = \int_{[0,1]} G_\lambda(t,\tau)\, h(\tau) d(F - F_n) \tag{2.6}$$

and

$$w_j(t) = g_j(t) + \sum_{\nu=1}^{\infty} R_{n\lambda}^{\nu}(g_j)(t). \tag{2.7}$$

(See the remarks at the end of Section 3 for a justification of this expansion.)

The leading term in (2.7), $G_\lambda(t,t_j)$, does not depend on the empirical distribution of $\{t_j\}$. The other terms in the infinite sum can be interpreted as corrections to this asymptotic expression based on the difference between F and F_n. Most of the asymptotic theory for splines is based on showing that these correction terms are in some sense negligible. If $\{t_j\}$ are equally spaced then these higher order corrections may be calculated using Fourier methods and this is one way of interpreting the analysis in Messer (1989).

Section 3 Assumptions, Preliminary Lemmas and the Proof of Theorem 1.1

Perhaps the most important property used to prove Theorem 1.1 is that bounds on the Green's function and its partial derivatives can be expressed in terms of a double exponential kernel. The basic idea behind the proof is elementary. Under the assumption that F_n is sufficiently close to F, an inductive argument will show that bounds on $R^{\nu}_{n\lambda}(g_j)$ may be inferred from bounds on g_j. This result in turn implies a similar bound on the infinite sum at (2.7).

For clarity we repeat the other key hypotheses in Theorem 1.1 .

Assumption B $\rho \leq \alpha$ *and* $\delta_n = 4KD_n/\rho^{2+\epsilon} < 1$

The following two lemmas will be needed in the proof of Theorem 1.1.

Lemma 3.1 *Let* $I(t,\tau,\rho) = \int_{[0,1]} e^{-\alpha|s-\tau|/\rho \; -\alpha|t-\tau|/\rho} d\tau.$

If $\rho < \alpha$ *then*

$$I(t,\tau,\rho) \leq 2\, e^{-\alpha|t-\tau|/\rho}, \qquad \text{for } t,\tau \in [0,1].$$

and

$$I(t,\tau,\rho) \leq (M+1/\alpha)\rho\, e^{-\alpha|t-\tau|/\rho}, \qquad \text{for } |t-\tau| \leq \rho M.$$

Proof Assume without loss of generality that $0 \leq s \leq t \leq 1$. Then

$$I(t,\tau,\rho) = [t-s]\, e^{-\alpha|t-\tau|/\rho}$$

$$-\ (\rho/2\alpha)\, e^{-\alpha|s+t|/\rho} + (\rho/2\alpha)\, e^{-\alpha|2-t-\tau|/\rho}$$

By assumption $\rho \leq \alpha$ and if w= min(u+v, 2−u+v) then $e^{-\alpha|w|/\rho} \leq e^{-\alpha|u-v|/\rho}$. The bound now follows. The second conclusion is proved using similar arguments.□

Lemma 3.2 *Under Assumptions A and B, for all* $\nu > 0$,

(3.1) $$\left|R^{\nu}_{n\lambda}(g_j)\right| < (\delta_n)^{\nu}\, (K/\rho^{1+\epsilon})\, e^{-\alpha|t-tj|/\rho}$$

(3.2) $$\left|\frac{\partial}{\partial t} R^{\nu}_{n\lambda}(g_j)\right| < (\delta_n)^{\nu} (K/\rho^{2+\epsilon})\, e^{-\alpha|t-tj|/\rho}$$

uniformly for $t,\ t_j \in [0,1]$.

