# On not being Rational

I.R. SAVAGE
*Yale University*

SUMMARY

A Bayesian decision-theoretic approach appears to me as a sensible idealization of a guide to behaviour. At the same time I would like to understand why my behaviour is not always of this form: I sometimes use randomization and I sometimes find confidence intervals acceptable. Not all of my problems have an explicit cost function. Am I lazy or irrational? Do I use non-Bayesian conventions to help communicate? Is the cost of rationality-computation missing from the Bayesian model?

*Keywords:* DECISION THEORY, UTILITY, COMPUTATION COST, RANDOMIZATION, IMPUTATIONS, CONCEPT FORMATION, NONPARAMETRIC STATISTICS, CONFIDENCE INTERVALS, MAXIMUM LIKELIHOOD, VAGUENESS, CONJUGATE PRIORS, STATISTICAL PACKAGE.

## 1. INTRODUCTION

Theories never apply perfectly. Serious use of statistical theory quickly runs into major problems. In this essay the emphasis is on difficulties faced by a Bayesian. Other approaches, Waldian or Fisherian, would raise similar problems. Since the decision-theoretic Bayesian model is the most complete of current theories it tends to create the most difficulties when used. I doubt that there is some eclectic statistical theory that can be used without difficulties.

If the range of application of a theory is very narrow then when the theory can be applied the application might be routine. For a short time I had a colleague who held the theory that a statistical analysis could make statements about the data in hand and nothing else. The applications that I have in mind in writing the following are often more complex and less well structured than the standard textbook examples.

The standard theories of statistics are involved in all of the issues to be raised here. In the story about $\pi$ (Section 2) either the non-Bayesian will suffer the same discomfort as the Bayesian or he will say it is not a statistical issue which shifts the problem to others. The inadequate treatment of cost of thinking (Section 3) appears to be common to all current statistical theories. The tension between the efficient use of standard statistical models and tailor-made procedures is common (Section 4).

Randomization (Section 5) in sampling and assignment of treatments is a very appealing process but it is seldom easy to show the need for it. Current statistical developments raise new topics and some discussion of imputation (Section 6) is appropriate.

Confidence intervals and maximum likelihood estimation (Section 7) have received an extensive non-Bayesian development. In application they often are given a Bayesian interpretation. This raises ethical and educational issues. In Section 8 a few annoying details are mentioned.

Three closely related topics not covered in this essay are data analysis, model making, and concept formation -see Suppes (1966)-. The lack of a mechanism of discovery in the Bayesian framework is crucial in the use of statistics in scientific research. In this conference Box has argued that the Bayesian framework must be inadequate in this respect. Also, the conference papers of Leonard and Dempster are much concerned with this point.

The author does not claim any new results. The references cover the discussion. Perhaps bringing these topics together with a minimum of technical distraction will be helpful. The presentation emphasizes the problems arising from the noninclusion of the cost of rationality in the Bayesian framework. Even if it should be argued that these costs are implicit in theory, it is clear they are not explicit in use.

## 2. SOME THOUGHTS ABOUT $\pi$

At no cost you can win a bottle of sherry if you correctly state the 29th digit of the decimal expansion of $\pi$; an incorrect statement yields nothing. In this situation I think I would pick my favourite digit, 7, and expect my chance is .1 to win the sherry. A moments thought tells me I am a bad Bayesian.

What I should do is think about the problem and compute the 29th digit. Since the Bayesian is rational he should be able to perform this task. Even if rational means something more restricted than perfect reasoning, it still must be noted that the usual Bayesian model does not include the cost of computing. So again in this situation the correct Bayesian action is to find the correct answer.

Some reading this story might know the required digit; it has often been computed and it is available in standard sources. In some situations it might be worthwhile to go to the library and look up the digit. If one bottle of sherry is replaced by a lorry load of sherry, I would come up with the correct digit.

