Title

A Probabilistic Evaluation of the Quality of Evidence

Author

Hoffman, Richard B.

Publication Date

1967

Peer reviewed

RICHARD B. HOFFMAN

A Probabilistic Evaluation of the Quality of Evidence

Center for Planning & Development Research: Institute of Urban & Regional Development
University of California, Berkeley

Reprint No. 28

Reprinted from: Law Enforcement Science and Technology,

Volume 1, 1967; pages 425 – 429

Proceedings of the First National Symposium on Law Enforcement Science and Technology sponsored by the U.S. Department of Justice, Office of Law Enforcement Assistance and I.I.T. Research Institute, Law Enforcement Science and Technology Center.

The Center for Planning and Development Research is a unit within the Institute of Urban and Regional Development. It is devoted to the study of a wide range of activities in environmental design and urban and regional development. Researchers in the Center are engaged in investigating the processes of urban and regional growth, the consequences of alternative patterns of growth, and the scientific bases for development policy and planning.

Toward these ends, current research is directed to improving theories and methods of planning, to building models capable of simulating urban growth processes, and to formulating improved criteria for urban and regional development policies. Fields of interest to the Center include the development of new and established communities, housing and renewal programming, the effects of technological development on cities and urban life, metropolitan planning, the economics of city-building in the United States and in developing nations, and the social planning and human development issues that accompany urbanization.

The Center contains the Architectural Design Laboratory, which is devoted to research in building technology, the behavioral bases for design, and design processes.

A probabilistic evaluation of the quality of evidence*

by RICHARD B. HOFFMAN
Center of Planning and Development Research
University of California
Berkeley, California

*This work was supported in part by National Aeronautics and Space Admistration Grant No. NGR-05-003-125 (Technology and Urban Management) to the Space Sciences Laboratory and the Center for Planning and Development Research, University of California, Berkeley. The author is also affiliated with the Department of Management Sciences, California State College at Hayward.

INTRODUCTION

The kind of knowledge which is supported only by observations and is not yet proved must be carefully distinguished from the truth; it is gained by induction, as we usually say. Yet we have seen cases in which mere induction led to error. Therefore, we should take great care not to accept as true such properties of the numbers which we have discovered by observation and which are supported by induction alone. Indeed, we should use such a discovery as an opportunity to investigate more exactly the properties discovered and to prove or disprove them; in both cases we may learn something useful.[1]

In every case wherein we rely upon observations to determine a causal relationship, we must be prepared to accept the possibility of being in error in this determination. The consideration of those observations which are to be used as evidence in terms of our existing legal system and the probability that the implications which are drawn from these observations are in error provides a framework under which we may evaluate the quality of a given set of observations which may alternatively be classified as evidence. To the degree that there exists a conceptual framework or theoretical mechanism to support the implications or deductions which are drawn from the set of observations, we shall have a basis to establish the possible validity of the deductions drawn from this set of observations. To the degree that we are able to determine the relative frequency of these observations in the "relevant population"* and to determine the appropriate measures of variability, we shall be able to find the probability that we are in error in assuming that the deductions from the evidence or observations are valid.

In any situation wherein we are forced to rely on observations, i.e., evidence, to determine the guilt or innocence of an individual or a group of individuals, there exists the probability that the implications drawn from the evidence may be in error. If such an implication is drawn, the error may be of two varieties, (1) an individual may be assumed innocent who is in fact guilty and (2) an individual may be assumed guilty who is in fact innocent. In both cases there are costs to society: (1) If an individual who was guilty of a crime is released, he may be likely to commit new crimes and additional effort may be expended by law enforcement agencies in trying to solve the crime, and (2) if an individual who is innocent is detained he may suffer loss of reputation, position, and income, and society will have to bear the cost of trial and detention, and if he is convicted the costs of incarceration, probation, and rehabilitation, and possible support of his family.[2] It is recognized by all law enforcement agencies that many "guilty" individuals are allowed to go free, due to an inadequate number of observations to ascertain the guilty individual, or insufficient quantity of evidence to obtain an indictment or conviction. If we desire to improve upon our system of gathering and evaluating evidence with regard to guilt or innocence it would seem first that we should determine the probability that an individual is in fact guilty given some set of observations

*This use of population is in the statistical sense of the totality of elements which are under discussion and about which information is desired, rather than in a strictly demographic sense. Mood, A. M. and Graybill, F. A., *Introduction to the Theory of Statistics* (McGraw-Hill Book Company, New York, 1963), 2nd ed., p. 141.

used as evidence and the probability that the same individual is innocent given the same set of observations used as evidence.