Proof By Assumption A, (3.1) and (3.2) both hold when $\nu=0$. Suppose that (3.1) and (3.2) hold for some $\nu=\mu$ then it will be shown that these inequalities must also hold for $\nu=\mu+1$.

$$R_{n\lambda}^{\mu+1}(g_j)(t) = \int_{[0,1]} G_\lambda(t,\tau)\, R_{n\lambda}^{\mu}(g_j)(\tau)d\Big(F(\tau)-F_n(\tau)\Big).$$

Integrating by parts,

$$R_{n\lambda}^{\mu+1}(g_j)(t) = \int_{[0,1]} \frac{\partial}{\partial\tau}\Big[G_\lambda(t,\tau)\, R_{n\lambda}^{\mu}(g_j)(\tau)\Big]\Big[F(\tau)-F_n(\tau)\Big]d\tau.$$

$$\Big|R_{n\lambda}^{\mu+1}(g_j)(t)\Big| \le \sup_{\tau\epsilon[0,1]}|F(\tau)-F_n(\tau)| \int_{[0,1]}\Big|\frac{\partial}{\partial\tau}\Big[G_\lambda(t,\tau)\, R_{n\lambda}^{\mu}(g_j)(\tau)\Big]\Big|d\tau$$

$$\le D_n\, (\delta_n)^{\mu}(2K^2/\rho^{3+2\epsilon})\, I(t,t_j,\rho).$$

From the first part of Lemma 3.1,

$$\le(\delta_n)^{\mu}(4D_nK/\rho^{2+\epsilon})\,(K/\rho^{1+\epsilon})\, e^{-\alpha|t-tj|/\rho}.$$

Thus (3.1) must hold for $\nu=\mu+1$. If the mixed partial for G_λ exists for $t=\tau$ then a similar argument is used to establish the induction hypothesis at (3.2). If (1.9) holds, however, then there is a slight difference in the proof.

$$(3.3)\qquad \Big|\frac{\partial}{\partial t} R_{n\lambda}^{\mu+1}(g_j)(t)\Big| \le \Big|\frac{d}{dt}\int_{[0,1]}\Big(\frac{\partial}{\partial\tau}G_\lambda(t,\tau)\Big)\, h(\tau)d\tau\Big| + \sup_{\tau\epsilon[0,1]}|F(\tau)-F_n(\tau)| \int_{[0,1]}\Big|G_\lambda(t,\tau)\,\Big(\frac{\partial}{\partial\tau}R_{n\lambda}^{\mu}(g_j)(\tau)\Big)\Big|d\tau$$

where $h(\tau)= (F(\tau)-F_n(\tau))\, R_{n\lambda}^{\mu}(g_j)(\tau)$. The first term of the LHS of this inequality can be bounded by using the conditions at (1.9) and (3.1) and the applying Lemma 3.1. The second term is handled using the same arguments as those given above. □

Proof of Theorem 1.1

$$(3.4)\qquad \Big|w_j(t)\Big| \le \Big|g_j(t)\Big| + \sum_{\nu=1}^{\infty}\Big|R_{n\lambda}^{\nu}(g_j)(t)\Big|$$

and by Lemma 3.2

$$\le \Big[1 + \sum_{\nu=1}^{\infty}\Big(4KD_n/\rho^{2+\epsilon}\Big)^{\nu}\Big](K/\rho^{1+\epsilon})\, e^{-\alpha|t-tj|/\rho}$$

$$\le \big[1 + \delta_n/(1-\delta_n)\big](K/\rho^{1+\epsilon})\, e^{-\alpha|t-tj|/\rho}$$

The second part of the theorem is proved by considering only the second term on the RHS at (3.4). □

This result is also a justification for the convergence of the infinite sum at (2.7) because it has been shown that $\sup\left|R^{\nu}_{n\lambda}(g_j)\right| < \omega^{\nu}M$ with $M<\infty$ and $\omega<1$.

Section 4 Asymptotic form for the bias

This section proves Theorem 1.2. Although the conclusion of this theorem is specific to first order splines most of the proof does not depend on this restriction. In fact proving the more general result:

$$E\,\hat{\theta}_{n\lambda}(t) - \theta(t) = (\lambda/f(t))\theta^{(2m)}(t) + o(\lambda)$$

where $\rho = \lambda^{1/2m}$, $\hat{\theta}_{n\lambda}$ is an m^{th} order smoothing spline and $\theta^{(2m)}$ satisfies the Holder condition only depends on verifying Assumption A. For this reason, most of the discussion in this section considers a general m.

