THE INSTITUTE OF STATISTICS

THE MEAN POSTERIOR VARIANCE OF A SMOOTHING SPLINE AND A CONSISTENT ESTIMATE OF THE MEAN SQUARED ERROR.

by

DOUGLAS NYCHKA

Institute of Statistics Mimeograph Series no. 1688.

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

The mean posterior variance of a smoothing spline and a consistent estimate of the mean squared error.

Douglas Nychka

Department of Statistics

North Carolina State University

Abstract

A smoothing spline estimator can be interpreted in two ways: either as the solution of a variational problem or as the posterior mean when a particular Gaussian prior is placed on the unknown regression function. In order to explain the remarkable performance of her Bayesian "confidence intervals" in a simulation study, Wahba (1983) conjectured that the mean posterior variance of a spline estimate evaluated at the observation points will be close to the expected mean squared error. The estimate of the mean posterior variance proposed by Wahba is shown to converge in probability to a quantity proportional to the expected mean squared error. This result is established by relating this statistic to a consistent risk estimate based on generalized cross validation.

Subject Classification: (Amos-Mos): Primary 62G05, Secondary 62G15

Key Words and Phrases

Nonparametric regression, smoothing spline, confidence intervals for a smoothing spline, risk estimate, cross validation.

1. Introduction

Consider the additive model:

$$Y_{kn} = f(t_{kn}) + e_{kn} \qquad k = 1,n.$$

where f is a smooth, unknown function evaluated at the points $0 \leq t_{1n} \leq t_{2n} \ldots \leq t_{nn} \leq 1$ and $\{e_{kn}\}_{k=1,n}$ are independent and identically distributed errors with $E(e)=0$, $E(e^2) = \sigma^2$.

This paper will study some of the large sample properties of a smoothing spline as a nonparametric estimate for f. For $\lambda > 0$ the m^{th} order natural smoothing spline estimate for f, f_λ, is the minimizer of

$$(1.1) \qquad \frac{1}{n}\sum_{k=1}^{n}\left[Y_{kn} - h(t_{kn})\right]^2 + \lambda \int \left(h^{(m)}(t)\right)^2 dt$$

for all $h \in W_2^m[0,1] = \left\{f\colon f^{(m)} \in L^2[0,1] \text{ and } f^{(k)},\ k=1, m-1 \text{ are absolutely continuous}\right\}$.

This nonparametric estimate for f has an intriguing stochastic interpretation (Wahba(1978)). Suppose $\underset{\sim}{e}_n \sim N(\underset{\sim}{0}, \sigma^2 I)$ and f is a realization from the Gaussian process

$$\sum_{j=1}^{m} \theta_j\, t^{j-1} + \frac{\sigma^2}{n\lambda} \int_0^t \frac{(t-s)^{m-1}}{(m-1)!} + dW(s)$$

where $\underset{\sim}{\theta} \sim N(0,\xi I)$ and $W(\cdot)$ is the standard Wiener process with $W(0) = 0$. If f_λ is a natural spline estimate then

$$f_\lambda(t) = \lim_{\xi \to \infty} E\left[f(t) \mid \underset{\sim}{Y}\right] .$$

When the smoothing parameter, λ, is fixed f_λ is a linear function of $\underset{\sim}{Y}$ and there is an matrix $A(\lambda)$, such that $\underset{\sim}{f}_\lambda = A(\lambda)\underset{\sim}{Y}$ with $\underset{\sim}{f}_\lambda' = \left[f_\lambda(t_{1n})\ldots, f(t_{nn})\right]$. The limiting conditional covariance is

$$\sigma^2 A(\lambda) = \lim_{\xi \to \infty} \mathrm{Var}\left\{ \underset{\sim}{f} \;\middle|\; \underset{\sim}{Y} \right\}.$$

A periodic smoothing spline is obtained by restricting the minimization of (1.1) to the set: $\left\{ h \varepsilon W^m[0,1]:\ h^{(k)}(0) = h^{(k)}(1),\ k = 0,\ m-1 \right\}$ and these conditional equations will also hold for a periodic spline provided the Gaussian process is modified slightly to reflect the periodicity of f.

Wahba (1983) used this correspondence between a smoothing spline estimator and the posterior distribution of f to motivate 95% pointwise confidence intervals for $f(t_{kn})$ of the form:

$$f_\lambda(t_{kn}) \pm 1.96\, \sigma \sqrt{[A(\lambda)]_{kk}}$$

Although these intervals were derived within a Bayesian framework the simulation results reported in Wahba (1983), indicate that they work well when evaluated by a frequentist criterion for fixed functions. Wahba's confidence procedure is one of the few approaches that has been developed for spline estimates, and, given the growing interest in the use of smoothing splines for data analysis (see Silverman (1984) for a review), it is important to understand the statistical properties of this method. In explaining the success of this confidence procedure, Wahba conjectured that the mean posterior variance (MPV) for f_λ at the points $\{t_{kn}\}_{k=1,n}$ is close to the expected mean squared error. Unfortunately Wahba's argument for her hypothesis has several places that are heuristic. Also, her analysis does not acknowledge the fact that the value for λ was not fixed in the simulations but rather was determined by generalized cross validation. In this paper we give a proof for a version of Wahba's conjecture that accounts for the adaptive choice of the smoothing parameter. To establish this result we consider a consistent estimator of the expected mean squared error obtained by subtracting an an estimate of $\frac{1}{n}\sum_{k=1}^{n} e_{kn}^{2}$ from the generalized cross validation validation function. Asymptotically, this estimator also turns out to be proportional to the estimated MPV proposed by Wahba. Given this relationship, our version of Wahba's conjecture is simple to establish.

Let

$T_n(\lambda) = \frac{1}{n}\sum_{k=1}^{n}\left[f_\lambda(t_{kn}) - f(t_{kn})\right]^2$ denote the mean squared error for f_λ and let λ^o be the minimizer of $E\,T_n(\lambda)$ for $\lambda \varepsilon [0,\infty)$. Wahba's conjecture is:

(1.2) *If* $f \varepsilon W_2^m[0,1]$ *then* $\sigma^2 \dfrac{\frac{1}{n}\,\mathrm{tr}\left[A(\lambda^o)\right]}{E\,T_n(\lambda^o)} = \alpha\left[1 + o(1)\right]$

as $n \to \infty$ *for some* $\alpha \varepsilon \left[(1+\frac{1}{4m})(1-\frac{1}{2m}),\ 1\right]$.

This conjecture is important because it links a frequency quantity with a functional of the posterior distribution. In fact, this correspondence provides a simple explanation for the accuracy of the confidence intervals in Wahba's Monte Carlo study. For a periodic spline, consider the random variable: $U = \left[f(\tau) - f_{\lambda^o}(\tau)\right]$ where τ is a random variable independent of $\underset{\sim}{e}$ taking on the values $\{t_{kn}\}_{k=1,n}$ with equal probability. U has a mean of zero, has variance $ET_o(\lambda^o)$ and for the test cases considered by Wahba, is approximately normally distributed. (see Nychka (1986)). Thus, we are lead to the 95% confidence interval.

$$f_{\lambda^o}(\tau) \pm 1.96\sqrt{E\,T_n(\lambda^o)} \qquad \text{for } f(\tau)\ .$$

Because Wahba evaluates her procedure using the average coverage probability of the pointwise confidence intervals at $\{t_{kn}\}_{,k=1,n}$ she in effect is computing the confidence level for the interval given above. Moreover, because the periodic splines used in this study were applied to equally spaced data, $A(\lambda)$ is a circulant matrix and thus $A_{kk}(\lambda^o) = \frac{1}{n} \mathrm{tr}A(\lambda^o)$. Therefore, for Wahba's Bayesian confidence intervals:

$$(1.3) \qquad f_{\lambda^o}(t_{kn}) \pm 1.96 \sqrt{\frac{\sigma^2}{n} \mathrm{tr}A(\lambda^o)} \quad ,$$

the average coverage probability will be close to the nominal level, provided that $\frac{1}{n} \mathrm{tr}A(\lambda^o)$ is close to $E\, T_n(\lambda^o)$.

Although (1.2) is stated in terms of expected values where λ^o is known, the actual confidence intervals computed in Wahba's Monte Carlo study were

$$(1.4) \qquad f_{\hat{\lambda}}(t_{kn}) \pm 1.96 \sqrt{\frac{\hat{\sigma}^2}{n} \mathrm{tr}A(\hat{\lambda})}$$

where $\hat{\lambda}$ is the minimizer of the generalized cross validation function:

$$V(\lambda) = \frac{\frac{1}{n} \left\| \left[I-A(\lambda)\right]Y \right\|^2}{\left[\frac{1}{n} \mathrm{tr}\left(I-A(\lambda)\right)\right]^2} \quad \text{and} \quad \hat{\sigma}^2 = \frac{\left\| \left[I-A(\lambda)\right]Y \right\|^2}{\mathrm{tr}\left[I-A(\lambda)\right]}$$

is an estimator for σ^2 . The frequency interpretation of the Wahba's confidence intervals that has been briefly outlined above can be extended to the situation when λ^o and σ^2 are estimated by $\hat{\lambda}$ and $\hat{\sigma}^2$. If $\frac{\hat{\sigma}^2}{n}\mathrm{tr}\left[A(\hat{\lambda})\right]$ is a consistent estimate for the MPV then the difference between the average coverage probabilities for the intervals in (1.3) and (1.4) will converge to zero in probability (Nychka 1986). This is the motivation for focusing on the "estimated" MPV: $\frac{\hat{\sigma}^2}{n} \mathrm{tr} \left[A(\hat{\lambda})\right]$ and our results are summarized by the theorem stated below.