Since an explicit discussion of the means of evaluating the effects of a given set of evidence or a particular item of evidence has not received wide attention in the literature, it should be necessary to devote the first section of this paper to setting up definitions and notations. In the later sections a discussion of the means by which a trade-off function for the two types of error involved in acceptance or rejection of a given unit of evidence are considered. A word should be said about the reasons for the use of this method of evaluating evidence. Implicitly, organizations which attempt to administer the system of criminal justice, be they oriented to law enforcement, or to judicial proceedings, do in fact accept and reject evidence with some probability of error without knowing (1) the probability of the various types of error or (2) the cost of making these errors. It is suggested by this paper that these errors be explicitly determined, and the appropriate costs be explicitly evaluated, and from this a determination of a best choice among the available actions be determined.

Definitions and notations

A basic problem which must be faced by one writing on a subject which is not commonly found in the literature of a particular subject, is that of finding a notation which will adequately represent the circumstances and still be understandable to the wide majority of readers in this area. Since this is but an introduction to the problem of evaluating evidence in a probabilistic framework, an attempt is made to provide a representation of the problem that is as simple as possible, and any attempt to treat the subject rigorously and still provide a general framework will fail to meet the basic objective of communicating the basic framework to the widest possible readership.

The initial definitions are those of (1) probability, (2) joint probability, (3) conditional probability, (4) marginal probability, (5) Type I error, and (6) Type II error.

(1) Probability: We say that P(a) is the probability that event (a) obtains (occurs) or that a set of evidence [a] is associated with a given individual and this probability can be expressed as the sum of the probabilities of a set of sample points that are associated with the event (a). Symbolically

$$P(a) = \sum \frac{P_i(a)}{p_i}.$$ [3]

(2) Joint probability: Often we are concerned with the probability of related events or bits of evidence rather than with just one event or bit of evidence. We may be interested in knowing whether a_1 and a_2 can both occur and what is the probability of a_1 and a_2 occurring. This joint event is denoted by the product $(a_1\ a_2)$ and its probability by $P(a_1, a_2)$. $P(a_1, a_2)$ can be interpreted as the probability that a bit or set of evidence (a_1) and a bit or set of evidence (a_2) is associated with a given individual. We are also interested in knowing whether at least one of the events (a_1) or (a_2) will obtain or alternatively whether the evidence (a_1) or the evidence (a_2) is associated with a given individual. This event will be denoted by the sum $(a_1 + a_2)$ and its probability by $P(a_1 + a_2)$. At least one of the two events will occur if (a_1) occurs and (a_2) does not, or if (a_2) occurs and (a_1) does not, or if both (a_1) and (a_2) occur. Then to derive the formula for the probability of $(a_1 + a_2)$ we must take into consideration (A) the probability that (a_1) occurs plus (B) the probability that (a_2) occurs less (C) the probability that $(a_1$ and $a_2)$ occur simultaneously. However, to derive a formula for $P(a_1a_2)$ in terms of the probabilities of single events, we must introduce the notion of a conditional probability.[4]

(3) Conditional probability: Suppose we are interested in determining whether (a_2) will occur subject to the condition that (a_1) is certain to occur. This is equivalent to determining whether evidence (a_2) is associated with an individual when we are sure that evidence (a_1) is associated with that individual. A specific example may be as follows: A crime was committed by a blonde male, the event or evidence (a_1), in the West Side of Chicago on February 10, 1967, the event or evidence (a_2). Our suspect is a blonde male, hence we know (a_1) has occurred with certainty or alternately we know evidence (a_1) is associated with the individual (the suspect) for certain. Hence we wish to determine whether (a_2) is satisfied, i.e., whether the suspect was in the West Side of Chicago on the date in question. Since (a_1) is certain to occur only when the sample space is restricted to those points lying in the region (a_1), it is necessary to consider how the probabilities assign the points in this newer sample space. If originally, a sample point in (a_1) had been assigned, twice as large a probability as another point in (a_1), then it should be assigned twice as large a probability in the new sample space also. Hence the probability of the new event equals the sum of the p_i associated with the (a_2) divided by the probability of (a_1). Now that we have determined the new sample space, we can calculate the probabilities in the usual manner by applying the basic definition of probability. Such probabilities will be conditional probabilities subject to the occurrence of (a_1). If the probability of (a_2) will occur subject to the restriction that (a_1) is certain to occur is denoted by $P(a_2 \mid a_1)$ then it follows from this definition that

