Why Isn't Everyone a Bayesian?

B. EFRON*

Originally a talk delivered at a conference on Bayesian statistics, this article attempts to answer the following question: why is most scientific data analysis carried out in a non-Bayesian framework? The argument consists mainly of some practical examples of data analysis, in which the Bayesian approach is difficult but Fisherian/frequentist solutions are relatively easy. There is a brief discussion of objectivity in statistical analyses and of the difficulties of achieving objectivity within a Bayesian framework. The article ends with a list of practical advantages of Fisherian/frequentist methods, which so far seem to have outweighed the philosophical superiority of Bayesianism.

KEY WORDS: Fisherian inference; Frequentist theory; Neyman–Pearson–Wald; Objectivity.

1. INTRODUCTION

The title is a reasonable question to ask on at least two counts. First of all, everyone used to be a Bayesian. Laplace wholeheartedly endorsed Bayes's formulation of the inference problem, and most 19th-century scientists followed suit. This included Gauss, whose statistical work is usually presented in frequentist terms.

A second and more important point is the cogency of the Bayesian argument. Modern statisticians, following the lead of Savage and de Finetti, have advanced powerful theoretical reasons for preferring Bayesian inference. A byproduct of this work is a disturbing catalogue of inconsistencies in the frequentist point of view.

Nevertheless, everyone is not a Bayesian. The current era is the first century in which statistics has been widely used for scientific reporting, and in fact, 20th-century statistics is mainly non-Bayesian. [Lindley (1975) predicts a change for the 21st!] What has happened?

2. TWO POWERFUL COMPETITORS

The first and most obvious fact is the arrival on the scene of two powerful competitors: Fisherian theory and what Jack Kiefer called the Neyman–Pearson–Wald (NPW) school of decision theory, whose constituents are also known as the frequentists. Fisher's theory was invented, and to a remarkable degree completed, by Fisher in the period between 1920 and 1935. NPW began with the famous lemma of 1933, asymptoting in the 1950s, though there have continued to be significant advances such as Stein estimation, empirical Bayes, and robustness theory.

Working together in rather uneasy alliance, Fisher and NPW dominate current theory and practice, with Fisherian ideas particularly prevalent in applied statistics. I am going to try to explain why.

*B. Efron is Professor, Department of Statistics, Stanford University, Stanford, CA 94305.

3. FISHERIAN STATISTICS

In its inferential aspects Fisherian statistics lies closer to Bayes than to NPW in one crucial way: the assumption that there is a *correct* inference in any given situation. For example, if $x_1, x_2, \ldots, x_{20}$ is a random sample from a Cauchy distribution with unknown center θ,

$$f_\theta(x_i) = \frac{1}{\pi[1 + (x_i - \theta)^2]},$$

then in the absence of prior knowledge about θ the correct 95% central confidence interval for θ is, to a good approximation,

$$\hat{\theta} \pm 1.96 \Big/ \sqrt{-\ddot{l}_{\hat{\theta}}},$$

where $\hat{\theta}$ is the maximum likelihood estimator (MLE) and $\ddot{l}_{\hat{\theta}}$ is the second derivative of the log-likelihood function evaluated at $\theta = \hat{\theta}$. The (mathematically) equally good approximation

$$\hat{\theta} \pm 1.96/\sqrt{10}$$

(10 being the expected Fisher information), is not correct (Efron and Hinkley 1978).

Fisher's theory is a theory of archetypes. For any given problem the correct inference is divined by reduction to an archetypal form for which the correct inference is obvious. The first and simplest archetype is that of making inferences about θ from one observation x in the normal model

$$x \sim N(\theta, 1). \tag{1}$$

Fisher was incredibly clever at producing such reductions: sufficiency, ancillarity, permutation distributions, and asymptotic optimality theory are among his inventions, all intended to reduce complicated problems to something like (1). (It is worth noting that Fisher's work superseded an earlier archetypical inference system, Karl Pearson's method of moments and families of frequency curves.)

Why is so much of applied statistics carried out in a Fisherian mode? One big reason is the *automatic nature* of Fisher's theory. Maximum likelihood estimation is the original jackknife, in Tukey's sense of a widely applicable and dependable tool. Faced with a new situation, the working statistician can apply maximum likelihood in an automatic fashion, with little chance (in experienced hands) of going far wrong and considerable chance of providing a nearly optimal inference. In short, he does not have to think a lot about the specific situation in order to get on toward its solution.

Bayesian theory requires a great deal of thought about the given situation to apply sensibly. This is seen clearly in the efforts of Novick (1973), Kadane, Dickey, Winkler, Smith, and Peters (1980), and many others to at least partially automate Bayesian inference. All of this thinking is admirable in principle, but not necessarily in day-to-day practice. The same objection applies to some aspects of
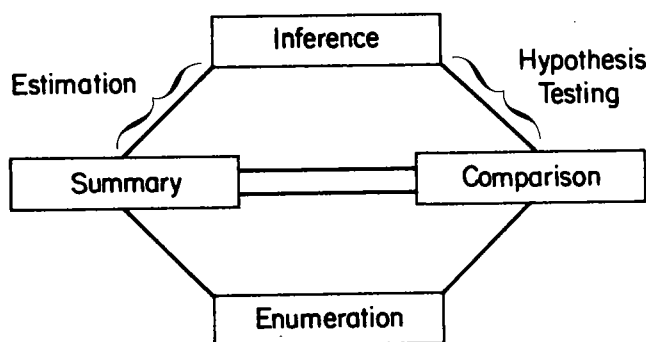
Figure 1. Four Basic Statistical Operations and How They Relate to Estimation. Source: Efron (1982b, fig. 2).

NPW theory, for instance, minimax estimation procedures, and with the same result: they are not used very much.

Not all of statistics is inference. The little diagram of all of statistics in Figure 1 (reprinted from Efron 1982b) starts at the bottom with "enumeration," the collecting and listing of individual datum. The diagram proceeds upward to the reduction of the raw data to more understandable form through the adversarial processes of summary and comparison. This is the part of the analysis where, usually, the statistician decides on a reasonable probabilistic model for the situation. At the top of the diagram is inference. This is the step that takes us from the data actually seen to data that might be expected in the future.

Bayesian theory concentrates on inference, which is the most glamorous part of the statistical world, but not necessarily the most important part. Fisher paid a lot of attention to the earlier steps of the data analysis. Randomization for instance, and experimental design in general, is a statement about how data should be collected, or "enumerated," for best use later in the analysis. Maximum likelihood is a provably efficient way to summarize data, no matter what particular estimation problems are going to be involved in the final inference (Efron 1982b). The NPW school has also contributed to the theory of enumeration, notably in the areas of survey sampling and efficient experimental design.