Apparently thought is very much like data. One has incentive to do more thinking (more data collecting) when the stakes are increased. Pure thought, stored data, and data not yet acquired are costly ways of removing uncertainty.

I.J. Good (1950, p. 49; 1968, pp. 125 and 129; 1976, pp. 135-136; 1977) has used the terms Type II rationality and dynamic probability in discussing the topics of this section and of Section 3. Also, de Finetti (1975, pp. 278,291) has discused $\pi$ in this context.

## 3. THE COST OF THINKING, ANALYZING OR COMPUTING

Recently, Watson and Brown (1978) have discussed the problem of how much value there is in doing an analysis —in the operations research context— *before* the analysis is performed. Their situation must include the analysis required to design a statistical investigation. Watson and Brown's references summarize the related work, including their efforts to find the value of analysis in several case studies.

The costs of analysis do not appear explicitly in statistical theories. In large scale statistical activities, such as a national census, there will be explicit budgeting for items such as planning, data handling, and publication. This process appears to be empirical; theory to help choose optimal amounts of these items is not used. Watson and Brown suggest that empirical evidence would be a good way to solve their operations research problem. It is not clear how well this process of learning can work because of the great variety of complex situations that one needs information about. In the public sector, it often appears that there is inadequate budget for analysis after data are collected. The problem might be that it is relatively difficult to obtain appropriate budgets for soft items like analysis in contrast to hard items, like data.

Economic theory perhaps could make a formal background for the optimal choice of amounts of thought and analysis. Those commodities are known to be valuable but they are hard to value. Most statistical consulting does not have the market mechanism to help establish value.

We will come back to this topic when we discuss standard models, randomization and imputations. To avoid giving the impression that the discussion does not relate to the usual activities of statisticians consider the cost of the following tasks:

(a) Limiting the scope of analysis, determining which variables need analysis.
(b) Specifying joint distribution of all variables, items to be measured and states of nature.
(c) Evaluating losses associated with decisions and states of nature.

(d) Searching for the best kinds and amounts of data to collect.

(e) Determining how much to spend on analysis and communication of results.

If these aspects of the problem are handled properly their costs might be comparable to the usually considered costs, such as sampling costs and terminal losses. It is my impression that we know very little about the correct expenditures on items (a)-(e). At this point statistical theory does not automatically help us to choose good levels for these activities. Subject specialists must be able to help with some of the choices such as limiting scope (a). Careful work on (a)-(e), even if the resulting sample sizes must be reduced, should provide a powerful mechanism to avoid superficial data collection and glib analysis. The difficulty is that we already think we know how to collect and analyse data and it is still challenging, if not frightening, to think about (a)-(e). One is put off from theorizing on these topics because of the unwieldy anticipated results. One fears an infinite regress.

Moore (1978, p. 72) considers some of the above costs are trivial but he incorrectly related them to total costs in contrast to costs due to uncertainty. (He does discuss many problems in applying Bayesian decision theory).

## 4. STANDARD MODELS

"Assume... are iid..." is an expression seen so often that one is tempted not to check its appropriateness. Although this set of assumptions is used by pushers of nonparametric statistics they are proud of their lack of use of assumptions. Some reflection could lead to different results:

1. The care for experimental detail to assure iid can often yield stronger assumptions such as normality.

2. In fact iid might be replaced by a weaker exchangeability assumption.

Two standard models of great importance are the packaged programs and the Raiffa-Schlaifer conjugate priors. These examples well illustrate the advantages of standardization: A great variety of problems can be handled at low costs. A convenient mode of communication is developed. Many people can use advanced technology. The problem is to make sure that these advantages greatly exceed the disadvantages: One can force a situation into the wrong model. One can be lazy and not take full advantage of the available options. One can unwisely restrict the kinds of data to be obtained just so a standard model can be used.

The ideal is to have the standard models available and that their use is supervised by skilled individuals. The net results are to increase resources for research and to make sure that unusual situations receive appropriate attention.