$P(a_2 \mid a_1) = \frac{\Sigma p_i(a_1 \text{ and } a_2)}{P(a_1)}$ and this reduces to the conditional probability or $P(a_2 \mid a_1) = \frac{P(a_1, a_2)}{p(a_1)}$.[5]

(4) Marginal probability: A comment should be made regarding marginal probabilities. They do not provide anything in terms of the meaning of probability, for it describes only the way in which a probability of an event is or may be computed.[6] The marginal distribution is then defined as the $P(a_1) = \Sigma P(a_1, a_{2i})$. In a similar manner the probability or marginal probability of (a_2) may be calculated.[7]

(5) Type I error or alpha (α) error is the error in rejecting a hypothesis H_0 the null hypothesis when H_0 is true. This can be viewed as assuming that a guilty individual is not guilty or assuming that a man is innocent when he is in fact guilty.

(6) Type II error or beta (β) error is the error in accepting H_0, the null hypothesis when H_0 is false. This can be interpreted as assuming that an innocent individual is not innocent or that a man is guilty when he is in fact innocent.

Although the latter two definitions may initially appear to be confusing, their basis is well founded. Historically statisticians have generally believed that in almost all two-action problems requiring statistical analysis, one of the two hypotheses will be such that the consequence of rejecting it when true will be far more serious than the consequence of accepting it when false. So that there will be a very strong tendency to accepting one of the two hypotheses under all but the most extreme conditions, and attention is so focused by calling this favored hypothesis the tested or the null hypothesis. Since it is considered good practice for a researcher to raise strong arguments against his own hypothesis, he sets up the null hypothesis in an attempt to prove that his research conjecture is false. He then tests his null hypothesis at some preassigned significance level alpha (α). Hence, he desires that the null hypothesis is rejected and his research conjecture will be thereby substantiated. The smaller the significance level chosen, the stronger are the results of the test, and the more impressive are his conclusions.[8]

The trade-off of alpha and beta error

Although it is unrealistic to assume that law enforcement or judicial agencies explicitly evaluate the probabilities of errors of the first kind and second kind, it does appear that the assumptions upon which each of these two systems base their actions are such that each group attempts to minimize one of the kinds of error.

In our present system the police agencies, as the enforcement arm of society, view the problem as one of minimizing the probability that an individual guilty of a crime is allowed to immediately return to society without that individual making some repayment for his transgressions to society. And alternately, the system of the courts view the problem as minimizing the probability that an innocent individual is forced to pay the penalty of public disgrace and/or incarceration for an alleged set of criminal acts which he may not have committed.

A feature of this system wherein two agencies have directly conflicting goals may operate in a manner analogous to a market mechanism. In that the system efficiently finds an optimum solution without an explicit calculation of the functional relationships of the parameters and the values these parameters assume. However, it does not appear necessary or even obvious that the existing system with its analogous inherent conflict of goals will be as efficient as a market pricing system is for economic allocation. It would appear that there are at least two basic problems not considered with this system and hence there may be a wide range of variability in this "solution" obtained by each of the jurisdictions, for (1) the system does not explicitly consider (A) the cost of incarceration, (B) the damage done to an individual who has been convicted of a crime of which he is not guilty in terms of his ability to exist in society in the long run and to be a productive member of that society, and (C) the precise effect of penalties as deterrents to crime, and (2) the system is unable to deal with the problems and costs to individuals who are arrested but not tried. (An additional problem beyond the scope of this paper is the determination of the effects of negotiations, adjustments, and the threat positions in a Nash bargaining sense[9] as reflected by variations in the crime(s) for which the individual is initially charged and the actual crime for which he receives a penalty.

