A THEORY OF SUBJECTIVE INFERENTIAL PROBABILITY

By

WILLIAM TALBERT TUCKER, JR.

Bachelor of Mechanical Engineering
Georgia Institute of Technology
Atlanta, Georgia
1957

Master of Science
Southern Methodist University
Dallas, Texas
1961

Master of Science
Southern Methodist University
Dallas, Texas
1963

Submitted to the faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
July, 1966

A THEORY OF SUBJECTIVE INFERENTIAL PROBABILITY

Thesis Approved:

____
Thesis Adviser

____

____

____

____

____
Dean of the Graduate College

ACKNOWLEDGEMENTS

I wish to express my gratitude to Dr. J. Leroy Folks for serving as chairman of my advisory committee, for his suggestions that led to the problem considered in this paper and for his guidance and criticism during the preparation of this paper.

I, also, appreciate Professors John E. Hoffman, Robert D. Morrison, Paul E. Torgerson, and David L. Weeks serving on my advisory committee.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

In this paper a subjective inferential theory of probability will be developed and investigated. In order to develop this concept of probability, the philosophical background of probability theory in general must be discussed.

Savage [33] discusses three types of probability: relative frequency or objective, logical or necessary, and personal or subjective. The relative frequency concept, championed by von Mises [35], and employed by Fisher [10], Neyman [25], and most modern statisticians explains probability as the relative frequency of a particular event in a certain type of sequence of events or in a set of events. In the logical concept, probability is considered to be basic to the extension of deductive logic and forms a basis of inductive logic. As such, probability is a measure of a logical relationship between one proposition regarded as known or given and another. Savage [33] attaches the title necessary since the probability of proposition a on the evidence b is a logical necessity to be deduced from the logical structure of the propositions a and b. One of the most extensive developments is that of Carnap [5]. In his formulation probability is a measure on the truth set of a proposition. In the developments of Keynes [18] and Jeffreys [15], the probability of a given b is taken as fundamental and the theory developed axiomatically. Cox [7] working with

Boolean algebras also develops a logical type of probability. One of the best known expositions on personal probability is that of Savage [32]. He generalizes a previous formulation due to de Finetti [9]. Savage and de Finetti interpret probability as a degree of belief of a given person concerning the occurrence of some event. De Finetti [9] gives two developments. One is based on a set of axioms and the other on the betting odds one would give on the occurrence of an event. One of the first developments of subjective probability is that of Ramsey [30] who used betting odds and utility in an axiomatic approach. Koopman [20] presents a system of personal probability that is somewhat of a compromise between the logical and personal concepts. He employs a conditional probability, that is, the probability of a given b but he allows his probability to change as the beliefs of a person change.

The development of these three types of probability has been influenced by the intended application or use of the probability system. The two primary uses of probability are in inference and decision making. In the inference viewpoint, probability is used as a weighting to be attached to various implications. In the decision viewpoint, a person has a set of actions which can be taken and a decision rule selects one of these actions. Probability theory is employed to aid the decision maker in the choice of a decision rule. Most writers on probability have been primarily interested in inference since inductive behavior (decision making) as a serious concept was first put forth by Neyman [24] in 1938. Statistical decision theory due primarily to Wald was first outlined by him in 1939 (Wald [37]) and was only fully developed in his book (Wald [38]) in 1950. Further, decision theory as related to hypothesis testing was first formulated by Neyman and Pearson in 1932 (Neyman and Pearson [23]). Thus, it is seen that decision making as an important concept is relatively young while the problem of inductive inference has been studied since the days of Jacob Bernoulli.

Certain criticisms can be leveled against each of the types of probability either in their use or in their philosophical principles. The rigidity imposed by the logical formulation vitiates its use in practice. Logical probabilities may well exist but they are seldom known and virtually impossible to estimate. If, as in the case of Carnap's system, probability is a measure of the truth set of a proposition, then the probability is extremely difficult to use in practice since rarely is the truth set of a proposition known. The trouble with the relative frequency approach comes in trying to apply it. While von Mises rejected the measure-theoretic concept of probability, H. Geringer in von Mises [35] has formulated the relative frequency theory of probability in measure-theoretic terms similar to the theories of Cramér [8] and Kolmogorov [19] so that the relative frequency viewpoint should now be acceptable to the strict theoreticians. However, since with certain restrictions the probability is the limit of the proportion of occurrences of a given event in an infinite sequence of events, the observation of no finite sequence can actually determine the true probability. This criticism is excellently put by N. R. Hanson in the introduction to Keynes [18] where he discusses the problem of how to find the true probability of the occurrence of a deuce when rolling a die. He writes:

> Two thousand dice rolls of two thousand deuces may not absolutely warrant the die's being loaded. Perhaps two thousand is an insufficiently short run, in which case it is irrelevant as an index of bias. ... However, if such a run cannot establish that the die *is* loaded, then a run of two thousand rolls out of which 333 *are* deuces can have no relevance for showing that the die is fair. And all the apparatus of *convergent* and *irregular* "collectives" fails to surmount this objection.[1]

Thus, in practice the relative frequency concept has serious drawbacks.

With the exception of the modern subjective writers, all of the proponents of the inference viewpoint have employed either relative frequency or logical probability as a basis of their theories. Fisher [10] gives an excellent discussion of inference employing relative frequency probability and Jeffreys [15] does the same while using logical probability. Since the logical concept has almost universally been rejected and the relative frequency theory is subject to serious philosophical criticisms, it is seen why the study of inductive inference gave way to the study of inductive behavior. However, the modern subjective writers have ventured into the field of inductive inference again but with a formulation somewhat different than that of either Fisher or Jeffreys. Since the modern subjective theory of inference stems indirectly from Wald's decision theory, (cf. Savage [32] who draws heavily from Wald) discussion of subjective inference will be deferred until an outline of decision theory has been given. It is reasonable to assume, however, that the widespread acceptance by both objectivists and subjectivists of Wald's ideas is due to the unfortunate

---

[1] N. R. Hanson, "Introduction to the Torchbook Edition," *A Treatise on Probability*, by J. M. Keynes (New York, 1962), p. vi.

position of inferential theory in either the relative frequency approach or as a partial implication in the logical approach.

The review of statistical decision theory will be a brief outline of the theory of Wald [38]. A comprehensive review of the statistical decision problem is given by Stewart [34]. In the structure of a statistical decision problem, a "decision maker" is faced with the choice among a set of "acts" such that the "consequence" of any act depends upon the unknown "state of nature." It is possible to gain knowledge about the unknown true state by performing an "experiment." The information contained in the "outcome" of the experiment then is employed in "selecting" the "best" act.

In the formal structure, let the set of acts be denoted by A and a particular act by a. The set of states of nature is designated by Ω with elements ω. A real-valued, bounded loss function defined on $A \times \Omega$ and denoted by $L(a, \omega)$ is the value of the consequence resulting from taking action a when nature is in state ω. This loss function is the negative of the utility. A definition of utility is given by von Neumann and Morgenstern [36]. Wald's formulation is a special case of game theory as introduced by von Neumann and Morgenstern. The possible outcomes of the experiment form a set X with elements x. The knowledge relating Ω and X is given by a conditional objective probability measure defined on $X \times \Omega$. For each $\omega \varepsilon \Omega$, $P_\omega(x)$ denotes this conditional probability measure. A function from X into A designated by $d(x)$ and called the decision function or rule is the mechanism by which an act is selected. The set of all decision rules is called D. It should be noted that in this formulation $P_\omega(x)$ is, in general, an approximation. If $P_\omega(x)$ is a "poor" approximation, then possibly many inopportune acts will be taken but that is all that can happen.

One of the most important contributions of Wald's theory is the systematic study of "best" decision rules. In comparing decision rules the expected loss or "risk" is employed. This risk is given by

$$R(d, \omega) = \int_X L[d(x), \omega] \; dP_\omega(x),$$

and is a function of only the decision rule, d, and the true state of nature, ω. Wald's use of risk was arbitrary but von Neumann and Morgenstern were able to show that if a decision maker can "order" his consequences, then a loss function (negative utility) exists and with respect to this order he behaves as if he were a minimizer of expected losses. Thus, it is reasonable to use risk to compare decision rules and given two decision rules d_1 and d_2 to consider d_1 to be at least as good as d_2 if

$$R(d_1, \omega) \leq R(d_2, \omega)$$

for every $\omega \varepsilon \Omega$. Further, d_1 is "better" than d_2 if it is at least as good as d_2 and for some $\omega \varepsilon \Omega$,

$$R(d_1, \omega) < R(d_2, \omega).$$

A decision rule d is called "admissible" if there is no decision rule better than d. A class of decision rules is "complete" if for every rule outside the class there is one in the class which is better. A complete class such that no proper subset is a complete class is called a "minimal" complete class. It is obvious that every complete class of rules includes the class of admissible rules. Further, it can be proven that if a minimal complete class exists, then it is identical to the class of admissible decision rules. The admissible class is of particular importance since the members of this class are rules that cannot be improved.

