E. L. Lehmann
Department of Statistics
University of California
Berkeley, USA

FOUNDATIONAL ISSUES IN STATISTICS: THEORY AND PRACTICE

Abstract. The foundational issues discussed by David Freedman are examined from the point of view of an investigator faced with the choices they imply. It turns out that in practice the boundaries between the various philosophies are less sharp than might be expected from the passion with which their differences have been argued for a long time. There is considerable common ground, and each approach gains by borrowing tools from the other.

1. Introduction

David Freedman's paper consists of two parts. The first describes the frequentist and Bayesian approaches to probability and statistics, the second is concerned with the validity of statistical models. Both emphasize the importance of putting realism above convenience and dogma. I shall here consider these issues from the point of view of an investigator faced with the choices implied by Freedman's distinctions, namely between

- data analysis without a model;

- a frequentist model involving unknown parameters;

- A Bayesian model which assigns a known distribution to these parameters. (I am restricting attention here to what Freedman calls the "classical" Bayesian position; bridge building to the "radical" position is more difficult).

2. Model-free data analysis

At this stage, a data set will typically be examined, analyzed, and organized in many different ways in an attempt to bring out its salient features and to pinpoint meaningful effects. However, the question then arises whether these effects are real. Unfortunately, except in very special circumstances (for an example see Freedman and Lane (1983)), this question cannot be answered without a probability model. The difficulty is greatly exacerbated by the multiple facets of such an approach. The suggested effects of interest, for example, will tend to be those that appear most significant out of large numbers that were examined, some in detail, some

only out of the corner of an eye, and many of which – in a phrase of Cournot (1843) – "have left no traces". The appearance of their significance may thus be greatly exaggerated. Cournot thought that the problem posed by this distortion was insoluble. However, often the situation can be saved by a second stage with new data at which a probability model is used to test the much smaller and more clearly defined set of effects found interesting at the first stage.

3. The modeling decision

The choice between an analysis with or without model depends largely on the extent of our knowledge concerning the underlying situation. Do we know enough from past experience and/or a theoretical understanding of the situation to be clear about the questions to be asked and to be able to formulate an adequately realistic model? Or do we have to examine the data to see what the problems are and to get an idea of the kind of model that might be appropriate?

In this connection it is important to realize that models used for different purposes need to meet different standards of realism. A well known classification (see for example Box, Hunter and Hunter (1978), Cox (1990) and Lehmann (1990)) distinguishes between on the one hand explanatory or substantive models which try to portray in detail the principal features of the situation and to explain their interrelations, and on the other empirical or off-the-shelf models, which employ the best fitting model from a predetermined standard family. The latter have little explanatory power but nevertheless can be very useful for limited practical purposes such as predicting school enrollments or forecasting elections. (Early astronomers were able to make very accurate predictions based on elaborate models [1] that bore little resemblance to reality. Their insistence on circular motions is reminiscent of the ubiquitous assumption of linearity in the kind of regression model discussed by Freedman in Section 5 of his paper.) As Freedman points out, the success of such predictions could and should be routinely monitored. A danger of empirical models is that once they have entered the literature, their shaky origin and limited purpose have a tendency to be forgotten.

4. Model-based inference

Suppose now that a model has been selected which we believe to be adequate, and that we know exactly which questions we want to ask. Let us assume that the problem has been formulated as a choice among a set of possible decisions and that we can assess the loss $L(\theta, d)$ that would result from taking decision d when θ is the true parameter value. The observations are represented by a random quantity X which has a probability distribution depending on θ. A statistical procedure is

---

[1] A recent account of the accomplishments and ultimate failures of these models can be found in North (1994).

a rule (or function) δ which specifies the decision $d = \delta(x)$ to be taken when X takes the value x. Finally, the performance of the decision rule δ is measured by the average loss resulting from its use,

$$R(\theta, \delta) = EL[\theta, \delta(X)] \tag{1}$$

This expected loss is called the risk function of δ.

A classical Bayesian has a prior distribution W for θ; for the sake of simplicity we suppose that W has a density $w(\theta)$. Within this Bayesian framework, the quantity of interest is the expected risk

$$r_\delta = \int R(\theta, \delta) w(\theta) d\theta \tag{2}$$

and the most desirable procedure is that for which r_δ is as small as possible. Determining this optimum procedure is in principle straightforward, although it may be computationally difficult.

The hardest part often is the choice of W. In practice, prior distributions, just like many probability models, frequently are of the off-the-shelf kind, that is, they are chosen from a convenient standard family which is sufficiently flexible to permit modeling the rough qualitative features desired. A popular choice is an uninformative prior that corresponds to a state of ignorance concerning θ.

