

**“BEHAVIORISTIC” VS. “INFERENTIAL” APPROACH
TO THE PROBLEM OF DISCRIMINATION**

BY

JERZY NEYMAN

A prepared discussion of the paper by Frederick Mosteller and David L. Wallace,
Inference in an authorship problem
presented at the Annual Meeting of the American Statistical Association
on September 9th, 1962, in Minneapolis

1. Introduction. I wish to begin by complimenting Professors Mosteller and Wallace on the very interesting and instructive paper they prepared and, even more heartily, on the manuscript of their exciting book under the same title of which the paper under discussion is a brief summary. I did not read all the sections of the book I kept receiving over the summer, but what I did read proved excellently written and I am sure the book will be a great success.

Quite apart from the high literary quality, there are quite a few scientific points that make Mosteller's and Wallace's book very attractive to me. Here are a few of them.

The tone of the book and of the paper indicates clearly that the two authors are emotionally partial to the approach to statistical problems through the application of the Bayes theorem. This is so not only in those cases where the problem considered implies the randomness of the unknown parameters and the prior distributions of these parameters, but also in other cases where the prior distribution has to be selected. Ordinarily, this kind of approach to statistics is combined with a dose of dogmatism and the selected *a priori* distribution is more or less pressed on the reader as the one he should adopt under penalty of being considered irrational. It is a pleasure to find that no such dogmatism is present in Mosteller's and Wallace's work. Their book contains a section under the title *Choosing the prior distributions* and in the paper I like the passage:

Your final odds — posterior odds — will differ from mine, if we choose differing values for p_1 . In some problems the choice of p_1 might be made on the basis of objective frequencies. But in others, personal degree of belief may be involved.

Not only do the two authors make no attempt to press upon the reader their chosen prior probabilities needed for the application of the Bayes formula, but also go to some lengths in order to present the solution of the discrimination problem through means not involving the Bayes theorem. Thus, the reader is given the opportunity to make up his own mind both as to the method of study and as to the subject matter, namely the authorship of the disputed papers.

The third point of excellence of Professor Mosteller's and Professor Wallace's study that I wish to mention is the attention they pay to the effect of selection of words that appear to be particularly suitable for discrimination. The basic assumption underlying the paper and the book is that, for a given author, the number of times a given word is used in an article is a random variable, say X . Thus, if the treatment of the problem of discrimination is based on words especially selected because the corresponding values of X in articles of known authorship are either large or small, then this circumstance must be subsequently taken into account.

There are many more characteristics of the paper and book that deserve hearty approval.

2. "Inferential" and "behavioristic" approach to problems of statistics. In their introductory section, the authors state that their interest in the controversy over *The Federalist Papers* is largely incidental and that what they are really after is statistical methodology with reference to problems of discrimination. Here they make a subdivision. On the one hand, there is the approach through the application of Bayes' theorem and, on the other, "the more usual methods of discrimination," presumably, without the use of the Bayes theorem. The rest of my discussion will be given to the categorization of statistical methods.

It is my impression that the use of Bayes' theorem or the abstention from using it is not the most revealing way of categorizing statistical studies. My own preference is to differentiate between pieces of statistical research according to their purpose. One possibility is that the purpose of a paper is to provide the reader with reasons for believing in certain statements. Another possibility is to provide the reader with a methodology of making decisions, whether combined with beliefs or not, that would insure something calculable about the frequency of errors. It seems to me that there is quite a fundamental difference in these two approaches and that this difference deserves attention and terminology. Thus, the first of the approaches mentioned, appealing to intensities of belief of the reader, might be called *inferential approach*. The second approach, characterized by assertions, or theorems, that a specified method of deciding will insure such and such a limit to the frequency of error, perhaps the minimum frequency of error (or some such), refers to actions or decisions, rather than to beliefs, and for this reason has been labeled *behavioristic*.

As is well known, each of these two approaches to statistical problems may well involve the use of Bayes' theorem. In particular, it is my impression that the work of Professor Mosteller and Professor Wallace is inferential in all of its parts, that using the theorem of Bayes and the one concerned with the more classical methods of discriminatory analysis. On the other hand, the Empirical Bayes' Procedure invented by Herbert Robbins is definitely behavioristic. In fact, even though it is concerned both with the *a priori* and the *a posteriori* distributions of the unknown parameters, it represents a solution of the problem of how to behave in order to approach the minimum mean square error in estimation. Professors Mosteller and Wallace were most successful in providing arguments for the belief that most of the disputed Federalist Papers have been authored by Madison. I assert this wholeheartedly because, after reading the paper and parts of the book, I personally feel completely convinced that, with the possible exception of Nos. 55 and 56, all the disputed papers were written by Madison. However, our thinking about statistics will gain in clarity if the achievement of a strong belief in a reader is distinguished from the application of a purely behavioristic theoretical result. With reference to the classification problem at hand this result might look as follows.

3. Behavioristic statistics. Let us visualize the lifetime of a statistician. I like to compare it with the activity of a gambling casino involving a number of different chance mechanisms that decide on winnings and losses. These mechanisms may include roulette, dice, cards, etc. Another plausible comparison is with a broad insurance establishment handling a great variety of risks, perhaps beginning with life contingencies and ending with fire and theft. In both cases, the prime concern of the management is that, whatever the kind of game or whatever the coverage of the insurance, the long-range frequency of losses or of claims does not exceed a preassigned limit.

