An Optimality Property of Smoothing Splines

by

R. L. Eubank

Technical Report No. SMU-DS-TR-188
Department of Statistics ONR Contract

December 1984

Research sponsored by the Office of Naval Research
Contract N00014-82-K-0207
Project NR 042-479

Department of Statistics
Southern Methodist University
Dallas, Texas 75275

An Optimality Property of Smoothing Splines

by

R. L. Eubank

Southern Methodist University

Abstract. It is shown that, under certain restrictions on the regression function, there always exists a smoothing spline which has smaller risk than the corresponding polynomial regression estimate or natural spline of interpolation. The method of proof is seen to imply a similar result for ridge regression, regarding estimation of the regression function, which parallels a property of the ridge regression coefficient estimates established by Hoerl and Kennard (1970).

Key words and phrases. Interpolation, minimum risk, nonparametric regression, polynomial regression, ridge regression.

AMS 1980 subject classification. Primary 62G05, 62J07, Secondary 65D07.

1. Introduction and Summary

Suppose that n observations $(t_1,y_1),\ldots,(t_n,y_n)$ are taken on a response variable y and independent variable t. The observations are assumed to follow the model

$$y_i = \mu(t_i) + \varepsilon_i \ , \quad i = 1,\ldots,n, \tag{1}$$

where μ is an unknown regression function, the ε_i are zero mean uncorrelated errors and the t_j satisfy $a \le t_1 < t_2 < \ldots < t_n \le b$ for finite constants a and b. In this paper we establish an optimality property of smoothing splines as estimators of μ.

When the true regression function is unknown in (1) a popular approach among statisticians has been to fit the data with a polynomial to provide an estimate of μ. Let m be fixed and define T as the $n\times m$ matrix with ijth entry t_i^j, $i=1,\ldots,n$, $j=0,\ldots,m-1$. Then,the mth order polynomial regression estimator of μ is

$$\mu_\infty(t) = (1, t,\ldots,t^{m-1})(T'T)^{-1}T'\underline{y} \quad , \tag{2}$$

where $\underline{y} = (y_1,\ldots,y_n)'$ (the meaning of the ∞ subscript will become clear momentarily). These estimators have received considerable attention in the literature and techniques for fitting polynomials to data are standard fare in first year graduate methods texts (see, e.g., Ostle and Mensing 1975). The usual motivation for their use is obtained by first assuming that μ admits m derivatives, then using Taylor's formula to write μ as a polynomial plus remainder, and finally lumping the remainders in with the random errors. However, if μ can be assumed to have m derivatives, another natural estimator of μ can be obtained by minimizing

$$\sum_{i=1}^{n}(y_i-f(t_i))^2 + \lambda\int_a^b(f^{(m)}(t))^2dt, \quad \lambda > 0 , \tag{3}$$

over all functions, f, having m-1 absolutely continuous derivatives and a square integrable m-th derivative. Provided $n \leq m$, (3) has a unique minimizer, μ_λ, that is known to be a polynomial spline of order 2m with knots at $t_1,\ldots,t_n$. It is, in fact, a generalization of polynomial regression since $\lim_{\lambda\to\infty}\mu_\lambda = \mu_\infty$ is the polynomial regression estimate in (2). General discussions of smoothing splines and their properties can be found in Wegman and Wright (1983) and Eubank (1984).

Another special case of smoothing splines is widely used in the mathematics community. This is the function which minimizes $\int_a^b f^{(m)}(t)^2dt$ subject to $f(t_i) = y_i$, $i=1,\ldots,n$. The solution is the unique natural spline that interpolates the data and coincides with μ_0. This particular function is typically used when the $\mu(t_i)$ are observed with little or no error since it tends to be quite wiggly with noisy data.

The objective of this note is to show that, under model (1), there exists values in $(0,\infty)$ for which μ_λ is superior to either the polynomial regression estimate or the interpolating spline. Define the estimation risk by

$$R(\lambda) = n^{-1}\sum_{i=1}^{n}E(\mu(t_i) - \mu_\lambda(t_i))^2$$

and note that $R(\infty) = \lim_{\lambda \to \infty} R(\lambda)$ and $R(0)$ are the risk from polynomial regression and spline interpolation, respectively. In the next section we establish the following.