Frequentists do not face the issue of determining a prior, but they have to contend with another difficulty. Rather than minimizing (2), they seek a procedure that minimizes $R(\theta, \delta)$ for all θ. Since, unfortunately, such a procedure does not exist, various devices are used to work around this problem, none of them really compelling. We mention here only one such criterion: minimizing the maximum risk. A procedure achieving this is said to be minimax.

5. Common ground

The distinctions between the three approaches: (i) model-free data analysis, (ii) frequentist and (iii) Bayesian model-based inference have been heatedly debated for a long time. However, in practice there is more contact – and the lines separating the three modes are less sharp – than this debate suggests.

For example, an exploratory data analysis may perform informal tests with ad hoc models, for guidance on which hypothesis suggested by the data to pursue; conversely, the subsequent formal-inference stage may utilize the first stage to search for the models it needs.

We shall in the remainder of this section restrict attention to some of the relationships between (ii) and (iii). Additional discussion of these and other points of contact between the two approaches can be found in Diaconis and Freedman (1986) and Lehmann (1985).

(a) Previous experience. For a large class of important situations in areas such as medicine, agriculture, business and education, a great body of related earlier experience is available to draw on. This experience provides an observed frequency distribution of θ's of which we have reason to believe that the present (unknown) θ-value is a typical member. A Bayesian is likely to accept this distribution as an appropriate prior (although possibly making slight modifications to allow for special features in which the present situation might differ from the past). A frequentist, even without a clearly defined random mechanism for producing the θ's, may analogously work with the empirical distribution as if it were the true frequency distribution of a random quantity. Although they will think very differently both of this distribution W and of the assumed distribution of X given θ (along the lines indicated in Sections 2 and 3 of Freedman's paper), both will be led to the procedure δ minimizing (2) and thus to the same decision.

(b) A fundamental connection. Even when no prior information is available, Bayes procedures are of importance to frequentists. This is a consequence of a fundamental theorem of Wald, which states that the only sensible (from a frequentist point of view) procedures are those that are Bayes solutions for some prior W or limits of such solutions. In the light of this result, it is often useful to examine a frequentist procedure from a Bayesian point of view to see whether it is Bayes with respect to some prior and if so, whether this prior seems at all plausible. In this connection it is interesting to note that minimax procedures are often Bayes solutions corresponding to an uninformative prior.

(c) Performance. In the opposite direction from (b), a Bayesian, after having computed the Bayes procedure δ corresponding to a preferred prior, can learn much about the procedure by studying its risk function $R(\theta, \delta)$. This latter step, although inconsistent with a Bayesian philosophy, is now acknowledged as useful by many Bayesians. (For a discussion from a contrary viewpoint, see Berger (1986).)

5. Conclusions

1. It seems clear that model-free data analysis, frequentist and Bayesian model-based inference and decision making each has its place. The question appears not to be – as it so often is phrased – which is the correct approach but in what circumstances each is most appropriate.

2. In practice, the three approaches can often fruitfully interact, with each benefiting from consideration of the other points of view.

3. Although the debates between adherents of Bayesian and frequentist philosophies have been carried on with much heat for a long time, in practice neither side very often lives up to its principles, since both models and priors frequently are of the off-the-shelf variety.

References

Berger, J. ,1986, Discussion of Diaconis and Freedman, *Annals of Statistics* **14**, 30-37.

Box, G. E. P., Hunter, W. G., and Hunter, J. S., 1978, *Statistics for Experiments.* New York, Wiley.

Cournot, A. A., 1843, *Exposition de la Théorie des Chances et des Probabilités.* Paris, Hachette

Cox, D. R., 1990, Models in Statistical Analysis, *Statistical Science* **5**, 169-174.

Diaconis, P. and Freedman, D., 1986, On the Consistency of Bayes Estimates, *Annals of Statistics* **14**, 1-67, (with discussion).

Freedman, D. A. and Lane, D., 1983, Significance Testing in a Nonstochastic Setting, pp.185-208 in P. J. Bickel, K. A. Doksum, and J. L. Hodges, Jr. (eds), *A Festschrift for Erich L. Lehmann in Honor of his Sixty-Fifth Birthday.* Belmont, Calif., Wadsworth.

Lehmann, E. L., 1985, The Neyman-Pearson Theory After Fifty Years, pp.1-14 in L. LeCam and R. Olshen (eds), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer.* Vol. 1. Monterey, Calif., Wadsworth.

Lehmann, E. L., 1990, Model Specification, *Statistical Science* **5**, 160-168.

North, J., 1994, *Astronomy and Cosmology.* London, Fontana.