Suppose that the knots, $\{t_{kn}\}_{k=1,n}$, are a random sample from a distribution with pdf g such that g is strictly positive on [0,1] and $g \epsilon C^{\infty}[0,1]$.

Theorem 1.1

If $E|e_{kn}|^8 < \infty$, $\hat{\lambda}$ *is the minimizer of* $V(\lambda)$ *restricted to* $[\lambda_n, \infty]$ *with* $\lambda_n \sim n^{-\frac{4m}{5}}$, f_λ *is a natural spline with* $m \geq 2$, $f \epsilon W_2^{2m}$, *and* f *satisfies the natural boundary conditions:*

(1.5) $f^{(k)}(0) = f^{(k)}(1) = 0$, $k = m, 2m-1$,

then

(1.6) $$\frac{\frac{\hat{\sigma}^2}{n} \operatorname{tr}A(\hat{\lambda})}{E\, T_n(\lambda^o)} \xrightarrow{P} K \quad \text{as } n \to \infty .$$

where

$$K = \left(1 - \frac{1}{2m}\right) \left(1 + \frac{1}{4m}\right)$$

The hypotheses of this theorem differ in several significant ways from Wahba's initial formulation. First, note that the frequency interpretation of the average coverage probability outlined above does not really depend on the errors being normally distributed. All that is required that is U be approximately normal. For this reason we considered the consistency of the estimated MPV where the normal assumption has been replaced by the boundedness of higher order moments of the error distribution. By placing more restrictions on f we obtain convergence to a particular constant that is actually the lower end point for the range of α given in (1.2). More will be said about the boundary conditions required of f at the end of this section. Finally, we note that Theorem 1.1 will also hold if f_λ is a periodic spline provided that the natural boundary conditions in (1.5) are replaced by the periodic ones: $f^{(k)}(0) = f^{(k)}(1)$, for $k = 0,2m-1$. The three smooth periodic test functions considered by Wahba in her simulations satisfy these boundary conditions.

Theorem 1.1 will be established by comparing $\frac{\hat{\sigma}^2}{n}$ tr $A(\hat{\lambda})$ to an estimator of $ET_n(\lambda^o)$ based on the generalized cross validation function. Besides giving a clear proof, we feel this risk estimate is of interest in its own right. From Speckman (1983) and Cox (1984), under the conditions of Theorem 1.1:

$$(1.7)\quad \frac{V(\hat{\lambda}) - S_n^2}{E\ T_n(\lambda^o)} \xrightarrow{P} 1 \text{ uniformly for } \lambda \in [\lambda_n, \infty) \text{ where } S_n^2 = \frac{1}{n}\sum_{k=1} e_k^2$$

Thus, given a consistent estimate of S_n^2 , say $\hat{S}_n^2$, a natural estimate of $E_nT\ (\lambda^o)$ is $\hat{T}_n = V(\hat{\lambda}) - \hat{S}_n^2$. The key point of our argument is to establish the following :

Theorem 1.2

Let

$$(1.8)\quad \hat{S}_n^2 = \frac{\| (I-A(\hat{\lambda}))Y \|^2}{tr(I-\mathcal{C}A(\hat{\lambda}))} \qquad \textit{and } \mathcal{C} = 2-\frac{1}{K}$$

then

$$\hat{S}_n^2 - S_n^2 = o_p\left(E\ T_n(\lambda^o)\right)$$

Combining this result with (1.2) we have $\hat{T}_n \big/ E\, T_n(\lambda^o) \xrightarrow{P} 1$ and Theorem 1.1 will follow easily from the consistency of $\hat{T}_n$ and the asymptotic behavior of tr $A(\hat{\lambda})$.

$$\hat{T}_n = \frac{1}{n} \left\| \left(I-A(\hat{\lambda})\right) Y \right\|^2 \quad \frac{(2-\mathcal{C}) \frac{1}{n} \operatorname{tr} A(\hat{\lambda}) + \frac{1}{n} \operatorname{tr} A(\hat{\lambda})^2}{\left[1 - \frac{1}{n} \operatorname{tr} A(\hat{\lambda})\right] \left[1 - \frac{\mathcal{C}}{n} \operatorname{tr} A(\hat{\lambda})\right]}$$

and after some algebra

$$(1.9) \qquad = (2-\mathcal{C}) \frac{\hat{\sigma}^2}{n} \operatorname{tr} A(\hat{\lambda}) \left[\frac{1 + \frac{1}{n} \operatorname{tr} A(\hat{\lambda})/(2-\mathcal{C})}{\left[1 - \frac{\mathcal{C}}{n} \operatorname{tr} A(\hat{\lambda})\right]} \right]$$

Thus, $\hat{T}_n$ is proportional to the estimated mean posterior variance. From Lemma 2.1 and the choice of λ_n, the bracketed term in (1.9) converges to 1 and noting that $K = \frac{1}{2-\mathcal{C}}$, Theorem 1.1 now follows.

Arguing from an analogy with linear regression, Wahba suggested $\hat{\sigma}^2$ as an estimate for σ^2 . Although $\hat{\sigma}^2 - S_n^2 = o_p(1)$,

$\dfrac{\hat{\sigma}^2 - S_n^2}{E\, T_n(\lambda_o)} \xrightarrow{P} \mathcal{E}$ where $\mathcal{E} \neq o$. This is the reason for the slight modification of $\hat{\sigma}^2$ in (1.8) through the insertion of the constant $\mathcal{C}$. This change is somewhat arbitrary and there other modifications that will also work. For example, one could obtain the same consistency results with

$$\hat{S}_n^2 = \frac{\left\| \left(I-A(\bar{\lambda})\right) Y \right\|^2}{\operatorname{tr} \left(I-A(\bar{\lambda})\right)} \quad \text{and} \quad \bar{\lambda} = \left(\frac{2m}{2m-1}\right)^{\frac{2m}{4m+1}} \hat{\lambda} \ .$$

Besides an estimate of σ^2 based on the residual sum of squares, consistent estimates for σ^2 can be constructed using first (or second) differences of $\underset{\sim}{Y}$. (see Rice (1983)). This latter estimate has the advantage that no assumptions need to be made about the functional form for the bias of f_λ . Unfortunately, estimators based on differencing $\underset{\sim}{Y}$ will not give consistent estimates of S_n^2 .

Risk estimators for a smoothing spline have been developed through the connection between Stein's unbiased risk estimate and generalized cross validation (Li (1985). One interesting feature of Li's work is that his risk estimate only requires an estimate of σ^2 rather than S_n^2 . His results, however, are limited to the case of normal errors and involve the asymptotic approximation of f_λ by a Stein estimate. Also, our work is related to Rice (1983) because in the special case when $t_{kn} = \frac{k}{n}$, a periodic smoothing spline is also a kernel estimate.

The next section states general versions of Theorems 1.1 and 1.2 while Section 3 contains the preliminary lemmas for the proofs of these theorems. The proofs of the theorems are given in Section 4. The last section establishes the form for the mean squared bias of a natural cubic spline on an interior interval of [0,1]. This technical result was included to support the discussion at the end of this first section.

The boundary conditions on f required in Theorem 1.1 are very restrictive and we end this section by discussing the background for this condition and suggesting a modification to avoid this limitation on f.

Let $\mathcal{S} = \left\{ f\colon f \in W_2^{2m}[o,1] \ , \ f^{(k)}(o) = f^{(k)}(1), \ k = m, \ 2m-1 \right\}$.

Members of $\mathcal{S}$ are the "very smooth" functions defined in Wahba (1977) or the case p=2 in the appendix of Wahba (1983) . This class was considered because E $T_n(\lambda^o)$ achieves its fastest rate of convergence for $f \varepsilon \mathcal{S}$ and the average squared basis for these "very smooth" functions has asymptotically an explicit functional form. This form is given in condition (F3) in the next section and is the key to establishing the consistency of $\hat{\lambda}$ and $\hat{S}_n^2$.

Although members of $\mathcal{S}$ must have more derivatives than is required by the penalty function in (1.1), the main objection to this class is that f must satisfy the natural boundary conditions in its higher derivatives. Probably for most applications this is not a reasonable assumption. Unfortunately, if $f \varepsilon W^{2m}$ [0,1] but does not satisfy these natural boundary conditions then condition (F3) will not hold. Rice and Rosenblatt (1984) considered a slightly different spline estimate and showed that if $f \varepsilon C^4[0,1]$, but f is not contained in $\mathcal{S}$ the bias of the spline in the neighborhoods of 0 and 1 will dominate the average squared bias. This has the unfortunate consequence that the mean squared error (and the cross validation function) will be dominated by the behavior of the estimate at the boundaries.

A simple solution to this problem is to restrict evaluation of f_λ to an interior interval: $[\delta, 1-\delta]$ for some $\delta > o$. Although this means that information about f at the boundary is lost, it is reasonable to believe that the boundary effects cited above will be eliminated and the mean squared error for f_λ in $[\delta, 1-\delta]$ will satisfy (F3).
In section 5 we show that this is indeed the case when $g \equiv 1$ and $m = 2$. If the results in Lemma 3.1 can be established when $V(\lambda)$, $T_n(\lambda)$ and and S_n^2 are appropriately restricted to $[\delta, 1-\delta]$ then Theorem 1.1 and Theorem 1.2 will still hold for all $f \varepsilon C^4[0,1]$. We feel this additional analysis is beyond the scope of this paper, however, because our main objective is to give a rigorous basis for Wahba's original

conjecture in the context of her simulation study.

