mean information in $\eta(\mathbf{X})$ about $\theta(Y)$. (This does not involve a Jacobian.)

ADDITIONAL REFERENCES

ALLISON, H. (1979). Inverse unstable problems and some of their applications. *Math. Sci.* **4** 9–30.

BREIMAN, L. (1977). Discussion of Consistent nonparametric regression, by C. J. Stone. *Ann. Statist.* **5** 621–622.

LUCE, R. D. and TUKEY, J. W. (1964). Simultaneous conjoint measurement: a new type of fundamental measurement. *J. Math. Psych.* **1** 1–27.

STONE, C. J. (1977). Consistent nonparametric regression (with discussion). *Ann. Statist.* **5** 595–645.

Comment

J. A. Nelder

I congratulate the authors on a fascinating piece of work and offer three comments.

1. In order to make smoothing work it is necessary to restrict it to one-dimensional covariate spaces, hence the strong assumption of additivity. In principle one could introduce cross-terms, e.g., have $x_{12} = x_1 x_2$, as well as x_1 and x_2, in the model; however, I suspect the convergence of the algorithm might now become immensely slow or even nonexistent because of the functional relations between the covariates. An alternative might be to include a term of the form $s_1(x_1) \cdot s_2(x_2)$, with coefficient to be estimated. Have the authors any comments on this problem?

2. To me it seemed intuitively surprising that the figures in Table 2 show the generalized additive model to have one parameter more than the original parametric one, but a deviance nearly 6 higher. I then realized that the latter has a cross-term in it, and this appears to be important. What would be the effect of adding a term in $s_1(x_1) \cdot s_2(x_2)$ to the former? Also it would help interpretation if the difference in deviance were given when each term in their model was replaced by a parametric form. This would give summary statistics for differences visible in Figures 3, 4, and 5.

3. The new version of GLIM (3-77) now available has a facility for inserting new code. I very much hope that the authors can be persuaded to exploit this in order to make available the fitting of generalized additive models in GLIM.

J. A. Nelder is Visiting Professor, Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7, England.

Comment

Charles J. Stone

Hastie and Tibshirani deserve commendation for the originality, significance, and interest of their approach and the excellent expository review in the present paper.

Recently I have been working on a different approach to fitting more or less the same class of models, but using polynomial cubic splines to model the component functions $s_j(\cdot)$ and the Newton–Raphson method to calculate the ordinary maximum likelihood estimate. In order to avoid artificial end effects of polynomial fits such as those shown in Figures 2 and 3, the splines are constrained to be linear to the left of the first knot and to the right of the last knot. To avoid multiple representations of the constant term, zero sum constraints are imposed on the individual terms (when $p \geq 2$), as is done in this paper. Thus, if there are N knots, there are $N + 4$ degrees of freedom for the unconstrained spline and $N - 1$ degrees of freedom for the constrained spline. There is also 1 degree of freedom for the constant term; so there are $(N - 1)p + 1$ degrees of freedom in total. This approach will be referred to as the parametric spline approach to distinguish it from the smoothing spline approach favored by Wahba and others in which smoothing is achieved by a roughness penalty instead of by confining attention to spline models with a modest number of degrees of freedom. In theory, N should tend to infinity as the sample size n tends to

Charles J. Stone is Professor of Statistics, University of California, Berkeley, California 94720.