First a lemma will be given that will be needed in the proof. To gain some insight into the asymptotic approximation of the Green's function to w(.,.) the reader should compare the representation of G_λ at (4.2) to (2.4).

Lemma 4.1 *Let* G_λ *be the Green's function defined in Section 2. Suppose that* $g\in\mathfrak{D}_m$ *and furthermore*

(4.1) $$\left|g^{(2m)}(s)-g^{(2m)}(t)\right| \le M|t-\tau|^{\beta} \quad \textit{for some } \beta>0 \textit{ and } M<\infty,$$

then

$$\frac{d^{2m}}{dt^{2m}}\int_{[0,1]} G_\lambda(t,\tau)\, g(\tau)dF(\tau) \rightarrow (-1)^m g^{(2m)}(t) \qquad \textit{as} \quad \lambda\rightarrow 0.$$

uniformly for $t\in[0,1]$.

Proof From the discussion in Section 2, Cox (1988) there exists a basis $\{\phi_\nu\}$ for $W^m_2[0,1]$ such that

$$\int_{[0,1]} \phi_\nu\phi_\mu\, dF = \delta_{\nu\mu}$$

and

$$\int_{[0,1]} \phi_\nu^{(m)} \phi_\mu^{(m)} \, dt = (-1)^m \gamma_\nu \, \delta_{\nu\mu} \qquad 0 \le \gamma_1 \le \gamma_2 \ldots .$$

Moreover,

(4.2) $$G_\lambda(t,\tau) = \sum_{\nu=1}^{\infty} \frac{\phi_\nu(t)\phi_\mu(\tau)}{1+\lambda\gamma_\nu}.$$

and

$$\phi_\nu(t)(1+\lambda\gamma_\nu) = \int_{[0,1]} G_\lambda(t,\tau)\phi_\nu(\tau)dF(\tau).$$

Thus we see that $\phi_\nu \in \mathfrak{D}_m$ and in fact, by induction $\phi_\nu \in C_\infty[0,1]$. Integrating parts and applying the boundary conditions from the definition of $\mathfrak{D}_m$,

$$\int_{[0,1]} \phi_\nu^{(m)} \phi_\mu^{(m)} \, dt = (-1)^m \int_{[0,1]} \phi_\nu \phi_\mu^{(2m)} \, dt.$$

Using the orthogonality relations and the fact that these functions are a basis for $W_2^m[0,1]$, it must follow that $\phi_\nu^{(2m)} = (-1)^m \gamma_\nu \phi_\mu$.

For $g \in W_2^m[0,1]$ we have the representation

(4.3) $$g = \sum_{\nu=1}^{\infty} c_\nu \phi_\nu \quad \text{where } c_\nu = \int_{[0,1]} g\phi_\nu dF.$$

Moreover, because of the condition at (4.1), and the boundary conditions for $g \in \mathfrak{D}_m$, the partial sums of this series expansion will actually converge to g in a norm that is stronger than the one for $C_{2m}[0,1]$. Specifically with respect to the scale of norms

$$\|g\|_\rho^2 = \sum_{\nu=1}^{\infty} c_\nu^2 (1+ \gamma_\nu^\rho)$$

(4.3) is convergent for all $\rho < 2 + 1/2m + \beta/2m$. We will also use the fact that this interpolation norm dominates supremum norm for $\rho > 1/2m$. The reader is referred to Cox 1988, Section 3 for a rigorous development of these norms and the related function spaces.