In developing his theory Wald employed prior distributions. He defined a prior distribution as a probability measure on some Borel field of subsets of Ω. However, a prior distribution was used by Wald only in finding complete classes of decision rules and minimax solutions. Wald made no attempt to interpret a prior distribution as reflecting the decision maker's prior beliefs or knowledge. A prior distribution will be denoted by $P(\omega)$. Further, the average over the prior of the risk, denoted by $B(d, P)$, will be called the "Bayes risk." That is,

$$B(d, P) = \int_{\Omega} R(d, \omega)\, dP(\omega).$$

The decision rule d_p is the "Bayes solution relative to P" if

$$B(d_p, P) \leq B(d, P)$$

for all $d \varepsilon D$. The decision rule d_0 is the "minimax solution" if

$$\operatorname*{Sup}_{\omega} R(d_0, \omega) \leq \operatorname*{Sup}_{\omega} R(d, \omega)$$

for all $d \varepsilon D$. Wald's chief objective was in finding minimax solutions. Unfortunately, a minimax solution is quite often impractical. This results from the overly pessimistic course of action of a minimax solution in attempting to insure against the worst that nature can possibly do. An excellent example reflecting this pessimism is given in Stewart [34]. One further definition is required and it is that of a least favorable distribution. A "least favorable prior distribution," $P_0(\omega)$, is a prior distribution such that

$$\inf_d B(d, P_0) \geq \inf_d B(d, P)$$

for all P.

Under general conditions Wald was able to show

(1) the existence of Bayes solutions (relative to any prior), minimax solutions, and least favorable prior distributions,

(2) that a minimax solution is also a Bayes solution relative to a least favorable prior distribution, and

(3) that the class of all Bayes solutions corresponding to all possible prior distributions is complete.

The class in part (3) becomes minimal complete when the Bayes solutions which are not admissible are excluded. For a special case Wald gave sufficient conditions for the admissibility of a Bayes solution. In some practical problems, it turns out that every Bayes solution is also admissible (cf. Anderson [1] page 132). In this case the class of all Bayes solutions is then minimal complete.

It is seen, then, that Wald's general solution to the decision problem gives only a class of decision rules. In order to select a specific rule, Wald employed minimax theory. As stated previously there are serious objections to a minimax solution. Thus, some other procedure must be devised to select a decision rule from the class of (minimal) complete decision rules.

Investigation of methods of selecting decision rules is covered in the literature under principles of choice or decision under uncertainty. One principle result stands out. In terms of the preceding developments it is this. If there exists a complete order relation in D that satisfies certain axioms representative of the decision maker's "rationality" or "coherence," then there exists a prior distribution, P, on Ω such that the preferred decision rule, if one exists, is also the Bayes solution relative to P. This result is due to Blackwell and Girshick [4]. In their formulation Ω is finite. Several other systems of axioms have also been presented. The best known system is the one due to Savage [32]. In Savage's development, the primary elements are the set, Ω, of the states of nature, the set of acts, A, a set of consequences, C, with elements c, and an assignment to each pair (a, ω) of $A \times \Omega$ of a consequence from C which is denoted by $a(\omega)$. Further, there exists a complete ordering defined on A which meets a set of axioms representing the decision maker's coherence. With these definitions and axioms, Savage is able to show the existence of a probability measure (finitely additive) defined on Ω, the existence of a utility function defined on C such that a_b is preferred to $a \in A$ if and only if a_b is the Bayes solution (employing the negative of the utility) relative to the prior on Ω. Thus, Savage's result is stronger than that of Blackwell and Girshick. Also, Savage is able to show that in finding the Bayes solution, Bayes Theorem may be employed; that is, the posterior distribution can be computed from the data and the prior on Ω and this posterior can be used to find the decision rule that minimizes the risk. This decision rule is also the Bayes solution. Blackwell and Girshick prove the same result in certain special cases. While Savage orders the acts and Blackwell and Girshick order the decision rules, it can be shown that both are equivalent to an ordering of the set of all real-valued, bounded functions defined on Ω. Thus, both orderings are equivalent. However, there is one major difference between the two formulations. Blackwell and Girshick employ utility theory and objective probability in defining the risk and then develop a subjective probability while Savage develops a subjective utility and a subjective probability from a single order relation and associated axioms.

Some of the other formulations are those of Anscombe and Aumann [2], Chernoff [6], Luce and Raiffa [22], and Pratt, Raiffa, and Schlaifer [28]. These formulations are similar to either that of Blackwell and Girshick or Savage. Also, the results are similar to the "weak" version of Blackwell and Girshick or to the "strong" version of Savage.

Except for the working out of the details in a practical problem, the decision problem has been solved to almost any decision maker's satisfaction. He can follow Wald and find the (minimal) complete class of decision rules and then employ the _ad hoc_ methods of most modern statisticians to select a particular decision rule. If this does not suit his inclination, then, provided he has a preference relation on the set D, he can follow Blackwell and Girshick. Finally, if neither of these courses of action is suitable to the decision maker, then he can "go all out" and follow Savage and interpret all probabilities as subjective probabilities and use the Bayes solution to his given problem.

While the decision problem has been very elegantly solved, the inference problem still remains open. Modern subjective writers (cf. Raiffa and Schlaifer [29]) following Savage have taken the solution to the inference problem from the decision problem in the following manner. It is possible to interpret the prior distribution on Ω as expressing the decision maker's degree of belief in the various states of nature. One way to do this is to think of $P(\Omega_1)$ as giving the odds at which one would bet on the occurrence of the event $\Omega_1 \subset \Omega$. For simplicity, assume that $P(\omega)$ and $P_\omega(x)$ both have densities given by $g(\omega)$ and $f(x|\omega)$, respectively. From Bayes Theorem it follows that

$$f(\omega|x) = \frac{g(\omega)f(x|\omega)}{h(x)},$$

where

$$h(x) = \int_\Omega f(x|\omega)g(\omega)d\omega.$$

Now $f(\omega|x)$, called the posterior distribution on ω, can be thought of as expressing the degree of belief in ω after having observed x. It is true that $f(\omega|x)$ is the conditional distribution of ω for given x and that as Savage showed, if $g(\omega)$ and $f(x|\omega)$ are subjective densities, then the Bayes solution d_p is such that (for the given x)

$$\int_\Omega L(d_p(x), \omega)f(\omega|x)d\omega$$

is a minimum. This is all that is true, however. Even if $g(\omega)$ represents the decision maker's degree of belief in the value ω, it does not follow that $f(\omega|x)$ should represent his degree of belief in ω after having observed x. All that any of the writers on decision making have shown (to this author's knowledge) is that $f(\omega|x)$ may be employed in obtaining the Bayes solution to the decision problem. (It should be noted, though, that if Jeffreys' logical approach is accepted, then $f(\omega|x)$ expresses the posterior logical probability density of ω given x.) Thus, even from the subjective viewpoint an inferential theory has not been successfully developed.

There is, however, one use of the posterior distribution on which both objectivists and subjectivists agree. If $g(\omega)$ represents "historical" data, then $f(\omega|x)$ represents the distribution of ω given the data of the present experiment as influenced by the prior historical data. The difficulty with this approach is in determining a $g(\omega)$ which reflects accurately the historical data. One possible solution is to employ a class of conjugate prior distributions as Raiffa and Schlaifer [29] do.

