THE BAYESIAN ALTERNATIVE TO STATISTICAL INFERENCE

Ronald C Wimberley, North Carolina State University
Louis E Dotson, University of Tennessee, Knoxville

BAYES' THEOREM

If sociologists were told that the knowledge of statistical inference accumulated throughout their careers was to become more or less obsolete in a decade or so, some would probably hope not. While this knowledge may not become obsolete so soon, major refinements and alternatives may pose challenges to vested statistical interests.

An alternative to conventional statistical testing is the Bayes theorem. Thomas Bayes was a minister in England whose ideas were published posthumously, but drew little attention until the middle 1900's (Bayes 1763). The Bayes theorem relates three probabilities, expressed in two equations:

$$1) \quad P(S_i|D) = \frac{P(D|S_i)\,P(S_i)}{P(D)}$$

$$2) \quad P(D) = \sum_{i=1}^{k} P(D|S_i)\,P(S_i)$$

$P(D)$ = probability of data observed.

$P(S_i)$ = hypothetical probability of a particular state of the world.

$P(D|S_i)$ = the hypothetical probability of the data, given the state of the world.

$P(S_i|D)$ is the conditional probability of a state of the world given the data.

The numerator in Equation 1 is merely the joint probability for the occurrence of the data and a particular world state. The denominator is the probability of the data for the possible states of the world, defined by Equation 2.

The theorem may be described as the ratio of the probability of a particular state of the world coinciding with observed data, as given in the numerator, to the sum of all joint probabilities of particular states of the world coinciding with observed data as given in the denominator. According to Bayes, this is the conditional probability that a certain state of an unobserved universe exists, based on information from limited data. The attempt is to acquire knowledge of an unknown population characteristic on the basis of a known sample characteristic where information from all available studies is utilized along with any background information an investigator may have.

Two types of distribution are involved in the theorem. The one containing background knowledge at disposal before examining a new data set is called a *prior distribution*, which is the hypothetical probabilities of particular states of the world as it is thus far revealed. The *posterior distribution* is a revised estimate of knowledge about the state of nature after new data have been collected and incorporated into prior knowledge. The posterior distribution is the conditional probabilities of certain states of the world, given the data which now include past and present data.

As a result, the Bayes approach is an *iterative* process. Today's posterior distribution becomes the prior distribution for tomorrow's research. In this process the true but unknown state of the world should become more narrowly confined. If fact, the standard deviation of a posterior distribution should become smaller than that of the prior distribution. Distributions should thus become more concentrated as the iterations continue. Therefore the probability intervals for values of an unknown population should continually shrink, allowing better predictions. Like the scientific method itself, the Bayesian approach tends to be self-correcting, if the model is appropriate and the method is used repeatedly.

The symbols used here are adapted from Iversen (1972 33) due to their simplicity and straightforwardness. There are many other discussions of the Bayes identity in statistical detail developed by Mosteller and Tukey (1968 167; Iversen 1970 186; Winkler 1972 40).

BAYESIAN AND CLASSIC CONTRASTS

The classical approach appears in virtually all statistics textbooks. Classical statistical inference is distinguished by the ideas of Type

I and Type II errors, and by testing for differences from parameter values using null hypotheses of no difference. The Bayesian approach does not engage in the sport of setting significance levels which must be met if an alternative hypothesis can be considered statistically sound. The level of significance is a moot point — a controversy which Bayesian inference ignores (Morrison, Henkel 1970)

Bayesian inference attempts to specify that a parameter value lies in an interval constructed from both past and present data, with a specified probability. Classical inference may also attempt to establish intervals, but in the classic approach, this is secondary to attempts to find whether the sample statistic is contrary to a null hypothesis of a parameter value. Bayesian inference does not posit a null value from which a sample statistic may indicate a significant difference, but rather *an interval containing the actual value.*

This suggests a third contrast which concerns the interpretation of statistical results. Bayes estimates give probability limits whereas classical statistics give confidence limits (Mosteller, Tukey 1968 182). For Bayesian results, a probability of, say .95 is the degree of certainty that an estimate within a specific interval is indeed true. In classical inference, however, such a statement cannot be justified since classical probability is based on relative frequencies of events in the long run. In this case, a .95 confidence interval means that in 95 samples out of 100, a statistic should fall within an interval containing a parameter. The classical probability of an outcome being in a given interval is either 0.00 or 1.00, absent or present. Therefore, Bayesian inference allows the prediction of particular events; classical inference only allows predictions in the long run.

Another difference is that Bayesian statistical inferences are made from data at hand, which are treated as knowns to unknowns in the parameter. The classical approach makes probability statements about the data at hand rather than about the unobserved parameters.