Thus, it follows that

(4.4) $$\frac{d^{2m}}{dt^{2m}} g(t) = \sum_{\nu=1}^{\infty} c_\nu \gamma_\nu \phi_\nu(t)$$

and

$$\psi(t)= \frac{d^{2m}}{dt^{2m}}\Big[(-1)^m g(t) - \int_{[0,1]} G_\lambda(t,\tau)g(\tau)dF(\tau).\Big]$$
$$= \sum_{\nu=1}^{\infty} c_\nu \frac{\lambda\gamma_\nu}{(1+\lambda\gamma_\nu)}\phi_\nu(t)$$

Now choose μ such $1/2m < \mu-2 < 1/2m + \beta/2m$. It follows that

$$\|\psi\|^2_{\mu\text{-}2} = \sum_{\nu=1}^{\infty} c_\nu^2\Big[\frac{\lambda\gamma_\nu}{1+\lambda\gamma_\nu}\Big]^2\Big(1+\gamma_\nu^{\mu-2}\Big) < \lambda^2 \sum_{\nu=1}^{\infty} c_\nu^2(\gamma_\nu^2+\gamma_\nu^\mu)$$
$$< \lambda^2\Big(\|g\|_2^2 + \|g\|_\mu^2\Big) = O(\lambda^2)$$

Since $\mu-2 > 1/2m$ there is an $M<\infty$ that does not depend on ψ such that $\sup|\psi| < M\,\|\psi\|^2_{\mu\text{-}2}$ and thus by (4.4), $\sup|\psi| = O(\lambda^2)$ as $\lambda\to\infty$. □

We now consider the approximation to the pointwise bias of a smoothing spline.

$$E\,\hat\theta_{n\lambda}(t)-\theta(t) = \frac{1}{n}\sum_{j=1}^{n} w(t,t_j)\theta(t_j) - \theta(t)$$
$$=\Big(\int_{[0,1]} G_\lambda(t,\tau)\,\theta(\tau)dF(\tau) - \theta(t)\Big)$$
$$+ \int_{[0,1]} r_{n\lambda}(t,\tau)\theta(\tau)dF_n(\tau) + \int_{[0,1]} G_\lambda(t,\tau)\theta(\tau)d(F_n-F)(\tau)$$

where $w(t,t_j)= G_\lambda(t,t_j) + r_{n\lambda}(t,t_j)$.

It will be convenient to refer to the three terms of (4.5) as

$$b_\lambda(t) + R_1 + R_2$$

The proof of Theorem 1.2 will consist of showing that $b_\lambda(s)$ converges to the functional form stated in the theorem while R_1 and R_2 are $o(b_\lambda(t))$ uniformly over the specified ranges for λ and t.

Proof of Theorem 1.2 First, assume that θ satisfies the natural boundary conditions in (2.5) and let $g(t)= \int_{[0,1]} G_\lambda(t,\tau)\,\theta(\tau)dF(\tau)$. By the definition of the Green's function

$$\lambda(-D^2)^m g + fg = f\theta,$$

so $b_\lambda(t)= g(t)-\theta(t) = \dfrac{\lambda(-1)^m}{f(t)} D^{2m}g$. From Lemma 4.1 it now follows that

(4.5) $$b_\lambda(t)=\frac{\lambda\theta^{(2m)}(t)}{f(t)}(1+o(1))$$

uniformly for $t\in[0,1]$ as $\lambda\to 0$.

We now deal with the case when θ does not satisfy the natural boundary conditions. Let $\overline{\theta}$ denote a function that agrees with θ on $[\Delta/2, 1-\Delta/2]$ but has been modified outside this interval so as to satisfy the natural boundary conditions.

$$g(t)=\int_{[0,1]} G_\lambda(t,\tau)\,\overline{\theta}(\tau)dF(\tau)+\int_{[0,1]} G_\lambda(t,\tau)\Big(\theta(\tau)-\overline{\theta}(\tau)\Big)dF(\tau).$$

Using the exponential bounds on G_λ, and the $\sup|\theta-\overline{\theta}|<\infty$, the second term will be $O(e^{-\alpha\Delta/2\rho}/\rho^{1+\epsilon})=o(\lambda)$ uniformly for $t\in[\Delta,1-\Delta]$ as $\lambda\to\infty$. (4.5) now holds for this version of θ and for t in this subinterval.