2. General Theorem

A general version of Theorems 1.1 and 1.2 depends on three conditions which we will refer to as (F1) - (F3).

Let G_n denote the empirical distribution for the knots, $\{t_{kn}\}_{k=1,n}$. Following Cox (1984), (subsequently abbreviated GA) we consider two cases:

A Designed knots: There is a distribution function G such that

$$\sup_{v \varepsilon [o,1]} | G_n(v) - G(v) | = o\left(\frac{1}{n}\right)$$

B Random knots: $\{t_{kn}\}$ is a random sample with distribution function G.

In either case we will assume that $g = \frac{d}{dv} G$ is strictly positive on $[0,1]$ and $g \varepsilon C^{\infty}[0,1]$.

(F1) $E \left| e_{nk} \right|^{2+\delta} < \infty$ with

Case A: $\delta > 4m-1$

Case B: $\delta > 2(8m-3)/5$

Unfortunately due to the difficulty in obtaining uniform asymptotic approximations for f_λ over all λ, we must restrict attention to values of the smoothing parameter in an interval $[\lambda_n, \infty)$ with $\lambda_n \to o$ at a particular rate. Thus, we will redefine the smoothing parameter estimate as

$$\hat{\lambda} = \underset{\lambda \varepsilon [\lambda_n, \infty)}{\operatorname{argmin}} V(\lambda) \quad .$$

(F2)

Case A: $\lambda_n \approx n^{-\frac{4m}{5}} \log(n)$

Case B: $\lambda_n \approx n^{-\frac{2m}{5}} \log(n)^m$

(F3) There is a $\gamma > 0$ such that

$$\frac{1}{n}\sum_{k=1}^{n}\left[E\, f_\lambda(t_{kn}) - f(t_{kn})\right]^2 = \gamma\lambda^2 \left[1+o(1)\right] \text{ uniformly for } \lambda\varepsilon\, [\lambda_{n,\infty}) .$$

Theorem 2.1

Under (F1)-(F3) *if* $\hat{S}_n^2$ *is given by* (1.8) *and* $\hat{T} = V(\hat{\lambda}) - \hat{S}_n^2$ *then*

2.1a) $S_n^2 - \hat{S}_n^2 = o_p\left[E\, T_n(\lambda^o)\right]$

2.1b) $\hat{T}\big/ E\, T_n(\lambda^o) \xrightarrow{P} 1$

2.1c) $$\frac{\hat{\sigma}^2 \frac{1}{n} \operatorname{tr} A(\hat{\lambda})}{E\, T_n(\lambda^o)} \xrightarrow{P} K$$

where $K = \left(\frac{2m}{2m-1}\right) \left(\frac{4m}{4m+1}\right)$ as $n \to \infty$.

3. Preliminary Results

The proof of Theorem 2.1 will follow from the five lemmas given below. To simplify notation, let r=2m and

$$\mu_k(\lambda) = \alpha \ell_k \frac{\lambda}{n}^{-\frac{1}{r}}, \text{ where } \ell_k = \left(\int_0^\infty \frac{dv}{(1+v^r)^k} \right) \quad k=1,2 \text{ and}$$

$$\alpha = \pi \Big/ \int_0^1 \left[g(v)\right]^{\frac{1}{r}} dv .$$

The first lemma is a collection of some well known asymptotic properties of a smoothing spline, while the second establishes the consistency of $\hat{\lambda}$. The remaining Lemmas can be motivated by examining the terms arising from the expansion of $\hat{S}_n^2$ and S_n^2 at lines (4.1) and (4.2) in the proof of Theorem 2.1. For all of the following lemmas it should be assumed that (F1)-(F3) are in force and $A(\lambda)$ may be interpreted to be the hat matrix for either a natural or periodic spline.

Lemma 3.1

Under F1-F3 , as $n \to \infty$

(3.1) $$\frac{1}{n} \operatorname{tr} \left[A(\lambda)^k\right] = \mu_k(\lambda) \left[1 + o(1)\right], \quad k=1,2$$

and

(3.2) $$E\left[T_n(\lambda)\right] = \left[\gamma\lambda^2 + \sigma^2 \mu_2(\lambda)\right] \left[1 + o(1)\right].$$

uniformly for $\lambda \in [\lambda_n, \infty)$

(3.3) a) $$\sup_{\lambda \in [\lambda_n, \infty)} \left| \frac{T_n(\lambda) - ET_n(\lambda)}{ET_n(\lambda)} \right| = o_p(1)$$

(3.3) b) $$\sup_{\lambda \in [\lambda_n, \infty)} \left| \frac{V(\lambda) - S_n^2 - T_n(\lambda)}{E\, T_n(\lambda)} \right| = o_p(1)$$

(3.4) $$\lambda^o = \left[\frac{\ell_2\, \sigma^2}{n\gamma 2r}\right]^{\frac{r}{1+2r}} \left[1 + o(1)\right]$$

(3.5) $$\frac{T_n(\hat{\lambda})}{E\, T_n(\lambda^o)} = 1 + o_p(1)$$

and

(3.6) $$\mu_k(\lambda^o) \Big/ E\, T_n(\lambda^o) = O(1)$$

Proof.

Examining the proof of Lemma 4.4 from GA, Cox actually shows that $\frac{1}{n} \operatorname{tr} A(\lambda)^k = \sum_{j=1}^{\infty} (1+\lambda\delta_j)^{-k} \left[(1+o(1)\right]$ for $\lambda \varepsilon \left[\lambda_n, \infty\right]$ where δ_j are the eigenvalues for the differential operator $L\varphi = \frac{(-1)^m}{g} \frac{d^{2m}}{dt^{2m}} \varphi$ with the domain for L consisting of $\varphi \varepsilon W_2^{2m}$ that either satisfy the natural or periodic boundary conditions depending on the choice of f_λ. From Naimark, 1967, pg. 78-79

$\delta_j = \alpha^{2m} j^{2m} \left[1 + o(1)\right]$ and it follows that

$$\sum_{j=1}^{\infty}(1 + \lambda\delta_j)^{-k} = \sum_{j=1}^{\infty} \left[1 + \lambda(\alpha j)^{2m}\right]^{-k} \left[1 + o(1)\right] \text{ as } \lambda \to 0.$$

Applying similar arguments to those in the proof of Lemma 2.1 in Cox (1983b), $\sum_{j=1}^{\infty} \left[1 + \lambda(\alpha j)^{2m}\right]^{-k} = \alpha\, \ell_k\, \lambda^{-\frac{1}{2m}} \left[1 + o(1)\right]$ as $\lambda \to 0$.

(3.2) follows trivally from (3.1) and (F3). Both (3.3a) and (3.3b) are proved by Speckman (1983) for normally distributed errors and are generalized in the proof of Theorem 5.1 of GA for error distributions that satisfy the moment conditions in (F2).

(3.4) is proved in Wahba (1977) and can be verified by computing the minimum of the RHS of (3.2). Note that $\lambda^o \varepsilon [\lambda_n, \infty)$ for λ_n given in (F2) and thus, these approximations are appropriate for studying the mean square convergence of $f_\lambda o$. Because $\lambda^o \varepsilon [\lambda_n, \infty)$ as $n \to \infty$ both (3.5) and (1.7) can be easily derived from the preceeding results. Finally, (3.6) may be verified by substituting the asymptotic form for λ^o into the RHS of (3.2) and into μ_k.

Lemma 3.2

If $\hat{\lambda}$ *is the minimizer of* $V(\lambda)$ *for* $\lambda \varepsilon [\lambda_n, \infty)$

then

$$\hat{\lambda} = \lambda^o \left[1 + o_p(1)\right]$$

as $n \rightarrow \infty$

Proof

Let $g(\lambda) = \gamma\lambda^2 + \frac{\sigma^2}{n}\alpha\ell_2\lambda^{-\frac{1}{r}}$

From (3.2) and (3.3a) we have $\sup_{\lambda \varepsilon [\lambda_n, \infty)} \left| \frac{T_n(\lambda) - g(\lambda)}{g(\lambda)} \right| = o_p(1)$

and thus $\frac{T_n(\hat{\lambda}) - g(\hat{\lambda})}{g(\hat{\lambda})} = o_p(1)$ or $\frac{g(\hat{\lambda})}{T_n(\hat{\lambda})} = 1 + o_p(1)$.

Combining this asymptotic equivalence with (3.2) and (3.5) it now follows that

$$\frac{g(\hat{\lambda})}{g(\lambda^o)} = \frac{g(\hat{\lambda})}{T_n(\hat{\lambda})} \quad \frac{T_n(\hat{\lambda})}{E\ T_n(\lambda^o)} \quad \frac{E\ T_n(\lambda^o)}{g(\lambda^o)} = 1 + o_p(1)$$

Set $\psi(u) = \dfrac{(au)^q + (bu)^{-(1-q)}}{a^q + b^{-(1-q)}}$.

