DOCUMENT RESUME

ED 069 707 TM 002 159

AUTHOR	Diamond, James
TITLE	Bayesian Statistics: A Place in Educational Research?
PUB DATE	[64]
NOTE	7p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Bayesian Statistics; Educational Research; Newsletters; *Prediction; *Probability Theory; *Research Methodology; *Statistical Analysis

ABSTRACT

The use of Bayesian statistics as the basis of classical analysis of data is described. Bayesian analysis is a set of procedures for changing opinions about a given phenomenon based upon rational observation of a set of data. The Bayesian arrives at a set of prior beliefs regarding some states of nature; he observes data in a study and then modifies his beliefs into posterior probabilities. The computational vehicle for determining posterior probabilities based upon observed data and prior beliefs is Bayes' theorem. Because of the problem of putting numerical values on prior information, the value of Bayesian ideas might ultimately be in the clarification of the classical approach rather than in substituting one for another. Application of the procedure to the concept of personal probability is used as an example of how the technique might apply to education. (DJ)

BAYESIAN STATISTICS : A PLACE IN EDUCATIONAL RESEARCH?

James Diamond

University of Pennsylvania

In the past few years, educational researchers have been introduced to the philosophy and concepts of Bayesian statistics. It is the judgment of this writer that exposure of this sort can aid one in the classical analysis of data. While in this newsletter format, great depth is not appropriate, perhaps this introduction will tempt those equipped with a basic statistics vocabulary to read further in this area.

Bayesian statistics are so named for the reason that in many instances the application of Bayes' theorem is made. This theorem is due to the Reverend Thomas Bayes and is actually a statement of a conditional probability. The essentials of Bayesian statistics were in existence before or developed within the framework of classical methodology. In one sense, then, Bayesian statistics are a sequel to the classical approach; in another sense they are an echo of 18th and 19th century statistical thought.

Bayesian analysis is similar to the Neyman-Pearson philosophy in that it is a system designed to help one make decisions about important problems. It is a set of procedures for changing one's opinion about a given phenomenon based upon the rational observation of a set of data. For example, "Given the data from my study on typewriting, should I invest money in the new typewriters and secretarial materials?" This may be a very important problem and one that needs to be acted upon; however one may be uncertain as to the appropriate course of action.

A Bayesian's uncertainty is expressed in terms of probability statements sometimes referred to as personalistic or subjective probabilities, although some take exception at the union of these two terms. The Bayesian attempts to show how one's initial or prior beliefs (personal probabilities) concerning an event can be modified by the observation of data. A Bayesian views his prior beliefs or prior probabilities in the same way as the layman, i.e. probability is simply one's <u>degree</u> <u>of</u> <u>belief</u> about an event. It has nothing to do with its relative frequency after a very large number of trials. (This latter view of probability as a limit can be called the frequentist point of view.) The Bayesian arrives at a set of prior beliefs regarding some states of nature; he observes data in a study; and then modifies his beliefs regarding those states into what are typically called posterior probabilities. The computational vehicle for determining posterior probabilities based upon observed data and prior beliefs is Bayes' theorem.

Let us look more closely at the concept of personal probability. Savage (1954) is usually credited with providing the initial impetus which resulted in the recent emphasis on Bayesian philosophy. Savage proposed that a subjective probability "measures the confidence that a particular individual has in the truth of a particular proposition (p. 3.)." It can also "refer to the opinion of a person as reflected by his real or potential behavior (p. 11.)." Even though this position on probability allows two rational individuals to have different degrees of belief when faced with the same evidence, this does not necessarily serve as a criterion for dismissing the concept. Differences of opinion are not the sole province of Bayesians by any means. One might bring up the point that in science these differences have no place. That may be, but as will be intimated below, the so-called "objective" methods in statistics are not as objective as they appear. Nevertheless, it is true that much of the criticism of the Bayesian method has centered around its use of personal probability. As Binder (1964) put it "... Bayesians reject the relative frequency interpretation of probability and substitute probability as personal opinion; ... they base their procedure of statistical inference upon various applications of Bayes' theorem; and ... they tend to use the language and concepts of decision theory in the formulation of their views (p. 279.)."

In addition, the basic character of a Bayesian analysis is the initial use of probability statements concerning various hypotheses, say, the null and alternative hypotheses. These statements reflect our accumulation of knowledge up to the point of the present experiment. They are referred to as prior probabilities and their distribution over various alternatives is called the prior distribution. It is prior because it comes before the experiment. Then data are collected. Following the collection of data, beliefs will presumably be altered. These new beliefs, the so-called posterior probabilities, have their own distribution, called the posterior probability distribution.

A Bayesian's prior distribution can take any form. It might be rectangular, bimodal, normal, or anything in between. Just what does this distribution reflect? It reflects everything the researcher knows about his research problem. This could come from past reports, from his own beliefs, or from any other source that helps him to form a rational opinion about the parameter in question.