Fisher's theory culminated in fiducial inference, which to me and most current observers looks like a form of objective (as opposed to subjective) Bayesianism. I will discuss the problems and promise of objective Bayesianism later, but it is interesting to notice that fiducial inference is alone among Fisher's major efforts in its failure to enter common statistical application. In its place, the NPW theory of confidence intervals dominates practice, despite some serious logical problems in its foundations.

4. THE NPW SCHOOL

Unlike Bayes and Fisher, the NPW school does not insist that there is a correct solution for a given inferential situation. Instead, a part of the situation deemed most relevant to the investigator is split off, stated in narrow mathematical fashion, and it is hoped, solved. For example, the correct Bayesian or Fisherian inference for θ in situation (1) leads directly to the correct inference for $\gamma \equiv 1/(1 + \theta)$, but this is not necessarily the case in the NPW formulation. (What is the uniform minimum variance unbiased estimate of γ?)

The NPW piecewise approach to statistical inference has

been justly criticized by Bayesians as self-contradictory, inconsistent, and incoherent. The work of Savage, de Finetti, and their successors shows that no logically consistent inference maker can behave in such a non-Bayesian way. The reply of the NPW school is that there is no reason to believe that statistical inference should be logically consistent in the sense of the Bayesians, and that there are good practical reasons for approaching specific inference problems on an individual basis.

As an example consider the following problem: we observe a random sample $x_1, x_2, \ldots, x_{15}$ from a continuous distribution F on the real line and desire an interval estimate for θ, the median of F. The experiment producing the x_i is a new one, so very little is known about F.

A genuine Bayesian solution seems difficult here, since it requires a prior distribution on the space of all distributions on the real line. Frequentist theory produces a simple solution in terms of a confidence interval based on the order statistics of the sample,

$$\theta \in [x_{(3)}, x_{(12)}]$$

with probability .963, no matter what F may be. The fact that this solution, unlike a Bayesian one, does not also solve the corresponding problem for say $\phi \equiv 50\%$ trimmed mean of F does not dismay the frequentist, particularly if a satisfactory Bayesian solution is not available.

The Bayesian accusation of incoherency of the frequentist cuts both ways: in order to be coherent Bayesians have to solve all problems at once, an often impossible mental exercise.

As another example consider "rejecting at the .05 level." The inconsistencies of this practice are well documented in the Bayesian literature (see Lindley 1982). On the other hand it is one of the most widely used statistical ideas. Its popularity is founded on a crucial practical observation: it is often easier to compare quantities than to assign them absolute values. In this case the comparison is between the amount of evidence against the null hypothesis provided by different possible outcomes of the data. For testing H_0: $x \sim N(0, 1)$ versus H_1: $x \sim N(2, 1)$, we know that a larger observed x provides greater evidential value against H_0 and in favor of H_1, even if we cannot absolutely quantify "evidence."

A Bayes solution to this problem, "the aposteriori odds ratio is 7 to 1 in favor of H_1," is more satisfactory than "the data are significant at the .05 level," but it also requires more input. In fact, it tacitly implies that we have assigned an *absolute* measure of evidence to every possible outcome. Absolute here means that the meaning of 7 to 1 is the same no matter what experiment it came from. [Good's (1965) Bayes–non-Bayes compromise suggests using Bayesian ideas in a comparative mode, but this is the only example I know.]

The heart attack decision tree (Fig. 2) illustrates another difficult situation for the honest Bayesian. The tree purports to predict coronary patients with high risk of dying (population 2) on the basis of variables observed at hospital admission. A series of dichotomous observations are made for example, high or low kinase level, which result in a final prediction. The nodes marked "2" on the tree predict death. Of the 389 patients classified by the tree, only 1 out
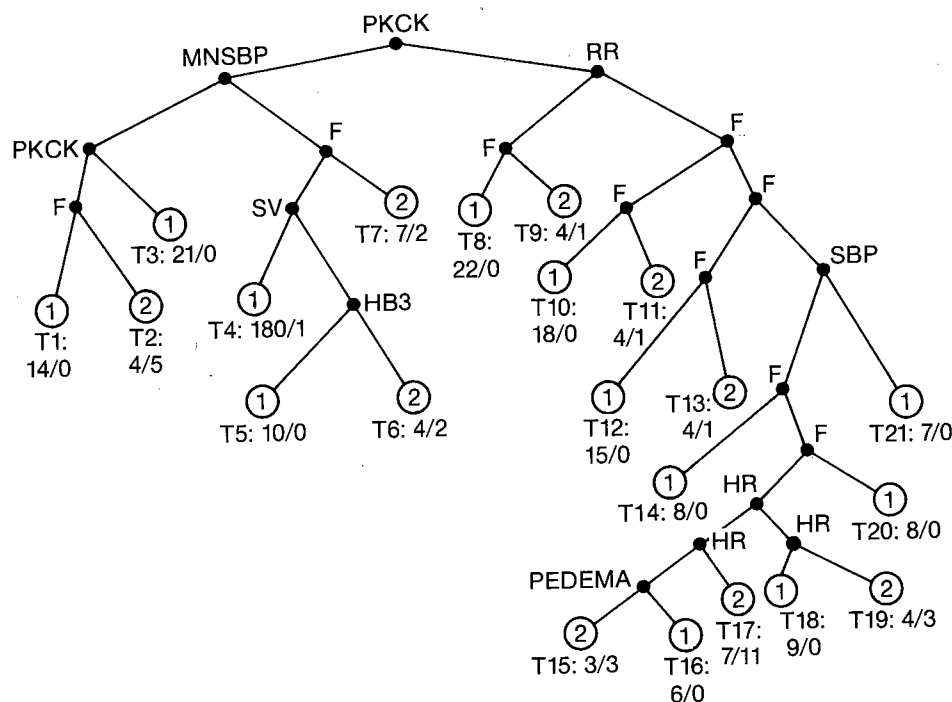
Figure 2. A Decision Tree for Classifying Heart Attack Patients Into Low Risk of Dying (population 1) or High Risk of Dying (population 2). Smaller values of the decision variables go to the left. Circled numbers at terminal nodes indicate population prediction. For example, 6 of the 389 patients in the training set end up at T6, 4 from population 1, 2 from population 2; these patients would all be predicted to be in population 2. Abbreviations: PKCK, peak creatinine kinase level; MNSBP, minimum systolic blood pressure; SBP, systolic blood pressure; FF, respiration rate; HR, average heart rate; SV, superventricular arrhythmia; HB3, heart block 3rd degree; PEDEMA, peripheral edema; F, Fisher linear discriminant function, differing from node to node. Source: Efron (1982a, fig. 7.1).

of 30 deaths was misclassified, that is, predicted to live. Can we believe that the tree has 96.7% probability of successfully predicting deaths?