Let $\delta = \frac{\sigma^2}{\gamma}$ and choose $\hat{\delta}$ such that $\hat{\lambda} = \left(\frac{\alpha\ell_2\ \hat{\delta}}{n2r}\right)^{\frac{r}{1+2r}}$

Then $g(\hat{\lambda}) \big/ g(\lambda^o) = \psi\left(\frac{\hat{\delta}}{\delta}\right)$ where $q = 2r \big/ 1 + 2r$,

$a = \left(\frac{\alpha\ell_2}{2r}\right)^q$ and $b = \left(\frac{\alpha\ \ell_2}{2r}\right)^{-(1-q)}$.

Thus, $\psi\left(\frac{\hat{\delta}}{\delta}\right) = 1 + o_p(1)$. Note that $\psi(u) = 1$ if and only if u=1. Because ψ (u) is continuous at 1, $\hat{\delta}/\delta \xrightarrow{p} 1$, otherwise, a contradiction will be obtained. From the definition of $\hat{\delta}$, the consistency of $\hat{\lambda}$ now follows.

Lemma 3.3

$$\frac{1}{n} e_n' A(\lambda^o)^k e_n = \sigma^2 \mu_k(\lambda^o)\left[1+o_p(1)\right], \quad k = 1,2 \quad \text{as } n \to \infty .$$

Proof

First we note that $A(\lambda)^{½}$ is well defined since $A(\lambda)$ is non-negative definite.

$$\text{Let } X = \frac{1}{n}\frac{\| Be \|^2}{\sigma^2\mu_k(\lambda)} \qquad \text{with } B = A(\lambda)^{\frac{k}{2}}$$

From (3.1) it follows $E(X) = 1 + o(1)$. For any matrix B, with $e_1,\ldots, e_n$ i.i.d. and $\nu = E(e_1^4)$,

$$\text{Var}\left(\|Be\|^2\right) = 2\sigma^4 \text{ tr } (B^2) + (\nu - 3\sigma^4)b'b \text{ , where } b = \text{Diag } (B).$$

Also $b'b \le \text{tr } (B^2)$. Therefore $\text{Var }\|Be\|^2 = O\left[\text{tr}(B^2)\right]$

$$\text{and} \quad \text{Var}(X) = O\left(\frac{\text{tr } A^k(\lambda^o)}{n^2\mu_k(\lambda^o)^2}\right) = O\left(\frac{1}{n\mu_k(\lambda^o)}\right)$$

From (3.4)

$$\frac{1}{n\mu_k(\lambda^o)} = \left(\frac{\lambda^{o\,-\frac{1}{r}}}{n^2}\right) = \left(n^{-\left(\frac{1+2r}{1+r}\right)}\right)$$

Thus $\text{Var}(X) \to 0$ and $X \xrightarrow{p} 1$

Lemma 3.4

$$(3.7) \qquad \hat{S}^2 - \frac{1}{n} \frac{\left\| \left[I-A(\lambda^o)\right] \underset{\sim}{Y} \right\|^2}{1-\mathcal{C}\mu_1(\lambda^o)} = o_p\left[E\ T_n(\lambda^o)\right]$$

Proof

Adding and subtracting

$$\frac{1}{n} \frac{\| I-A\ (\lambda^o)\ \underset{\sim}{Y} \|^2}{1 - \frac{\mathcal{C}}{n} \operatorname{tr} A(\lambda^o)} ,$$

the LHS of (3.7) can be rewritten as

$$(3.8) \qquad V(\hat{\lambda})\ \varphi\ (\hat{\lambda}) - V(\lambda^o)\ \varphi(\lambda^o)$$

$$(3.9) \qquad + \ \frac{1}{n} \left\| \left[I-A(\lambda)\right]\underset{\sim}{Y} \right\|^2 \tau$$

$$\text{with } \varphi(\lambda) = \frac{\left[1- \frac{1}{n} \operatorname{tr} A(\lambda)\right]^2}{1-\mathcal{C}\frac{1}{n} \operatorname{tr} A(\lambda)} \text{ and } \tau = \frac{1}{1 - \frac{\mathcal{C}}{n} \operatorname{tr} A(\hat{\lambda})} - \frac{1}{1 - \mathcal{C}\ \mu_1(\hat{\lambda})}$$

Expanding (3.8) using the identity:

(ab−cd) = ½ (a−c)(b+d) + ½ (a+c) (b−d) yields

$$(3.10) \qquad \left[V(\hat{\lambda}) - V(\lambda^o)\right] \left[\varphi(\hat{\lambda}) + \varphi(\lambda^o)\right] / 2$$

$$(3.11) \qquad + \left[V(\hat{\lambda}) + V(\lambda^o)\right] \left[\varphi(\hat{\lambda}) - \varphi(\lambda^o)\right] / 2$$

Now from Lemma 3.1 ,

$$\left[V(\hat{\lambda}) - V(\lambda^o)\right] = \left[T_n(\hat{\lambda}) - T_n(\lambda^o)\right] \left[1 + o_p(1)\right]$$
$$= o_p\left[E\ T_n(\lambda^o)\right] .$$

From (3.1) it is clear that $\varphi(\lambda)$ is bounded and therefore (3.10) converges to zero at the necessary rate.

To simplify notation, let $\mu_{1n}(\lambda) = \frac{1}{n} \operatorname{tr} A(\lambda)$.

Using the algebraic identity given above

$$\varphi(\hat{\lambda}) - \varphi(\lambda^{o}) = \mathfrak{c}\left[\mu_{1n}(\hat{\lambda}) - \mu_{1n}(\lambda^{o})\right]\xi$$

$$+ \left[\mu_{1n}(\lambda^{o}) - \mu_{1n}(\hat{\lambda})\right] \left[2 - \mu_{1n}(\lambda^{o}) - \mu_{1n}(\hat{\lambda})\right]\zeta$$

where $\xi = \dfrac{\left[1 - \mu_{1n}(\hat{\lambda})\right]^{2} + \left[1-\mu_{1n}(\lambda^{o})\right]^{2}}{2\left[1-\mathfrak{c}\mu_{1}(\hat{\lambda})\right] \left[1-\mathfrak{c}\mu_{1}(\lambda^{o})\right]}$

and

$$\zeta = \frac{2-\mathfrak{c}\left[\mu_{1}(\hat{\lambda}) + \mu_{1}(\lambda^{o})\right]}{2\left[1-c\mu_{1}(\hat{\lambda})\right] \left[1-c\mu_{1}(\lambda^{o})\right]}$$

Note that

$$\mu_{1}(\hat{\lambda}) - \mu_{1}(\lambda^{o}) = o_{p}\left[\mu_{1}(\lambda^{o})\right] = o_{p}\left[E\ T_{n}(\lambda^{o})\right]$$

from Lemma 3.2 and (3.6). From (3.1) by similar arguments we also have $\mu_{1n}(\hat{\lambda}) - \mu_{1n}(\lambda^{o}) = o_{p}\left[E\ T_{n}(\lambda^{o})\right]$.
Because both ξ and ζ are bounded in probability $\dot{\varphi}(\hat{\lambda}) - \varphi(\lambda^{o}) = o_{p}\left[T_{n}(\lambda^{o})\right]$ and it now follows that (3.11) also converges to zero at the correct rate.

Finally, we consider τ in (3.9)

$$\tau = \frac{\mathfrak{c}\left[\mu_{1n}(\lambda^{o}) - \mu_{1}(\lambda^{o})\right]}{\left[1-\mathfrak{c}\mu_{1n}(\lambda^{o})\right] \left[1-\mathfrak{c}\mu_{1}(\lambda^{o})\right]}$$

in view of (3.1) and (3.6), $\tau = o_{p}\left[E\ T_{n}(\lambda^{o})\right]$

Using arguments similar to those in lemma 3.3, one can show $\frac{1}{n} \left\|\left[I-A(\lambda^{o})\right]\underset{\sim}{Y}\right\|^{2} \xrightarrow{P} \sigma^{2}$. Therefore (3.9) is $= o_{p}\left[E\ T_{n}(\lambda^{o})\right]$ and the lemma now follows.

Lemma 3.5

Let $\underset{\sim}{f}' = \left[f(t_1),\ldots,f(t_n)\right]$, *then*

$$(3.12) \qquad \Gamma = \frac{1}{n} \left\| \left[I-A(\lambda^o)\right] \underset{\sim}{f} \right\|^2 - \sigma^2(2-\mathfrak{C})\mu_1(\lambda^o) + \sigma^2\mu_2(\lambda^o)$$

$$= o\left[E\ T_n(\lambda^o)\right] \quad \text{as } n \to \infty \quad .$$

Proof

Adding and subtracting $\frac{1}{n}$ tr $A(\lambda^o)^2$ from Γ we have

$$\Gamma / E\ T_n(\lambda^o) = 1 + \frac{\sigma^2(2-\mathfrak{C})\mu_1(\lambda^o)}{E\ T_n(\lambda^o)} - \frac{\sigma^2\left[\mu_2(\lambda^o) - \frac{1}{n}\text{ tr } A(\lambda^o)^2\right]}{E\ T_n(\lambda^o)}$$

$$(3.13) \qquad = 1 - \frac{\sigma^2(2-\mathfrak{C})\mu_1(\lambda^o)}{\gamma\lambda^{o2} + \sigma^2\mu_2(\lambda^o)} + o(1) \qquad \text{as } n \to \infty \ ,$$

by the asymptotic equivalences given in Lemma 3.1.