The union of sample information with the prior distribution takes place in Bayes' theorem:

If events $E_1, E_2, \ldots, E_k$ are a set of mutually exclusive and exhaustive events and A is defined as some observation, the probability of some event E_i given A is

$$P(E_i/A) = \frac{P(E_i) \cdot P(A/E_i)}{\sum_i P(E_i) \cdot P(A/E_i)} \qquad (1)$$

where $P(A) \neq 0.0$ and $i = 1, 2, \ldots, k$. $P(E_i)$ is the prior probability of event E_i and $P(E_i/A)$ is the posterior probability of E_i given that event A has been observed. Equation (1) is Bayes' theorem.

Let us assume that event E_1 is a statement of the null hypothesis, E_2 is a statement of the alternative, and A is the observed data from a study. For a Bayesian, the two probability statements $P(H_o/A)$ and $P(H_1/A)$ are basic to describing what happened in an experiment. However, in taking action, i.e. in acting "as if" H_o or H_1 were true, it may be useful to consider losses. In many problems we can put dollar signs on these losses, especially in the business sector where, incidentally, Bayesian methods receive a great deal of attention. At other times this is not so straightforward. Do we have losses in classical statistics? Surely we have them. Where do we see them? Bayesians would propose that the fixing of α and β levels implicitly sets losses or at least tries to control them in some manner. That is, it might be worse in one researcher's judgment to fail to reject H_o when H_o was false than to reject H_o when it was true. What that researcher is saying, then, is that he needs a powerful test. How might he raise power? One way would be to raise α. Another researcher might be much more conservative owing to a lack of funds and/or other considerations and set α at a very small value. Who is correct?

Either of them! Which one of them is considering losses? Both of them!

Perhaps Bayesian statistics is, then, nothing more than an advocacy of higher α levels. While this strategy might be appropriate in some instances, the situation is more complex. As Meyer (1964) stated: "The posterior probabilities and odds calculated ... have a conceptual basis in terms of ... (one's) ... personal beliefs about the true value of ... (a parameter), whereas in the classical method these quantities cannot be computed. ... the classical approach rests on two conditional probabilities, α and β, which are determined without regard for the sample result and which may or may not be relevant to the particular problem being studied. The probability of H_o being rejected if H_o is true is set at α; but if H_o is rejected, it does not follow that the probability that H_o is true is α. Unfortunately, this latter statement concerning probability is often made or implied by the users of the classical method. For example, if alpha is .05 and H_o is rejected, it is sometimes said that the odds against H_o are 19 to 1 or that the probability of H_o being true is .05. Statements of this type within the classical framework have no conceptual meaning (pp. 225-6)."

One might also note that the posterior probability statements $P(H_o/A)$ and $P(H_1/A)$ are probably closer to the kind of information researchers want. Many students in statistics ask, "Is H_o true?" They can attempt to understand the classical approach, but still are dissatisfied in that direct information such as that encompassed in the posterior probability distribution is not available.

We have given a very brief overview of Bayesian philosophy. We should, however, in fairness mention what the critics of Bayesian statistics have to say. Most of the criticism has focused on prior probabilities.

Pearson indicated that he and Neyman thought about many of the aspects of the Bayesian approach but decided that since it was rarely possible to place numerical values on prior information, they would leave personal judgment to issues such as selecting the appropriate significance level. Pearson believed that the value of Bayesian ideas might ultimately be in the clarification of the classical approach and not in the substitution of one for the other. The present writer puts great weight on this statement. By being exposed to the Bayesian philosophy, one obtains a clearer understanding of the advantages and limitations of the classical method. This in turn should lead to more appropriate application of these techniques.

Arnold Binder delivered a telling blow designed to temper the missionary zeal of Bayesians when he pointed out that, although Bayesians argue for a unified logic and consistency of approach, many problems which can be readily handled by conventional methods cannot be handled at all by Bayesians or can be dealt with only by resorting to what one writer called "a welter of makeshifts and approximations."

A fitting final statement was supplied by Edwards et al (1963): "The Bayesian outlook is flexible, encouraging imagination and criticism in its everyday applications. Bayesian experimenters will emphasize suitable chosen descriptive statistics in their publications, enabling each reader to form his own conclusions. Where an experimenter can easily foresee that his readers will want the results of certain calculations, he will publish them. Adoption of the Bayesian outlook should discourage parading statistical procedures, Bayesian or other, as symbols of respectability pretending to give the imprimatur of mathematical logic to the subjective process of empirical inference (p.240)."

REFERENCES

Binder, A. Statistical theory. Annual Review of Psychology, 1964, 15, 277-310.

Edwards, W., Lindman, H., and Savage, L. J. Bayesian statistical inference for psychological research. Psychological Review, 1963, 70, 193-242

Meyer, D. L. A Bayesian school superintendent. American Educational Research Journal, 1964, 1, 219-228.

Savage, L. J. The Foundations of Statistics. New York: Wiley, 1954.

Savage, L. J. The Foundations of Statistical Inference. New York: Wiley, 1962.