Because the medical investigators had little prior knowledge of the situation, the tree was constructed by an elaborate data-fitting procedure, which in fact was designed to maximize the apparent success rate. At each stage the dichotomous variable to be used and the splitting point defining "high" or "low" were chosen to give the maximum apparent difference between populations 1 and 2. A bootstrap analysis, much like a cross-validation, gave an unbiased estimate of successful prediction of death of about 70%, rather than 96.7%, for this tree. (Details appear in sec. 7.6 of Efron 1982a.)

The fact that the observed data were used to construct the tree, and how they were used, makes no difference to the Bayesian, since it has no effect on the likelihood function. This is similar in spirit to the fact that the stopping rule used in a sequential procedure has no Bayesian consequence. It makes a world of difference to the frequentist. If exactly the same tree had been constructed by a less flexible rule, the unbiased estimate would move closer to the observed value 96.7%. This is incoherent behavior. The Bayesian estimate, whatever it is, would not change.

"Ad hoc" is a pejorative adjective in Bayesian descriptions of frequentist statistics. On the other hand, ad hoc reasoning produces a reasonable answer here, in a problem that seems far too complicated for a full Bayesian solution. The right to split off the simple part of a complicated inference problem should not be the exclusive property of the frequentists, but I am not aware of much Bayesian activity

along these lines. The coherency approach of Savage and de Finetti seems to have discouraged it. (For a counterexample to this statement see Boos and Monahan, in press.)

The NPW school invented decision theory, but it is not decision theory that separates them from the Bayesians. In fact, Bayesians have made good use of decision theory. The parting of the ways occurs on the crucial issue of θ averages, expectations taken with the state of nature θ fixed. In other words, frequentist calculations. Controlling, or at least computing, θ averages is central to the NPW approach and irrelevant to the Bayesians. This brings us to the topic of objectivity, in my opinion the linchpin of non-Bayesian success with statistical practitioners.

5. OBJECTIVITY

So far I have been careful not to define the kind of Bayesian theory under criticism. The dominant Bayesian school, and the one with the legitimate claim to philosophic coherency, is the subjective Bayesianism of de Finetti and Savage. Now by definition one cannot argue with a subjectivist, so I will just state the obvious fact: though subjectivism is undoubtedly useful in situations involving personal decision making, for example, business and legal decisions, it has failed to make much of a dent in scientific statistical practice. The nature of scientific communication makes me doubt that it ever will.

"Scientific objectivity" is more than a catch-phrase. Strict objectivity is one of the crucial factors separating scientific thinking from wishful thinking. Complete objectivity about one's own work is a little much to expect from a human being, even a scientist, but it is not too much to expect from

one's colleagues. A prime requirement of any statistical theory intended for scientific use is that it reassures oneself *and others* that the data have been interpreted fairly.

With this in mind, it is not surprising that intuitively "fair" statistical ideas, like unbiasedness, confidence intervals, and .05 significance, are immensely popular with the statistical public. Ideas that do not pass the test of objectivity are not much used. This includes NPW ideas as well as Bayesian ones, for example, James–Stein estimation. (An interesting borderline case, which could go either way depending on how it develops, is robust estimation.)

Of course there is no scientific law that says that objectivity must be interpreted in a frequentist sense, and in fact, there is another line of Bayesian thought that attempts to deal directly with the issue of fairness. I call this "objective Bayes theory" to differentiate it from the Savage–de Finetti approach. Bayes and Laplace were objective Bayesians, and in this century, Jeffreys (1961) wrote a famous book on the subject. The goal of objective Bayesianism is to produce prior distributions that capture the idea of objectivity.

Consider situation (1) again. The obvious prior here is the improper one, spreading probability mass for θ uniformly from $-\infty$ to $+\infty$, often denoted simply by "$d\theta$." A Bayesian using this prior obtains good frequentist results. The central 90% aposteriori interval for θ, for example, agrees exactly with the standard 90% confidence interval.

Next consider the situation where we observe

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, I \right), \qquad (2)$$

which is just two independent copies of (1). In some recent work it was necessary for me to make inferences about the parameter $\lambda = \theta_1 \theta_2$. It seemed intuitively reasonable, and objective, to use the improper prior $d\theta_1 \, d\theta_2$, which spreads probability mass for (θ_1, θ_2) uniformly over the entire plane.

As Table 1 shows, the aposteriori central 90% probability interval for λ derived from the prior $d\theta_1 \, d\theta_2$ does not have good frequentist properties. For values of (θ_1, θ_2) in the first quadrant, it gives overly low probabilities of missing λ on the left and overly high probabilities of missing λ on the right. From a frequentist viewpoint we have not been very objective at all, having biased the interval estimates toward the origin.

Recently Charles Stein has given a method of constructing priors that have better frequentist properties (Stein 1982).

For the parameter λ, Stein's theory suggests using the improper prior $(\theta_1^2 + \theta_2^2)^{1/2} \, d\theta_1 \, d\theta_2$, and Table 1 shows that this does indeed give better frequentist coverage probabilities. [In fact, these figures were derived for the bias-corrected percentile intervals, sec. 10.7 of Efron (1982a), and it was then verified that these intervals were almost the same as Stein's.]

The point of the example is that the theory of Bayesian objectivity cannot be a simple one. The correct objective prior seems to depend on which parameter we want to estimate. In higher dimensions when we have several parameters rather than just two, these problems become acute (Efron 1982b). This does not mean that the situation is hopeless. Even a partial solution to the problem of Bayesian objectivity would likely be a valuable contribution to statistical theory and practice. As a hopeful prototype, the Bayesian explanation of the James–Stein estimator has deepened our understanding of this potentially wonderful tool. The whole subject of empirical Bayes can be thought of as an exercise in Bayesian objectivity—trying not to put more information than necessary into the prior—and more progress in this area can be expected.

6. SUMMARY

A summary of the major reasons why Fisherian and NPW ideas have shouldered Bayesian theory aside in statistical practice is as follows:

1. Ease of use: Fisher's theory in particular is well set up to yield answers on an easy and almost automatic basis.

2. Model building: Both Fisherian and NPW theory pay more attention to the preinferential aspects of statistics.

3. Division of labor: The NPW school in particular allows interesting parts of a complicated problem to be broken off and solved separately. These partial solutions often make use of aspects of the situation, for example, the sampling plan, which do not seem to help the Bayesian.

4. Objectivity: The high ground of scientific objectivity has been seized by the frequentists.

None of these points is insurmountable, and in fact, there have been some Bayesian efforts on all four. In my opinion a lot more such effort will be needed to fulfill Lindley's prediction of a Bayesian 21st century.