Many of the problems of uncertainty and variations in individual preferences which make for indeterminancy in the relatively well behaved economic markets exist in the "system" which is used by society to administer justice. The uncertainty is derived from the individual's imperfect knowledge as to whether he will be apprehended, on what crime will he be charged, how his conflict with the law enforcement agencies will affect his public and private life, what type and kind of penalty will be assessed if he is found guilty (incarceration in terms of jurisdiction, level of security, years to parole, and length of sentence), and what is the probability of acquittal. The individual preferences and values assigned to each of the above will, in addition, lead to even greater possibilities for indeterminancy in the solution reached by the existing system for the administration of justice.

Given the above which indicates the wide range of alternative forms in which society extracts its repayment, the probability is indeed small that the system will obtain anything approximating an optimal solution in terms of any acceptable criteria of social welfare for society.

If the implicit process is believed to be less than effective, the next question is: Would explicit consideration of the parameters be (1) an improvement? and/or (2) effective in itself?

Although there might be some improvement on the first question, there would be a new problem in that it would be necessary to then specify "appropriate" (statistical) critical regions in terms of the error being measured, and the explicit consideration of the errors would not seem to be a sufficient condition for us to assume that society would be effectively served by the "system of justice."

It does seem that we might attempt to determine the expected costs of each of the types of error and then perhaps attempt to minimize the total expected costs of errors made by the system.

If (α) is the probability that allows an error of the first kind and a "criminal" is released to society with a cost (C_1) and (β) is the probability that an individual is convicted of a crime he did not commit with cost (C_2) there are a number of approaches to the development of measures for the expected from this decision process.

We can attempt to minimize $E(\alpha)$, the expected cost of (α) error, given some maximum allowable level of (β) or $E(\beta)$, the expected cost of (β) error, and alternatively we can attempt to minimize $E(\beta)$ given some maximum allowable level of (α) on $E(\alpha)$. This method is equivalent to that described by Hoel[10] as the "best" test with one type of error agreed upon, and wherein the method of constructing the best test depends on the use of the Neyman-Pearson Lemma.[11]

Savage[12] suggests that the problem consists of not two but three kinds of error. In which the "type three error" is to make a decision on the basis of the existing evidence or observations and to discontinue the search for additional evidence or information, when in fact search "should" be continued.

The hypothesis testing problem, according to Savage, is an observational decision problem derived from two basic acts, F_0 and F_1. These two basic acts are called accepting and rejecting the null hypothesis respectively. If there is a set S, partitioned into the three events H_0, H_1, and H_n where (n) is unimportant and often either ignored or made empty by some manipulation of the definitions, then H_n cannot truly be called a hypothesis, and accepting H_n can be only a temporary decision for we must continue to search until (n) is in fact empty.[13] This then leads us to sequential analysis which refers in principle to the theory of observational programs wherein the selection of the observations to make in later phases of the program is dependent upon what has been observed in the earlier phases. Such behavior is commonplace in everyday life. For example, we look for something until we find it, and then we discontinue the search.[14] The implication for the process in which evidence is evaluated is that if the total expected cost of errors of both types cannot be reduced to acceptable levels we should continue to search for additional observations or evidence. It would appear that it may be possible for continued search for evidence to yield an (n) which is not empty and yet acceptable levels of (α) and (β) will not have been attained. Hence we should, before engaging in additional search for evidence, calculate the expected value of the information and the net gain from the search for additional evidence. This can be accomplished using the methods proposed by Raiffa and Schlaiffer wherein the conditional and expected value of perfect information and the expected cost of search for evidence are the important variables.[15]

Given that we feel that the search for additional evidence is desirable in terms of the expected net gain of search for new evidence, the question is in what areas is search likely to be most successful. We must realize that a positive expected net gain does not guarantee improvement for a specific search but only indicates that in the "long-run" similar endeavors will, on the average, lead to improvement.

However, the work of Marschak on the economics of information has an interesting interpretation in terms of the search for evidence.