Substituting the asymptotic form for λ^o into (3.12) yields

$$(3.14) \qquad \Gamma / E\ T_n(\lambda^o) = 1 - \frac{(2 - \mathfrak{C})\ \ell_1}{\ell_2\left[\frac{1}{4m} + 1\right]} + o\ (1)$$

Referring to Selby (1979), pg. 461 $\frac{\ell_1}{\ell_2} = \frac{2m}{2m-1}$ and recall that

$$(2-\mathfrak{C}) = \frac{1}{K} = \left[\frac{2m-1}{2m}\right]\left[\frac{4m+1}{4m}\right] \quad .$$

Thus by the choice of $\mathfrak{C}$ it is straight forward to verify that the middle term on the LHS of 3.14 above equals 1.

4. Proof of Theorem 2.1

To simplify notation, let

$$\mu_k = \mu_k(\lambda^o) \ k=1,2 \quad ,$$

$A = A(\lambda^o)$, and $\underset{\sim}{f}' = \left[f(t_{1n}),\ldots, f(t_{nn})\right]$.

$$\text{(4.1)} \qquad \hat{S}_n^2 - S_n^2 = \left[\hat{S}_n^2 - \frac{1}{n}\frac{\|(I-A)\underset{\sim}{Y}\|^2}{1 - \mathcal{C}\mu_1}\right] + \left[\frac{1}{n}\frac{\|\left[I-A\right]\underset{\sim}{Y}\|^2}{1-\mathcal{C}\mu_1} - S_n^2\right]$$

The first term in (4.1) converges in probability at the necessary rate by Lemma 3.4. Noting that $\underset{\sim}{Y} = \underset{\sim}{f} + \underset{\sim}{e}$, by expanding the second term and adding and subtracting

$- \sigma^2\mathcal{C}\mu_1 + 2\sigma^2\mu_1 - \sigma^2\mu_2$ we have the expression

$$\text{(4.2)} \quad \left[R_1 + R_2 + R_3 + \Gamma\right] \Big/ \left[1 - \mathcal{C}\mu_1\right]$$

where $\quad R_1 = \mathcal{C}\mu_1\left[S_n^2 - \sigma^2\right]$

$$R_2 = -2\left[\frac{1}{n}\|A^{\frac{1}{2}}\underset{\sim}{e}\| - \mu_1\sigma^2\right] + \left[\frac{1}{n}\|A\underset{\sim}{e}\|^2 - \mu_2\sigma^2\right]$$

$$R_3 = \frac{1}{n}\underset{\sim}{f}'(I-A)^2\underset{\sim}{e}$$

and Γ is defined in the statement of Lemma 3.5.

Because $1 - \mathcal{C}\mu_1 \longrightarrow 1$ to complete the proof it is sufficient to show that each term in the numerator of (4.2) is $o_p\left[E\, T_n(\lambda^o)\right]$. Moreover, in view of the asymptotic relations between $E\, T_n(\lambda^o)$ and $\mu_k(\lambda^o)$ in Lemma 3.1, it is enough to establish a convergence rate to zero of the form $o_p(\mu_1 V \mu_2)$ for each term.

The convergence of R_1 follows from the fact that $S_n^2 \xrightarrow{P} \sigma^2$. $\qquad R_2$ is $o_p\,(\mu_1 V\mu_2)$ by Lemma 3.3

Note that $E(R_3) = 0$ and

$$\operatorname{Var}(R_3) = \frac{1}{n}\, 2\|(I-A)^2\underset{\sim}{f}\|^2 \le \frac{1}{n}\, 2\|(I-A)\,\underset{\sim}{f}\|^2 \le \frac{1}{n}\, E\, T_n(\lambda^o)$$

where the first inequality follows because the eigenvalues of

I-A lie in [0,1]. Now $\mathrm{Var}\left[R_3/ET_n(\lambda^o)\right] \leq \frac{1}{n}\left[E\ T_n(\lambda^o)\right]^{-1}$

and $\frac{1}{n}\left[E\ T_n(\lambda^o)\right]^{-1} = o(1)$ in view of the optimal convergence rate using (3.2) and (3.4). Thus, by Chebyshev's inequality, R_3 converges at the appropriate rate.

Finally, note that $\Gamma = o_p\left[E\ T_n(\lambda^o)\right]$ by Lemma 3.5 and thus having considered all the terms in (4.2), 2.1a follows.

The second of the theorem part is a consequence of (2.1a) and (1.7). For 2.1c) we refer to the discussion in the introduction and note that from (3.1), $\frac{1}{n}\ \mathrm{tr}\ A(\hat{\lambda}) = \frac{1}{n}\mu_1(\hat{\lambda})\left[1 + o(1)\right]$.

Because $\frac{1}{n}\mu_1(\hat{\lambda}) \leq \frac{1}{n}\mu_1(\lambda_n) = O\left[n^{-\frac{1}{5}}\right]$, $\frac{1}{n}\ \mathrm{tr}\ A(\hat{\lambda}) \to 0$ as $n \to \infty$ and the bracketed term in (1.8) converges to 1.

5. Bias of a natural spline in a interior interval.

In this section we will show that when a natural spline is restricted to an interior interval the average bias will satisfy condition (F3). This result complements the discussion at the end of the first section.

Theorem 5.1

Let f_λ *be a natural cubic spline* (m=2) *and for* $\delta > 0$ *let*

$$b(\lambda,\delta)^2 = \frac{1}{n_\delta} \sum_{t_{kn}\,\epsilon\,[\delta,1-\delta]} \left[f(t_k) - E\ f_\lambda(t_k)\right]^2$$

where $n_\delta = \#\left\{t_{kn}:\ t_{kn}\ \epsilon[\delta,1-\delta]\right\}$.

Assume that (F1) *and* (F2) *hold and* $G_n(u) \to u$ as $n \to \infty$.

If $f \varepsilon C^4$ [0,1] *and* $D^4 f$ *is not identically zero, then*

$$b(\lambda,\delta)^2 = \gamma\lambda^2\left(1 + o(1)\right) \quad \textit{uniformly for } \lambda\varepsilon\left[\lambda_n,\infty\right] \textit{ as } n\to\infty$$

where $\gamma = \int_{[\delta,1-\delta]} \left[D^4 f\right]^2 dt$

This theorem will be shown using the three lemmas developed in this section and for each of these results it should be assumed that the hypotheses of Theorem 5.1 are in force.

Let $\left[B_{n\lambda} f\right]\left[t\right] = f_\lambda(t) - f(t)$. It will be useful to approximate $B_{n\lambda}$ using an operator that does not depend on n. Let L be the linear differential operator: $\lambda D^4 + I$ with domain: $\mathcal{D}=\left\{\varphi\varepsilon W_2^4\ [0,1]:\ \varphi^{(k)}(0) = \varphi^{(k)}(1)=0\ k=2,3\right\}$ and let $h_\lambda(\cdot,\cdot)$ be the Green's function for L. That is, if $\psi \varepsilon L^2$ [0,1] then $\varphi(s)=\int_{[0,1]} h_\lambda\ (s,t)\ \psi(t)dt$ solves $\lambda \frac{d^4}{ds^4}\varphi + \varphi = \psi$ with $\varphi\varepsilon\mathcal{D}$. The form for h_λ can be computed (using symbol manipulation software) and is given in the Appendix. Also in the appendix is a lemma giving some upper bounds on this integral kernel. One way of interpreting the Green's function is that it is a limiting form for $A(\lambda)$ as the knots become dense in the interval. (Cox (1983a) for more details of this interpretation). If we let $(B_\lambda f)\ (s) = \int_{[0,1]} h_\lambda(s,t)\ f(t)\ dt - f(s)$ then Cox (1983b) shows that

(5.1) $$\left\|B_{n\lambda} f - B_\lambda f\right\|_{L^2[0,1]} = o\left(\left\|B_\lambda f\right\|\right) \to 0 \text{ as } n\to 0 \text{ for}$$

$\lambda\varepsilon[\lambda_n,\infty)$ under (F1) and (F2).

To simplify notation, let χ_δ be the indicator function for $[\delta,1-\delta]$ and unless otherwise labeled $\|\cdot\|$ should be taken to be the usual norm for $L^2[0,1]$.

Lemma 5.1

Let $f, \mathcal{g} \in C^1[0,1]$ *and suppose that* $f = \mathcal{g}$ *for* $s \in \left[\frac{\delta}{2}, 1-\frac{\delta}{2}\right]$ *then for any* $p>0$ *there is an* $N < \infty$

a) $\| \chi_\delta B_\lambda(\mathcal{g}-f)\|^2 = o(\lambda^p)$

b) $\| \chi_\delta B_{n\lambda}(\mathcal{g}-f)\|^2 = o(\lambda^p)$

for $\lambda \in [\lambda_n, \infty)$ *and* $n \geq N$.

Proof

$$(5.2) \qquad B_\lambda(\mathcal{g}-f)(s) = \int_{\left[0,\frac{\delta}{2}\right]} h_\lambda(s,t) \left[\mathcal{g}(t)-f(t)\right] dt + \int_{\left[1-\frac{\delta}{2},1\right]} h_\lambda(s,t) \left[\mathcal{g}(t) - f(t)\right] dt$$

For $s \in [\delta, 1-\delta]$ we have $|s-t| > \frac{\delta}{2}$ and thus from Lemma A.1

$$\left| \int_{\left[0, \frac{\delta}{2}\right]} h_\lambda(s,t) \left[\mathcal{g}(t) - f(t)\right] dt \right| \leq \rho M e^{-\frac{\delta}{2}\rho} \sup_{t\in[0,\delta]} \left|\mathcal{g}(t) - f(t)\right|$$

where $\rho = \lambda^{-\frac{1}{4}}$ and $M < \infty$. A similar bound holds for the second term in (5.2) and thus part a) of the lemma follows.