## 5. RANDOMIZATION

A Bayesian is about to sample a finite population. Should he take a random sample? If a random sample costs no more than a grab then why not random sample? If the Bayesian acted as if the population elements were exchangeable then random sampling has no disadvantage. (For a discussion see Ericson [1969].) In this situation there is an advantage to the Bayesian in taking the random sample, even if he is unhesitating about the exchangeability. In particular, random sampling gives others confidence that the work has been done properly.

This confidence might be at two levels. The use of random sampling is an indicator that the whole job has been done with professional care. For some, there will be increased acceptability of the results because they feel random sampling is a necessary part of a valid procedure. How much the Bayesian should pay to buy confidence of others is not clear.

When would the Bayesian have an aversion to random sampling in the above situation? This would happen if he did not really accept the exchangeability assumption; he might really prefer some form of stratified sample. In fact, the acceptability of random sampling can be used as a form of self examination of the Bayesian to tell if he strongly believes in exchangeability.

The Bayesian might use randomization as a technique to avoid expensive activities such as thinking. Thus in the current example let us assume his interest centres on the total income of the population. He might well have many variables that he could use for stratification, such as age, sex, address, education, profession, number of children, etc., etc. Regardless of the costs of the various types of samples the Bayesian may conclude it is more economical to ignore the other variables and just work with income. He saves thinking about the complex multivariate distribution of all of the variables. This averted task is one where there is limited experience. He might have other substantial savings in data collection and analysis.

If the Bayesian followed this simple path then he might even be willing to pay for the randomization for his own peace of mind. To put this into a formal analysis could be awkward.

Rubin (1978 a, Sect. 5) contains a technical discussion of some of these comments. His 1974 article is also relevant. Savage (1962, pp. 34, 88-89) vividly describes a Bayesian's problem with randomization.

## 6. IMPUTATIONS

It is common in handling large data sets to have missing values. Those are replaced by imputed values. The data set is then ready for analysis. No matter how much or what kinds of analysis are to be performed there usually is just one imputation process applied to a set of data.

The great advantage of doing the imputation is that statistical methods for complete data sets are much simpler than those for data sets with missing values. So if there is going to be much analysis it is less expensive to do one costly imputation and many routine procedures than many complex missing data procedures.

If computing costs did not dominate then presumably imputations would not be used. For it is hard to believe that one set of imputations will be correct for a variety of problems involving different loss functions and different prior distributions. Again, the Bayesian probably cannot afford to think it out in detail —thinking is expensive— so that he is inclined to use the single imputation process. (See Rubin —1978b— for a more technical discussion).

## 7. CONFIDENCE INTERVALS AND MAXIMUM LIKELIHOOD ESTIMATION

The first few times I was told about these procedures it sounded like gibberish. The instructors knew the correct definitions and attempted to present them. The definitions are awkward and appear to be about the wrong thing. Compare:

(a) The *mle* is that value of the parameter which would have maximized the probability of the data.

(b) Given the data the *mle* is the most probable parameter value.

Or:

(a) In the long run the procedure for 95% confidence intervals will create intervals including the parameter 95% of the time. On any particular occasion the probability of coverage is either 0 or 1 but there is not evidence from the data to say which value is correct.

(b) A 95% confidence interval includes the parameter with probability 0.95.

Both of the (b) statements are false. Many users of the procedures have (b) and not (a) statements in mind. In a Bayesian framework the (b) statements are good approximations when the samples are large or the prior is diffuse. Since these are important and commonly used procedures how should the statistical community reduce the large number of errors? I am convinced the non-Bayesian can do nothing. They have had little success in fifty years of expositing; their message is useless. Perhaps the Bayesian should help the users of the (b) statements to understand the implications. This also might not be useful for most people don't want to expand their formal knowledge of statistics. The Bayesian can often let well enough alone.