From the second assertion of Theorem 1.1,

$$\begin{aligned}|R_1| &< \sup_{[0,1]}|\theta(\tau)f(\tau)|\int_{[0,1]}|r_{n\lambda}(t,\tau)|d\tau\\ &< \sup_{[0,1]}|\theta(\tau)f(\tau)|\left[\frac{\delta_n}{1-\delta_n}\right]\int_{[0,1]}(K\gamma(\tau)/\rho)e^{-\alpha|t-\tau|/\rho}d\tau\\ &= O(D_n/\rho^2) \qquad \text{as } n\to\infty.\end{aligned}$$

The last remainder term may be handled by integrating by parts and using Assumption A. Thus,

$$|R_2|< D_n\int_{[0,1]}\left|\frac{d}{d\tau}\Big(G_\lambda(s,\tau)\,\theta(\tau)\Big)\right|d\tau=O(D_n/\rho) \qquad \text{as } n\to\infty.$$

Finally, note that by the choice of λ_n, for $\rho=\sqrt{\lambda_n}$, $(D_n/\rho^4)\to 0$ as $n\to\infty$. □

Section 5 Exponential bounds on the Green's function.

The main result in this article depends on Assumption A. This condition is verified for the case when m=1 and we discuss the symbolic computation of G_λ for the case m=2. Both of these calculations are done, however, under the assumption that F is the uniform distribution function. This section ends by extending the results for the uniform case when m=1 to all strictly positive, continuously differentiable densities.

When the roughness penalty depends on the first derivative (m=1) it is possible to derive a fairly simple expression for the Green's function, G_λ. This integral kernel will be the solution of

$$-\lambda \frac{d^2}{ds^2}G_\lambda(t,\tau) + G_\lambda(t,\tau) = 0 \quad \text{for } s\neq t$$

subject to the natural boundary conditions

$$\frac{d}{ds}G_\lambda(0,t)= \frac{d}{ds}G_\lambda(1,t)=0,$$

and continuity conditions

$$G_\lambda(t,\tau)\Big|_{s=t^+} - G_\lambda(t,\tau)\Big|_{s=t^-} = 0$$

and

$$\frac{d}{ds}G_\lambda(t,\tau)\Big|_{s=t^+} - \frac{d}{ds}G_\lambda(t,\tau)\Big|_{s=t^-} = 1/2\lambda.$$

Setting $\rho=\sqrt{\lambda}$, the solution has the form

$$G_\lambda(t,\tau)=\begin{cases} \psi_1(t,\tau) & s<t \\ \psi_2(t,\tau) & s\geq t \end{cases}$$

$$\psi_1(t,\tau)= a_1(t)e^{s/\rho} + a_2(t)e^{-s/\rho} \text{and} \quad \psi_2(t,\tau)= b_1(t)e^{s/\rho} + b_2(t)e^{-s/\rho} .$$

Note that by symmetry $G_\lambda(t,\tau)=G_\lambda(t,s)$ and also $G_\lambda(t,\tau)= G_\lambda(1-s,1-t)$. Thus it is sufficient to find just ψ_1. Solving for the coefficients yields

$$a_1(t)=\frac{e^{-t/\rho}}{2\rho(1-\delta(t))}, \qquad a_2(t)= a_1(t),$$

where

$$\delta(t)= e^{-2t/\rho}- \omega(t)\Big(e^{-2(1-t)/\rho}-1\Big), \qquad \omega(t)= \frac{1 + e^{-2t/\rho}}{1+ e^{2(1-t)/\rho}}.$$

Therefore, for $s\leq t$

$$G_\lambda(t,\tau)= \frac{e^{-|t-s|/\rho} + e^{-(t+s)/\rho}}{2\rho(1-\delta(t))}$$

It is tedious, but straight forward to verify that G_λ given above satisfies parts (1.6), (1.7) and (1.8) of Assumption A. The mixed partial does not exist when $t=\tau$ and so it is necessary to consider (1.9). Right and left derivatives with respect to t do exist for $(\partial/\partial\tau)G_\lambda$ when $t=\tau$, however, and the difference between these two functions is uniformly bounded. These properties along with (1.8) imply (1.9). Since G_λ is a symmetric function of t and τ (1.9) will also hold with the roles of t and τ interchanged:

$$(5.0)\quad \left| \frac{d}{d\tau} \int_{[0,1]} \left(\frac{\partial}{\partial t} G_\lambda(t,\tau)\right) h(t)dt \right| \leq (K/\rho^{3+\epsilon}) \left[\int_{[0,1]} \frac{1}{2} e^{-\alpha|t-\tau|/\rho} |h(t)| dt + |h(t)| \right]$$

This version will be needed in the proof of Theorem 5.1.