At this point, the desirability of an inferential theory, especially when the prior is not objective, could be questioned. Might not decision theory be employed universally? This writer believes that the answer to this question is no. In the fields of engineering, management, economic analysis, business, and so forth, decision theory is unquestionably the proper tool of analysis. In all of these endeavors certain decisions must be made and a form of loss is inherent when the "wrong" decision is made. It is true that to analytically order all of the possible decisions in a practical problem is a formidable task, but the successful engineer or business man somehow does this. However, he more than likely does it in a very inefficient manner which could be improved upon if decision theory were circumspectly employed. Consider now the scientist. His problems are somewhat different from those of the engineer. Broadly, his task is to formulate or investigate physical "laws." These laws should approximately describe given physical phenomenon mathematically. The scientist's problem can be placed in the decision theory framework, however. The possible decisions could be accept the hypothesized law, reject the hypothesized law, or continue experimentation. This process has certain drawbacks. If the hypothesized law is rejected what law does the scientist accept? Seldom in scientific research are all alternatives known. If this were the case science would not exist. In order to apply decision theory the scientist must have a loss function. It does not seem reasonable for the scientist to assume a loss function when he does not know all of the purposes for which his theory might be used. Thus, serious difficulties are encountered when the scientist tries to use decision theory.

In selecting a law or formulating a theory, the scientist generally relies on his beliefs about the law and the physical phenomenon it is to describe. Through theoretical calculations or experimentation, he alters his previous beliefs. It appears that some form of inferential theory would be of use to the scientist. A suitable theory might help the scientist to systematically order his degrees of belief.

In this paper the concept of a subjective inferential theory of probability will be studied. The term plausibility will be used to name this form of subjective probability. This use of the word plausibility comes from Polya [27] where he uses the phrase plausible reasoning. An attempt will be made to keep a sharp distinction between the "real world" and a person's beliefs about this real world. Objective probability will be denoted by probability and is assumed to exist in the real world. For example, in a coin tossing experiment it would be assumed that a limiting relative frequency for the proportion of occurrences of heads exists. Call this limit p. Then a person might "believe" that the true value of p is one half. The value of this belief suitably normed will be the plausibility of the statement that p equals one half. Plausibilities will be real numbers in the interval [0, 1] and have most of the same properties that probability measures have. Carnap [5] also considers two forms of probability in his logical development of probability theory. His probability_1 is similar to the plausibility concept considered in this paper and his probability_2 is relative frequency probability or what is termed probability in this paper.

In attempting to develop an inferential theory, three major problems should be considered. First, there is the problem of how to evaluate the prior or original plausibility. This is a vexing problem to which no really acceptable solution has been obtained. Jeffreys [15], Jaynes [14], and Raiffa and Schlaifer [29] have proposed three different techniques that have been used in certain cases to obtain prior plausibilities. In this paper this problem is not of primary interest and will not be discussed further.

The second problem to be considered is that of determining how data or "outside" information should change a prior plausibility to a posterior plausibility. This problem is the one to which most writers in inference have addressed themselves. As previously stated the solution has been obtained by employing Bayes Theorem. However, there are objections to the so obtained solution. In this paper this problem will be reformulated and the solution obtained as a special case of the solution to the third major problem of interest in inference. The question of how outside information should change directly the plausibilities of a given proposition will not be considered further in this paper.

The third major problem is this. If the posterior plausibilities of certain propositions are given, how should a person change the plausibilities of other propositions? The adjectives prior and posterior will not be employed in the developments of this paper, since the process of changing plausibilities will be considered as a discrete process in time. In general, all plausibilities will be denoted by a lower case p, q, and so forth, subscripted by a natural number which will indicate at what stage in time the plausibilities are being considered. In general, the subscript one will denote the initial plausibilities which will be assumed to be known in any problem. Plausibilities will be attached to propositions which are elements of an arbitrary Boolean algebra. These propositions will be denoted by lower case letters from the first of the alphabet, that is, by a's, b's, c's and so forth. Suppose that $p_1(a)$, $p_1(b)$, ... are given. If $p_2(a)$ is given, then how should $p_2(b)$, $p_2(c)$, ... be determined? In any reasonable scientific theory the propositions are, in general, related to each other. Thus, if it is known that the plausibility of the proposition, a, has changed from $p_1(a)$ to $p_2(a)$, then if a person is reasoning in a "rational" manner how should the plausibilities of propositions related to a be changed? In this paper this problem will be considered in some detail. A concept of "information" based on the changes in plausibilities will be developed axiomatically and then this information will be employed to answer the question of how to change plausibilities in certain special cases.

In Chapter II the axioms which determine the algebra of plausibilities will be given. Certain assumptions concerning the class of propositions, which will be made, are also discussed. The concept of information is developed in Chapter III and in Chapter IV this information function is employed in solving the problem of how to change the plausibilities of certain propositions given that the plausibilities of certain other related propositions have changed from one value to another. The solution obtained will be compared to the present solution of the inference problem in some special cases. The summary will be presented in Chapter V.

CHAPTER II

THE ALGEBRA OF PLAUSIBILITIES

Introduction

In this chapter the axioms are presented from which the algebra of plausibilities is derived. Also, the derivation of this algebra is given. Before stating the axioms certain definitions and assumptions must be made. Also, the general problem of how to determine plausibilities must be discussed.

The axiomatic approach in this paper is similar in some respects to that given by Jeffreys [15]. Thus, the usefulness of another presentation might be questioned. There is, however, one major difference between the derivation given in this paper and that of Jeffreys. Jeffreys only considers conditional probabilities while in this paper the concept of a conditional plausibility as such is not considered directly. The philosophical principle of conditioning the probability of a proposition on all previous known or given knowledge is important. However, incorporating the conditioning directly into the probability system places serious restrictions on the use of the system. In conditional probabilities the given proposition must be taken as true. In general, in scientific experimentation the only things that are known are the data observed. Virtually all of scientific theory is formulated in terms of unobservables. Relations among these unobservables are assumed to hold and from these relations it is deduced that certain experimental results should hold. In general, the experimental results that are actually observed do not agree perfectly with the results that the theory predicted. Thus, the data observed should only increase or decrease a person's previous beliefs in the validity of the theory. Further, for the important propositions considered to be given it is not possible to state that these propositions are true. If conditioning is considered only indirectly, then it is possible to cope with the situation in which a conditioning proposition is not known to be true.

In this paper no direct formulation of conditional plausibilities will be made. Rather, the process of incorporating outside information (either experimental data or further theoretical formulations) will be considered to be a discrete process in time. The outside information will change the plausibilities of certain known propositions by an assumed known amount. The question of how the plausibilities of propositions related to those whose plausibilities have changed will then be investigated. In order to start the process it will always be assumed that an initial set of plausibilities for all propositions under consideration is given. Thus, the only question to be considered is that of how to change plausibilities internal to the set of propositions under consideration. The answer to this question will be considered in Chapter IV.

In formulating the axioms it will be assumed that subject to certain conditions a plausibility (a real number in the interval $[0, 1]$) can be assigned to any proposition under consideration. This approach will be employed since it avoids the assumptions concerning an order relation made by most modern writers. The order approach will be avoided since the intuitive counterpart of a plausibility is a person's degree of belief in the truth or falsity of a proposition. It is true that if the plausibilities of two propositions a and b are known, then a and b can be ordered by the values of their plausibilities. However, this order in no way reflects the preference of either a over b or b over a.

One further important point should be made. The theory, while relying on certain concepts of deductive logic, is not a strictly necessary concept in the terms of Savage [33]. The degrees of belief expressed by plausibilities are personal in the sense that they reflect the reasoning process of some "ideal" person who always thinks "rationally." Here rationality is expressed by the set of axioms that define plausibility. It is possible to assign to any proposition any plausibility that is consistent with the axioms. Thus, the theory presented here is a sort of compromise between the logical and personal approaches.

