The Statistician (1993) 42, pp. 317–319

Smoothing splines—a tutorial

BRIAN S. YANDELL

Departments of Horticulture and Statistics, University of Wisconsin, 1210 West Dayton Street, Madison, WI 53706, USA

Abstract. This tutorial is a non-technical overview of smoothing splines, drawing connections to kriging and kernel smoothing approaches. Technical and historical material can be found in Wahba's (1990) monograph, with additional ideas and alternative approaches in Wegman and Wright (1983) and the books of Eubank (1988) and Hastie and Tibshirani (1990).

1 Types of spline

The archetypical spline is the polynomial spline, a piecewise polynomial function joined at knots. The original mechanical spline was a thin, flexible piece of wood curved to desired shape and tacked down at selected knots. The curved segment between knots can usually be well approximated by a cubic equation, with segments joined seamlessly by appropriate side conditions on the slope and values to insure continuity—an interpolating cubic spline.

One can relax from interpolating to smoothing spline by attaching springs at the knots. If the springs are sufficiently weak, the spline reverts to a straight piece of wood. This mechanical smoothing spline provides a trade-off between fidelity to the data and smoothness, depending on the strength of the spring. This is formalized as the criterion

$$\text{criterion} = (\text{mean squared deviation}) + \text{trade-off} \times (\text{smoothness penalty})$$

Cubic splines can be generalized to higher-order polynomial splines by placing more conditions on higher-order derivatives. There is a common misconception that these are the only types of splines, since computer software generally builds on polynomial splines. In fact, general smoothing splines cover a broad class of functions.

One can model the response as

$$\text{response} = (\text{smooth function}) + (\text{white noise})$$

finding the 'best' estimate as that function which minimizes the penalized criterion.

2 Characterizing smoothness

There are two main ways to characterize smooth functions. A function may be assumed to be fixed but unknown. One can think of it as having a parametric component (such as a straight line) plus a smooth component. The smoothness of the latter is measured by a (usually global) penalty functional, such as the intregrated squared second derivative. Essentially, the penalty has a quadratic form in the smooth function evaluated at the knots.

A Bayesian approach is to consider the smooth function as a random realization of a stochastic process, and view the criterion as a log likelihood. In this case, the smoothness penalty is a quadratic form corresponding to a normal variate. Thus one determines the smoothness of the function by specifying the covariance of a stochastic process. The trade-off measures the ratio of variances of the stochastic process and the error noise.

The latter characterization shows the intimate connection between smoothing splines and kriging, a common methodology in geostatistics and other fields. Essentially (subject to certain conditions), splines and kriging amount to the same thing. However, with splines one tends to set a penalty function, while with kriging one tends to set the covariance or correlogram. Kriging may be easier to use if the focus is on association of response among the knots rather than on estimation of the function itself.

A third approach is to approximate the smooth function at any point by a weighted average of responses at nearby positions. Weights should drop off as distance increases, which can be characterized by a moving window, or 'kernel'. Ideally, the width of the kernel should vary, depending on properties of the smooth function. The smoothing spline estimate can be closely approximated by such a method, in theory and in practice. Buja *et al.* (1989) show that many different smoothing approaches have similar estimation behavior.

The smoothing spline idea can be readily generalized to 2, 3 or more dimensions. Usually, there are more conditions on smoothness, and a corresponding requirement for more data to get reasonable estimates. In addition, computer algorithms are much slower. One-dimensional algorithms are $O(n)$, increasing linearly as sample size increases. Higher-dimensional algorithms are $O(n^3)$. This can be reduced to $O(np^2)$ by artificially restricting the dimensionality of the smooth function space to p, but one may want p to be rather large (100) in practice.

One way to reduce the dimensionality of higher-dimensional problems is to consider additive functions (Hastie and Tibshirani, 1990). The idea is much the same as additive linear models in analysis of variance or regression. Once the interaction(s) are assumed away, the problem reduces to several one-dimensional estimations, which can be accomplished rapidly. It is also much easier to explain additive relationships, and to present graphical summaries of same. Recent work by Gu and Wahba (1992) show that one can develop tests for interaction and construct Bayesian confidence intervals for the smooth function, with or without interactions. Various researchers have considered ways to represent such interval estimates using shading, color or overlays.

3 Generalized linear model and smoothing splines

Many problems involve categorical responses, which bring into question assumptions of normality. The natural extension of the penalized criterion above is to consider a penalized likelihood

$$-2 \times (\text{log likelihood}) + \text{trade-off} \times (\text{smoothness penalty})$$

which has been examined in some detail for binomial and Poisson response. Essentially, one combines ideas about generalized linear models (GLMs) and smoothing splines. This has been variously called general splines, semi-parametric GLMs, penalized likelihood, generalized additive models and iteratively reweighted least squares.

The basic idea is to approximate locally the log likelihood (or deviance) by a quadratic form. It works very well in practice, but odd things can happen with some data sets. For instance, if one has binomial data, and the true proportion is near 0 or 1 over some region (or worse, at the boundary), then the estimation procedure can blow up. Those familiar with GLMs note that the smooth function is trying to approximate infinity! This problem may or may not disappear as sample size increases.

4 The current situation

There are still many holes in formal inference with splines, especially for penalized likelihood methods. The most promising developments seem to be associated with Bayesian 'confidence interval' estimates, which seem to have the right frequentist coverage probabilities, at least in a certain average sense.

One of the chief criticisms of the use of smoothing splines concerns the attention to global smoothing rather than local adaptation. In other words, cubic spline estimates tend to oversmooth in areas of high curvature, missing sharp peaks, and subsequently tend to undersmooth in linear regions. Current research by a number of individuals (notably Joan Staniswallis, Doug Nychka and David Steinberg) offer suggestions of how one might adapt methods to address this problem.

Software is largely still in the research/development stage. The S (or S-plus) package has one-dimensional splines as one possible choice of smoother. The Core Mathematics Library, available through the Internet library 'netlib' as cmlib, contains the B-spline one-dimensional code as well as the multi-dimensional GCVPACK. Chong Gu's RKPACK is available through the Internet 'statlib' library, as is GCVPACK and S drivers for GCVPACK and RKPACK.

Various other authors who have published in the field have public domain or proprietary software in various modes of availability. Anyone interested in software should contact the author(s) of their choice and/or monitor statlib and netlib for new developments.

References

BUJA, A., HASTIE, T. J. & TIBSHIRANI, R. J. (1989) Linear smoothers and additive models, *Annals of Statistics*, 17, pp. 453–555.

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

GU, C. & WAHBA, G. (1993) Semiparametric analysis of variance with tensor product thin plate splines, *Journal of the Royal Statistical Society, Series B*, 55, pp. 353–368.

HASTIE, T. J. & TIBSHIRANI, R. J. (1990) *Generalized Additive Models* (London, Chapman & Hall).

WAHBA, G. (1990) *Spline Models for Observational Data* (SIAM/CBMS).

WEGMAN, E. J. & WRIGHT, J. W. (1983) Splines in statistics, *Journal of the American Statistical Association*, 78, pp. 351–365.