ELSEVIER

Journal of Computational and Applied Mathematics 64 (1995) 3–10

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Bayesian and non-Bayesian approaches to statistical inference and decision-making[1]

Arnold Zellner*

Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637, United States

Received 8 May 1993

Abstract

Bayesian and non-Bayesian approaches to statistical inference and decision-making are discussed and compared. A pragmatic criterion, success in practice, as well as logical consistency are emphasized in comparing alternative approaches. Attention is given to learning models, probability concepts, prior information, axiom systems and selected technical aspects of Bayesian and non-Bayesian approaches.

Keywords: Bayesian inference; Bayesian decision-making; Bayes–non-Bayes foundations

Since there are so many varieties of Bayesians (B's) and non-Bayesians (NB's), it would be presumptuous to claim that the issues between B's and NB's as I see them are one-to-one with those perceived by my colleagues in Statistics, Econometrics and many other fields which employ statistical methods. Thus what I shall present are my own personal views which began to take shape in the early 1960's and developed subsequently in connection with my program of research to evaluate the comparative merits of B and NB statistical approaches, not merely philosophically but also in actual statistical applications. In this connection, the theoretical and applied work of Jeffreys [14, 15] has had a great impact on my thinking and work as will become evident in what follows — see articles by S. Geisser, I.J. Good and D.V. Lindley in [33], in which Jeffreys' contributions to Statistics are discussed. Many others, including participants in our semiannual meetings of the NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics since 1970 and my students have shaped my thinking on B and NB issues. Also, what I have written below has benefited from listening to the fine Bayesian papers presented at this year's Annual

* E-mail: fac_azellner@gsbacd.uchicago.edu.

[1] Invited address presented at the 1992 Annual Conference of the American Statistical Association Harrisburg Chapter, 6 April 1992. Research financed in part by the National Science Foundation and by income from the H.G.B. Alexander Endowment Fund, Graduate School of Business, University of Chicago.

SSDI 0377-0427(95)00002-X

Conference of the ASA Harrisburg Chapter, entitled Applications of Bayesian Methods and from discussions with authors of these papers and other participants.

To begin with, everyone probably recognizes that part of what is done in Statistics, as well as in any other field, is art and part is science. Some examples of art in Statistics include statistical graphics, exploratory data analysis, multivariate model formulation, etc. Also, years ago, and even today, some use the free hand method, a very artful method, to fit straight lines and curves to data. Further, the production of new statistical methods is more an art than a science. On the other hand, much in Statistics is science, that is logically sound, operational procedures that yield reproducible, dependable results. Artful work in Statistics and many other fields is hard to reproduce. For example, an artful forecaster may state that next year's growth rate of the US economy will be 5%. If asked how he arrived at his 5% forecast, he may reply, "Trust my judgment". A second forecaster who forecasts a 2% growth rate may be able to indicate explicitly how he computed his forecast and an interval for it; thus anyone using this forecaster's model, data and other information can reproduce his 2% forecast, a necessary condition, in my opinion, for it to be termed a scientific forecast. Of course, it is impossible to prove deductively which of these two forecasts is better. However, in many fields scientific, reproducible procedures have been shown to be superior in explanation, prediction and problem-solving to informal, artful procedures in a large number of cases and for a wide range of problems and hence our confidence in science and its methods. Further, as Jeffreys [15] emphasizes, there is an unavoidable uncertainty associated with procedures and propositions. It is impossible to prove deductively that a procedure will work well in practice or that a proposition about the world is 100% true. This revelation has great implications for those in Statistics who claim that a particular approach is absolutely correct. What is needed, in my opinion, is evidence that a particular approach is free from logical error and is more effective in applications and treats a broader range of problems successfully than do competing approaches. I consider this pragmatic approach to evaluating alternative statistical philosophies to be fruitful and have pursued it for many years.

