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 investigate the interrelationships between these variables and to see what connection they had with classifying patients as normal, chemical diabetic, or overt diabetic.

The data set was examined with the PRIM-9 program at the Stanford Linear Accelerator Computation Center. This program permits any three variables to be selected and then displayed as a two-dimensional image of the projection of the points along any direction. By continuously moving the direction, the three-dimensional configuration of the points is revealed.

From a study of the various combinations of three variables, the configuration displayed in Figure 1 (an artist's rendition of the data cloud) emerged. This figure was considered to be very interesting to medical investigators. The plump middle of the points roughly corresponds to normal patients, the right arm to chemical diabetics, and the left arm to overt diabetics.

After this unusual configuration was discovered through the use of the PRIM-9 program, it was learned through ordinary two-dimensional plots of pairs of variables that the same figure could be seen in these simpler plots. However, it is not clear that merely examining the three plots of SSPG vs. Glucose Area, SSPG vs. Insulin Area, and Glucose Area vs. Insulin Area would have revealed the figure whereas it leaped out of the display screen for the viewers with the PRIM-9 projections and rotations.

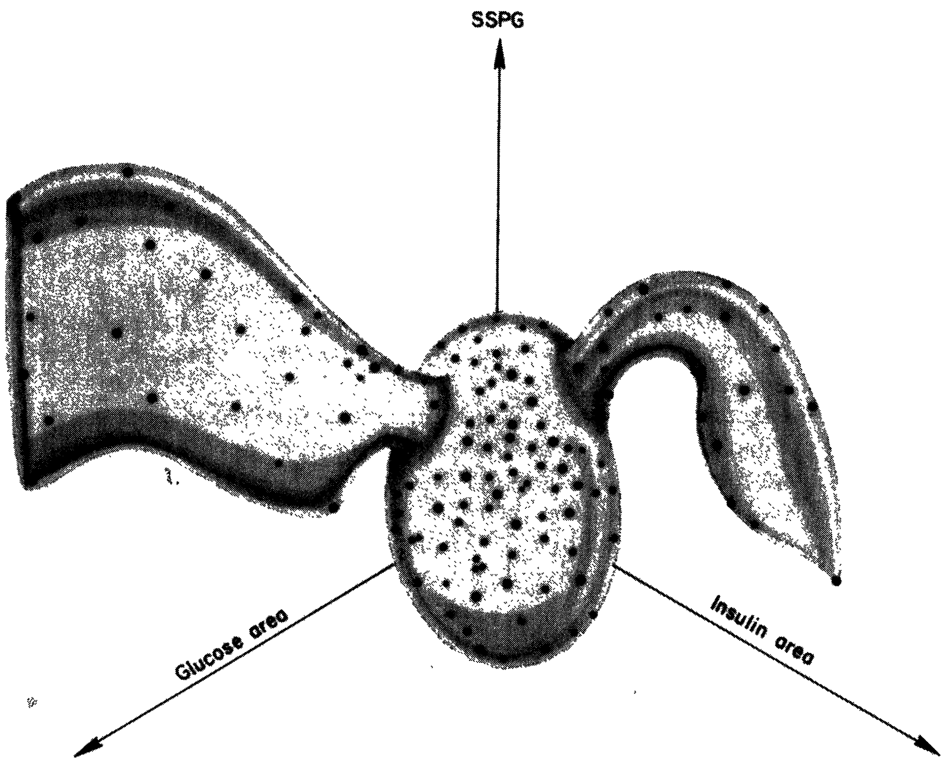

FIG. 1

This work, along with a simple clustering analysis separating the patients into the three groups, was published in Reaven and Miller (1979). This application of PRIM-9 was referred to in a *Science* article by Kolata (1982). Different clustering algorithms were applied to this data set by Symons (1981). Also, Diaconis and Friedman (1983) investigated these data by M and N plots and found a weak association between relative weight and SSPG, which was not discovered in our original work. This data set has also been referred to in Diaconis (1985).

During the last decade and a half, advances in computer technology have revolutionized the practice of statistics. PRIM-9 and its newer versions have enabled investigators to visualize multidimensional data as never before. Other specialized hardware and software have provided the statistician with a new variety of graphical and computational procedures for handling multivariate data. These new techniques are being applied to data sets, and conclusions are being drawn on the basis of these analyses.