Table 1. Theoretical Versus Actual Probability of Not Covering $\lambda = \theta_1\theta_2$ Using the Central 90% Aposteriori Interval Based on the Improper Prior $d\theta_1 \, d\theta_2$

(θ_1, θ_2)	Theoretical	
	.050	.050
(0, 0)	002 (006)	002 (006)
(2, 2)	023 (036)	087 (065)
(3, 3)	029 (041)	074 (053)
(4, 4)	031 (046)	070 (047)
(5, 5)	036 (048)	064 (043)
(1, 10)	045 (053)	051 (053)

NOTE: Figures in parentheses are, essentially, the same probabilities using the improper prior $(\theta_1^2 + \theta_2^2)^{1/2} \, d\theta_1 \, d\theta_2$.

REFERENCES

Boos, D., and Monahan, J. (in press), "The Bootstrap for Robust Bayesian Analysis: An Adventure in Computing," *14th Symposium on the Interface.*

Efron, B. (1982a), *The Jackknife, the Bootstrap, and Other Resampling Plans,* CBMS-NSF Monograph 39, Philadelphia: SIAM.

—— (1982b), "Maximum Likelihood and Decision Theory," *Annals of Statistics,* 10, 340–356.

Efron, B., and Hinkley, D. V. (1978), "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information," *Biometrika,* 65, 457–487.

Good, I. J. (1965), *The Estimation of Probabilities: An Essay on Modern Bayesian Methods,* Cambridge: MIT Press.

Jeffreys, H. (1961), *Theory of Probability* (3rd ed.), Oxford: Clarendon Press.

Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S., and Peters, S. C. (1980), "Interactive Elicitation of Opinion for a Normal Linear Model," *Journal of the American Statistical Association*, 75, 845–854.

Lindley, D. V. (1975), "The Future of Statistics—A Bayesian 21st Century," *Supp. Adv. Appl. Prob.*, 7, 106–115.

—— (1982), "Scoring Rules and the Inevitability of Probability," *ISI Review*, 50, 1–26.

Novick, M. (1973), "High School Attainment: An Example of a Computer-Assisted Bayesian Approach to Data Analysis," *ISI Review*, 41, 268–271.

Stein, C. (1982), "The Coverage Probability of Confidence Sets Based on a Prior Distribution," Technical Report 180, Stanford University.

Comment

HERMAN CHERNOFF*

This discussion of Efron's elegant paper reminds me of the story of the reaction of a jealous mathematician to a new theorem. "This theorem is irrelevant. Moreover, it is wrong. Besides, I derived it long ago." I wish (a) to add some reasons why not everyone is a Bayesian, (b) to explain why everyone is, should be, or will soon be a Bayesian, and (c) to claim that (objective) Bayesianism is wrong.

1. WHY NOT EVERYONE IS A BAYESIAN

Teaching

Statistical practice changes slowly because teaching of elementary statistics changes slowly. Most statistical practice is at a rather elementary level, although frequently applied to complex problems. It is dedicated to communicate with people with meager statistical training and constrained by conventions in elementary texts. Therefore, Bayes's theorem is relatively abstract and less intuitive than applications of proportions, means, and standard deviations.

Once sociologists and physicians have learned about significance levels well enough to use them, a major reorganization of the thought processes is required to adapt to decision theoretic or Bayesian analysis. Guttman (1978) has railed for years about the stargazing habits (associated with significance levels .01 and .05) of his fellow sociologists, which he attributes to poor teaching on the part of the statistical profession.

Practice Versus Theory

The popularity of Fisherian theory stems in part from the byproducts of the analysis suggested by the theory. Analysis of variance is one of the most important data analytic tools in statistics. This tool stems from a hypothesis-testing formulation that is difficult to take seriously and would be of limited value for making final conclusions. Its importance stems from the fact that most scientific and statistical practice is concerned, not with grand final conclusions, but with many small steps gradually contributing clarity and order in moderately confused situations. Theory serves as a guide and enhances intuition, and it should not be used as a prescription requiring optimal behavior. Fisherian theory provides relatively simple, effective tools, the robustness of

which seems apparent in circumstances in which the Bayesian and decision theorist may find it difficult to operate.

By the time clarity is attained, there is little additional need for sophisticated statistics or for statisticians. By then the issues are clear, the appropriate experiments are evident, the noise factors are reduced to minor significance, and the conclusions are obvious to any intelligent *observer*. At this point, it is usually forgotten that a statistician ever played a useful role.

2. WHY EVERYONE IS, SHOULD BE, OR WILL BE A BAYESIAN

How I Am a Bayesian

In doing applied statistics, I feel relatively uninhibited and engage in many practices that might be frowned on by careful dogmatists. With the help of theory, I have developed insights and intuitions that, I believe, prevent me from giving undue weight to generalizations drawn from excessive data dredging or other forms of statistical heresy. This feeling of ease and freedom, however, does not exist until I have formulated some decision theoretic and Bayesian view of the problem. Until then, there is discomfort and the feeling that the problem is not well stated or understood. In this sense, I am a Bayesian and decision theorist in spite of my use of Fisherian tools. Is this a unique position or are most of us closet Bayesians and decision theorists in this sense?

When Bayesianism Is the Only Way

Whereas most applications of statistics are to small science and technology, there are cases in which great issues are involved. Makers of public policy are sometimes forced to make decisions on issues for which the scientific foundations are unclear and fundamental data are lacking (e.g., acid rain, ozone depletion, the safety of nuclear power, and carcinogenic effects of diesel fumes). Decisions made now in considerable ignorance have important consequences. Delaying such decisions is also potentially costly. Should we rely on the vagaries of coin tossing, or uninformed public discussions, or the possibly slightly better scientific analysis? But scientific analysis in nuclear safety may, for lack of any real alternative, require a Bayesian fault tree analysis with priors based more on imagination than on real data. One hopes that numbers so derived will not have undue influence on a public impressed by quantification.

*Herman Chernoff is Professor, Department of Statistics, Harvard University, Cambridge, MA 02138.

3. BAYESIANISM IS WRONG

I shall not review the standard objections to Bayesian practice. Bayesian philosophy is rather well founded. But this foundation does not support objective Bayesianism. Agreement with frequentist theories may be interesting but is no justification. An attempt at an axiomatization of objective ignorance leads to the Laplace equally likely prior, to which there are serious objections (Chernoff 1954).

4. CRITICAL COMMENTS

It is difficult to agree with the notion that empirical Bayes is an objective form of Bayesian practice.

It is not clear what a Bayesian estimate in Efron's heart attack example will be. The 96.7% estimate is suggested by naive frequentist intuition. The Bayesian estimate would be based on a prior that would affect the construction of the tree and hence should not be separated from the choice of tree.

Efron's discussion of the 5% significance level test is unsatisfying on two levels. First, he interprets the strength of the statement about a posteriori odds as a requirement for more input. Certainly the ability to make strong statements is linked to the coherence requirements. But the tacit assumption of coherence demands very little overt input in this problem. Second, the NPW practitioner should be uncomfortable with the use of 5% significance tests for a simple hypothesis against a simple alternative. For example, an observation of $x = 10$ in testing $N(0, 1)$ versus $N(20, 1)$ not only is significant but suggests that the statistician's model is wrong. Who would use a 5% test for testing $N(0, 1)$ versus $N(.1, 1)$ with a single observation? Efron may not have wished to raise the issue of the universality of 5% and 1% significance levels, but it seems to intrude and has not been well addressed.