When we talk about the science of Statistics or Statistical Science, what do we mean? I take these terms to mean the study, creation and use of methods for producing and employing data for description, measurement, explanation, prediction, control and decision-making. These methods include design of experiments and surveys, model-building, estimation, testing, prediction, graphical analysis, data analysis, and so on. Now different statistical approaches are based on different concepts, procedures and justifications. Clearly, it is important to compare different approaches conceptually and in practice. To let the cat out of the bag, it is my view that the B approach can produce almost all reasonable NB results and others that are difficult, if not impossible to produce using NB methods. As will be seen, it is also the case that B's can use a much broader range of probability statements than is the case for NB's and this capability is most valuable in applications. Finally, B's can justify procedures not only with respect to a single application with given data but also in terms of their "operating characteristics" in relevant repeated trials.

Let me consider a number of specific B and NB issues in relation to the above view of Statistics. First there is the following issue: In B and NB approaches is a formal learning model employed? The answer is that B's use Bayes' theorem as a learning model while NB's appear to learn informally without the explicit use of a learning model. While it is true that the Bayesian learning model, Bayes' theorem has produced fine results in many problems, as with all models there is the possibility that it can be generalized. After all, in the auto industry the Model T Ford was replaced by the Model A, the Ford V8 and later improved models. It is a fact that some researchers are currently working to extend and generalize the Bayesian learning model, Bayes' theorem — see, e.g., [2, 9, 37]. And indeed Jeffreys [15, p. 25] indicates that he has attempted to generalize the product rule of probability. Thus Bayesians have a formal, reproducible learning procedure that combines prior information with new sample information to yield posterior information whereas NB's learn informally in an artful fashion.

A second issue has to do with whether one associates probabilities with hypotheses, models, and propositions. Concepts of probability used by B's are such that it is meaningful to associate probabilities with hypotheses, models, and propositions and make it possible to use Bayes' Theorem to update such probabilities to reflect new sample information and to test hypotheses in a well-defined decision theoretic framework — see, e.g., [15], a pioneering work, and [1, 10, 16, 21, 22, 31, 34, 38] for many examples. NB's who utilize a long-run frequency or other definitions of probability (see [15, Ch. 7] for a critique of various concepts of probability) do not or cannot formally associate probabilities with models, hypotheses and probabilities with models, hypotheses and propositions even though they often do so informally. Further, some NB practitioners associate p-values with degrees of belief in null hypotheses, an error that is often encountered. However, as indicated in [34, Ch. 3.7], for many testing problems with a given sample, the posterior odds expression for a null hypothesis versus a composite alternative hypothesis is a monotonic function of the usual NB p-value which proves a B interpretation of a NB nefarious practice. Since so many users of Statistics put probabilities on all kinds of propositions, hypotheses and models, it is indeed fortunate that the Reverend Thomas Bayes and others have given us formal methods for using and updating such probabilities in a formal, reproducible scientific manner. This is a very important range of considerations because policy-makers often want to know what is the probability that a measurable effect or a future outcome is between a and b, two given numbers. Or scientists and others may wish to have measures of confidence associated with alternative models which can enable them to take account of model uncertainty in making decisions — see, e.g., [13]. That B's who associate probabilities with hypotheses, models and propositions can deal with these important problems in an operational, reproducible manner is indeed fortunate. Also, with such probabilities available, it is possible to average results over alternative models, hypotheses and forecasters — see, e.g., [19, 23, 29, 30, 40].

Third, another issue that arises in B and NB comparisons is whether it is meaningful to introduce a prior distribution for parameters in a statistical analysis. Generally B's do so but NB's generally claim that they do not. Here as Drèze [6], Drèze and Richard [7], Good [11], Zellner [34] and other have pointed out, NB's use much prior information in identifying parameters in such models as the errors-in-the-variables model (EVM), structural econometric models, etc. For example, assuming that the ratio of error term variances is known in the EVM can be regarded as a dogmatic, degenerate prior density for the ratio of variances. Also, as Swamy [25] and Good [11] have emphasized, use of random parameter models by NB's involves the use of an assumed distribution for parameters which is quite like a prior density. Similarly, in a filtering problem, putting a distribution on an initial state is much like the introduction of a prior density. The same can be said for distributions placed on random effects and unobserved trend, cyclical and seasonal components of time series variables. Indeed this controversy about whether to put distributions on parameters resembles the 19th century controversy about whether to introduce unobserved error terms with assumed properties as vividly described in [24]. This range of considerations is for

Good [11] a possible basis for a B/NB compromise. However, there are still the above issues regarding an appropriate concept of probability and whether to associate probabilities with hypotheses, models and propositions.