Preliminary Definitions and Theorems

The assumptions and definitions that affect directly the development of the axioms will now be considered. To begin with, it will be assumed that the set of all propositions under consideration form a Boolean algebra. In ordinary language, complex sentences are formed from elementary sentences by the operations of disjunction, conjunction, and negation. By assuming that the set of all propositions under consideration is a Boolean algebra, it is insured that the abstract operations corresponding to disjunction, conjunction, and negation are defined for all propositions involved in these operations and that the result of performing any of the operations gives again a proposition in the Boolean algebra. The definition of a Boolean algebra employed is due to Birkhoff and MacLane [3]. A Boolean algebra is a set B of elements a, b, c, ... with equality that for any $a, b, c \in B$ has the following properties:

(1) B has two binary operations . and V which satisfy

$a.a = aVa = a$

$a.b = b.a \qquad aVb = bVa$

$a.(b.c) = (a.b).c \qquad aV(bVc) = (aVb)Vc$

(2) These operations satisfy the absorption laws:

$a.(aVb) = aV(a.b) = a$

(3) These operations are mutually distributive, that is,

$a.(bVc) = (a.b)V(a.c)$

$aV(b.c) = (aVb).(aVc)$

(4) B contains two distinguished elements O and I which satisfy

$O.a = O \qquad OVa = a$

$I.a = a \qquad IVa = I$

(5) B has a unary operation a' of negation which obeys the laws

$a.a' = O \qquad aVa' = I$

Using these axioms it is possible to prove (cf. Birkhoff and MacLane [3]) that

$$O' = I \qquad I' = O, \tag{2.1}$$

$$a'' = a, \text{ and} \tag{2.2}$$

$$(a.b)' = a'Vb' \qquad (aVb)' = a'.b'. \tag{2.3}$$

Further, a Boolean homomorphism (cf. Halmos [13]) is a mapping α from a Boolean algebra B, say, to a Boolean algebra A, such that

$$(a.b)\alpha = a\alpha . b\alpha , \qquad (2.4)$$

$$(a \vee b)\alpha = a\alpha \vee b\alpha , \text{ and} \qquad (2.5)$$

$$a'\alpha = (a\alpha)', \qquad (2.6)$$

whenever a and b are in B. It is an easy consequence of the definition of homomorphism that (cf. Halmos [13])

$$O\alpha = O, \text{ and} \qquad (2.7)$$

$$I\alpha = I. \qquad (2.8)$$

There exists in B a natural order relation defined by

$$a \leq b \text{ if and only if } a.b = a \qquad (2.9)$$

This order is a partial order (cf. Halmos [13]). This partial order is not to be confused with the complete order relation that is postulated to exist in order to develop a modern subjective theory. The order relation $a \leq b$ is natural in that it is defined in terms of operations in the Boolean algebra. The complete order necessary for a subjective theory must come from outside of the Boolean algebra and, in general, would have no connection with the relation $a \leq b$. The importance of the order relation $a \leq b$ is in its connection with logical implication. In order to define logical implication the definition of the conditional is needed. The conditional is an operation defined for pairs of propositions by

$$a \rightarrow b = a' \vee b \qquad (2.10)$$

The conditional is read "if a then b." Proposition a implies b if and only if $a \rightarrow b = I$. Any proposition that is equal to I is said to be logically true or a tautology and any proposition that is equal to O is said to be logically false or a self-contradiction. Therefore, if $a \rightarrow b$ is a tautology, then a implies b.

It is now possible to prove Theorem 2.11. As far as this writer knows Theorem 2.11 has no counterpart in other theories.

Theorem 2.11. The proposition a implies the proposition b if and only if $a \leq b$.

Proof: (i) Suppose that $a \leq b$. Then $a.b = a$ from (2.9) so that $a' = (a.b)' = a'Vb'$ from (2.3). Therefore, $a'Vb = a'Vb'Vb$ and $a'Vb = a'V\,I = I$ which implies that $a \rightarrow b = I$ from (2.10).

(ii) Now suppose that $a \rightarrow b = I$. Then from (2.10) it follows that $a'Vb = I$. Thus, $a = a.I = a.(a'Vb) = (a.a')V(a.b) = O\,V(a.b)$ and $a = a.b$. Therefore, from (2.9) $a \leq b$.

In view of Theorem 2.11, $a \rightarrow b = I$ will always be written in the shorter form $a \leq b$.

In any reasonable scientific theory, each proposition may be thought of as being either true or false. This concept of truth or falsity must now be abstracted and defined for the propositions that are elements of a given Boolean algebra. Let $U = \{O, I\}$ be the Boolean algebra consisting of only O and I. The algebra U is the smallest Boolean algebra in the sense that U is a subalgebra of any Boolean algebra.

Definition 2.12. A homomorphism from a Boolean algebra B onto

$\overline{U} = \{\overline{O}, \overline{I}\}$ is said to be an interpretation of the Boolean algebra B where the bar is used to indicate that $\overline{U}$ is the image Boolean algebra under the homomorphism.

Definition 2.13. If $a\alpha = \overline{I}$, then a is true in the interpretation α and if $a\alpha = \overline{O}$, then a is false in the interpretation α whenever $a \varepsilon B$.

In general, for brevity the phrase "a is true in the interpretation α" will be shortened to "a is true" and similarly "a is false in the interpretation α" will be shortened to "a is false." It should be noted that being true is not equivalent to being logically true. For a proposition a to be true, all that is required is that $a\alpha = \overline{I}$. However, a is logically true only if a = I. Thus, if a is logically true, then in view of (2.8) it is true for all interpretations and hence the title tautology.

The set of all possible interpretations will be denoted by G, that is,

$$G = \left\{\alpha \,\middle|\, \alpha : B \xrightarrow{\text{onto}} \overline{U}\right\}.$$

It will be assumed that in any problem that there exists a unique $\alpha \varepsilon G$ for the Boolean algebra B under consideration. This means that for no a in B is both a true and a false and that for every a in B, either a is true or a is false. In logic this first condition is called consistency and the second is called completeness. For a different formulation of these concepts in terms of ideals in B see Halmos [12].

It is now possible to prove two theorems whose content is not new but whose formulation in terms of the concepts of this paper is somewhat novel.

Theorem 2.14. If a is true, then a' is false and if a' is false, then a' is true.

Proof: It must be shown that if $a\alpha = \overline{I}$, then $a'\alpha = \overline{O}$ and conversely.

(i) Assume that $a\alpha = \overline{I}$. Then, from (2.6) it follows that $a'\alpha = (a\alpha)' = \overline{O}$.

(ii) Now assume that $a\alpha = \overline{O}$. Then, again, from (2.6) it follows that $a'\alpha = (a\alpha)' = \overline{I}$.

Theorem 2.15. Let h be any homomorphism from a Boolean algebra B into a Boolean algebra A. Let $a, b \varepsilon B$. Then, if $a \leq b$, $ah \leq bh$.

Proof: Suppose $a \leq b$ and let h be any homomorphism from B into A. From (2.9) it follows that $(a.b) = a$ so that $(a.b)h = ah$. Thus, from (2.4) $ah.bh = ah$ and, therefore, $ah \leq bh$ from (2.9).

Corollary 2.15.1. If $\alpha \varepsilon G$ and $A = \overline{U}$ then if $a \leq b$, then $a\alpha \leq b\alpha$.

An immediate consequence of Corollary 2.15.1 is that whenever a implies b then if $a\alpha = \overline{I}$, then $b\alpha = \overline{I}$ since $O \leq a \leq I$ for any Boolean algebra. The usefulness of implication is in deductive reasoning. For, if a implies b and a is true, then b must also be true.

In general, in any scientific theory it is not known for all propositions whether they are true or not. Thus, the particular interpretation of the theory is unknown. It is the purpose of scientific inquiry to find out the truth value of those propositions whose truth value is unknown. One way to do this is to employ implication. That is, it is shown that $a \rightarrow b = I$ or, in other words, it is shown that a implies b. Then a is shown to be true. Normally a is known to be true from previous work. Then it follows that b must be true. However, not always can it be shown that a implies b. In this case it is possible to achieve the same result if it can be shown that just $a \rightarrow b$ is true. That this is so is the content of Theorem 2.16. It is obvious that as propositions are found to be either true or false (but not both) then the set G_0 of possible interpretations is a subset of G. To explain G_0 consider an example. Suppose that a is shown to be true. Now G comprises all homomorphisms from B onto $\overline{U}$. But if a is true, then the possible interpretations of B must be such that for any α that is a possible interpretation it follows that $a\alpha = \overline{I}$. If G_0 denotes the set of possible interpretations, then

$$G_0 = \left\{\alpha \mid \alpha \in G \text{ and } a\alpha = \overline{I}\right\}$$

so that $G_0 \subset G$. The ultimate aim of scientific inquiry is to restrict G_0 so that it contains only one α. Suppose now that $a \rightarrow b$ is also true, then

$$G_0 = \left\{\alpha \mid \alpha \in G \text{ and } a\alpha = (a \rightarrow b)\alpha = \overline{I}\right\}.$$

It follows, then, that b is also true and if G_1 denotes the set of possible interpretations of B when a, b, and $a \rightarrow b$ are all true, that is,

$$G_1 = \left\{\alpha \mid \alpha \in G \text{ and } a\alpha = b\alpha = (a \rightarrow b)\alpha = \overline{I}\right\},$$

then $G_0 = G_1$. Thus, in reducing G, to find out that $a \rightarrow b$ is true is as an effective tool as finding out that a implies b. That $G_0 = G_1$ is a consequence of

<u>Theorem 2.16.</u> If α is such that $(a \rightarrow b)\alpha = \overline{I}$, then either $a\alpha = b\alpha = \overline{I}$ or $a\alpha = \overline{O}$ and $b\alpha = \overline{I}$ or $a\alpha = b\alpha = \overline{O}$.