Calculating the Green's function for the case when m=2 involves much more algebra. For example, there are eight linear equations of coefficients that need to be solved rather than four. For this reason a symbol manipulation program was used to calculate G_λ. A MACSYMA batch program is included in the Appendix and the reader is referred to Nychka (1985) for more details on the form of this expression. From this expression, however, one may show that Assumption A holds. In this case the mixed partial exists for all t and τ and thus one can omit verifying (1.9).

The two Green's functions described above are limited to the case when the marginal distribution of $\{t_k\}$ converges to a uniform distribution. In general, the Green's function associated with the differential operator $\lambda(-D^2)^m + f$ and domain $\mathfrak{D}_m$ is much more difficult to derive. An alternative approach is to use the form of G_λ for a uniform density to infer some of the properties of more general Green's functions. Restricting attention to the case when m=1, we show that Assumption A will also hold for the class of densities described in Section 1. This result is the content of the following theorem and is based on an inductive argument similar to that given in the proof of Lemma 3.2.

Theorem 5.1 *Let* f *be a continuously differentiable, strictly positive density function. Let* $\mathbb{L}_f = -\lambda D^2 + f$ *with domain* $\mathfrak{D}_1$ *and take* $G_{\lambda,f}$ *to denote the corresponding Green's function. Let*

$$\zeta = \sup_{[0,1]} \left|(\gamma'/\gamma^2)\right|,$$

$$\beta_1 = \sup_{[0,1]} |\gamma| \text{ and } \beta_2 = \sup_{[0,1]} |1/\gamma|$$

where

$$\gamma = f^{1/2} \Big/ \int_0^1 f^{1/2}(w)dw.$$

For any $\epsilon>0$, *if* $\omega=2K\zeta\rho^{\epsilon} < 1$ *and* $\rho<\alpha$ *then* $G_{\lambda,f}$ *will satisfy Assumption A provided that* K *is replaced by* $K\beta_2/(1-\omega)$ *and* α *is replaced by* $\alpha\beta_1$.

Before giving the proof of this theorem it will be helpful to present a transformation that relates $\mathbb{L}_f$ to $\mathbb{L}_1$. Let

$$\Gamma(t)= \int_0^t \gamma(w)\, dw \ .$$

Note that Γ is a 1-1 differentiable map of [0,1] onto itself. Now suppose that h solves $\mathbb{L}_f h=g$ for some $g\in L^2[0, 1]$. If q is a function satisfying $q(\Gamma(t))=h(t)$ then from the chain rule for differentiation it follows that

$$\psi_1(u)\,(\mathbb{L}_1 q)(u) \ + \ \lambda\psi_2(u)\, Dq(u)= g(\,\Gamma^{-1}(u)\,)$$

$$q'(0)=q'(1)=0 \tag{5.1}$$

with

$$\psi_1(u)=\left[\gamma(\,\Gamma^{-1}(u)\,)\,\right]^2 \quad \text{and} \quad \psi_2(u)=\gamma'(\,\Gamma^{-1}(u)\,).$$

Thus if $H_\lambda(u,v)$ is the Green's function associated with the differential operator $\psi_1\mathbb{L}_1+\lambda\psi_2 D$ and domain $\mathfrak{D}_1$ then

$$G_{\lambda,f}(t,\tau)= H_\lambda(\Gamma(t)\ ,\ \Gamma(\tau))\ \gamma(\tau). \tag{5.2}$$

In this way bounds on $G_{\lambda,f}$ may be inferred from bounds on H_λ. This later kernel is easier to analyze because it is involves $\mathbb{L}_1$.