Theorem 1. *Assume that μ satisfies condition (5) below. Then there is a value $\lambda_\infty \in (0,\infty)$ such that $R(\lambda) < R(\infty)$ for all $\lambda > \lambda_\infty$. In addition, there exists $\lambda_0 \in (0,\infty)$, with $R(\lambda) < R(0)$ for all $\lambda < \lambda_0$.*

Theorem 1 has the consequence that there are always smoothing splines which improve upon polynomial regression or spline interpolation. Unfortunately, it does not indicate how these estimators can be found. Note, however, that as a result of the theorem any value, $\tilde{\lambda}$, which minimizes $R(\lambda)$ will satisfy both $R(\tilde{\lambda}) < R(\infty)$ and $R(\tilde{\lambda}) < R(0)$ and will therefore provide a single estimator with smaller risk than either μ_0 or μ_∞. Thus, in practice, λ should be selected to minimize an estimate of the risk to obtain a good value for λ. One procedure for accomplishing this is provided by generalized cross validation (GCV). Optimality properties of GCV and the relationship between GCV and estimation of the risk are discussed, for example, in Craven and Wahba (1979), Speckman (1984), Cox (1984), and Li (1983).

2. Proof of Theorem.

The proof of Theorem 1 is elementary but requires the introduction of a specific form for $\underline{\mu}_\lambda = (\mu_\lambda(t_1),\ldots,\mu_\lambda(t_n))'$. Smoothing splines are linear estimators and hence there is an $n \times n$ matrix $H(\lambda)$ such that $\underline{\mu}_\lambda = H(\lambda)\underline{y}$. It follows from Demmler and Reinsch (1975) that $H(\lambda)$ has the form

$$H(\lambda) = VD(\lambda)V', \tag{4}$$

where

$$V = [\underline{v}_1, \ldots, \underline{v}_n]$$

is a $n \times n$ unitary matrix and $D(\lambda)$ is a diagonal matrix with diagonal entries $(1+\lambda d_j)^{-1}$, $j=1,\ldots,n$, for constants $d_1,\ldots,d_n$ satisfying $0=d_1=\ldots=d_m<d_{m+1}\leq\ldots\leq d_n$. The first m columns of V provide a basis for the column space of T.

Using (4), $R(\lambda)$ is seen to have the explicit form

$$R(\lambda) = n^{-1}(\underline{\mu}-E\underline{\mu}_\lambda)'(\underline{\mu}-E\underline{\mu}_\lambda) + n^{-1}\sigma^2 E(\underline{\mu}_\lambda-E\underline{\mu}_\lambda)'(\underline{\mu}_\lambda-E\underline{\mu}_\lambda)$$

$$= n^{-1}\sum_{j=m+1}^{n} c_j^2(1 + 1/\lambda d_j)^{-2} + n^{-1}\sigma^2\sum_{j=1}^{n}(1+\lambda d_j)^{-2} ,$$

where $\underline{c} = (c_1,\ldots,c_n)' = V'\underline{\mu}$. The first statement of the theorem will be established if $R(\lambda)$ can be shown to be increasing for λ sufficiently large. Since $R(\lambda)$ is continuously differentiable it therefore suffices to show that there exists $\lambda_\infty \varepsilon (0,\infty)$ with $dR(\lambda)/d\lambda > 0$ for all $\lambda>\lambda_\infty$.

Differentiation of $R(\lambda)$ gives

$$dR(\lambda)/d\lambda = n^{-1}\lambda\sum_{j=m+1}^{n} c_j^2 d_j^2(1+\lambda d_j)^{-3} - n^{-1}\sigma^2\sum_{j=m+1}^{n} d_j(1+\lambda d_j)^{-3} .$$

Assuming that

$$c_j = \underline{v}_j'\underline{\mu} \neq 0, \quad j=m+1,\ldots,n \quad , \tag{5}$$

the first term in this expression is always positive whereas the second is always negative on $(0,\infty)$. Thus the choice $\lambda_\infty = \sigma^2/\min_{m+1\leq j\leq n}(c_j^2 d_j)$ will suffice. Condition (5) is unlikely to be violated by most functions of interest. It does exclude regression functions which are polynomials of order m. This however is not surprising since, in this case, $\underline{\mu}_\infty$ is the minimum variance unbiased estimator of $\underline{\mu}$ and, hence, $\lambda = \infty$ minimizes $R(\lambda)$.

