unkindest remark (which is not so unkind): "It is difficult to see how these questions can even be posed within the frequentist framework." This seems wrong. There is no difficulty in posing the questions, in either the frequentist or Bayesian framework; Hodges just did it. The problem is finding answers.

Now there comes a shade of difference between us. He is a little more optimistic than I am about the potential usefulness of Bayesian techniques for properly integrating judgments about uncertainty. For example, he discusses predictive distributions starting from (i) a prior on models and their parameters and (ii) a likelihood function for the data given the model and parameters.

This is quite sensible, provided there is a sound basis for choosing the prior and the likelihood. Unfortunately, Bayesian policy analysts can be just as slaphappy in such matters as us frequentists. For discussion of this issue, see Freedman and Navidi (1986) or Hill (1985).

Good statistical analysis can be done in either the frequentist or the Bayesian framework. However, for either approach to succeed, the analyst has to get the model right, or close enough. That idea may seem ridiculously old fashioned. As policy analysts can be heard to sputter, "Models be right? How can they be right? They're all approximations. Even Newton was wrong. And a mystic besides." Because nothing is perfect, anything goes.

Hodges wants "to bring de Finetti to . . . practitioners." As I understand him, for de Finetti a prior represents a major intellectual commitment to be adopted only after serious investigation of the subject at issue. If policy analysts followed that percept, we would all be better off. The real issues here are of science, not statistical technique.

ADDITIONAL REFERENCES

FREEDMAN, D. (1987). As others see us: A case study in path analysis (with discussion). *J Educational Statist.* To appear.

FREEDMAN, D. and ZEISEL, H. (1987). From mouse to man: The quantitative assessment of cancer risk (with discussion). Technical Report, Dept. Statistics, Univ. California, Berkeley. *Statist. Sci.* To appear.

HILL, B. (1985). Some subjective Bayesian considerations in the selection of models (with discussion). *Econometric Rev.* **4** 191–288.

KOLATA, G. (1986). Asking impossible questions about the economy and getting impossible answers. *Science* **234** 545–546.

LEAMER, E. (1983). Taking the 'con' out of econometrics. *Amer. Econom. Rev.* **73** 31–43.

NATIONAL ACADEMY OF SCIENCES. (1984). *Improving Energy Demand Analysis.* Washington.

ZEISEL, H. (1981). Race bias in the administration of the death penalty: The Florida experience. *Harvard Law Rev.* **95** 456–468.

Comment

Seymour Geisser

Now comes James Hodges to inform us on some of the larger issues of statistics. And what are these issues? They are the ones that statisticians have dealt with—lo these many years—uncertainties from various sources. And there are other issues besides—is it an observational study? a controlled experiment? a retrospective investigation? a haphazard collection of items? Is what is measured or observed actually what one defines as measured? Are there flawed observations? Was the experiment or trial carried out according to the protocols? Is there a temporal imperative with regard to an action or a decision? There is, to say the least, limited interest (other than procedural validation perhaps) in the prediction of events that already have occurred and been observed.

What is the point then? The point is that we have here a lucid and trenchant exposition vividly reminding us of three of the principal sources of uncertainty or variation. What is more novel than most previous explications is that the sources are related to predictivism, which is stressed as the penultimate aim when taking an action is the ultimate goal. Hence, from my point of view, there is really nothing to quarrel with. But it is the job of a discussant if not to be quarrelsome to be at least quibblesome—to coin a neologism.

Hodges intimates that for proper application of statistical methods, the implementation of de Finetti's approach is required. He also states that the approach "lacks a crucial connection to real problems." I would like to quibble with both these points. In regard to the latter point, we have only to realize that de Finetti was involved in applications especially in finance,

Seymour Geisser is Professor and Director, School of Statistics, University of Minnesota, 270 Vincent Hall, 206 Church Street, S.E., Minneapolis, Minnesota 55455.

actuarial mathematics, censuses, football pools, lotteries, etc.—and these problems are just as real in their own realm as spare parts is to the Air Force. In fact I would surmise that de Finetti's views were molded to some extent by his actuarial experience.