ADDITIONAL REFERENCES

Chernoff, H. (1954), "Rational Selection of Decision Functions," *Econometrica*, 22, 422–443.

Guttman, L. (1978), "Cyril Burt and the Careless Star Worshippers," technical report, Hebrew University, Department of Sociology, pp. 1–7.

Comment

D. V. LINDLEY*

Judging from the general thrust of the article, I presume that the "everyone" of the title is "every statistician" and the question will be answered in that sense. The answer is simply that statisticians do not know what the Bayesian paradigm says. Why should they? There are very few universities in the world with statistics departments that provide a good course in the subject. Only exceptional graduate students leave the field of their advisor and read for themselves. A secondary reason is that the subject is quite hard for someone who has been trained in the sampling-theory approach to understand. It may be that we Bayesians are poor writers, and certainly the seminal books by Jeffreys (1961) and de Finetti (1974, 1975) are difficult reading, but it took Savage (see the preface to Savage 1972) several years to understand what he had done; naturally, it took me longer. The subject is difficult. Some argue that this is a reason for not using it. But it is always harder to adhere to a strict moral code than to indulge in loose living.

What most statisticians have is a parody of the Bayesian argument, a simplistic view that just adds a woolly prior to the sampling-theory paraphernalia. They look at the parody, see how absurd it is, and thus dismiss the coherent approach as well. Efron has studied the Bayesian argument more than have most statisticians, but it is still only a parody that is presented in this article. Many of the arguments he produces are distortions of the thing he is attacking.

It is not true that "everyone used to be a Bayesian." Laplace would use any sensible argument that came along, and it is often hard to see whether it is Bayesian, in the modern sense of the term, or not. It would be nearer the truth to say "everyone used to use some Bayesian ideas." But this is true today: we have all heard of Bayes procedures—they are admissible.

The advantage of Fisherian methods being automatic is correctly emphasized. But Bayesian methods are even more automatic. What is unknown and of interest? X. What is known? H. Calculate $p(X|H)$. How? The probability calculus is the only tool. The Bayesian argument provides a recipe (Lindley 1983). Of course, you have to think, to link reality with the mathematical model. Sampling-theory statistics takes place in a Greek hinterland (see below) that diminishes this connection with reality. It is surprising to find Efron defending Fisherian ideas when they have been so carefully investigated and found inadequate by Basu (1975, 1977, 1978). Of course, sampling theorists do not read this brilliant, lucid writer. His results discomfit them.

It is not true that "Bayesian theory concentrates on inference." If it concentrates on anything, it is decision analysis and has often been unjustly accused of marketplace philosophy. It embraces all of the topics mentioned, including randomization and experimental design. It is a way of "thinking about things"; it is relevant to everyone (Lindley 1985).

A gross parody is provided in the discussion of the heart attack decision tree. "The fact that the observed data were used . . . makes no difference to the Bayesian, since it has no effect on the likelihood function." But we Bayesians

have a prior, remember? The discussion of this example is bedevilled by the fact that we do not have a neat way of handling multivariate distributions in the unit cube of chances, so let me illustrate by a similar case. In a trial of n varieties to select the best, the merit of the variety that did best in the trial is exaggerated if judged by the sample mean, just as the 96.7% exaggerates the worth of the diagnostic procedure. But a reasonable prior compensates for this and "shrinks" (to use the popular phrase) the merit of the apparently best variety toward the common mean. Educational psychologists and actuaries have been Bayesians in this regard since the 1920s. Efron knows this. Why does he not mention it?

I should have thought that any serious student of the literature would have said that the objective Bayesian view is not doing very well at the moment and that it is the subjective approach that is developing. It is not true that "strict objectivity is one of the crucial factors separating scientific thinking from wishful thinking." The objective element is the data: interpretation of data is subjective, as anyone who has interacted with scientists knows. Furthermore, the Bayesian view accepts the data, whereas the sampling-theory approach has to make a subjective embedding of them in a sample space.

The last example with $\lambda = \theta_1 \theta_2$ is typical of a sampling theorist's impractical discussions. It is full of Greek letters, as if this unusual alphabet was a repository of truth. The situation presumably referred to real things, and the discussion is almost pointless without them. Even if it were

useful, why should we consider the particular class of repetitions used here and unstated? Lindley and Phillips (1976) show how any level can be produced, and I expect that any coverage probability could be managed. Why is a choice of sample space objective? A Bayesian, faced with repetitions, would learn as he repeated. Thus the repetitions cease to be identical.

In short, this article is an attack on a parody of a serious argument, and because it is a parody it is easily abused. Perhaps the author has been falling over all those bootstraps lying around. Every statistician would be a Bayesian if he took the trouble to read the literature thoroughly and was honest enough to admit that he might have been wrong.

ADDITIONAL REFERENCES

Basu, D. (1975), "Statistical Information and Likelihood," *Sankhya*, Ser. A, 37, 1–71.
——— (1977), "On the Elimination of Nuisance Parameters," *Journal of the American Statistical Association*, 72, 355–366.
——— (1978), "On Partial Sufficiency: A Review," *Journal of Statistical Planning and Inference*, 2, 1–13.
De Finetti, B. (1974), *Theory of Probability* (Vol. 1), London: John Wiley.
——— (1975), *Theory of Probability* (Vol. 2), London: John Wiley.
Lindley, D. V. (1983), "Theory and Practice of Bayesian Statistics," *The Statistician*, 32, 1–11.
——— (1985), *Making Decisions*, London: John Wiley.
Lindley, D. V., and Phillips, L. D. (1976), "Inference for a Bernoulli Process (a Bayesian View)," *The American Statistician*, 30, 112–119; Discussion (1977), 31, 181–183.
Savage, L. J. (1972), *The Foundations of Statistics*, New York: Dover Publications.

Comment

C. N. MORRIS*

I do not know whether there will be a Bayesian 21st century, but Efron (1978) once gave that proposition probability .15. Actually, what Bayesian analysis is, and hence the answer to this question, depends on the Bayesian with whom one talks. There seems to be increasing acceptance of Bayesian ideas. Even if this continues, however, the 21st century certainly will not ignore or discard the enormous gains made by the frequency approach to statistics. For example, in nonparametric problems and in randomization, Bayesian theory has had relatively little to contribute. Rather, a frequency–Bayes compromise seems likely. Bayesian theory does contribute much to inference for parametric models. If by a "Bayesian 21st century" one means that all statisticians will be taught something of Bayesian ideas and applications in combination with the frequentist approach they now learn, then the statistics profession will gain, for all statisticians will possess a powerful tool not now available to many of them. For example, Bayes's theory, which is the correct theory for updating statistical information, often suggests the appropriate analysis, even if that analysis is

*C. N. Morris is Professor, Department of Mathematics and Center for Statistical Sciences, University of Texas, Austin, TX 78712.

later justified in repeated-use frequentist terms. Furthermore, statistics serves many purposes, and the frequency, objective Bayesian, and subjective Bayesian viewpoints are all needed for the variety of situations that arise.