With reference to this new technology, I feel that there is a gap developing between practice and theory. While interesting shapes, clusters, and separations are now more discoverable in high-dimensional data, how do we evaluate whether these are real or merely random variation? The theory for separating real from random seems somewhat limited at this point in time.

As an example, consider the simple problem of comparing two populations with high-dimensional data. Are the two populations the same, or is there a boundary which can be effectively used to separate data points into the two populations with small error rates? There exist the classical multivariate tests of equality of two populations and some newer tests like those of Friedman and Rafsky (1979, 1981) using the minimal spanning tree, but these tests do not describe a separation boundary. The classical Fisher linear discriminant function, or the logistic function, with all or a subset of the variables and possibly some of their interactions are typically used in practice to determine a boundary. This analysis, however, assumes a very specific parametric form for the boundary which may not be the best. Recursive partitioning, introduced by Friedman (1977), Gordon and Olshen (1978), and Breiman et al. (1984), is sometimes used, but at times this technique can produce somewhat bizarre multiple noncontiguous regions for classification.

Suppose the statistician through his/her visualization of the multidimensional data on the computer thinks that he/she sees some sort of smooth contour which effectively separates the populations. How does he/she evaluate whether this separation is real or random? The theory and critical constants to measure the significance of the separation into a 2×2 table, or the significance of coefficients in an equation defining the boundary, are unavailable at the present. The multiple comparisons problem involved in this evaluation is enormous.

This article by Peter Huber is therefore a welcome addition to the literature. It brings together for the first time some theoretical results on a set of diverse but related new techniques. Much additional work in these directions should appear in the statistical literature over the forthcoming decades because that is where the computer is taking us.

REFERENCES

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., and STONE, C. J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont, CA.

DIACONIS, P. (1985). Theories of data anlysis from magical thinking through mathematical statistics. In *Exploring Tables, Trends and Shapes* (D. C. Hoaglin, F. Mosteller, and J. W. Tukey, eds.). Wiley, New York.

DIACONIS, P. and FRIEDMAN, J. H. (1983). *M* and *N* plots. *Recent Advances in Statistics* (M. H. Rizvi, J. Rustagi, and D. Siegmund, eds.). 425–447, Academic, New York.

FRIEDMAN, J. H. (1977). A variable metric decision rule for nonparametric classification. *IEEE Trans. Comput.* **C-26** 404–408.

FRIEDMAN, J. H. and RAFSKY, L. C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7** 697–717.

FRIEDMAN, J. H. and RAFSKY, L. C. (1981). Graphics for the multivariate two-sample problem. *J. Amer. Statist. Assoc.* **76** 277–287.

GORDON, L. and OLSHEN, R. A. (1978). Asymptotically efficient solutions to the classification problem. *Ann. Statist.* **6** 515–533.

KOLATA, G. (1982). Computer graphics comes to statistics. *Science* **217** 919–920.

REAVEN, G. M. and MILLER, R. G. (1979). An attempt to define the nature of chemical diabetes using a multidimensional analysis. *Diabetologia* **16** 17–24.

SYMONS, M. J. (1981). Clustering criteria and multivariate normal mixtures. *Biometrics* **37** 35–43.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

L. A. SHEPP

AT&T Bell Laboratories

I would like to discuss the many interesting similarities and differences between projection pursuit (PP) and computed tomography (CT) as emphasized by Professor Huber.

First, just in case PP becomes a successful tool in exploratory data analysis (EDA), I hasten to point out that CT has indeed had some influence on PP. Indeed, the basic (to PP) concept of ridge function is an idea which was introduced to CT already in Logan's paper in the *Duke Math. J.*, 1975. The idea of superposing filtered projections, which is the basis of all algorithms used in commercial CT, is analogous to superposing filtered projections in PP even though "filtered" in PP is used in the sense that all but the "interesting few of the projections" are discarded while in CT "filtered" means convolution of each projection with a "filter function" (this is also due to Logan). Actually PP seems even closer to emission CT than to transmission CT—see Vardi, Shepp and Kaufman (1985), *J. Amer. Statist. Assoc.* **80** 8–37.

Emphasizing next the differences between PP and CT for accuracy of discussion, we note:

(a) Parallel linear projections in PP are not fundamental—one could easily imagine nonparallel projections, fan-beam, or even curvilinear projections rather than straight line projections. Thus in PP the density is sampled,