Tech. Rept. 734, Dept. of Statistics, Univ. of Wisconsin, Madison.

OWEN, A. (1983). The estimation of smooth curves. Unpublished manuscript.

REINSCH, C. (1967). Smoothing by spline functions. *Numer. Math.* **10** 177–183.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. Ser. B* **47** 1–52.

STONE, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14** 590–606.

STONE, C. J. and KOO, C.-Y. (1986a). Additive splines in statistics. *Proc. Statist. Comp. Sect. Amer. Statist. Assoc.*

TIBSHIRANI, R. (1984). Local likelihood estimation. Stanford technical report and unpublished Ph.D. dissertation, Dept. of Statistics, Stanford Univ.

WAHBA, G. (1980). Spline bases, regularization, and generalized cross-validation for solving approximation problems with large quantities of noisy data. *Proc. Conf. on Approximation Theory in Honour of George Lorenz, Jan. 8–10, Austin, Texas* (W. Chaney, ed.). Academic, New York.

WAHBA, G. and WOLD, S. (1975). A completely automatic French curve: Fitting spline functions by cross-validation. *Comm. Statist.* **4** 1–7.

YOUNG, F. W., TAKANE, Y. and DE LEUUW, J. (1978). The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika* **43** 279–282.

Comment

David R. Brillinger

> "All considered, it is conceivable that in a minor way, nonparametric regression might, like linear regression, become an object treasured for both its artistic merit as well as usefulness."
>
> L. Breiman (1977)

This paper by Hastie and Tibshirani lays bare the insight of the above remark of Leo Breiman made in the course of the discussion of a seminal work on regression with smooth functions (Stone, 1977). Here Hastie and Tibshirani increase the store of both artistic merit and usefulness by plugging nonparametric regression into the generalized linear model and by alluding to a variety of possible further extensions. It all makes being a statistician these days a joy—it seems approaches are now available to attack most any applied problem that comes to hand. (Understanding the operational performance of those approaches is clearly another matter however.)

It was nice to be asked to comment on such a stimulating paper. I have divided my comments into several sections, striving to focus on individual strains present in the paper, believing that future research on those strains will proceed at different rates.

1. STRUCTURE OF A BASIC PROBLEM

One has data $(Y_i, \mathbf{X}_i)$, $i = 1, \cdots, n$, with n moderately large. One is willing to consider a model for the individual Ys wherein: i) the conditional distribution of Y given $\mathbf{X}$ belongs to an exponential family, ii) it involves $\mathbf{X}$ only through $\eta = \sum s_j(X_j)$ with the $s_j(\cdot)$ unknown, but smooth, and iii) $E\{Y \mid \mathbf{X}\} = h(\sum s_j(X_j))$, with $h(\cdot)$ known. The parameter of the model is $\theta = \{s_j(\cdot), j = 1, \cdots, p\}$, and possibly a scale. The two key elements of the model are a) that the $s_j(\cdot)$ are smooth and b) that $\sum s_j(X_j)$ is additive.

It is to be noted that this model continues the contemporary statistical trend to eliminate distinctions between the cases of finite and infinite dimensional θ or between discrete and continuous data.

The problem is of interest, for one may wish to make inferences from the data via the model or one may wish to validate a model with a low dimensional parameter by imbedding it in a broader model, for example.

2. CONSTRUCTION OF ESTIMATES

To begin, focus on estimating $\eta = \eta(\mathbf{X})$, via a relationship that characterizes the true value η_0. Suppose one has a function $\rho(Y \mid \eta)$ such that $E_0\{\rho(Y \mid \eta) \mid \mathbf{X}\}$ is maximized at $\eta = \eta_0$. An example would be $\log f(Y \mid \eta)$, $f(\cdot)$ denoting the conditional density of Y. Alternately, suppose one has a function $\psi(Y \mid \eta)$ such that $E_0\{\psi(Y \mid \eta) \mid \mathbf{X}\} = \mathbf{0}$ at $\eta = \eta_0$. An example would be $\partial \log f(Y \mid \eta)/\partial\eta$. Estimates of the true η_0 may be constructed by paralleling these relations on the data. For example, given weights $W_{ni}(\mathbf{X})$ such as in Stone (1977) one might take $\hat{\eta}$ to maximize

$$\sum_i \rho(Y_i \mid \hat{\eta}) W_{ni}(\mathbf{X})$$

or to satisfy

$$\sum_i \psi(Y_i \mid \hat{\eta}) W_{ni}(\mathbf{X}) = \mathbf{0}.$$

The estimate of Hastie and Tibshirani based on (26) takes this form. One can expect such estimates to be

David R. Brillinger is Professor of Statistics, University of California, Berkeley, California 94720.