Proof of Theorem 5.1 First we derive bounds on the integral kernel H_λ associated with the integral operator $(\psi_1\mathbb{L}_1 \ + \ \lambda\psi_2 D)^{-1}$. To simplify notation set $\mathbb{G}=\mathbb{L}_1^{-1}$. Letting $\mathbb{A}=\mathbb{G}(\psi_2/\psi_1)D$ then at least in a formal sense

$$(\psi_1\mathbb{L}_1 \ + \ \lambda\psi_2 D)^{-1} = (I+ \ \lambda\mathbb{A})^{-1}\,\mathbb{G}(1/\psi_1)$$

$$= \mathbb{G}(1/\psi_1) + \sum_{\nu=1}^{\infty}(-\lambda\mathbb{A})^{\nu}\mathbb{G}(1/\psi_1)$$

This expansion will be justified through the following analysis of the operators $(\lambda \mathbb{A})^{\nu}\mathbf{G}$.

Let R_{λ}^{ν} denote the integral kernel associated with $(\lambda \mathbb{A})^{\nu}\mathbf{G}$. It will be argued that for all $\nu \geq 0$ and $u,v \in [0,1]$

$$(5.3) \qquad \left|R_{\lambda}^{\nu}(u,v)\right| < K(\omega^{\nu}/\rho^{1+\epsilon})\, e^{-\alpha|u-v|/\rho}.$$

Clearly if (5.3) is true then the series expansion of $(I+\lambda \mathbb{A})^{-1}\mathbf{G}$ is valid.

Suppose that (5.3) holds for some $\nu = \mu$.

$$\left|R_{\lambda}^{\mu+1}(u,v)\right| < \int_{[0,1]} \left|R_{\lambda}^{\mu}(u,w)\, (\rho^{2}\psi_{2}(w)/\psi_{1}(w))\, \frac{\partial}{\partial w} G_{\lambda}(w,v)\right| dw$$

$$< \sup_{[0,1]} \left|(\psi_{2}/\psi_{1})\right| \int_{[0,1]} (\omega^{\mu}K^{2}/\rho)\, e^{-\alpha(|w-u|+|w-v|)/\rho} du$$

Applying the first inequality in Lemma 3.2,

$$(5.4) \qquad < K\omega^{\mu}(2\zeta K/\rho)\, e^{-\alpha|u-v|/\rho} = K(\omega^{\mu+1}/\rho^{1+\epsilon})\, e^{-\alpha|u-v|/\rho}.$$

Clearly (5.3) holds when $\nu = 1$ and thus by induction (5.3) must hold for all ν. Summing these bounds on the kernels it now follows that

$$H(u,v) < K/(1-\omega)\, (1/\psi_{1}(v))\, (1/\rho^{1+\epsilon})\, e^{-\alpha|u-v|/\rho}$$

or by (5.2),

$$G_{\lambda,f}(t,\tau) < K/(1-\omega)\, (1/\gamma(t))(1/\rho^{1+\epsilon})\, e^{-\alpha|\Gamma(t)-\Gamma(\tau)|/\rho}$$

Now, $|\Gamma(t)-\Gamma(\tau)| < \beta_{1}|t-\tau|$ and so

$$G_{\lambda,f}(t,\tau) < K\beta_{2}/(1-\omega)\, (1/\rho^{1+\epsilon})\, e^{-\alpha\beta_{1}|t-\tau|/\rho}$$

Thus, the first inequality of Assumption A now holds for $G_{\lambda,f}$ given the appropriate identifications of the constants K and α and the other parts of Assumption A can be proved using similar inductive arguments. Note that establishing (1.9) will depend on the bound at (5.0). □