Proof: Suppose α is such that $(a \rightarrow b)\alpha = \overline{I}$. Then, since $a \rightarrow b = a' \vee b$ it follows that $\overline{I} = (a'Vb)\alpha$. But from (2.6) it then follows that $\overline{O} = (a'Vb)'\alpha$. However, $(a'Vb)' = a.b'$ and thus, $\overline{O} = (a.b')\alpha = a\alpha . b'\alpha$ from (2.4). Therefore, either $a\alpha = \overline{O}$ or $b'\alpha = \overline{O}$ and the conclusion of the theorem follows.

Since the conclusion of the theorem is that there are only three possible pairs of truth values for a and b and only one of them has a true, then if a is known to be true, b must be true also. This follows since the truth pair for which a is true has b also true.

There is one final definition that must be made and it is that of the dependence and independence of two propositions. The definition will involve the possible pairs of truth values that two propositions can take on. There are four combinations of truth values for a pair of propositions a and b. Denoting truth by T and falsity by F the combinations are:

a	b
T	T
T	F
F	T
F	F

In view of the assumptions made about the Boolean algebra B, one of the four combinations must hold. Suppose that it is not known which pair holds. Then, in intuitive terms, the concept to be defined is this. If a is found to be true, does this indicate anything about the truth or falsity of b? Obviously, if a implies b, then the answer is in the affirmative. Suppose that a does not imply b. There are still cases in which a person might believe that finding a is true would lead him to change his previous beliefs about the truth or falsity of b. For example, a person might believe that $a \rightarrow b$ is true. If it is not known if $a \rightarrow b$ is true or not, then from the viewpoint of deductive logic nothing can be said about the truth or falsity of b. However, it is the purpose of this paper to develop a theory to deal with a person's beliefs so that the definition of dependence must somehow cover this example.

Before the definitions of dependence and independence can be formulated, the definition of logical dependence and independence must be given. A pair of propositions are logically dependent if they are logically related; otherwise, they are independent. There are two kinds of dependence, namely, trivial and nontrivial which stem from the trivial and nontrivial logical relations. As defined by Kemeny, Schleifer, Snell, and Thompson [17] there are six nontrivial logical relations. They are inconsistent, subcontraries, first implies second, second implies first, equivalents, and contradictories. However, all of these relations can be defined in terms of implication (cf. Kemeny et al. [17]). For example, if $a \leq b$ and $b \leq a$, then $a = b$ or a and b are equivalent. The trivial relations are when either a or b is either logically true or logically false.

Definition 2.17. Let a and b be elements of B and suppose that neither are logically true nor false. Then a and b are said to be dependent if either they are nontrivially logically dependent or if any one or combination of the following four cases holds:

(i) There exists a $c_1 \varepsilon B$ such that if c_1 is true, then $a \rightarrow b'$ is true where c_1 does not equal $a'.b$ or $a.b'$ or $a'.b'$ or any disjunction of these and $c_1.(a'.b) \neq O$, $c_1.(a.b') \neq O$, and $c_1.(a'.b') \neq O$.

(ii) There exists a $c_2 \varepsilon B$ such that if c_2 is true, then $a \rightarrow b$ is true where c_2 does not equal $a.b$ or $a'.b$ or $a'.b'$ or any disjunction of these and $c_2.(a.b) \neq O$, $c_2.(a'.b) \neq O$, and $c_2.(a'.b') \neq O$.

(iii) There exists a $c_3 \varepsilon B$ such that if c_3 is true, then $a' \rightarrow b'$ is true where c_3 does not equal $a.b$ or $a.b'$ or $a'.b'$ or any disjunction of these and $c_3.(a.b) \neq O$, $c_3.(a.b') \neq O$, and $c_3.(a'.b') \neq O$.

(iv) There exists a $c_4 \varepsilon B$ such that if c_4 is true, then $a' \rightarrow b$ is true where c_4 does not equal $a.b$ or $a.b'$ or $a'.b$ or any disjunction of these and $c_4.(a.b) \neq O$, $c_4.(a.b') \neq O$, and $c_4.(a'.b) \neq O$.

Otherwise, a and b are independent.

Certain comments are in order. The condition that, for example, in part (i) that, say, $c_1 \neq a'.b$ is necessary since $a'.b \leq (a'.b)V(a.b')V(a'.b')$ which equals $a \rightarrow b'$. Thus, if $c_1 = a'.b$, then $c_1 \leq a \rightarrow b'$ and if c_1 is true, then $a \rightarrow b'$ is true. However, if $c_1 = a'.b$ is true, then there is no interest in $a \rightarrow b'$ since the truth values of a and b are known. Similar comments hold for $a.b'$, $a'.b'$, and any of the disjunctions. A c_i, $i = 1, 2, 3, 4$, that either equals one of the four pairs $a.b$, $a.b'$, $a'.b$, $a'.b'$ or any disjunction of these pairs will be called a trivial function of a and b. Further, the condition that, say, $c_1.(a'.b) \neq O$ is necessary. For suppose that $c_1.(a'.b) = O$. Then if c_1 is true, then $a'.b$ is false so that a is true and b is false which again leads to an excessively restrictive c_1. Further, to show that a and b are independent is not trivial. It must be shown that a and b are not nontrivially logically dependent and that none of the c's exist that result in one of the four cases stated in the definition. Finally, it should be noted that a and b are independent when they are trivially logically related.

In order to demonstrate how the part of the definition that involves the four cases can be used, an example will be considered. The example concerns three propositions denoted by a, b, and c. They are:

a ≡ Life as it exists on Earth also exists on Mars.

b ≡ The conditions for life on Earth are also present on Mars.

c ≡ There is an analogy between Earth and Mars in that the same physical and biological processes occur on both.

Now it is known on Earth from experimentation that if life exists, then the conditions for life are present. If the same processes that exist on Earth also exist on Mars then if Earth life exists on Mars, the conditions for Earth life must be present on Mars. Now c is not a trivial function of a and b and $c.(a.b) \neq O$, $c.(a'.b) \neq O$, and $c.(a'.b') \neq O$. Thus, the c here plays the role of c_2 in the definition of dependence and there exists a c_2 which is not a trivial function of a and b such that if c_2 is true, then $a \rightarrow b$ is true. Therefore, a and b are dependent. This example while possibly trivial in nature has many nontrivial counterparts in scientific endeavor since the technique of analogy is employed repeatedly. Further, consider how this example relates to the possible T-F combinations for a and b. If c is true, then $a \rightarrow b$ is true. But from Theorem 2.16 if $a \rightarrow b$ is true, then the combination a true and b false is not possible. However, it is not known that c is true. If c is not true, then $a \rightarrow b$ may or may not be true. A scientist investigating Mars believes very strongly in c and, thus, he thinks that $a \rightarrow b$ is true. He realizes, however, that c could be false and that then $a \rightarrow b$ may not be true.

Now suppose that the proposition b is replaced by the proposition b_1 where

$b_1 \equiv$ The writer of this paper will eventually graduate.

May not some of the c's in the definition exist? They may. But in any reasonable theory relating to Mars there will exist no c's such that any of the four cases hold. Further, a and b_1 are not logically related. For example, there exist interpretations in which a is true and b_1 is false. But if $a \leq b_1$, then the T-F combination cannot occur. Similar arguments hold for the other logical relations. Therefore, a and b_1 are independent.

Axioms for Plausibilities

The axioms that define the algebra of plausibilities will now be presented. The axioms will be stated and then a brief discussion of each will be given.

Axiom 2.18. To any element of a given Boolean algebra B (satisfying the previously stated assumptions in this chapter) that is not logically true or false, it is possible to assign any number in the interval $[0, 1]$ subject to the restrictions determined by the remaining axioms. This number will be called the plausibility of the proposition and will be denoted by $p(a)$ where $a \varepsilon B$.

Axiom 2.19. If $a \leq b$, then $p(a) \leq p(b)$ whenever $a, b \varepsilon B$.

<u>Axiom 2.20</u>. If $a.b = O$, then $p(aVb) = f[p(a), p(b)]$ where $f(x, y)$ is such that any change in $f(x, y)$ for fixed $y \varepsilon (0, 1)$ is proportional to the change in x and similarly for y.

