REFERENCES

BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practise.* MIT Press, Cambridge, MA.

BREIMAN, L. and FRIEDMAN, J. H. (1982). Estimating optimal transformations for multiple regression and correlation. Dept. of Statist. technical report, Orion 16, Stanford University.

CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836.

DEMING, W. E. and STEPHAN, F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* **11** 427–444.

DIACONIS, P. (1983). Projection pursuit for discrete data. Stanford University technical report no. 198.

EFRON, B. (1979). Bootstrap methods: another look at the jacknife. *Ann. Statist.* **7** 1–26.

FIENBERG, S. E. (1977). *The Analysis of Cross-Classified Categorical Data.* MIT Press, Cambridge, MA.

FRIEDMAN, J., STUETZLE, W. and SCHROEDER, A. (1984). Projection pursuit density estimation. *J. Amer. Statist. Assoc.* **79** 599–608.

HASTIE, T. and TIBSHIRANI, R. (1984). Generalized additive models. Stanford University, Lab. for Comput. Statist. report no. 2.

MCCULLAGH, P. and NELDER, J. (1983). *Generalized Linear Models.* Chapman Hall, London.

NELDER, J. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135** 370–384.

TIBSHIRANI, R. and HASTIE, T. (1984). Local likelihood estimation. Stanford technical report 97 and unpublished Ph.D. thesis (1st author), Dept. of Statist. Stanford University.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

M. C. JONES

University of Birmingham

Professor Huber presents a most interesting paper reviewing the broad area within multivariate data analysis now encompassed by the term "projection pursuit." My own comments relate to recent research in this field undertaken at the University of Bath, UK, by myself and Professor Robin Sibson. Our work focussed on the basic projection pursuit algorithm thought of as an exploratory tool applied to point clouds—as a method for finding "interesting" low-dimensional "views" of a multivariate data set—in the spirit of Friedman and Tukey (1974); as such, these comments are most relevant to Section II of the current paper.

Initially, we had access only to Friedman and Tukey's pioneering paper and during much of the course of our work remained unaware of the more recent work by Professors Huber, Friedman and others. Bearing this in mind, the close agreement between many of Professor Huber's ideas and our own, which are outlined below, seems quite remarkable.

The particular implementation of the projection pursuit method described by Friedman and Tukey allowed considerable scope for improvement on both theoretical and practical grounds. Consequently our aim was to provide a new version of the algorithm which was based on firmer mathematical foundations (like the "abstract" version of projection pursuit in the current paper) and was more efficient computationally, remembering the needs of many potential users with rather limited computational resources.

On the theoretical (abstract) side, many of our ideas coincide with those in Professor Huber's paper. Our initial step was to ensure that the data were sphered—in the sense of zero mean and unit covariance matrix—hence bypassing the need to worry about affine invariance of our chosen projection index (Professor Huber's "Class III" functionals). We first noticed that Friedman and Tukey's original projection index was essentially an estimate of $\int f^2$ (see Section 5 of the current paper) and, as such, is a monotone function of order-2 entropy (for a definition of order-α entropy see Rényi (1961)) which is minimised, over densities with given mean and variance, by a parabolic density; it was a natural step to divert attention to the more usual (order-1) Shannon entropy, the negative of which is minimised by the normal density. In this way, we independently decided to focus on projection indices which, in some sense, measured nonnormality of projections of the data; by doing so, we replaced the ad hoc notion of seeking out clusters (as used by Friedman and Tukey) by the device of searching for projections which looked least like the normal distribution. Further agreement between ourselves and the current paper lies in the use of higher order cumulants in alternative projection indices measuring nonnormality. Indeed, a heuristic argument due to Professor Sibson links entropy with a quantity involving the (squares of) third- and fourth-order cumulants; the corresponding projection index has considerable computational advantages in certain circumstances and has proved to work well in practice (albeit with the expected marked preference for projections indicating possible outlying observations).

These ideas have been incorporated into a new implementation of the method. To obtain sample versions of those projection indices which are functionals of the (marginal) density of a projection, kernel density estimates have been employed (as alluded to by Professor Huber). At least for one-dimensional projection pursuit (i.e. $k = 1$ in the notation of this paper), much use has been made of an efficient algorithm for computing such density estimates on a grid of points, due to Silverman (1982); this takes a time virtually independent of the number of data points and hence is almost equally applicable to large as well as small data sets. Of course, use of kernel density estimates requires the user to choose a sensible value for the smoothing parameter, or window-width. Appropriate choice of window-width is a matter for considerable investigation; however, empirical evidence suggests that such choice is not as critical as in many other contexts since there appears to be a broad band of values for which the method behaves in much the same manner. For projections onto $k = 2$ dimensions, indices based on higher order cumulants come into their own because all computations may be based on a set of summary statistics which is calculated once and for all (namely the appropriate multivariate higher order cumulants), rather than using the entire data set at each step.

Each of the projection indices under consideration varies smoothly as the projection axis or plane varies. Derivatives of the indices are available at little extra computational cost and may be employed to great advantage in the course of the numerical optimisation of the chosen index. As Professor Huber points out in a slightly different context "it does not matter very much if a particular direction ... is determined inaccurately", so a conceptually simple and computationally efficient steepest ascent algorithm has been used to good effect.

If we were to take issue with any of Professor Huber's remarks, it would only be to doubt the usefulness of three-dimensional projections in this exploratory setting, particularly bearing in mind the additional computational burden such projections would impose. Representation of three-dimensional data in a single informative picture (on two-dimensional paper!) is not readily achieved in an immediately meaningful way. Two-dimensional projections, via scatter plots or bivariate density estimates, are readily interpretable, however, and, as Professor Huber points out, may often show interesting features of the data which are not apparent in any one-dimensional projection. For these reasons, we have restricted our attention to both one- and two-dimensional projection pursuit, even, on occasion, for application to three-dimensional data.

Finally, practical experience with the resulting version of the projection pursuit algorithm has proved to be most encouraging. Considerable discussion of the practical advantages and limitations of the technique, together with many further details of the work outlined briefly above, may be found in the thesis of Jones (1983) and in a forthcoming paper to be written jointly with Professor Sibson.

REFERENCES

JONES, M. C. (1983). The projection pursuit algorithm for exploratory data analysis. Ph.D. thesis, University of Bath.

RÉNYI, A. (1961). On measures of entropy and information. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** (J. Neyman, ed.) 547–561. Univ. California Press.

SILVERMAN, B. W. (1982). Algorithm AS176. Kernel density estimation using the Fast Fourier Transform. *Appl. Statist.* **31** 93–99.

DEPARTMENT OF STATISTICS
UNIVERSITY OF BIRMINGHAM
PO BOX 363
BIRMINGHAM, ENGLAND R15 2TT

RUPERT G. MILLER, JR.

Stanford University

In 1976 Dr. Gerald Reaven and I, with the assistance of M. A. Fisherkeller, successfully applied projection pursuit to some diabetes data. The data consisted of the (1) relative weight, (2) fasting plasma glucose, (3) area under the plasma glucose curve for the three-hour glucose tolerance test (OGTT), (4) area under the plasma insulin curve for the OGTT, and (5) steady state plasma glucose response (SSPG) for 145 subjects at the Stanford Clinical Research Center, who volunteered for a study of the etiology of diabetes. The goal of the study was to