Let $h_{\lambda n}(s,t)$ be the kernel such that:

$f_\lambda(s) = \int_{[0,1]} h_{\lambda n}(s,t)\, dG_n$ and we first derive a bound similar to (A.3) for $h_{\lambda n}(s,t)$. This estimate is obtained using the techniques in the proof of Theorem 3.2 in Cox (1983a).

If $R_{n\lambda}\, g(s) = \int_{[0,1]} h_\lambda(s,\tau)\, g(\tau)\, d(G-G_n)(\tau)$ and $h_j = h_\lambda(s,t_j)$

then given the bounds for h in Lemma A.1 and (F1) and (F2) Cox (1983a) shows that

$$\sum_{\nu=0}^{\infty} R_{n\lambda}^{\nu} h_j(s) = h_{\lambda n}(s,t_j) \qquad \text{for n sufficiently large.}$$

We claim that for $\mu = 0,1$,

$$(5.3) \qquad \frac{d^{\mu}}{ds^{\mu}} R_{n\lambda}^{\nu} h_j(s) \leq \left(6B\rho^2 D_n \Delta\right)^{\nu} \rho^{1+\mu} e^{-\rho\Delta}$$

where $D_n = \sup_{u\varepsilon[0,1]} \left| G_n(u) - G(u) \right|$, $\Delta = \left|s-t_j\right|$ and $\rho = \lambda^{-\frac{1}{4}}$.

The proof is by induction. Clearly (5.3) holds when $\nu = 0$ by Lemma A.1.

$$\frac{d^{\mu}}{ds^{\mu}} R_{n\lambda}^{\nu+1} h_j(s) = \int_{[0,1]} \frac{d^{\mu}}{ds^{\mu}} h_{\lambda}(s,t) R_{n\lambda}^{\nu} h_j(\tau)\, d(G-G_n)(\tau)$$

and integrating by parts,

$$= \int_{[0,1]} \frac{d}{d\tau} \left[\frac{d^{\mu}}{ds^{\mu}} h_{\lambda}(s,\tau) R_{n\lambda}^{\nu} h_j(\tau) \right] \left(G(\tau) - G_n(\tau)\right) d\tau$$

Now applying a Hölder inequality,

$$\left| \frac{d^{\mu}}{ds^{\mu}} R_{n\lambda}^{\nu+1} h_j(s) \right| \leq D_n \int_{[0,1]} \left| \frac{d}{d\tau} \left[\frac{d^{\mu}}{ds^{\mu}} h_{\lambda}(s,t) R_{n\lambda}^{\nu} h_j(\tau)\right] \right| d\tau$$

and using (Lemma A.1) and the induction hypothesis

$$\leq \left(6\Delta\right)^{\nu} \left(B\, \rho^2 D_n\right)^{\nu+1} \rho^{1+\mu} 2\, I(\rho,s,t_j)$$

where $I(\rho,s,t) = \int_{[0,1]} e^{-\rho|s-\tau| - \rho|t-\tau|} d\tau$

Integrating $I(\rho,s,t)$ and identifying the dominating terms we have $I(\rho,s,t_j) \leq 3\,\Delta\, e^{-\rho\Delta}$. and thus (5.3) will now hold when $\nu+1$ replaces ν. Given this bound note that

$$\left| h_j(s) \right| \leq \left[\sum_{\nu=o}^{\infty} \left(B\, \rho^2 D_n \Delta\right)^{\nu} \right] \rho^2 e^{-\rho\Delta}$$ and by (F2) using the fact

that $D_n = \mathcal{O}\left(n^{-1/4}\log(n)\right)$, $B\rho^2 D_n \to 0$ as $n \to 0$ for $\lambda \in [\lambda_n, \infty)$.

Thus, for n sufficiently large, the bracketed term above is a convergent geometric series with a limit of 1 as $n \to \infty$. With this estimate for $h_j(s)$, part b) now of the lemma now follows in the same manner as part a).

Lemma 5.2

(5.4) $$\left\| \chi_\delta B_\lambda f \right\|^2 = \gamma\lambda^2\left[1 + o(1)\right] \text{ as } \lambda \to 0 \text{ and } \gamma = \left\|\chi_\delta D^4 f\right\|^2 .$$

Proof

First, we will consider $g \in C^4\,[0,1]$ such that $g = f$ for $s \in \left(\frac{\delta}{2}, 1-\frac{\delta}{2}\right)$ and $g^{(k)}$ is a periodic function on [0,1] for k=0,4. We will argue that

(5.5) $$\left\|\chi_\delta\, B\, g\right\|^2 = \gamma\lambda^2\left[1 + o\,(1)\right]$$

and given the results of lemma 5.1a, the extension to nonperiodic f will follow.

From the discussion in Cox (1983a), Section 5, h_λ has the decomposition:

$h_\lambda(s,t) = K_\lambda(s-t) + r_\lambda(s,t)$ where K_λ is a periodic kernel with two continuous derivatives and the form for r_λ is given in the appendix. Note that

$$B_\lambda\, g(s) = g(s) - \int_{[0,1]} K_\lambda(s-t)\, g(t)dt + \int_{[0,1]} r_\lambda(s,t)\, g(t)dt$$

From the definition of $K_\lambda(s-t)$ as a Green's function (see appendix) ,

$$g(s) - \int_{[0,1]} K_\lambda(s,t)\, g(t)dt = \lambda D^4 \int K_\lambda(s,t)\, g(t)\, dt$$

and extending g periodically outside $[0,1]$

$$= \lambda D^4 \int_{[0,1]} K_\lambda(u)\, g\,(s-u)\, du$$

$$= \lambda \int_{[0,1]} K_\lambda(u) \frac{d^4}{ds^4}\, g(s-u)\, du$$

$$\rightarrow \lambda D^4 g(s) \text{ as } \lambda \rightarrow 0.$$

By the construction of r_λ reported in the appendix it is clear that $\frac{d^4}{ds^4} r_\lambda(s,t) = -\lambda r_\lambda(s,t)$. Moreover, from Lemma A.1 we have the rough bound: $\left|r_\lambda(s,t)\right| < M \rho^4 e^{-\frac{\delta}{2}\rho}$ for $s \in [\delta, 1-\delta]$

where $0 < M < \infty$ and $\rho = \lambda^{-1/4}$. Therefore,

$$\left|\frac{d^4}{ds^4} \int r_\lambda(s,t)\, g(t)dt\right| \leq M \sup\left|g\right| \rho^4 e^{-\frac{\delta}{2}\rho} = o(\lambda^p)$$

for any p, uniformly for $s \in [\delta, 1-\delta]$.

Taking $p = 1$, we have $B_\lambda g = \lambda D^4 g + o(\lambda)$ and (5.5) now follows.

Finally, we relate these results to nonperiodic f. By Lemma 5.1 with $p > 1$

$$\left| \, \|\chi_\delta B_\lambda f\| - \|\chi_\delta B_\lambda \mathfrak{g}\| \, \right| \le \|\chi_\delta B_\lambda (f-\mathfrak{g})\|$$

$$= o(\lambda^p) = o\left(\|\chi_\delta B_\lambda \mathfrak{g}\|\right)$$

By construction, f and $\mathfrak{g}$ agree on $[\delta, 1-\delta]$ and

$$\int_{[\delta,1-\delta]} (D^4 f)^2 \, dt = \int_{[\delta,1-\delta]} (D^4 \mathfrak{g})^2 \, dt \quad .$$

The lemma now follows.

Lemma 5.3

If $\mathfrak{g} \in \mathcal{G}$ then

$$\text{(5.6a)} \qquad \|\chi_\delta B_{n\lambda} \mathfrak{g}\| = \|\chi_\delta B_\lambda \mathfrak{g}\| \left(1 + o(1)\right)$$

and

$$\text{(5.6b)} \qquad \frac{1}{n_\delta} \sum_{t_k \in [\delta,1-\delta]} B_{n\lambda} \mathfrak{g} (t_k)^2 = \|\chi_\delta B_\lambda \mathfrak{g}\|^2 \left(1 + o(1)\right)$$

uniformly for $\lambda \in [\lambda_n, \infty)$ as $n \to \infty$.

Proof

$$\left| \, \|\chi_\delta B_{n\lambda} \mathfrak{g}\| - \|\chi_\delta B_\lambda \mathfrak{g}\| \, \right|$$

$$\le \|\chi_\delta (B_{n\lambda} - B_\lambda) \mathfrak{g}\| \le \|(B_{n\lambda} - B_{n\lambda}) \mathfrak{g}\|$$

$$= o\left(\|B_\lambda \mathfrak{g}\|\right) = o(\lambda)$$

by (5.1) and the assumption that $\mathfrak{g} \in \mathcal{G}$.

Given the convergence rate for $\|\chi_\delta B_\lambda \mathfrak{g}\|$ in Lemma (5.2) the first part of the lemma now follows.