<u>Axiom 2.21</u>. If a and b are independent, then $p(a.b) = h[p(a), p(b)]$ where $h(x, y)$ is such that any change in $h(x, y)$ for fixed $y \varepsilon (0, 1)$ is proportional to the change in x and similarly for y.

Axiom 2.18 postulates the existence of a number between zero and one that represents an ideal person's degree of belief in the truth or falsity of a given proposition. This is not completely unreasonable if it is remembered that, for any $a \varepsilon B$, a is either true or false. Intuitively, the number assigned to a represents how true a is. Axiom 2.19 is perhaps the least innocuous of the axioms. It establishes an order in the plausibilities. Intuitively, if a implies b, then the belief in a should not be greater than the belief in b. This is reasonable since if a implies b, then it cannot happen that a is true and b is false. The assumption of proportionality in Axioms 2.20 and 2.21 is quite strong. The assumption of proportionality determines the rate at which either $p(aVb)$ or $p(a.b)$ will change in certain special cases. The only justification of the proportionality assumption is that it makes the rate of change simple. If $a.b = O$, then both a and b cannot be true at the same time. One or the other (but not both) is true or both are false. Thus, in considering what plausibility to assign to aVb it is not unreasonable to assume that it is a function of only $p(a)$ and $p(b)$. Finally, if a and b are independent, then the plausibility of a.b should only change as the plausibilities of a and b change. This is not unreasonable since the intuitive motivation of independence is that

knowledge of the truth or falsity of a does not affect the knowledge of the truth or falsity of b and conversely.

Derivation of Plausibility Theory

Certain theorems that stem from the axioms will now be given. As it turns out, plausibilities enjoy most of the same properties that elementary (objective) probabilities do. The exception is the properties of conditional probabilities. Recall that no definition of conditional plausibility will be made and, therefore, no theorems corresponding to conditional probabilities can be proven.

Theorem 2.22. If $a = b$, then $p(a) = p(b)$.
Proof: If $a = b$, then $a \leq b$ and $b \leq a$. From Axiom 2.19 it follows that $p(a) \leq p(b)$ and $p(b) \leq p(a)$. Therefore, $p(a) = p(b)$.

Theorem 2.23. The function $f(x, y)$ in Axiom 2.20 is given by $f(x, y) = x + y + kxy$ where k is an undetermined (as yet) constant.
Proof: From the proportionality assumption of Axiom 2.20 it follows that $\partial f/\partial x$ and $\partial f/\partial y$ both exist and are given by

$$\frac{\partial f}{\partial x} = g(y) \qquad (2.23.1)$$

$$\frac{\partial f}{\partial y} = h(x) \qquad (2.23.2)$$

Integrating (2.23.1) with respect to x yields

$$f(x, y) = g(y)x + g_1(y).$$

But $\partial f/\partial y = h(x)$ from (2.23.2) so that

$$\frac{\partial}{\partial y}[g(y)x] + \frac{\partial}{\partial y} g_1(y) = h(x)$$

or

$$x\frac{\partial}{\partial y} g(y) + \frac{\partial}{\partial y} g_1(y) = h(x).$$

Thus, $\partial g/\partial y$ and $\partial g_1/\partial y$ are constants and f(x, y) may be written as

$$f(x, y) = (k_1 y + k_2)x + k_3 y + k_4$$

$$= k_1 xy + k_2 x + k_3 y + k_4 \qquad (2.23.3)$$

where the k's are constants. Now aVO = a and a.O = O so that p(aVO) = p(a) = f[p(a), p(O)] = f[p(a), 0]. From Axiom 2.18 if a is not logically true or false it is possible for p(a) to take on any number in the interval [0, 1]. Let $x \varepsilon [0, 1]$. Then x = f(x, 0) and from (2.23.3) it follows that

$$x = k_2 x + k_4$$

for every $x \varepsilon [0, 1]$. Thus $k_4 = 0$ and $k_2 = 1$. Now aVO = OVa so that from Theorem 2.22 it follows that also $k_3 = 1$ and the theorem is proved.

<u>Theorem 2.24</u>. If a and b are independent, then p(a. b) = p(a)p(b). Proof: The theorem will be proved by proving that h(x, y) = xy. Similarly to the first part of the proof of Theorem 2.23 it can be shown that Axiom 2.21 implies that

$$h(x, y) = k_1 xy + k_2 x + k_3 y + k_4 \qquad (2.24.1)$$

where the k's are constants. Now, a. b = b. a so that from Theorem 2.22 p(a. b) = p(b. a) and h(x, y) = h(y, x). Now, a and O are independent and also, a and I are independent for any a. Since a.O = O and a.I = a and from Axiom 2.18 if a is not logically true or false, then p(a) can take on any value in the interval [0, 1], it follows that $0 = h(x, 0) = h(0, x)$ and $x = h(x, 1) = h(1, x)$. Then, from (2.24.1) it follows that

$$0 = k_2x + k_4 = k_3x + k_4$$

for every $x \varepsilon [0, 1]$. Thus, $k_2 = k_3 = k_4 = 0$ and $h(x, y) = k_1xy$. But $x = h(x, 1)$ so that $x = k_1x$ for every $x \varepsilon [0, 1]$ and, therefore, $k_1 = 1$ and the theorem is proved.

Theorem 2.25. If a.b = O, then p(aVb) = p(a) + p(b).

Proof: Now a.(bVc) = (a.b)V(a.c) for any a, b, c ε B. Suppose that none of a, b, or c are either logically true or false. Further, assume that b.c = O and that a is independent of both b and c and bVc. From Theorem 2.22 p[a.(bVc)] = p[(a.b)V(a.c)]. Consider

$$(a.b).(a.c) = a.(b.c) = a.O = O.$$

Therefore,

$$p[(a.b)V(a.c)] = p(a.b) + p(a.c) + kp(a.b)p(a.c)$$

from Theorem 2.23. From Theorem 2.24 it follows that p[a.(bVc)] =p(a)p(bVc) and applying Theorem 2.23 to p(bVc) it further follows that p(a)p(bVc) = p(a)[p(b) + p(c) + kp(b)p(c)]. Finally, it is deduced that

$$p(a)[p(b) + p(c) + kp(b)p(c)] = p(a)p(b) + p(a)p(c) + kp(a)p(b)p(a)p(c)$$

or that

$$p(a)p(b) + p(a)p(c) + kp(a)p(b)p(c) = p(a)p(b) + p(a)p(c) + kp^2(a)p(b)p(c).$$

Therefore, $kp(a) = kp^2(a)$. But p(a) can take on any value in the interval [0, 1]. Thus, $k = 0$ and $f(x, y) = x + y$ and the theorem is proved.

<u>Theorem 2.26</u>. $p(I) = 1$ and $p(O) = 0$.

Proof: Let $a \varepsilon B$ and suppose that a is not logically true or false. Now I and a are independent and $I.a = a$. Thus, from Theorems 2.22 and 2.24 it follows that $p(a) = p(I.a) = p(I)p(a)$. But from Axiom 2.18 p(a) can be any number in the interval [0, 1] so that $p(I) = 1$. Further, $O.a = O$ and $OVa = a$ so that from Theorems 2.22 and 2.25 it follows that $p(a) = p(OVa) = p(O) + p(a)$. However, p(a) can take on any value in the interval [0, 1]. Thus, $p(O) = 0$.

<u>Theorem 2.27</u>. For any $a \varepsilon B$, $p(a) + p(a') = 1$.

Proof: Now $a.a' = O$ so that $p(aVa') = p(a) + p(a')$. But $aVa' = I$ and $p(aVa') = 1$.

<u>Theorem 2.28</u>. For any $a, b \varepsilon B$, $p(aVb) = p(a) + p(b) - p(a.b)$.

Proof: Now $b = (a.b)V(a'.b)$ and $(a.b).(a'.b) = O$ so that $p(b) = p(a.b) + p(a'.b)$ from Theorem 2.25. Therefore, $p(a'.b) = p(b) - p(a.b)$. But $aVb = aV(a'.b)$ and $a.(a'.b) = O$ so that again from Theorem 2.25 it follows that $p(aVb) = p(a) + p(a'.b)$ and, therefore, $p(aVb) = p(a) + p(b) - p(a.b)$.

Theorem 2.28 is the last theorem of this chapter. With the proof of Theorem 2.28 it is seen that plausibilities enjoy all of the properties of elementary objective probability except those that involve conditional probabilities. One further comment is in order. The algebra of plausibilities developed herein does not have the property of countable additivity. This is a desirable property as advanced theories of probability indicate. However, the subjective systems of Savage and de Finetti do not have the countable additivity property either and as de Finetti points out it is difficult to conceive of a practical situation in which countable additivity is needed.