> *Given (that there exists) a probability measure P defined on the set X of states, with P given the probability distribution over any partition Z of X is determined. In particular, P determines for a given pay-off condition (Ω), the probability distribution over the set Z^{Ω} of pay-off-relevant state descriptions. Also, given a channel h, P determines the probability distribution over the set Y^h of potential messages that may be received by the decider. . . . In particular the measure P on X determines the bivariate distribution of (Z^{Ω},Y^h), for every pay-off condition Ω over every channel h.*[16]

One way of looking at the problem of gathering evidence statistics is in terms of the choice of that channel or that source of observations which will allow the decision-maker to end his search in the shortest possible time if he so desires. The implication is that there exists a real state of the world which is unknown and that the decision-maker can determine this state of the world with varying cost depending upon the choice of channel chosen. If in fact he chooses inappropriate means to pursue the collection of new or additional evidence this can be viewed as the choice of poor channels of communication so that the problem may be viewed as choosing the channel which will minimize the combined cost of information (search for and assimilation of new evidence) and the errors in decision-making.

Are there legal, ethical and moral questions to be considered in the use of probabilities in determining guilt or innocence of individuals? Will not individuals feel that effectively the government and/or society is gambling with their lives in determining their guilt or innocence? Yes, this is true, but it can be demonstrated that the existing system (1) convicts more innocent people and (2) allows more guilty to go free and (3) given the cost structure earlier described, we should find that the costs to society under the existing system are far higher for both reasons, i.e., the cost in assuming an innocent individual is guilty may be higher than the costs from guilty individuals being released into society. Both (α) and (β) error may be higher under the existing conditions. Hence, we should find that the proposed system can be a considerable improvement. Second, there can be a rather great improvement in the quality and level of police service to society in that resources will be more effectively allocated (1) by reducing the cost of the search for the evidence needed in those cases of obvious guilt, (2) by allowing certain individuals who are obviously innocent to go free at a much earlier stage with far lower costs including that of temporary incarceration, and (3) only providing increased search for evidence for those individuals for whom it appears the probability of his (a) being innocent or (b) being guilty is already relatively high. So that when the alpha error is not low enough to permit the release of the individual immediately, and yet in terms of the information implied by the beta error there exists a high probability that the suspect is innocent, the officer would seek primarily for information to confirm this innocence. In those cases in which the alpha error were relatively high, and the beta error fairly low implying the greater likelihood of guilt, alternatively the investigating officer would primarily seek information which would tend to confirm the suspect's guilt.

REFERENCES

1. EULER
   Specimen de usu observationum in mathesi pura
   Opera Omnia series 1 vol. 2 p 459 in Polya, G Mathematics and Plausible Reasoning: Induction and Analogy in Mathematics vol. I Princeton University Press Princeton, New Jersey 1954
2. R. B. HOFFMAN
   The transfer of space and computer technology to urban security
   Fall Joint Computer Conference 1966 AFIPS Conference Proceedings vol. 29 p. 525 1966
3. P. G. HOEL
   Introduction to mathematical statistics
   Wiley New York 3d ed. p. 7ff 1962
4. *Ibid.*
5. *Ibid.*
6. J. W. PRATT, H. RAIFFA and R. SCHLAIFFER
   Introduction to statistical decision theory (a preliminary edition.
   McGraw-Hill Book Company, Inc. New York chap. 5 p. 20 1965
7. P. G. HOEL *op. cit.* p. 7ff
8. J. W. PRATT *op. cit.* appendix 3 pp 18-20
9. R. D. LUCE and H. RAIFFA
   Gains and decisions
   Wiley New York 1957
10. *Op. cit.* pp. 214-15
11. J. NEYMAN and E. S. PEARSON
    Testing statistical hypotheses in relation to probabilities a priori
    Proceedings of the Cambridge Philosophical Society vol. 29 pp. 492-516 1933
12. L. J. SAVAGE
    The foundations of statistics
    Wiley New York pp. 247 ff. 1954
13. *Ibid.*
14. *Ibid.* pp. 142ff
15. H. RAIFFA and R. SCHLAIFFER
    Applied statistical decision theory
    Harvard University Boston pp. 89-91 1961
16. J. MARSCHAK
    Problems in information economics
    C. P. BONINI, R. K. JAEDICKE,- and H. M. WAGNER Management Control McGraw-Hill Book Company, Inc. New York pp. 38-80 1964