This section ends by discussing some modifications to Theorem 5.1. Although this paper does not focus on approximations to the shape of w(. , .) the proof of this theorem does yield some insight into an approximation based on the Green's function for uniform densities. If one restricts attention to t,τ such $|\Gamma(t)-\Gamma(\tau)| \leq M\rho$ then one may apply the second inequality of Lemma 3.1 (rather than the first) at line (5.3). This second inequality eliminates a factor of $(1/\rho)$ and implies that R_ν for $\nu>0$ is of higher order than R_0. The final result is that

$$\left|G_{f,\lambda}(t,\tau) - G_\lambda(\Gamma(t), \Gamma(\tau))\,\gamma(\tau)\right| < (C\omega/\rho^\epsilon)\, e^{-\alpha\beta_2|t-\tau|/\rho}$$

where $C= (M+1/\alpha)\, K\beta_1/(1-\omega)$. Thus by a judicious transformation of t, $G_{f,\lambda}$ may approximated locally by using the Green's function from a uniform density. Note that when $t-\tau$ is small $\Gamma(t)-\Gamma(\tau)\approx(t-\tau)\gamma(t) = (t-\tau)\, f^{1/2}(t)$. This rescaling agrees with Silverman's results and implies a variable amount of smoothing where the local bandwidth is proportional to $f^{1/2}(t)$.

The extension of Theorem 5.1 to higher order Green's functions has the same general form as that in the proof given above. The correct choice for γ involves replacing the exponent 1/2 by 1/2m. It will be necessary to assume that f has $2m-1$ continuous derivatives along with some boundary conditions that guarantee that q will be in $\mathfrak{D}_m$. Like the case for m=1 a change of variables based on Γ will yield a differential operator with $\psi_1\mathbb{L}_1$ as the first term plus a differential operator of order $2m-1$. This second operator will be more complicated but can be handled using the same techniques given above. The proof will require that the Green's function for the uniform density satisfy Assumption A. In addition, one will need exponential bounds similar to those at (1.7) for the partials $(\partial^\nu/\partial s^\nu)G_\lambda$ $1\leq\nu\leq(2m\text{-}1)$ and a bound similar to (5.0) to bound the derivative of

$$\frac{d}{d\tau}\int_{[0,1]}\left(\partial^\nu/\partial t^\nu\, G_\lambda(t,\tau)\right) h(t)dt.$$

References

Cox, D. D. (1983). Asymptotics for *M*-type smoothing splines. *Annals of Statistics* 11 530-551.

Cox, D. D. (1984a). Gaussian Approximation of Smoothing Splines. *University of Wisconsin /Madison Department of Statistics Technical Report* #743. Madison , WI.

Cox, D.D. (1984b). Multivariate smoothing spline functions. *SIAM Journal of Numerical Analysis.* 21 789-813.

Cox, D. D. (1988). Approximation of method of regularization estimates. *Annals of Statistics* 16 694-712.

Eubank, R. L. (1988). *Spline smoothing and nonparametric regression.* Marcel Dekker, Inc., New York.

Härdle, W. and Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Annals of Statistics* 13 1465-1481.

Härdle, W., Hall, P. and Marron, S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association* 83 86-95.

Li, K. (1986). Asymptotic optimality of C_L and generalized cross-validation in ridge regression with an application to spline smoothing. *Annals of Statistics* 14 1101-1112.

Marron, J. S. and Härdle, W. (1986). Random approximations to some measures of accuracy in nonparametric curve estimation. *Journal of Multivariate Analysis* 20 91-113.

Messer, K. (1989). A comparison of a spline estimate to its equivalent kernel estimate. to appear in *Annals of Statistics.*

Nadaraya, F. A. (1964). On estimating regression. *Theory of Probability Applications* 9 141-142.

Nychka, D. W. (1986). The mean posterior variance of a smoothing spline and a consistent estimate of the mean squared error. North Carolina Institute of Statistics Mimeograph Series No. 1688, Raleigh, North Carolina.

Silverman, B. W. (1984). Spline smoothing: The equivalent variable kernel method. *Annals of Statistics* 12 898-916.

Silverman, B. W. (1985). "Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting," *Journal of the Royal Statistical Society* B 47, pp. 1-52.

Speckman, P. (1983). Efficient nonparametric regression with cross-validated smoothing splines. Unpublished manuscript.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Series A* 26 359-372.