CHAPTER III

THE INFORMATION FUNCTION

Introduction

In this chapter a concept of information will be presented and discussed. An information function will be developed axiomatically which will measure the change in information when the plausibility of a changes from, say, $p_1(a)$ to $p_2(a)$. The approach is similar to that of Good's [11] in developing Shannon's entropy function. Since the idea here is to develop a measure of change in information the resulting information function differs somewhat in form from that of Shannon's entropy function. The axioms are similar to those given by Rényi [31]. They are, however, much simpler than those of Rényi and do not employ his concept of a generalized probability distribution. According to Rényi a generalized distribution is a measure function whose measure on the whole space is less than or equal to one. Once the concept of information for individual propositions has been developed then a measure of total information in a special set of propositions will be defined. The total information function is the same as the mean information given by Kullback [21]. However, Kullback makes no attempt to derive axiomatically an information function and his interpretation and use of the total information function is different from that made in this paper.

As pointed out by Luce and Raiffa [22], if the purpose of scientific investigation is the search of information in some sense, then this information should be formalized and introduced into the problem. In order to do this some concept of what information should be is required. Since the objects under consideration are abstract propositions of an arbitrary Boolean algebra, information as such cannot reflect the meaning of these propositions. The only abstract interpretation a proposition can have is in its truth value. Thus, any formulation of information must of necessity be expressed in terms of the truth or falsity of a proposition. The abstract extension of truth or falsity is the plausibility of a proposition so that in this paper information will be expressed in terms of plausibilities or rather, in terms of changes in plausibilities. Let a be an element of a Boolean algebra B and suppose that in going from stage 1 in time to stage 2 in time $p_1(a)$ changes to $p_2(a)$. Since these plausibilities are, intuitively, degrees of belief, if $p_2(a)$ is greater than $p_1(a)$ it is reasonable to assume that the information has increased and if $p_2(a)$ is less than $p_1(a)$ that the information has decreased. That is, if a is now believed more strongly, then the knowledge about a has increased and if a is now believed less strongly, then the knowledge about a has decreased. Following this line of reasoning, the assumption that information is some function of plausibilities will be made.

Information Axioms

In this section the axioms that lead to the information function will be presented. It is assumed that the change in information is a real-valued function of the respective plausibilities. Consider again $p_1(a)$

and $p_2(a)$. The change in information or for brevity the information in knowing that $p_1(a)$ has changed to $p_2(a)$ will be denoted by $i[p_2(a), p_1(a)]$. This notation will in general be shortened to $i(p_2, p_1)$. If either p_1 or p_2 is zero, then $i(p_2, p_1)$ will not be assumed to exist. The reason for this is that if $p_1(a) = 0$, then it is possible that $a = O$ and if $a = O$, then it should not be possible to change the plausibility of a from zero. If a has been assigned a nonzero plausibility, then $a \neq O$ and it is unreasonable for the plausibility of a to change to zero since this could indicate that $a = O$.

<u>Definition 3.1.</u> Let $a_j \varepsilon B$ for $j = 1, 2, \ldots, n$ and suppose that $a_j . a_k = O$ if $j \neq k$. Then the set of propositions is said to be mutually exclusive.

<u>Definition 3.2.</u> Let $a_j \varepsilon B$ for $j = 1, 2, \ldots, n$ and suppose that $a_1 V a_2 V \ldots V a_n = I$. Then the set of propositions is said to be exhaustive.

Let a_j, $j = 1, 2, \ldots, n$ be a mutually exclusive and exhaustive set of propositions, that is, $a_j . a_k = O$ for $j \neq k$ and $a_1 V \ldots V a_n = I$. Further, let 1 denote some stage in time and 2 some later stage in time and put

$$p_{1j} = p_1(a_j), \text{ and}$$

$$p_{2j} = p_2(a_j).$$

Denote by P_1 the set $\{p_{1j} | j = 1, 2, \ldots, n\}$ and by P_2 the set $\{p_{2j} | j = 1, 2, \ldots, n\}$. The total information in knowing that p_{1j} has changed to p_{2j} for $j = 1, 2, \ldots, n$ is then denoted by $I(P_2 | P_1)$.

Axiom 3.3. Let p, q be any elements of the interval $(0, 1]$. Then if $p \leq q$, $i(q, p) \geq 0$ and if $p \geq q$, $i(q, p) \leq 0$.

Axiom 3.4. Let p, q, r, s be any elements of the interval $(0, 1]$. Then $i(pq, rs) = i(p, r) + i(q, s)$.

Definition 3.5. Let $a_j \varepsilon B$, $j = 1, 2, \ldots, n$ be a mutually exclusive and exhaustive set of propositions. Then $I(P_2|P_1) = \sum_{j=1}^{n} p_{2j}\, i(p_{2j}, p_{1j})$.

Axiom 3.3 serves to establish an order for the information function. If p and q refer to a particular proposition, then $i(q, p)$ can be interpreted in this manner. If $q = p_2$ and $p = p_1$, then if $p_2 \geq p_1$, the information is nonnegative and if $p_2 \leq p_1$, the information is nonpositive. Axiom 3.4 follows from considering independent propositions and may be interpreted in this manner. Let $p = p_2(a)$, $q = p_2(b)$, $r = p_1(a)$, and $s = p_1(b)$. Then Axiom 3.4 postulates that the information for a.b is given by the sum of the information for a and the information for b. Definition 3.5 defines the total information as a function of the information for the individual propositions when those propositions comprise a mutually exclusive and exhaustive set.

Derivation of the Information Functions

The derivation of the information function will be given in a series of lemmas that follow from Axioms 3.3 and 3.4. Then the form of the total information function is given immediately by Definition 3.5.

Theorem 3.6. The function $i(q, p)$ for $p, q \varepsilon (0, 1]$ is given by

$$i(q, p) = k \log \frac{q}{p}$$

where $k > 0$ is an arbitrary constant.

Proof: The completion of the proof will be deferred until certain lemmas which are necessary have been proved.

Lemma 3.7. If $p = q$, then $i(q, p) = i(p, p) = 0$.

Proof: Now if $p = q$, then $p \leq q$ and $p \geq q$ so that from Axiom 3.3 it follows that $i(q, p) \geq 0$ and $i(q, p) \leq 0$. Therefore, $i(q, p) = 0$ and the lemma follows.

Lemma 3.8. The function $i(q, p)$ may be written as

$$i(q, p) = i(1, p) - i(1, q). \qquad (3.8.1)$$

Proof: Consider Axiom 3.4 and let $p = s = 1$ so that $i(q, r) = i(1, r) + i(q, 1)$. If $q = r = p$, then $i(q, r) = i(p, p) = i(1, p) + i(p, 1)$. But from Lemma 3.7 $i(p, p) = 0$ so that $i(p, 1) = -i(1, p)$. Therefore, $i(q, r) = i(1, r) - i(1, q)$ and the lemma follows by renaming r.

Lemma 3.9. If $p \geq q$, then $i(1, p) \leq i(1, q)$.

Proof: Suppose that $p \geq q$. Then there exists an r such that $p \geq r \geq q$. Thus, from Axiom 3.3 it follows that $i(r, p) \leq 0$ and $i(r, q) \geq 0$. Now $i(r, p) \leq 0$ implies $i(1, p) - i(1, r) \leq 0$ from (3.8.1) and $i(r, q) \geq 0$ implies $i(1, q) - i(1, r) \geq 0$ from (3.8.1). Thus, $r(1, p) \leq i(1, r)$ and $i(1, q) \geq i(1, r)$ so that $i(1, q) \geq i(1, r) \geq i(1, p)$ and the lemma is proved.

Lemma 3.10. The following formula holds:

$$i(1, pq) = i(1, p) + i(1, q). \qquad (3.10.1)$$

Proof: Consider Axiom 3.4 and let $p = q = 1$. Then $i(pq, rs) = i(1, rs) = i(1, r) + i(1, s)$. The lemma follows by renaming r and s.

Lemma 3.11. The function $i(1, p)$ is given by

$$i(1, p) = k \log \frac{1}{p} \tag{3.11.1}$$

where $k > 0$ is an arbitrary constant and $p \varepsilon (0, 1]$.