Fourth, as regards the issue as to whether B's are subjective while NB frequentists are not, it is relevant to point out, as Jeffreys [15, Ch. 7] does, that frequentists cannot prove that long-run frequencies exist; they have to make subjective assumptions which allow them to prove certain convergence theorems which they cannot check empirically since it is impossible to observe an infinite sequence of outcomes. Further, by introducing assumptions about the real world, it is possible that elements of circularity appear in axiom systems designed to guide researchers in their efforts to learn about real world phenomena, as Jeffreys has indicated. Thus it is accurate to state that both B's and NB's use many subjective elements in sorting out hypotheses, formulating models, choosing distributions for models' error terms and random coefficients, providing identifying assumptions and in defining probabilities. And further, in evaluating models, B's consider the quality of their predictions in many different cases. Nothing in the B approach precludes evaluations in terms of the relative frequency of successful predictions as well as other criteria.

Fifth, on defining probabilities, there is an interesting issue involving B's. Probability as a measure of the degree of confidence one has in a proposition can be introduced as a axiom, as Jeffreys [15] did, or as Bayes, Ramsey, Savage and others have done through betting and utility considerations. Although Jeffreys [15, pp. 30–33] was aware of the Bayes–Ramsey betting/utility approach, he had misgivings about it and decided to introduce his definition of probability as an axiom and add other axioms and theorems to give him the laws of probability. Then he recognized that his probabilities could of course be employed in betting and utility considerations. In view of recent developments in utility theory, see e.g., [5, 18], it appears to me that Jeffreys made a wise decision in this matter.

Sixth, to return to B and NB issues, there are some formal axiomatic systems which underlie the B approach, namely those of Jeffreys, Savage, de Finetti and others. In my opinion, none of these axiom systems is perfect in the sense of meeting all of our needs in the context of learning from our data and making decisions. There is room for improvement. On the NB side, there are axiom systems for probability theory but none, so far as I know, for learning from data and making decisions. Wald's [28] work, which started out as an NB approach to decision-making turned out to yield the B approach as optimal. As regards these axiom systems, while they are impressive intellectual achievements, as stated above, they are not perfect. Thus, I do not tell others that they are incoherent because they do not follow one or another of the available axiom systems. As recognized above, researchers can behave as artists and sometimes get rather good results without rigorously following a particular axiom system. On the other hand, formal, reproducible results are to be desired and that they can be produced by the B approach for many problem is noteworthy.

Seventh, another important issue that is critical in appraising alternative approaches to inference and decision-making is what I call the conditional/unconditional probability issue. That is, in the B approach, one can make probability statements about values of parameters, future values of random variables, and alternative hypotheses or models conditional on given sample and prior information. For example, for most estimation problems, it is easy to evaluate the probability that a parameter's value lies between two given values, as Bayes showed more than two centuries ago in connection with a binomial parameter. Also, conditional on the data, it is usually straightforward to compute the probability that a future outcome will lie in a particular given interval or region.

These conditional probability statements are extremely useful in practice to scientists and decision-makers and yet they cannot be made in NB approaches based on fixed parameter models. It should be noted however that R.A. Fisher's fiducial inference procedures were designed to produce such conditional statements. However, as Savage put it, Fisher tried to make Bayesian omelettes without the eggs (prior distributions).

Eighth, B's can make "unconditional" probability statements about the operating characteristics of B procedures. For example, B's can show that certain 95% Bayesian intervals have 95% coverage in repeated trials. In addition, as is well known the operating characteristics of B point estimates and predictions have been studied in terms of risk functions and admissibility considerations, concepts which NB's also utilize extensively. However, B's go one step further to show that B estimators and predictors are those that minimize average or Bayes risk. NB's do not use the concept of average or Bayes risk since they generally do not use prior densities which are critical in defining average or Bayes risk.