Let G_n^δ denote the empirical distribution for $t_{kn} \varepsilon\ [\delta, 1-\delta]$ and G^δ its limit as $n \to \infty$. With this notation, the discrete norm in (5.6b) can be reexpressed as: $\int_{[0,1]} \left(B_{n\lambda}\beta\right)^2 d\, G_n^\delta$ and we have

$$\left| \int_{[0,1]} \left(B_{n\lambda}\beta\right)^2 d\, G_n^\delta - \left\| \chi_\delta\, B_{n\lambda}\beta \right\|^2 \right| = \left| \int_{[0,1]} \left(B_{n\lambda}\beta\right)^2 d\left[G_n^\delta - G\right] \right| .$$

Intergrating by parts and applying the Cauchy-Schwarz inequality:

$$= \left| \int_{[0,1]} 2 \left(\frac{d}{ds} \left(B_{n\lambda}\beta\right)^2\right) \left(G_n^\delta - G\right) ds \right|$$

$$\leq 2 \sup_{s \varepsilon [0,1]} \left| G_n^\delta (s) - G^\delta (s) \right| \left\| \frac{d}{ds} B_{n\lambda}\beta \right\| \left\| B_{n\lambda}\beta \right\|$$

Let $D_n = \sup \left| G_n^\delta - G^\delta \right|$ and taking the usual norm for $W_2^1\,[0,1]$ there is an $M < \infty$ such that the above expression

$$\leq M\, D_n \left\| B_{n\lambda}\beta \right\|^2_{W_2^1\,[0,1]} .$$

Now following the arguments given in the proof of Lemma 4.1, pg. 38 of GA.

$$D_n \left\| B_{n\lambda}\beta \right\|^2_{W_2^1 [0,1]} = o\left(\left\| B_{n\lambda}\beta \right\|^2\right) \text{ uniformly } \lambda\ \varepsilon\ [\lambda_n, \infty).$$

Moreover, by the analysis given above and Lemma 5.2 we know that

$\left\| \chi_\delta\, B_{n\lambda}\beta \right\|^2 \sim \lambda^2$. Thus, we can conclude that

(5.7) $$\int \left(B_{n\lambda}\beta\right)^2 dG_n^\delta = \left\| X_\delta\, B_{n\lambda}\beta \right\|^2 \left(1 + o(1)\right) \text{ as } n \to \infty$$

uniformly for $\lambda\ \varepsilon\ [\lambda_n, \infty]$ and given (5.6a) the second part of the lemma now follows.

Proof of Theorem 5.1

The proof consists of extending the convergence results for $\mathscr{g} \varepsilon \mathscr{S}$ in Lemma 5.3 to functions that do not satisfy the natural boundary conditions. For $f \varepsilon C^4[0,1]$ choose $\mathscr{g} \varepsilon \mathscr{S}$ so that $\mathscr{g}$ agrees with f on $\left[\frac{\delta}{2}, 1 - \frac{\delta}{2}\right]$.

By a slight modification of the proof of part b in Lemma 5.1, it is possible to show that $\int_{[0,1]} \left[B_{n\lambda}\left(f-\mathscr{g}\right)\right]^2 d\, G_n^\delta = o(\lambda^p)$, for any $p > 0$. while combining Lemmas 5.2 and 5.3 we have $\int \left(B_{n\lambda}\mathscr{g}\right)^2 dG_n^\delta \quad \gamma\lambda^2$. Thus,

$$b(\lambda,\delta) \equiv \int_{[0,1]} \left(B_{\lambda n} f\right)^2 dG_n^\delta = \int_{[0,1]} \left(B_{\lambda n}\mathscr{g}\right)^2 d\, G_n^\delta \left(1 + o(1)\right) = \gamma\lambda^2\left(1 + o(1)\right)$$

APPENDIX

This appendix gives an explicit formula for the Greens function, $h_\lambda(s,t)$, for the differential operator L defined in Section 5. Lemma A.1 following these formulae basically states that $|h_\lambda(s,t)|$ decreases exponentially as s diverges from t. A MACSYMA batch program is included at the end of this appendix for those that wish to reproduce these results.

Let $\ell = \dfrac{\lambda^{1/4}}{\sqrt{2}}$.

For $s \leq t$ $h_\lambda(s,t)$ is given below. By the symmetry in the operator L if $s>t$, then $h_\lambda(s,t) = h_\lambda(1-s,1-t)$

(A.1) $$h_\lambda(s,t) = a2(t)\, e^{\ell s} \sin(\ell s) + a3(t)\, e^{-\ell s} \sin(\ell s) + a1(t)\, e^{\ell s} \cos(\ell s) + a4(t)\, e^{-\ell s} \cos(\ell s)$$

To shorten the expressions make the substitutions:

$[ss = \sin(\ell s),\ cs = \cos(\ell s),\ es = e^{\ell s}]$

$[ct = \cos(\ell t),\ st = \sin(\ell t),\ et = e^{\ell t}]$

$[c = \cos(\ell),\ s = \sin(\ell),\ e = e^{\ell}]$

then

$$[a1 = \frac{e20}{e21},\ a4 = \frac{e23}{e21}]$$

and

(A.2) $$a2 = a3 = (a1-a4)/2$$

where

$$e20 = 2\ ce^2\ et^2\ \ell s\ st + 2\ ce^2\ \ell s\ st + 2\ c^2\ e^2\ et^2\ \ell\ st - e^2\ et^2\ \ell\ st$$

$$- et^2\ \ell\ st + e^4\ \ell\ st + 6\ c^2\ e^2\ \ell\ st - 7\ e^2\ \ell\ st - 2\ c\ ct\ e^2\ et^2\ \ell s$$

$$- 6\ c\ ct\ e^2\ \ell s + 2\ c^2\ ct\ e^2\ et^2\ \ell + ct\ e^2\ et^2\ \ell - 3\ ct\ et^2\ \ell$$

$$+ ct\ e^4\ \ell + 2\ c^2\ ct\ e^2\ \ell - 3\ ct\ e^2\ \ell$$

$$e23 = - 2\ ce^2\ et^2\ \ell s\ st - 2\ ce^2 \ell s\ st + 6\ c^2\ e^2\ et^2\ \ell\ st - 7\ e^2\ et^2\ \ell\ st$$

$$+ et^2\ \ell\ st - e^4\ \ell\ st + 2\ c^2\ e^2\ \ell\ st - e^2\ \ell\ st - 6\ c\ ct\ e^2\ et^2\ \ell s$$

$$- 2\ c\ ct\ e^2\ \ell s - 2\ c^2\ ct\ e^2\ et^2\ \ell + 3\ ct\ e^2\ et^2\ \ell - ct\ et^2\ \ell$$

$$+ 3\ ct\ e^4\ \ell - 2\ c^2\ ct\ e^2\ \ell - ct\ e^2\ \ell \quad ,$$

and

$$e21 = 2\ e^4\ et + 8\ c^2\ e^2\ et - 12\ e^2\ et + 2\ et$$

As mentioned in Section 5,

$$h_\lambda(s,t) = K_\lambda(s,t) + r_\lambda(s,t)$$

where K is a Greens function for the operator $\lambda D^4 + I$ with periodic boundary conditions and r_λ must have the form:

$$r_\lambda(s,t) = K_\lambda^{(2)}(t)\ \varphi_1(s) + K_\lambda^{(2)}(1-t)\ \varphi_2(s)$$

$$+ K_\lambda^{(3)}(t)\ \varphi_3(s) + K_\lambda^{(3)}(1-t)\ \varphi_4(s)$$

and φ_ν solves $(\lambda D^4 + I)\ \varphi_\nu = 0$ with boundary conditions:

$$\varphi_\nu^{(2)}(0) = \delta_{\nu 1}\ ,\ \ \varphi_\nu^{(2)}(1) = \delta_{\nu 2}\ ,\ \varphi_\nu^{(3)}(0) = \delta_{\nu 3}\ ,\ \varphi_\nu^{(3)}\ (1) = \delta_{\nu 4}\ .$$

with $\delta_{\nu\mu}$ being Kronecker's delta function.

The forms for K_λ (s,t) , $\varphi_1(s)$ and $\varphi_2(s)$ are given below. φ_3 and φ_4 can be found by the relations: $\varphi_3(s) = \varphi_1(1-s)$ and $\varphi_4(s) = \varphi_2(1-s)$.

We note that given the periodicity of $K_\lambda(s,t)$ with respect to (t-s), r_λ simplifies

to $r_\lambda(s,t) = \frac{\partial^2}{\partial s^2} K_\lambda(0,t) \left[\varphi_1(s) + \varphi_1(1-s)\right] + \frac{\partial^3}{\partial s^3} K_\lambda(0,t) \left[\varphi_2(s) + \varphi_2(1-s)\right]$.

K_λ will have the same form as h_λ in (A.1) although the coefficients $a_1(t) - a_4(t)$ will be different.

$$[a1 = \frac{e41}{e42},\ a2 = \frac{e43}{e42},\ a3 = \frac{e44}{e45},\ a4 = \frac{e46}{e45}]$$

$e41 = e\ell\ \ell\ s\ell\ st - e\ell^2\ \ell\ st + c\ell\ e\ell\ \ell\ st - ct\ e\ell\ \ell\ s\ell - ct\ e\ell^2\ \ell + c\ell\ ct\ e\ell\ \ell$

$e42 = 8\ e\ell^2 et - 16\ c\ell\ e\ell\ et + 8\ et$

$e43 = -\ e\ell\ \ell\ s\ell\ st - e\ell^2 \ell\ st + c\ell\ e\ell\ \ell\ st - ct\ e\ell\ \ell\ s\ell + ct\ e\ell^2 \ell - c\ell\ ct\ e\ell\ \ell$

$e44 = -\ e\ell\ et\ \ell\ s\ell\ st - c\ell\ e\ell\ et\ \ell\ st + et\ \ell\ st + ct\ e\ell\ et\ \ell\ s\ell - c\ell\ ct\ e\ell\ et\ \ell + ct\ et\ \ell$

$e45 = 8\ e\ell^2 - 16\ c\ell\ e\ell + 8$

$e46 = -\ e\ell\ et\ \ell\ s\ell\ st + c\ell\ e\ell\ et\ \ell\ st - et\ \ell\ st - ct\ e\ell\ et\ \ell\ s\ell - c\ell\ ct\ e\ell\ et\ \ell + ct\ et\ \ell$

Finally, we give the forms for φ_1 and φ_2 .