> It seems reasonable that one should be able to use probabilities to measure the degree to which one is uncertain. We are uncertain about what the true value is of a parameter ... (Iversen 1972 31)

BAYES AND SOCIAL THEORY

How can Bayesian probability be used with existing sociological and psychological theories? If applicable, it should enhance precision in predictions. Good social theories should fit everyday experience. Since Bayes' assumptions allow us to build on past and present experience in the form of more and more refined posterior distributions, Bayesian notions may be more appropriate than classical notions for modeling everyday social life. Bayesian types of predictions could serve to guide behavior. Conversely, classical usage start us with the same normal distribution each morning, and then complicate matters with long-term odds on the outcomes of particular situations. The Bayesian approach explicitly allows for cumulative past experience whereas classic inference does not.

In the interactionist perspective, social interaction is seen as a process of continually interpreting meanings from others, and defining meanings to others. The Bayesian approach is likewise a continual process of refining distributions of expected outcomes in order to predict certain occurrences. The Bayesian approach provides a closely corresponding instrument for such theory.

The same can be said of *exchange, conflict,* and *behavioral* theories. Thus, persons will emit behaviors in present situations which are similar to behavior which brought rewards in past situations (Homans 1961 53). Bayesian conditional probabilities provide an appropriate modeling tool for this process. If two children or two nations have a prior history of fighting when they meet, it would be difficult based on prior distributions to expect that they could interact without assault. But if there were an instance where no assault occurred, however atypical that might be, the new posterior distribution would be modified to some extent so that the probability of fighting on the next exhange would be somewhat reduced. Thus, prior distributions are iteratively modified by experience into posterior distributions which alter expectations from one time to the next, allowing for both regularity and change.

Bayesian inference can be used in the modeling of normative and deviant behavior at the micro level. Here, research topics might include the process of socialization, career patterns, and personality development. Similarly, at the macro level, developmental change in family life cycles, social differentiation patterns, or community development might lend themselves to analysis by Bayesian probability models. And theoretical problems in social consensus and conflict exist at the macro level and would represent areas for the possible use of Bayesian predictions.

BAYES AND RESEARCH PROBLEMS

Sampling. In classical significance testing, random sampling is often assumed. This is not to say that the assumption is often satisfied. However, Bayesian inference would not seem to be hindered by nonrandom samples as would classical statistics. This is because biased samples would be used cumulatively in Bayesian inference. Unless there was systematic bias in most of the samples, inferences should gain in precision as more are incorporated into the Bayes distributions. The distributions may be kept gentle, or somethat flat, by biased samplings, but they should provide better estimates than could be achieved by a classical test using a single biased sample.

Research Design. The Bayesian approach seems especially suited for longitudinal, panel, and experimental designs in which data are collected at two or more periods. When used in conjunction with causal regression models in such designs, emphasis could be placed on probabilities for change rather than amounts of change in succeeding variables. As the Bayesian approach becomes more familiar to researchers, such uses should receive more attention.

Descriptive Statistics. Application of the Bayes theorem of statistical inference are found in Bayesian oriented textbooks (Raiffa, Schlaifer 1961; Lindley 1965, 1972; Pratt, Raiffa, Schlaifer 1965; Schmitt 1969; Hayes, Winkler 1971; Press 172; Box, Tiao 1973; Phillips 1974; Novick, Jackson 1974a). Feinberg and Zellner (1975) use the Bayesian approach in survey research.

Other works illustrate use of Bayes' assumptions in many types of statistical analyses conducted by sociologists. Contingency tables are a recurrent object (Lindley 1964; Altham 1969; Gunel, Dickey 1974). Bayesian treatments of the least squares principle are applied in analysis of variance (Box, Tiao 1966), linear modeling (Smith 1973) and multiple regression (Tiao, Zellner 1964 Hoadley 1970; Jackson, Novick, Thayer 1971; Lindley, Smith 1972; Novick et al. 1972; Novick, Jackson 1974b.) A Bayesian approach is applied to analysis of Markov chains (Martin 1967). For measurement problems, the Bayes theorem is directed to reliability (Bhattacnarya 1967) paired comparisons (Davidson, Solomon 1973), and Guttman scale error (Proctor 1971)

DRAWBACKS

Difficulties in Bayesian statistics deal with distribution problems. If there is no prior research, there is no prior distribution. This problem is associated with the lack of replication studies so common in the social sciences. Different researchers may develop different prior distributions. And there will be epistemic and empirical problems as to whether the results of apparently similar distributions are comparable enough to be placed together. If the discrepancy is minor and operational definitions are compatible, this should not be a major problem.