## 8. FINE TUNING

In practice it is hard to even begin to be a Bayesian. One generates inconsistent prior distributions. Computation of exact probabilities would be meaningless. Utilities are often not even approximated. Since the theory does not provide for the expense of these costly activities it is not surprising that the behaviour required by the theory does not occur.

Lindley, Tversky and Brown (1979) present a mechanism for the resolution of inconsistencies (they assume no cost for this mechanism). Without giving any details here, it seems appropriate to suggest that in investigating ways to remove inconsistencies one should not discard the possibility of the Bayesian doing further introspection.

Many discussions of axiom systems for the Bayesian have appeared. Suppes (1974) specifically evolves theories which do not require the Bayesian to give exact values. Again, the cost of accuracy is not in the model.

Absence of utility measurement in much of applied Bayesian statistics can reflect a variety of causes, such as lack of interest, excess cost, or unable to produce at any cost. It does seem possible that a statistician could present probabilities that would be moderately acceptable to all interested parties. On the other hand the interested individuals might have widely varying utilities. An example of some interest is the allocation of funds from central to local governments. This is now often done by formulas using social and economic data. The utilities of the civil servant statistician, the executive, the legislative body, the local governments, pressure groups, and the people might all be different and all hard to approximate. Even for such major activities this work is seldom begun. It would be costly and it would be, technically, hard to justify. At this time the evidence regarding the usefulness of such analysis is ambiguous.

''The coherent individual is supposed to assess his probabilities and utilities for everything. Of course, taken literally, this is absurd; but it does not invalidate the theory any more than the failure of the claim to predict the whole future of the universe, given the position and velocities of particles now, invalidates Newton's theory'' Lindley's discussion of Suppes (1974, p.181).

## REFERENCES

DE FINETTI, B. (1975). *Theory of Probability: A Critical Introductory Treatment,* Vol. 2. London: Wiley.

ERICSON, W.A. (1969). Subjective Bayesian models in sampling finite populations (with discussion) *J. Roy. Statist. Soc. B* **31**, 195-223.

GOOD, I.J. (1950). *Probability and the Weighing of Evidence.* London: Griffin.

— (1968). Corroboration, explanation, evolving probability, simplicity, and a sharpened razor. *British J. Philos. Sci.* **19**, 123-143.

— (1976). The Bayesian influence, or how to sweep subjectivism under the carpet. In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science,* Vol. 2 (C.A. Hooker and W. Harper, eds.), 125-174. Dordrecht, Holland: D. Reidel.

— (1977). Dynamic probability, computer chess, and the measurement of knowledge. In *Machine Intelligence 8* (E.W. Elcock and D. Michie eds.), 139-150. New York: Wiley.

LINDLEY, D.V., TVERSKY, A. and BROWN, R.V. (1979). On the reconciliation of probability assessments (with discussion). *J. Roy. Statist. Soc. A* **142**, 146-180.

MOORE, P.G. (1978). The mythical threat of Bayesianism. *Int. Statist. Rev.* **46**, 67-73.

RUBIN, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educat. Psycol.* **66**, 688-701.

— (1978a). Bayesian inference for causal effects: the role of randomization. *Ann. Statist.* **6**, 34-58.

— (1978b). Multiple imputations in sample surveys a phenomenological Bayesian approach to nonresponse. *Proceedings of the Survey Research Section of the Amer. Statist. Assoc.*, 183-188.

SAVAGE, L.J. *et. al.* (1962). *The Foundations of Statistical Inference — a Discussion*. London: Methuen.

SUPPES, P. (1966). Concept formation and Bayesian decisions. In *Aspects of Inductive Logic* (Hintikka J. and Suppes P. eds.). Amsterdam: North-Holland.

— (1974). The measurement of belief (with discussion). *J. Roy. Statist. Soc. B* **36**, 160-191.

WATSON, S.R. and BROWN, R.V., (1978). The valuation of decision analysis. *J. Roy. Statist. Soc. A* **141**, 69-78.