Let $\psi(s)$ = a1*es*cs + a2*es*cs + a3*ss/es + a4*cs/es then $\varphi_1 = \psi$ where

$$[a1 = \frac{e14}{e15},\ a2 = \frac{e16}{e15},\ a3 = \frac{e17}{e15},\ a4 = \frac{e18}{e15}]$$

and

$$e14 = -2\ ce^2\ s - 2c^2e^2 + e^2 + 1$$

$$e16 = -2\ ce^2\ s + 2c^2e^2 - 3e^2 + 1$$

$$e17 = -2\ ce^2\ s - e^4 - 2c^2\ e^2 + 3e^2$$

$$e18 = 2\ ce^2\ s + e^4 - 2c^2\ e^2 + e^2$$

$$e15 = 2\ e^4\ \ell^2 + 8c^2\ e^2\ \ell^2 - 12\ e^2\ \ell^2 + 2\ell^2$$

$\varphi_2 = \psi$ where

$$a1 = \frac{e14}{e15},\ a2 = \frac{e16}{e17},\ a3 = \frac{e16}{e17},\ a4 = \frac{e18}{e15}]$$

and

$$e14 = -2\ ce^2\ s + e^2 - 1$$

$$e16 = c^2e^2 - e^2$$

$$e18 = -2\ ce^2\ s + e^4 - e^2$$

$$e15 = 2\ e^4\ \ell^3 + 8c^2\ e^2\ \ell^3 - 12\ e^2\ \ell^3 + 2\ell^3$$

$$e17 = e^4\ \ell^3 + 4c^2\ e^2\ \ell^3 - 6\ e^2\ \ell^3 + \ell^3$$

Lemma A.1

For p = 0,1 and $\rho = \lambda^{-1/4}$ there is a $B<\infty$ such that

(A.3) $$\frac{d^p}{ds^p} h_\lambda (s,t) \leq B\rho^{(1+p)} e^{-\rho |s-t|}$$

(A.4) $$r_\lambda(s,t) \leq B\rho\left[e^{-\rho s} + e^{-\rho(1-s)}\right]$$

for all $s,t \in [0,1]$ and $\rho > 0$.

Proof

First assume that $s \leq t$ and p=0. Note that $e21 \geq 2\, e^{4\rho + t}$ and by examining the separate terms in e20 we have

$$\left|e20\right| \leq \rho\, e^{4\rho}\left[1 + \varepsilon(\rho)\ \right]$$

and for, $t \in [0,1]$, $\varepsilon(\rho) < B_o e^{4\rho}$ for some $B_o < \infty$. Thus,

$$\left|al(t)\right| \leq B_o\rho\ e^{-\rho t}$$

and $$\left|al(t)\ \cos(\rho t)\ e^{\rho s}\ \right| \leq B_o\rho\ \ e^{\rho(s-t)}$$

repeating these steps for a4(t) yields a similar bound for $\left|a4(t)\ \sin\ (\rho s)\ \exp\ (-\rho s)\right|$. Using (A.2) similar analyses can be applied to $a_2(t)$ and $a_3(t)$ and the same bounds may be obtained for the two components of h_λ involving these coefficients. (A.3) now follows for p=0 and s<t. When s>t the bound is obtained by substituting 1-s and 1-t for s and t. This has the effect of simply interchanging s and t in the exponential. For ρ=1 (A.3) follows in a straight forward manner using the the bounds on al(t) - a4(t) described above.

For (A.4) using similar arguments to those given above it is straight forward but tedious to show that φ_1 is bounded by $\frac{B}{\rho^2} e^{-\rho s}$ for some $B<\infty$.

By examining the form of K, it is also possible to establish that

$$\left|\frac{\partial^j}{\partial s^j} K(0,t)\right| \leq B' \rho^{j+1} e^{-\rho t} \quad \text{for some } B' < \infty \quad \text{and thus,}$$

$$\left|\frac{\partial^2}{\partial s^2} K(0,t)\, \varphi_1(s)\right| \leq BB' \rho^3 e^{-\rho(t+s)}$$

$$\leq BB' \rho^3 e^{-\rho s}$$

Applying this argument to the other terms in $r_\lambda(s,t)$ yields similar bounds and (A.4) now follows.

```
" MACSYMA batch program to compute the greens function for the operator
"$
"                     4                                                  "$
"  Lg =    (lambda* D   + I ) with boundary conditions:                  "$
"                              2      2                                  "$
"                             D g(0)= D g(1) = 0                         "$
"                        and                                             "$
"                              3      3                                  "$
"                             D g(0)= D g(1) = 0                         "$
"                                                                        "$
"   In the expressions below l= 1/ ( sqrt(2) * lambda**(1/4) )           "$
"                                                                        "$
"       The form for the Greens function is :                            "$
"                                                                        "$
"         h = a1*cs*es + a2*ss*es +a3*ss/es + a4*cs/es    for s<=t       "$
"                                                                        "$
"         h = b1*cs*es + b2*ss*es +b3*ss/es + b4*cs/es    for s>t        "$
"                                                                        "$
"         where the a's and b's are functions of t and l that must be    "$
"         computed and the other variables are defined in subs1 below.  "$
"         The a and b coefficents will be found by solving a set of      "$
"         linear equations prescribed by the continuity of h and its     "$
"         derivatives at s=t and the boundary conditions.                "$
"                                                                        "$
"****** define general form for Greens function                          "$

h : a1*cs*es + a2*ss*es +a3*ss/es + a4*cs/es $

subs1 : [ss=sin(l*s),cs=cos(l*s),es=exp(l*s)]$
subs2: [cos(l*t)=ct,sin(l*t)=st,exp(l*t)=et]$
subs3: [cos(l)=c,sin(l)=s,exp(l)=e]$

" ******* part when s<=t                                                  "$
ha : ev(h,subs1)$

" ******* part when s>t                                                   "$
hb : ev(ha,a1=b1,a2=b2,a3=b3,a4=b4)$


" ****** set up system of 4 equations due to the continuity              "$
"          conditions at s=t                                             "$

differ: ev(hb-ha,s=t)$
cont-cd : [differ=0 ,diff(differ,t)=0,
  diff(differ,t,2)=0,diff(differ,t,3)=4*(l**4)]$

" ****** solve for b in terms of a and simplify                          "$

bcoef:   linsolve(cont-cd,[b1,b2,b3,b4]) $
bcoef : ev(bcoef,subs2)$
bcoef : ratsimp( expand( ev(bcoef,st**2= 1-ct**2) ) ) $

" ****** system of 4 equations due to boundary condtions at 0 and 1      "$
```

```
temp :  [ diff(ha,s,2)=0 , diff(ha,s,3)=0]$
temp2:  [ diff(hb,s,2)=0 , diff(hb,s,3)=0]$
nbc: append( ev(temp,s=0), ev(temp2,s=1) )$
nbc: ev(nbc, subs3)$

" ****** express equations just in terms of a to give 4 equations in    "$
" ****** 4 unknowns                                                       "$

nbca : ev(nbc,bcoef)$

" ****** solve for a                                                    "$
acoef: combine( expand( linsolve(nbca,[a1,a2,a3,a4]) ) ) $
acoef : combine( expand( ev( acoef, s**2= 1-c**2) ) ) $


" ****** display solution                                               "$
pieces :pickapart( acoef,3);

" ****** Note: hb can be found by the relation: ha(s,t)= hb( 1-s,1-t)   "
```

REFERENCES

Cox, D. D. (1983a). Asymptotics for M-type smoothing splines. Ann Statist 11, 550-551.

Cox, D. D. (1983b). Approximation of method of regularization estimators. Technical Report 723, University of Wisconsin/Madison, Department of Statistics.

Cox, D. D. (1984). Gaussian approximation of smoothing splines. Technical Report 743, University of Wisconsin/Madison, Department of Statistics.

Naimark, M. A. (1967). Linear differential operators Part I. Frederick Ungar Publishing Co., Inc.; New York.

Nychka, D. W. (1986). Manuscript in preparation.

Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. Ann Statist 11, 141-156.

Rice, J. (1984). Bandwidth choice for nonparametric regression. Ann Statist 12 1215-30.

Selby, S. (ed.) (1979). CRC Standard Mathematical Tables. Central Rubber Company, Cleveland, Ohio.

Silverman, B. W. (1984). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. J. R. Statist Soc. B. 47 1-52.

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

Wahba, G. (1977) Practical approximate solutions to linear operator equations when the data are noisy. SIAM J. NUMER ANAL. 14, 661-667.

Wahba, G. (1978) Improper, priors, spline smoothing and the problem of guarding against model errors in regression. J.R. Statist. Soc. B 49 364-372 .

Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. J. R. Statist Soc. b. 45 133-150.

Dr. Douglas Nychka
Department of Statistics
North Carolina State University
Campus Box 8203
Raleigh, North Carolina 27695-8203