Ninth, the treatment of nuisance parameters, that is, parameters not of special interest to an investigator has been emphasized as a point of difference between B and NB approaches. In the B approach nuisance parameters can be readily integrated out of a joint posterior density to obtain the marginal posterior density of the parameter or parameters of interest — see [3, 31] for many examples of this procedure. In NB approaches, it is not easy to deal with the nuisance parameter problem. For example, when an "optimal estimator", say a weighted least squares estimator, depends on parameters with unknown values, investigators often insert estimates of these parameters in the formula defining the "optimal estimator" and give the resulting estimator an asymptotic, approximate justification. Many times such an approximate estimate can be produced as the mean of a conditional posterior density — see [31] for illustrations of this point in several central estimation problems and for marginal posterior means for these problems which are optimal estimates relative to quadratic loss functions.

Tenth, in the B approach it is generally possible to produce finite sample inference results for time series problems, nonlinear problems, and many others either analytically or by use of numerical integration and other computational procedures. That is, complete posterior densities for parameters and predictive densities for future observations are generally available for use. In NB approaches, such posterior densities and predictive densities are not utilized. Usually, asymptotic probability statements about the operating characteristics of estimators, intervals and predictors are employed. In the B approach it is generally possible to show that with proper prior densities B estimators and predictors are admissible and minimize Bayes risk. If asymptotic properties of B procedures are desired, they are available for standard iid and stochastically dependent observations — see e.g., [4, 12].

Further, B's can derive finite sample estimates, intervals and point predictions which are optimal relative to a wide range of loss functions. See e.g., [1, 27, 31, 36] for discussion of B results for a variety of symmetric and asymmetric loss functions. That estimates, intervals and point predictions can be chosen so as to minimize expected loss or maximize expected utility is appealing to many on theoretical and applied grounds. NB approaches have some estimators that have optimal operating characteristics, e.g., minimal variance, unbiasedness, etc. which can often be produced by B's using diffuse prior densities. In the case of maximum likelihood (ML) estimation, the ML estimate has a conditional optimality property, namely it is the estimate that makes the probability density function of the observations as high as possible. This property is many times also shared by the modal value of a posterior density based on a diffuse prior density. Also, while modal values of posterior densities which are often equivalent to ML estimates have good operating characteristics vis-a-vis a zero–one loss function, in finite samples the ML estimator can often be suboptimal with respect to other loss functions, for example squared error loss. Particularly interesting cases are encountered in estimation of reciprocals and ratios of population means and regression coefficients in which ML estimators fail to possess finite moments and can have bimodal distributions in small samples. Also in these cases posterior distributions for these reciprocals and ratios based on diffuse or natural conjugate priors do not possess finite posterior moments and can be bimodal. See [32, 35, 39] for Bayesian analyses of these problems which provide complete posterior densities and optimal point estimates relative to certain generalized quadratic loss functions.

From what has been presented above, it is clear that introduction of a prior density to represent information about the possible values of parameters is a distinctive feature of the B approach. Given that its introduction permits use of Bayes' Theorem to obtain exact finite sample posterior densities, predictive densities, posterior probabilities associated with alternative hypotheses and models, etc., it appears that there are large net benefits associated with the introduction of prior densities in theoretical and applied statistical inference and decision-making procedures.

On the basic issue of the introduction of subjective prior information in analyses, the following quotations from the writings of several leading NB's are illuminating. Freedman [8] has written as follows:

"When drawing inferences from data, even the most hardbitten objectivist has to introduce assumptions and use prior information. The serious question is how to integrate that information into the inferential process and how to test the assumptions underlying the analysis" (p. 127).

Thus Freedman recognizes the fundamental problem but does not offer a solution to it. Without a solution, NB Statistics, as viewed by Freedman, can hardly be said to be "objective". Rather, it is somewhat of an art, as discussed above. Some NB's have come to appreciate that B methods deal with prior information in a formal and reproducible manner and have been used to produce solutions to many statistical problems.