In regard to the first point there is a fundamental principle in the de Finetti canon that statisticians (even Bayesian predictivists) often ignore. To be fully coherent one's view about the prior distribution should not depend on the likelihood. This straightjacket is almost always flouted when searching for models in analyzing data. I believe the attitude that many reasonable statisticians take is, if I assume this then I will believe that, and decide, after examining the data in various ways, what it is they eventually believe about values as yet unobserved. Also the de Finetti canon precludes testing one's predictive methodology against realized further observations or using some other validation technique. But these methods are far too useful in searching for and selecting appropriate models to be summarily dismissed.

Of course the Neyman-Pearson approach is further beset by even greater stringencies. My view of it is to paraphrase a boxing manager's lament over his fallen fighter "N–P has some good qualities but its bad qualities ain't so good."

With regard to the use of predictive distributions for testing and as a diagnostic tool, I believe its most appropriate application is in discordancy testing. In such instances the bulk of the observations are assumed to concord with the postulated model although possibly a few may not and they are to be tested in the absence of any discernible alternative, e.g., Geisser (1980b, 1985, 1987), Johnson and Geisser (1983).

In discussing estimation risk and predictive risk the prediction of future values from a $N(\theta, 1)$ distribution is presented in the frequentist framework. Because $X_{n+1} - \bar{X}$ has variance $1 + n^{-1}$, Hodges asserts that in terms of variance $1/n$ is the estimation risk and 1 the prediction risk. This is not so, although it is true that n^{-1} is the estimation risk, $1 + n^{-1}$ is the prediction risk. Estimation risk is inevitably embedded in prediction risk. In fact a deeper interpretation of this example indicates clearly the inappropriateness, even in the classic frequentist setup, in using estimation procedures for predictive purposes, i.e., using $N(\bar{x}, 1)$ as the "best" estimate of the distribution function clearly leads to predictive confidence intervals for X_{N+1} that are too short for the stated confidence level (Geisser, 1980a). For further discussion of these points see Geisser (1982) and in particular for the relative importance of the two risks see my rejoinder to the discussion by DeGroot (1982) of that paper.

Although it is true that actual applications of predictive distributions to new data sets are relatively rare—much the same could be said of Bayesian applications. And predictive distributions properly exist only in Bayesian contexts. But to state that Duncan and Lambert (1986) is the only application he can find, indicates either myopia or tunnel vision, because there are many others. For example, every Bayesian model selection or classification procedure uses predictive distributions.

In other parts of this otherwise perceptive essay there are some comments alluding to the use of mixing models. It is true that in the context of Bayesian prediction, when the number of entertained models is exhaustive and prior probabilities for them are determined, then the predictive distribution is a mixture, cf., Clayton Geisser, and Jennings (1986) and Geisser and Eddy (1979). Aside from the fact that exhaustiveness is often not the case, it appears that most scientific workers tend to opt for a single preferred model, operating on some principle of parsimony, or some latent loss function that heavily penalizes mixing discrete models. I must admit to exhibiting ambivalence about this issue, or more precisely, I have not really formulated an appropriate loss function that I find satisfactory. There is certainly something elegant about a single reasonable model that appears to be adequate for the data. On the other hand if the data are such that no single model seems to adequately capture all of the features in the data, then I would be more inclined to mixing because it is more likely that my predictions will be better (on some defined measure) than using a single model. A situation not altogether different presents itself in the potential infinite regress on hyperparameters and hyperpriors (Geisser, 1980c, page 466)—when does one stop? True solutions to problems of this sort require a belief in the potential reality of the models and the specified parameters. Serious doubt concerning these and other aspects of the modeling situation should, to a degree, relieve one of the necessity of trying to conjure up distributions for nonexistent entities especially if one is reasonably satisfied with something workable and adequate.

The spread of penetrating predictive applications to social affairs and policy analyses is certainly an important and welcome development and one that requires much thought and considerable effort in dealing with these action-oriented problems. Hodges has made a fine start in delineating some of the principal issues.