We all have faced many inference problems for which we hold vague prior information about the parameter, so vague that we readily abandon it when presented with sharper information from the experiment. In such cases we do not attempt to update our prior distribution; we simply take directly the likelihood $L_y(\theta)$ as our posterior distribution for θ after observing y. Or we might improve this slightly by multiplying $L_y(\theta)$ by a factor $\pi(\theta)$ to account for the form of θ, perhaps approximating π by Jeffreys's prior $\pi(\theta) \propto i^{.5}(\theta)$, where $i(\theta)$ is Fisher's information for θ, and claiming that $L^*(\theta) = L_y(\theta)\pi(\theta)$ is proportional to the posterior density for θ. For parametric problems with known likelihood and no prior information to be accounted for, I find the claim, ascribed to Jeffreys, that $L^*(\theta)$ is the posterior density for θ to be the best jackknife of parametric statistics. It explains when the MLE is reasonable [if the mode is near the center of $L^*(\theta)$]; when the MLE is nearly normal [if $L^*(\theta)$ is nearly normal], and what to do with nuisance parameters (integrate them out). It tells us to use observed

information, not expected information, to approximate variances (the Efron–Hinkley result), and it yields inferences conditional on the data. Efron's $\theta_1\theta_2$ estimation example reminds us that this method does not work perfectly in multiparameter problems; but for a range of parametric problems, I believe this objective Bayesian method generally will outperform Fisher's maximum likelihood paradigm, or any other automatic frequentist competitor.

Empirical Bayes inference, to which Efron has contributed significantly, provides a perspective from which the frequency–Bayes controversy may be viewed. Empirical Bayes (Morris 1983) involves *two families* of distributions: one, $f_\theta(y)$, for the data Y given the parameters $\theta \in \Theta$, and another family, Π, of possible distributions for the parameters. If risk evaluations of a procedure t are to be made, they are to include (at least) the double integral expectation of the loss function over both the data and the parameters for each distribution, $\pi \in \Pi$, thus producing risk function $r(\pi, t)$. The key is that Π is assumed known. In empirical Bayes modeling, just as when choosing a family of sampling distributions, the statistician ordinarily would limit Π to be substantially smaller than all possible distributions, perhaps choosing a parametric family of distributions. Then the data are used to learn about the correct π in Π via the marginal likelihood $f_\pi^*(y)$ of the data.

Some Bayesians will object to the frequency-like empirical Bayes averaging with respect to the data. Such averages, however, reassure statisticians developing a procedure or program for repeated use because these averages provide a standard of objectivity by calibrating the procedure over a range of applications. Rubin (1984) elaborates on these points. Of course, Bayesians are right in that the derived standards do not necessarily apply to the procedure for any particular application or data value.

The empirical Bayes model shows us that Bayesian and frequentist statistics include only the two most extreme cases of all possible models for the parameters in a statistical analysis. Bayesian theory is the case for which the statistician can limit Π to have exactly one member. Frequentist theory represents the other extreme, when Π corresponds to all possible distributions on Θ. [Then all "atom" distributions on Θ, each putting point mass at some $\theta \in \Theta$, are in Π, and the double integral risks $r(\pi, t)$ for these atom distributions become the single integral risk functions $R(\theta, t)$ used in frequentist evaluations.] There are practical problems for which intermediate information about Π can be identified, however, neither at the Bayesian nor the frequentist extreme. Since the data then can be used to evaluate the adequacy of this choice via the marginal likelihood, a Bayes–frequency compromise can be achieved via the empirical Bayes model. Thus it seems inappropriate that statistics in the 21st century should settle exclusively at either pole, Bayesian or frequentist.

Bayesian statistics provides the cornerstone for empirical Bayes theory. For example, consider the simple case of estimating k normal means from k sufficient statistics $Y_i \sim N(\theta_i, V_i)$ independently, $i = 1, 2, \ldots, k$, with $V_i = \text{var}(Y_i)$

known. Then a rule of the form $\hat{\theta}_i = (1 - B_i)Y_i + B_i\mu$ is desired, with B_i shrinking Y_i toward a guess at the prior mean of the θ_i and with B_i larger for large V_i than for small V_i. The frequentist minimax formulation, which so successfully produced Stein's estimator in the "equal variances case," $V_i = V$ for all i, gives the incorrect shrinking pattern for unequal variances by requiring B_i to decrease, not increase, as V_i increases. Bayesian theory, assuming exchangeable prior distributions with $\theta_i \sim N(\mu, A)$ independently, $A = \text{var}(\theta_i)$, suggests the quite reasonable shrinkage pattern $B_i = V_i/(V_i + A)$.

Bayesian methods also provide the only reasonable answer to the question of accuracy of shrinkage estimators, revealing in the equal variances case the variability $s_i^2 = \text{var}(\theta_i|\text{data})$ of the estimator $\hat{\theta}_i = (1 - \hat{B})Y_i + \hat{B}\bar{Y}$ to be

$$s_i^2 = V(1 - \hat{B}) + V\hat{B}/k + v(Y_i - \bar{Y})^2$$

as the (posterior) variance of θ_i given the data (Morris 1983). Here $\hat{B}$ is the estimated shrinkage factor of the common value $B = V/(V + A)$. The first term in s_i^2 is an estimate of the residual variance $V(1 - B)$ of the Bayes rule; the second is a penalty for not knowing the prior mean μ and instead using $\bar{Y}$; and the third is an additional penalty for not knowing the prior variance A and therefore using $\hat{B}$ instead of B. Using $\hat{B}$ introduces the error $(\hat{B} - B)(Y_i - \bar{Y})$, which has variance, given the data, $v(Y_i - \bar{Y})^2$, $v = \text{var}(B|\text{data})$. This formula for s_i follows easily from Bayesian moment calculations, but frequentist methods have given no hint at an appropriate value. Thus it is an excellent example not only of how Bayesian reasoning provides a way for modeling distributions for parameters, as empirical Bayes does, but also of how it leads to the proper form for parametric procedures. Paradoxically, however, frequentist calculations, not Bayesian calculations, have suggested the appropriate noninformative distribution for the prior variance A needed to complete these calculations [taking A to be uniformly distributed on $(0, \infty)$ works well].