Tukey [26] in commenting on the above issues has written as follows:

"It is my impression that rather generally, not just in econometrics, it is considered decent to use judgment in choosing a functional form, but indecent to use judgment in choosing a coefficient. If judgment about important things is quite all right, why should it not be used for less important ones as well? Perhaps the real purpose of Bayesian techniques is to let us do the indecent thing while modestly concealed behind a formal apparatus. If so, this would not be a precedent. When Fisher introduced the formalities of the analysis of variance in the early 1920's, its most important function was to conceal the fact that the data was being adjusted for block means, an important step forward which if openly visible would have been considered by too many wiseacres of the time to be "cooking the data". If so, let us hope that the day will soon come when the role of decent concealment can be freely admitted .... The coefficient may be better estimated from one source or another, or, even best, estimated by economic judgment. It seems to me a breach of the statistician's trust not to use judgment when that appears to be enough better than using data" (p. 52).

Along with Tukey, I hope that the day will soon come when "decent concealment can be freely admitted" because that will help to make statistical analyses more easily reproduced. Also, it is safe to say that the wise use of judgment will lead to better analyses in many cases.

The following quotation from Lehman [17] provides evidence that NB's are hardly "objective" in testing hypotheses:

"Another consideration that frequently enters into the specification of a significance level is the attitude toward the hypothesis before the experiment is performed. If one firmly believes the hypothesis to be true, extremely convincing evidence will be required before one is willing to give up this belief and the significance level will accordingly be set very low" (p. 62).

While Lehman's advice is probably sound, it is not what most would call an "objective" approach to hypothesis testing.

It is clear that these leading NB's see the need for judgment and prior information in statistical inference and decision-making. The Bayesian paradigm is one that allows such judgment and prior information to be introduced formally in a reproducible, unconcealed manner. That the B approach is being more widely developed and used — see, e.g., [20] for a report on Bayesian publications in leading journals over the years, is probably an indication that a growing number of "customers" are satisfied with its results.

In conclusion, I hope that I have provided you with a useful summary of B and NB approaches to statistical inference and decision-making that will help you in your deliberations on the relative merits of B and NB approaches.

References

[1] J.O. Berger, *Statistical Decision Theory* (Springer, New York, 2nd ed., 1985).
[2] R.F. Bordley, A Bayesian extension of Bayes' rule, manuscript, Operating Sciences Department, General Motors Research Labs, Warren, MI, 1991.
[3] G.E.P. Box and G.C. Tiao, *Bayesian Inference in Statistical Analysis* (Addison-Wesley, Reading, MA, 1973).
[4] C.-F. Chen, On asymptotic normality of limiting density functions with Bayesian implications, *J. Royal Statist. Soc. Ser. B* **47** (1985) 540–546.
[5] M. Cohen and J.Y. Jaffray, Rational behavior under complete ignorance, *Econometrica* **48** (1980) 1281–1300.
[6] J. Drèze, The Bayesian approach to simultaneous equations systems, ONR Research Monograph No. 67, Technological Research Institute, Northwestern Univ., Evanston, IL, 1962.
[7] J. Drèze and J.-F. Richard, Bayesian analysis of simultaneous equation systems, in: Z. Griliches and M.D. Intriligator, Eds., *Handbook of Econometrics, Vol.* 1 (North-Holland, Amsterdam, 1983) 517–598.
[8] D.A. Freedman, Reply, *J. Business Econom. Statist.* **4** (1986) 126–127.
[9] M. Goldstein, Separating beliefs, in: P.K. Goel and A. Zellner, Eds., *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti* (North-Holland, Amsterdam, 1986) 197–216.
[10] I.J. Good, *Good Thinking: The Foundations of Probability and Its Applications* (Univ. of Minnesota Press, Minneapolis, MN, 1983).
[11] I.J. Good, The Bayes/non-Bayes compromise: a review, paper presented at the American Statistical Society Meetings, Atlanta, GA, August, 1991.
[12] C.C. Heyde and I.M. Johnstone, On asymptotic posterior normality for stochastic processes, *J. Roy. Statist. Soc. Ser. B* **41** (1978) 184–189.
[13] J.S. Hodges, Uncertainty, policy analysis and statistics, *Statist. Sci.* **2** (1987) 259–275.
[14] H. Jeffreys, *Scientific Inference* (Cambridge Univ. Press, Cambridge, 2nd ed., 1957).
[15] H. Jeffreys, *Theory of Probability* (Oxford Univ. Press, Oxford, 3nd reprinted ed. 1988; 1st ed., 1939).
[16] G. Koop, 'Objective' unit root tests, *J. Appl. Econometrics* **7** (1992) 65–82.
[17] E. Lehman, *Testing Statistical Hypotheses* (Wiley, New York, 1959).
[18] M.J. Machina, Choice under uncertainty: problems solved and unsolved, *J. Econom. Perspectives* **1** (1987) 121–155.