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation Grant DMS-86-01314 and National Institutes of Health Grant GM-25271.

ADDITIONAL REFERENCES

Clayton, M. K., Geisser, S. and Jennings, D. E. (1986). A comparison of several model selection procedures. In *Bayesian*

Inference and Decision Techniques: Essays in Honor of Bruno de Finetti (P. K. Goel and A. Zellner, eds.) 425–439. North-Holland, Amsterdam.

GEISSER, S. (1980b). Discussion of "Sampling and Bayes' inference in scientific modeling and robustness" by G. E. P. Box. *J. Roy. Statist. Soc. Ser. A* **143** 416–417.

GEISSER, S. (1980c). Predictive sample reuse techniques for censored data (with discussion). In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 433–468. University Press, Valencia.

GEISSER, S. (1982). Aspects of the predictive and estimative approaches in the determination of probabilities (with discussion). *Biometrics Suppl.* **38** 75–93.

GEISSER, S. (1985). On the prediction of observables: A selective update. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds.) 203–230. North-Holland, Amsterdam.

GEISSER, S. (1987). Influential observations, diagnostics and discordancy tests. *J. Appl. Statist.* **14** 133–142.

GEISSER, S. and EDDY, W. F. (1979). A predictive approach to model selection. *J. Amer. Statist. Assoc.* **74** 153–160. Corrigendum **75** 765 (1980).

Comment

Peter J. Huber

This is a very stimulating paper, and the issues raised and discussed in it—how to deal with three distinct types of uncertainty: structural, stochastic and technical—clearly are important not only in applied statistical work, but also beyond.

I shall confine my comments to two central issues of this paper: the problem of the infinite regress and the question of whether and when to combine different kinds of uncertainty.

The main and obvious difficulty one faces with the structural type of uncertainty is an infinite regress: once one has quantified the structural uncertainty, one also should quantify the uncertainty of this quantification, and so on. The customary (perhaps: the only?) way to cut this regress is to act as if at a certain level there was certainty. Often (although not necessarily), this means that one assumes some parametric family of structural models; if one is a Bayesian, one also posits a fixed prior on the space of parameters. It is somewhat awkward in the case of the Bayesian approach that at this stage of modeling the prior will not reflect a reasonably accurate, objective or subjective probability; it rarely is anything more than a conventional substitute for ignorance (e.g., a flat or a conjugate prior). But what is much worse, and this equally affects all approaches, is that the true structure with practical certainty will lie outside of the parametric family. I am always surprised how glibly a majority of statisticians (especially Bayesians!) are able to talk around these difficulties. Roughly speaking, what happens is that in large samples the procedure will pick some member of the parametric family close to the true structure (whatever that means) and then try to do the best possible for that member. It depends on the parametric family, on the type of procedure, on the true situation and on an unspecified kind of closeness whether the "best possible" for the model is any good at all for the true situation. Mere intuition can be very misleading.

It would be a delusion to think that a Bayesian approach, that is, the opportunity to choose also a prior, in such situations provides more security than the Neyman-Pearson version—if the family of models chosen by the statistician (i.e., the support of his prior) does not contain the true underlying situation, one has to go outside of the Bayesian framework in order to justify the use of a Bayes procedure.

For me, this infinite regress was a major conceptual difficulty when I first got into statistics in 1961; the stumbling block then turned out to be the stepping stone leading into a theory of robustness. Some personal reminiscences about the struggle preceding my 1964 paper may help to illustrate the point. Somehow, I then wanted to capture situations describable by statements like: "With this kind of data I would expect about 2% grossly wrong observations, but probably not more than 10%; these values could be anywhere." After some stillborn attempts with a nonparametric version of maximum likelihood, I naturally tried Bayesian approaches next, since by then I knew that even large data samples (I had had experiences with nonlinear least squares problems with a few thousand observations) would not allow me to assess distribution tails reliably without using outside information. However, I was unable to invent believable priors. After a while I realized that the problem was not

Peter J. Huber is Professor of Statistics, Harvard University, One Oxford Street, Cambridge, Massachusetts 02138.