The remainder of the theorem follows in a similar fashion. $R(\lambda)$ is found to be decreasing for all $\lambda < \sigma^2/\max_{m+1\leq j\leq n}(c_j^2 d_j)$. Notice that a restriction such as (5) is not necessary in this case.

Observe that the bias squared component of $R(\lambda)$, $n^{-1}\underline{\mu}'(I-H(\lambda))^2\underline{\mu}$, is an increasing function of λ, vanishing when $\lambda=0$, whereas the variance term decreases to $\sigma^2 m/n$ at $\lambda = \infty$. Thus the interpolating spline and polynomial regression estimator minimize bias and variance, respectively. Theorem 1 can be paraphrased as stating that it is best to balance these two components rather than trying to minimize either separately.

3. Application to ridge regression.

Suppose now that $\underline{\mu}$ admits the parametric form $\underline{\mu} = X\underline{\beta}$ for some known $n\times m$ matrix X of rank $m\leq n$ and unknown parameter vector, $\underline{\beta}$. The least-squares estimators of $\underline{\beta}$ and $\underline{\mu}$ are

$$\tilde{\underline{\beta}}_0 = (X'X)^{-1}X'\underline{y}$$

and

$$\tilde{\mu}_0 = X\tilde{\beta}_0.$$

An alternative to least-squares estimators of β and μ is provided by the ridge regression estimates

$$\tilde{\beta}_\lambda = (X'X+\lambda I)^{-1}X'y, \quad \lambda>0,$$

and

$$\tilde{\mu}_\lambda = X\tilde{\beta}_\lambda.$$

These estimators were introduced by Hoerl and Kennard (1970) as a cure for multicollinearity ills of least-squares estimators. Hoerl and Kennard showed that there exist ridge estimators of β with smaller risk than $\tilde{\beta}_0$. By following the proof of Theorem 1 and using the singular value decomposition of X it is easy to establish an analog of their result applicable to the estimation of μ.

Theorem 2. *There is a value* $\lambda_0 \in (0,\infty)$ *such that if* $\lambda<\lambda_0$ then $E(\tilde{\mu}_\lambda-\mu)'(\tilde{\mu}_\lambda-\mu) < E(\tilde{\mu}_0-\mu)'(\tilde{\mu}_0-\mu)$.

For references on the selection of λ to minimize risk in ridge regression see Golub, Heath and Wahba (1979).

REFERENCES

Cox, D. D. (1984). Gaussian approximation of smoothing splines. Tech. Rep. No. 743, Dept. of Statistics., Univ.of Wisconsin-Madison.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline function. Numer. Math. 31, 377-403.

Demmler, A. and Reinsch, C. (1975). Oscillation matrices with spline smoothing. Numer. Math. 24, 375-382.

Eubank, R. L. (1984). Approximate regression models and splines. Commun. Statist.-Theor. Meth. A13(4), 433-484.

Golub, G., Heath, M. and Wahba, G. (1979). Generalized cross-validation as a method of choosing a good ridge parameter. Technometrics 21, 215-223.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12, 55-67.

Li, K. C. (1983). From Stein's unbiased risk estimates to the method of generalized cross-validation. Tech. Rep. No. 83-34, Dept. of Statist., Purdue Univ.

Ostle, B. and Mensing, R. W. (1975). Statistics in Research, 3rd edition. Iowa State Press; Ames.

Speckman, P. (1983). Efficient nonparametric regression with cross-validated smoothing splines. Ann. Statist., to appear.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: SMU-DS-TR-188	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): An Optimality Property of Smoothing Splines	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): R. L. Eubank	8. CONTRACT OR GRANT NUMBER(s): N00014-82-K-0207
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Southern Methodist University, Department of Statistics, Dallas, Texas 75275	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 042-479
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Arlington, Va 222]7	12. REPORT DATE: December 1984
	13. NUMBER OF PAGES: 8
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any putpose of The United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Interpolation, minimum risk, nonparametric regression polynomial regression, ridge regression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is shown that, under certain restrictions on the regression function, there always exists a smoothing spline which has smaller risk than the corresponding polynomial regression estimate or natural spline of interpolation. The method of proof is seen to imply a similar result for ridge regression, regarding estimation of the regression function, which parallels a property of the ridge regression coefficient estimates established by Hoerl and Kennard (1970).

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)