[19] F.C. Palm and A. Zellner, To combine or not to combine? issues of combining forecasts, *J. Forecasting* **11** (1992) 687–701.
[20] D.J. Poirier, A report form the battlefront, *J. Business Econom. Statist.* **7** (1989) 137–140.
[21] D.J. Poirier, Ed., *Bayesian Empirical Studies in Economics and Finance*, *J. Econometrics* (Annals Issue) **49** (1991).
[22] S.J. Press, *Bayesian Statistics* (Wiley, New York; 1989).
[23] R.A. Reynolds, Posterior odds for the hypothesis of independence between stochastic regressors and disturbances, *Internat. Econom. Rev.* **23** (1982) 479–490.
[24] S.M. Stigler, *The History of Statistics* (Harvard Univ. Press, Cambridge, MA, 1986).
[25] P.A.V.B. Swamy, *Statistical Inference in Random Coefficient Regression Models* (Springer, New York, 1971).
[26] J.W. Tukey, Discussion of granger on seasonality, in: A. Zellner, Ed., *Seasonal Analysis of Economic Time Series*, US Department of Commerce, Bureau of the Census (US Government Printing Office Washington, DC, 1978) 50–53.
[27] H.R. Varian, A Bayesian approach to real estate assessment, in: S.E. Fienberg and A. Zellner, Eds., *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage* (North-Holland, Amsterdam, 1975) 195–208.
[28] A. Wald, *Statistical Decision Functions* (Wiley, New York, 1950).
[29] R.L. Winkler, Combining probability distributions from dependent information sources, *Management Sci.* **27** (1981) 479–488.
[30] R.L. Winkler, Combining forecasts; a philosophical basis and some current issues, *Internat. J. Forecasting* **5** (1989) 605–609.
[31] A. Zellner, *An Introduction to Bayesian Inference in Econometrics* (Wiley, New York, 1971; reprinted by Krieger, Malabar, FL, 1987).
[32] A. Zellner, Estimation of functions of population means and regression coefficients including structural coefficients: a minimum expected loss (MELO) approach, *J. Econometrics* **8** (1978) 127–158.
[33] A. Zellner, Ed., *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys* (North-Holland, Amsterdam, 1980).
[34] A. Zellner, *Basic Issues in Econometrics* (Univ. of Chicago Press, Chicago, IL, 1984).
[35] A. Zellner, Further results on Bayesian minimum expected loss (MELO) estimates and posterior distributions for structural coefficients, *Adv. in Econometrics* **5** (1985) 171–182.
[36] A. Zellner, Bayesian estimation and prediction using asymmetric loss functions, *J. Amer. Statist. Assoc.* **81** (1986) 446–451.
[37] A. Zellner, Bayesian methods and entropy in economics and econometrics, in: W.T. Grandy Jr. and L.H. Schick, Eds., *Maximum Entropy and Bayesian Methods* (Kluwer Academic Publishers, Dordrecht, 1990) 17–31.
[38] A. Zellner, Time series analysis, forcasting and econometric modeling: the structural econometric modeling, time series analysis (SEMTSA) approach, *J. Forecasting* **13** (1994) 215–233.
[39] A. Zellner and S.B. Park, Minimum expected loss (MELO) estimators for functions of parameters and structural coefficients of econometrics models, *J. Amer. Statist. Assoc.* **74** (1979) 185–193.
[40] A. Zellner and W. Vandaele, Bayes–Stein estimators for k-means, regression and simultaneous equation models, in: S.E. Fienberg and A. Zellner, Eds., *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage* (North-Holland, Amsterdam, 1975) 317–343.