To conclude, practical statisticians encounter a variety of problems, and frequency, objective Bayes, subjective Bayes, and empirical Bayes methods provide a range of possible responses. There can be no clear victory for any approach for all applications; rather, we should train statisticians for a frequency–Bayes compromise so that they can more flexibly respond to new situations. In the context of the title's rhetorical question, even if everyone is not a Bayesian, statisticians need greater exposure to Bayesian ideas than they now receive.

ADDITIONAL REFERENCES

Efron, B. (1978), "Controversies in the Foundations of Statistics," *American Mathematical Monthly*, 85, 231–246.

Morris, C. N. (1983), "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78, 47–65.

Rubin, D. (1984), "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician," *Annals of Statistics*, 12, 1151–1172.

Comment

S. JAMES PRESS*

The title of this paper is so provocative, and the content is so well handled, that I could not resist Efron's invitation to add my own comments.

First, I must point out that I have also attempted to address this question (Press 1985, sec. 2.6.1). The answers I have offered in that monograph focus on three areas: (a) objectivism versus subjectivism; (b) choice of the "public policy prior"; and (c) the vested interest some statisticians have in frequentist methodology.

The first two of my reasons have also been given by Efron; so we agree that these two issues are part of why everyone is not (yet) a Bayesian.

Efron points out in his summary (Sec. 6) that "the high ground of scientific objectivity has been seized by the frequentists." I agree with this conclusion, although I deplore the underlying rationale. It is my view that the Bayesian paradigm admits both objectivity and subjectivity, with the former appropriate for many situations involving scientific reporting and public policy decision making and the latter appropriate for all other situations. There are subjective, or personalistic, prior distributions that must be assessed for the subjective inference or decision making problem, and there are "vague" or "indifference" priors for the objective inference or decision making problem. But because so much attention has been focused on the subjective aspects of Bayesian statistical procedures, we have tended to lose sight of the objectivity inherent in the approach, which was how Laplace saw the paradigm in the early 1800s, and which is how many Bayesian statisticians of the 20th century see the objectivity problem (e.g., see Jaynes 1983 and Jeffreys 1961).

Of course there are questions about which prior to use to express objectivity. Efron seeks the best method for accomplishing this. I applaud him for focusing on an issue that is so important to the field of statistics. Bayesians will traditionally use the improper uniform distribution for unknowns lying on the whole real line (or they will use the uniform distribution for the logs of the unknowns when they lie on the positive axis). But we may be able to do better with an entirely new probability system. In Press (1985, sec. 1.2.1) I assert:

It is clear that our field is in a transitional state. What is needed is a formal, axiomatic theory of probability which is conditional in the sense of Renyi, and which permits probabilities on the entire real line.

In Section 3 Efron claims that "Bayesian theory concentrates on inference," although he acknowledges in Section 5 that "subjectivism is undoubtedly useful in situations involving personal decision making, for example, business and legal decisions." I must point out that the basic statistics course now typically presented in many business schools across the country is a course in Bayesian decision making. Students are taught that much public policy decision making on a day-to-day basis does not involve NPW-type analysis;

there is no time to collect any data; the decisions are made subjectively, using formal or informal priors. When data are collectible, students are taught to make decisions in a Bayesian way. Many well-read articles and books attest to this fact (e.g., see Pratt, Raiffa, and Schlaifer 1964, Raiffa 1968, Raiffa and Schlaifer 1961, Schlaifer 1969, and Winkler 1972). But because Bayesian decision making involves averaging only over the prior, or only over the posterior density conditional on the observed data, and not averaging over observations not yet taken (the entire sample space), and (therefore) because Bayesians are generally not concerned with admissibility, it is assumed that Bayesian procedures are largely inferential rather than decision theoretic. This impression should be revised.

In Section 3 Efron discusses the automated nature of Fisher's theory, pointing out that the MLE "is the original jackknife, in Tukey's sense of a widely applicable and dependable tool." In the same sense that maximizing the likelihood functions is mechanical, so is the simple application of the Bayesian paradigm. In either the frequentist or the Bayesian approach, however, the statistician cannot use his tools mindlessly and mechanically. If he merely wishes to turn a crank, the frequentist can summarize his data with a likelihood function, and the Bayesian can, analogously, use a vague prior and summarize his beliefs with the induced posterior density. But what frequentist does not have to approach his problem gingerly when he knows in addition that $2 \leq \theta \leq 5$, and in the context of the problem, $\theta > 3$ makes little sense! The careful statistician does not apply his procedures mechanically, be he frequentist or Bayesian. The Bayesian has a formal introspection procedure for introducing his prior information, however, which for many problems is a great advantage. As Efron implies, however, automating the procedure of Bayesian analysis (in the subjective case) has proved to be difficult.

Finally, I would like to comment on the hypothesis testing problem. Efron's interpretation of Bayesian hypothesis testing (Sec. 4) "tacitly implies that we have assigned an *absolute* measure of evidence to every possible outcome." There are several methods of testing hypotheses from a Bayesian point of view (just as the frequentist might use likelihood ratio testing, union-intersection methods, robust procedures, etc.). I will restrict myself to the use of the posterior odds ratio procedure (Jeffreys 1961, chapters 5 and 6; see also Zellner 1971). This procedure does not involve "every possible outcome"; it involves only the observed data. Specifically, the log of the posterior odds is the sum of the logs of the prior odds and the likelihood ratio, and the likelihood ratio depends only on the observed data. Efron's use of the term "absolute measure of evidence" begs the question, disputed among Bayesians, of whether experimental design parameters should be included in the prior. Should the posterior odds ratio differ depending on the nature of the experiment? For example, Bayesian purists claim we should make the same posterior inferences

*S. James Press is Professor, Department of Statistics, University of California, Riverside, CA 92521.

about the parameter p in binomial sampling as we do in negative binomial sampling, for which the experiment is different but the p is the same. That is, strict adherence to the "likelihood principle" is what Efron claims for the Bayesian approach. Other Bayesians who would admit experimental design parameters into the prior distribution would have the posterior odds ratio reflect these experimental design differences.

In summary, it is only in actively seeking answers to the basic question Efron has raised that our field can begin to understand the real theoretical and applied advantages of one approach over another and ultimately arrive at a unified foundational approach.

ADDITIONAL REFERENCES

Jaynes, E. (1983), *Papers on Probability, Statistics, and Statistical Physics*, ed. R. D. Rosenkrantz, Dordrecht, Holland: D. Reidel.

Pratt, J. W., Raiffa, H., and Schlaifer, R. (1964), "The Foundations of Decision Under Uncertainty: An Elementary Exposition," *Journal of the American Statistical Association*, 59, 353–375.

Press, S. J. (1985), *Introduction to Applied Bayesian Statistical Inference*, unpublished manuscript.

Raiffa, H. (1968), *Decision Analysis*, Reading, MA: Addison-Wesley.