Acute and extensive changes in social or behavioral phenomena may make prediction difficult. Generally, the bias in Bayesian predicitons may be toward future stability, or toward limited variation from the past. However the prediction of severe departures from normality may be handled no better by classical inference.

Bayesian techniques that have been developed are rarely simplified to cookbook dimensions. The lack of easy instructions is not an absolute disadvantage, but is it a practical disadvantage. At present, many Bayesian techniques are accessible only to more advanced students of statistics.

Bayesian ideas create controversy among statisticians who tend to be more in agreement among themselves on the parameters of classical statistical inference.

REFERENCES

Altham P 1969 Exact Bayesian analysis of 2x2 contingency table and Fisher's exact significance test. *J Royal Stat Soc* 31 261-269

Bayes Thomas 1763 An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions, Royal Society, London 53 370-418; Biometrika* 45 296-315

Bhattacnarya S 1967 Bayesian approach to life testing and reliability estimation. *J Amer Stat Assn* 62 48-62

Box E, G Tiao 1973 *Bayesian Inference in Statistical Analysis.* Reading Mass. Adddison Wesley

_______ 1966 Bayesian analyhsis of random -effect models in the analysis of variance II. Effect of autocorrelated errors. *Biometrika* 53 477-495

Davidson R, D Solomon 1973 A Bayesian approach to paired comparison experimentation. *Biometrika* 60 477-487

Feinberg S, A Zellner 1975 *Studies in Bayesian Econometrics and Statistics.* Amsterdam. North Holland Press

Gunel E, J Dickey 1974 Bayes factors of independence in contingency tables. *Biometrika* 61 545-557

Hayes, W, R Winkler 1971 *Statistics: Probability, Inference and Decision.* New York. Holt Rinehart Winston

Homans George C 1961 *Social Behavior: Its Elementary Forms.* New York. Harcourt Brace World

Hoadley A B 1970 A Bayesian look at linear regression. *J Amer Stat Assn* 65 356-369

Iversen G 1972 *Statistics and Sociology.* Indianapolis Indiana. Bobbs Merrill

_______ 1970 Statistics according to Bayes. 85-199 E Borgatta, G Bohrnstedt eds. *Sociological Methodology 1970.* San Francisco. Jossey Bass

Jackson P, M Novick, D Thayer 1971 Estimating regression in *m* groups. *British J of Math & Stat Psych* 24 129-153

Lindley D 1972 *Bayesian Statistics: A Review.* Philadelphia. Soc for Industrial & Applied Math.

1965 *Introduction to Probability and Statistics.* Harvard U Press

1964 The Bayesian analysis of contingency tables. *Annals of Math Stat* 35 1622-1643

Lindley D, A Smith 1972 Bayes estimates for the linear model. *J Royal Stat Soc* 34 1-18

Martin J 1967 *Bayesian Decision Problems and Markov Chains.* New York. Wiley

Morrison D, R Henkel 1970 *The Significance Test Controversy.* Chicago. Aldine

Mosteller F, J Tukey 1968 Data analysis including statistics. 160-183 G Lindzey, E Aronson eds. *Handbook of Social Psychology.* Vol 2 2nd ed. Reading Mass. Addison Wesley

Novick M, P Jackson 1974a *Statistical Models for Education and Psychological Research.* New York. McGraw Hill

_______ 1974b Further cross-validation analysis of the *M*-group regression method. *Amer Educ Res J* 11 77-85

Novick M, P Jackson, D Thayer, N Cole 1972 Estimating multiple regression in *m* groups: a cross-validation study. *British J Math & Stat Psych* 25 33-50

Phillips L 1974 *Bayesian Statistics for Social Scientists.* New York. Crowell

Pratt J, H Raiffa, R Schlaifer 1965 *Introduction to Statisitcal Decision Theory.* New York. McGraw Hill

Press S 1972 *Applied Multivariate Analysis.* New York. Holt Rinehart Winston

Proctor C 1971 Reliability of a Guttman scale score. *Proc Amer Stat Assn* 348-350

Raiffa H, R Schlaifer 1961 *Applied Statistical Decision Theory.* Boston. Harvard Sch of Business Admin

Schmidt S 1969 *Measuring Uncertainty.* Reading Mass. Addison Wesley

Smith A 1973 A general Bayesian linear model. *J Royal Stat Soc* 35 67-75

Tiao A, A Zellenr 1964 Bayes' theorem and the use of prior knowledge in regression analysis. *Biometrika* 51 219-230

Winkler R 1972 *Introduction to Bayesian Inference and Decision.* New York. Holt Rinehart Winston