In his lifetime the statistician is faced with a great variety of problems which may refer to authorship of articles, the complicity in crimes, the effectiveness of treatment of cancer or the expansion or stability of the universe. Each time, as in a gambling casino or in an insurance office, there is involved a chance mechanism the nature of which changes from problem to problem and is only partially known. For me, apart from the varying complexities of these mechanisms, the difference between them is no more than the difference between the mechanism of roulette and the mechanism of dice. If the two mechanisms can produce certain happenings with the same probabilities (long-range relative frequencies), the physical differences between them (roulette or dice) can be ignored.

Each of the problems that the statistician is called upon to handle requires, in its final stage, a decision. (Do we decide that Madison wrote paper No. 55? Do we administer X-rays to patient A ill with cancer?) This decision

corresponds to the acceptance of a bet and to the issuance of a policy covering a given risk.

The operations of the gambling casinos and the insurance institutions are based on certain rules intended to guarantee low frequencies of unsatisfactory outcomes. These rules are subject to continuous study and alterations on the precept, more or less: however, the gambler may choose to set his stakes within the framework of the rules of the house, in the long run the house shall be the winner. More precisely, the long-run frequency of my winnings (or the long-run average of my winnings, etc.) shall be this: owing to the basic nature of chance mechanisms, the outcome of each particular bet or game or insurance contract must remain uncertain until the decisive moment. However, the particular game or particular policy are secondary to the long-range effect.

Quite similarly, the practising statistician may set for himself the problem of devising rules that he will later apply to all the variety of applied problems with which he will be called to deal. This dealing with official problems will inevitably involve committing errors of judgment or of decision. According to the statistician's personal attitudes, all the errors may be of the same or of varying importance. To be satisfactory, a rule devised by the statistician must control the frequency of the more important errors to a properly low level. At the very least, the rule should have known limits of the frequency of errors.

When the rules in question are devised, they can be applied to relevant problems. As in the case of games and insurance, the outcome of a particular application of the rules will remain unknown at least for some time. However, if the statistician's deductions were correct, he may be sure of the long-range frequencies of errors.

4. Behavioristic treatment of discriminatory analysis. From the behavioristic point of view the problem of discrimination has two facets. First, there is the question of the optimal discrimination criterion. The question is, what arithmetic should be performed on the observational data in order to arrive at the conclusion that, say, a given paper has been authored by Hamilton rather than by Madison. The natural behavioristic requirement regarding this criterion is that its application minimizes the relative frequency of misclassification.

The second facet of the problem is the calculation of the performance characteristic of the optimal (and perhaps also of certain alternative) discrimination procedure.

Both facets are very important. Both are difficult, particularly in those cases where, as in the case of authorship of the Federalist Papers, the functional form of the distribution of the observable random variables is at least uncertain. Professors Mosteller and Wallace emphasize that their numerical results depend to a considerable extent on whether one assumes the Poisson or the Negative Binomial Distribution.

Ten years ago the problem of optimal nonparametric discrimination was treated, with a considerable amount of success, in two joint papers by Evelyn Fix and Joseph L. Hodges, Jr. Unfortunately, the publication of these two papers is not quite satisfactory. They appeared as Project Reports of the School of Aviation Medicine, Nos. 4 and 11, respectively, and few statisticians are aware of them. If Professor Mosteller's and Professor Wallace's work on their book can still stand some extension, they may wish to take the Fix-Hodges papers into account.

The problem of the performance characteristic of the classical discrimination function was the subject of study by Professor T. W. Anderson and by Professor Rosedith Sitgreaves.

Two more points relevant from the behavioristic point of view may be mentioned. Both refer to the circumstance that in the problem of the Federalist Papers there is not one but several papers to assign either to Hamilton or to Madison. The first behavioristic question to arise is as follows: Suppose we choose a limit of probability of error, say 5 per cent. Which of the disputed papers can be assigned to either of the two authors, using the classical Fisher discrimination function so that the risk of at least one error does not exceed 5 per cent? It should be remembered that this is a question of multiple comparisons dealt with by Tukey and Scheffé.

The second question relates to the compound statistical problems recently studied by Blackwell, Robbins and Stein. It is this. Instead of judging each of the disputed papers separately, on its own merits, one might attempt a simultaneous joint classification, so that the verdict on any one paper depends upon scores of other papers. Can one use this device to decrease the expected overall frequency of error?

5. Concluding remarks. Upon reading what has been written above I find that the reader may obtain the impression that the paper and the book of Professor Mosteller and Professor Wallace do not contain behavioristic elements at all. I wish to emphasize that it is not my intention to give any such impression. The behavioristic analysis is given in Section 7 of the paper. Also I find that this analysis is excellent. The main purpose of my remarks is to emphasize the difference between "inferential" and "behavioristic" approaches to the problem. It appears to me that this distinction is more informative than that based on the application or nonapplication of Bayes' theorem.

I conclude by wishing the two authors the best of success with their very interesting book.