Raiffa, H., and Schlaifer, R. (1961), *Applied Statistical Decision Theory*, Boston: Harvard University Press.

Schlaifer, R. (1969), *Analysis of Decisions Under Uncertainty*, New York: McGraw-Hill.

Winkler, R. L. (1972), *Introduction to Bayesian Inference and Decision*, New York: Holt, Rinehart & Winston.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley.

Comment

ADRIAN F. M. SMITH*

In a recent review presented at the 150th Anniversary Meeting of the Royal Statistical Society (Smith 1984), I wrote that "Bayesian thinking requires a fairly formal, structured frame of discourse, and is thus not directly applicable to every kind of activity in which statisticians are involved." I am very happy, therefore, to accept those parts of Efron's analysis that are predicated on the recognition that "not all of statistics is inference." This much is surely very common ground.

A great deal of theoretical and applied statistics *is* concerned, however, with analysis in the context of more-or-less formal models, and here Efron himself acknowledges the "powerful theoretical reasons for preferring Bayesian inference" and the "disturbing catalogue of inconsistencies in the frequentist point of view." Why, then, in this inferential context, is everyone not a Bayesian? Efron's answer appears to be that the vast majority of statisticians are committed to an intuitive notion of statistical "fairness," nowhere defined but apparently common to concepts such as unbiasedness and confidence intervals and underpinning the "key" (but also undefined!) concept of "scientific objectivity."

Well, insofar as I understand this notion of "fairness," it appeals to long-run frequencies, violates an equally "intuitive" conditionality principle, and is thus directly responsible for those "disturbing inconsistencies" to which Efron refers. As for "scientific objectivity," I personally am only able to make sense of the concept in the context of a Bayesian philosophy that predisposes one to seek to report, openly and accessibly, a rich range of the possible belief mappings induced by a given data set, the range being chosen to reflect and potentially to challenge the initial perceptions of a broad class of interested parties. If a fairly sharp consensus of views emerges from a rather wide spread of initial opinions, then, and only then, might it be meaningful to refer to "objectivity." In the article previously referred to, I remarked that

one of the most attractive features of the Bayesian approach is its recognition of the legitimacy of a plurality of (coherently constrained) responses to data. Any approach to scientific inference which seeks to legitimize *an* answer in response to complex uncertainty is, for me, a totalitarian parody of a would-be rational human learning process. (p. 250)

An understanding of why this kind of philosophy does not hold sway continues to elude me but seems in any case to lie largely outside the purview of Efron's rather narrowly focused analysis. Maybe it has more to do with Kuhnian scientific cultural continuities? Or computational constraints? Or educational and institutional inertia? Or perhaps most statisticians now *are* Bayesians (when it matters!), but they do not want to spoil the fun by admitting it.

ADDITIONAL REFERENCE

Smith, A. F. M. (1984), "Present Position and Potential Developments: Some Personal Views. Bayesian Statistics," *Journal of the Royal Statistical Society*, Ser. A, 147, 245–259.

*Adrian F. M. Smith is Professor of Mathematical Statistics, Department of Mathematics, University of Nottingham, University Park, Nottingham, NG7 2RD, England.

B. EFRON

This article was actually the text of a talk delivered at a conference on Bayesian statistics held at the Virginia Polytechnic Institute and State University. The question its title asks was not intended to be rhetorical or sarcastic. Despite the considerable philosophical advantages of the Bayesian approach, most scientific data analysis is carried out in a non-Bayesian framework. Why? The talk attempted to answer this question, arguing mainly from a set of examples in which Bayesian analysis is difficult while Fisher/frequentist solutions are relatively easy.

The examples are genuine ones, suggested by my own picaresque adventures as an applied statistician in the Stanford Medical School. For instance, the product-of-normal-means problem arose in the context of comparing two non-nested linear models (Efron 1984). There is nothing particularly striking or unusual about these examples, but of course that was the point of my talk.

Here are a few brief comments on the commentaries. I was surprised at how little support objectivity aroused. Bringing some degree of consensus to the interpretation of noisy data is certainly one of our profession's principal accomplishments. Perhaps, as Smith suggests, we have purchased consensus at a high price in intellectual tyranny, but our scientific clientele seem happy to pay this price.

Understanding the true meaning of objectivity has occupied statistical thinkers from Bayes, Laplace, and Gauss to Fisher, Neyman, and Jeffreys. Fisher's phrase, "the logic of uncertain inference," is particularly evocative of a theory that goes from the data and a family of possible probability models to a consensus agreement of reasonable conclusions.

Such a theory does not yet fully exist, and may never exist, but I hope we have not given up looking for it. In the discussion, only Press had specific good words to say for objectivity. On the other hand, the empirical Bayes approach, very nicely stated by Morris, has strong objective Bayes connections. These connections are made explicit in Good's theory of Type II maximum likelihood (Good 1965).

Lindley makes the serious complaint that I have criticized only a parody of the true Bayesian argument. The heart attack decision tree was particularly offensive. I chose this example because it shows how a simple frequentist technique, cross-validation or the bootstrap in this case, can make headway even in very complicated situations. Devices like cross-validation violate the likelihood principle; they ask what would happen for data sets other than the one actually observed and so are non-Bayesian according to Lindley's strict definitions. What is the "true Bayesian argument" in this example? It is not that I do not think such an argument exists, but Lindley has not given us the slightest idea of what it might be. This is a practical point, not a philosophical one, but practical points are crucial in trying to understand why Bayesian theory is not much used in scientific practice.

Chernoff can be excused a feeling of deja vu, since parts of my talk, in particular the decision tree example, were based on memories of his own lectures. Section 1.b of his discussion is a particularly neat statement of the role of the statistician in scientific practice. It was not my intention to separate decision theory from statistical inference in general (the top box in Fig. 1), as both Chernoff and Press thought I was doing. The relationship between Fisherian and decision-theoretic inferential systems is discussed in Efron (1982b).

My talk was intended as an argument for more Bayesian research, not less. The problems of *practical* Bayesian inference should be given top priority. Morris's discussion is a promising model of how Bayesian theory could be aimed at the actual needs of applied statisticians.

Finally, I must warn Professor Lindley that his brutal, and largely unprovoked, attack has been reported to FOGA (Friends of the Greek Alphabet). He will be in for a *very nasty time indeed* if he wishes to use as much as an epsilon or an iota in any future manuscript.

ADDITIONAL REFERENCE

Efron, B. (1984), "Comparing Non-Nested Linear Models," *Journal of the American Statistical Association*, 79, 791–803.

ADDITIONAL COMMENT BY LINDLEY

Brad Efron asks what *the* true Bayesian argument might be in the heart attack case. *A* (nearly) Bayesian approach was given by David J. Speigelhalter and Robin P. Knill-Jones [*Journal of the Royal Statistical Society*, Ser. A (1984) 147, 35–77, with discussion].