Proof: From Lemma 3.9, $i(1, p)$ is a monotone decreasing function of $p \varepsilon (0, 1]$. Since $i(p, p) = 0$ from Lemma 3.7 it follows that $i(1, 1) = 0$ and therefore, $i(1, p) \geq 0$ for $p \varepsilon (0, 1]$. Now consider Lemma 3.10 and let $p = e^{-x}$ and $q = e^{-y}$. Then,

$$i(1, rs) = i(1, e^{-x}e^{-y}) = i[1, e^{-(x+y)}], \tag{3.11.2}$$

$$i(1, r) = i(1, e^{-x}), \text{ and} \tag{3.11.3}$$

$$i(1, s) = i(1, e^{-y}) \tag{3.11.4}$$

where $x, y \geq 0$. Define $f(x)$ as $f(x) = i(1, e^{-x})$ for $x \geq 0$. Then from (3.11.2), (3.11.3), and (3.11.4) it follows that $f(x)$ is a real-valued function satisfying

$$f(x + y) = f(x) + f(y) \tag{3.11.5}$$

Since $i(1, p) \geq 0$ for $p \varepsilon (0, 1]$, it follows that $f(x) \geq 0$ for $x \geq 0$. The relation (3.11.5) defines a functional equation whose solution is

$$f(x) = kx, \quad x \geq 0 \tag{3.11.6}$$

where k is an arbitrary constant. That (3.11.6) is the solution to

(3.11.5) is shown by Parzen [26] page 123. From (3.11.6) it follows that $i(1, e^{-x}) = kx$ and letting $p = e^{-x}$ it further follows that

$$i(1, e^{-x}) = i(1, p) = -k \ln p = k \ln \frac{1}{p}$$

where $p \in (0, 1]$ and ln denotes the natural logarithm. Since $i(1, p) \geq 0$ it follows that k must be greater than or equal to zero. If $i(1, p)$ is to be nontrivial then $k \neq 0$ and the lemma is proved.

From Lemma 3.11 and Lemma 3.8 it follows that

$$\begin{aligned} i(q, p) &= k \log \frac{1}{p} - k \log \frac{1}{q} \\ &= k[\log \frac{1}{p} - \log \frac{1}{q}] \\ &= k \log \frac{q}{p} \end{aligned}$$

where k is an arbitrary constant greater than zero and the proof of Theorem 3.6 is complete. Notice that $i(q, p)$ is not defined if either p or q is zero. In fact, as $p \to 0$, $i(q, p) \to \infty$ and as $q \to 0$, $i(q, p) \to -\infty$. This is reasonable since it should take a "large" amount of information to change the plausibility of a proposition from zero to some number greater than zero. Similarly, a "large" amount of information should be lost if a nonzero plausibility goes to zero.

<u>Theorem 3.12.</u> The total information is given by

$$I(P_2|P_1) = k \sum_{j=1}^{n} p_{2j} \log \frac{p_{2j}}{p_{1j}} \qquad (3.12.1)$$

where $k > 0$ is an arbitrary constant.

Proof: From Theorem 3.6, $i(p_{2j}, p_{1j}) = k \log \frac{p_{2j}}{p_{1j}}$ so that the theorem follows immediately from Definition 3.5.

The proof of Theorem 3.12 concludes the results that are to be presented in this chapter. Reference to Kullback [21] is made for some interesting results concerning the total information $I(P_2|P_1)$.

CHAPTER IV

DETERMINATION OF CHANGES IN PLAUSIBILITIES

Introduction

In this chapter the total information function will be used to solve the specific problem of how to change the plausibilities of certain propositions given that the plausibilities of certain other related propositions have changed from one known value to another known value. The solution obtained will be compared to the present solution of the inference problem in some particular cases. Now $I = I(P_2|P_1)$ is defined to within an arbitrary constant k. For use in this chapter it will be assumed that k is such that

$$I = \sum_{j=1}^{n} p_{2j} \ln \frac{p_{2j}}{p_{1j}}$$

where ln denotes the natural logarithm. The general problem to be treated is this. Let 1 denote some stage in time and 2 denote some later stage in time. Let $p_{1j} = p_1(a_j)$ and $p_{2j} = p_2(a_j)$ where $a_j \varepsilon B$ for $j = 1, 2, \ldots, n$. Suppose that p_{1j} for $j = 1, 2, \ldots, n$ and certain linear combinations of the p_{2j} are known. Then, how should the p_{2j} be determined? The criterion of minimizing I subject to the linear restraints on the p_{2j} is proposed as a solution; that is, the p_{2j}, $j = 1, 2, \ldots, n$ will be determined such that I is a minimum subject to the restraints on the p_{2j}. Since I was only defined for a set of mutually exclusive and exhaustive propositions there is one restraint that the p_{2j} must always meet and that is that the sum of the p_{2j} over all j must equal one.

Kullback [21] also considers minimizing I subject to side conditions. His treatment of the problem is more general than the treatment given here in that his results hold for any set of dominated probability measures but less general (for the discrete case) in another sense because he considers only one side condition. The general problem treated here allows for more than one side condition. Kullback's use of the minimization technique is different than that made in this paper. The application in this paper is similar to the maximum-entropy estimates of Jaynes [14]. However, Jaynes maximizes the entropy in order to obtain a prior distribution. This problem is not considered in this paper.

Two theorems will be proved which will be employed to find solutions for the problems investigated. The first theorem is a special case of a theorem proved by Kullback [21]. However, the method of proof is different than that employed by Kullback. The second theorem is (in the discrete case) a generalization of a theorem given by Kullback.

Minimization of the Total Information

In order to facilitate the proof of the theorems a change in notation will be made. Thus, p_{2j} will be written as q_i and p_{1j} will be written as p_i where $i = 1, 2, \ldots, n$. Then the total information will be

$$I = \sum_{i=1}^{n} q_i \ln \frac{q_i}{p_i}$$

for $i = 1, 2, \ldots, n$.

<u>Theorem 4.1.</u> If the only restriction on the q_i is that they sum to one, then the q_i that minimize I are $q_i = p_i$ for $i = 1, 2, \ldots, n$.

Proof: The method of Lagrange multipliers will be employed in proving the theorem. Let the Lagrange multiplier be denoted by λ and set

$$F = I + \lambda \sum_{i=1}^{n} q_i.$$

Then,

$$\frac{\partial F}{\partial q_r} = \ln \frac{q_r}{p_r} + 1 + \lambda, \quad r = 1, 2, \ldots, n. \qquad (4.1.1)$$

Setting the equations (4.1.1) equal to zero yields,

$$\ln \frac{q_r}{p_r} = -(\lambda + 1)$$

so that

$$\frac{q_r}{p_r} = e^{-(\lambda + 1)}$$

and

$$q_r = p_r e^{-(\lambda + 1)}, \quad r = 1, 2, \ldots, n. \quad (4.1.2)$$

But $\sum_{i=1}^{n} q_i = 1$ and, therefore, $e^{-(\lambda + 1)} = 1$ and from the equations (4.1.2) $q_r = p_r$, $r = 1, 2, \ldots, n$, is a critical point. The second partials are

$$\frac{\partial^2 F}{\partial q_r^2} = \frac{1}{q_r} > 0$$

and

$$\frac{\partial^2 F}{\partial q_s \partial q_r} = 0, \quad s \neq r.$$

Since all of the mixed partials are zero and the second partials with respect to the same variable are all positive, from the second directional derivative test (cf. Kaplan [16] pp. 128-129) the critical point is a minimum and the theorem is proved.

Theorem 4.2. Suppose that $\sum_{i=1}^{n} T_j(i)\, q_i = k_j$, $j = 1, 2, \ldots, p < n - 1$, are p linearly independent equations where $T_j(i)$ is an arbitrary function of $i = 1, 2, \ldots, n$ and the k_j are given constants. Recall that $\sum_{i=1}^{n} q_i = 1$. Then, subject to these p + 1 restraints the minimum of I is given by

$$q_r = p_r e^{-\sum_{j=1}^{p} \lambda_j T_j(r) - \mu - 1}$$

where $r = 1, 2, \ldots, n$ and the λ_j and μ are Lagrange multipliers which must be solved for by using the restraining equations.

Proof: Denote by λ_j the Lagrange multiplier for the j th restraint and by μ the Lagrange multiplier for the restraint $\sum_{i=1}^{n} q_i = 1$ and set

$$F = I + \sum_{i=1}^{n} \sum_{j=1}^{p} \lambda_j T_j(i)\, q_i + \mu \sum_{i=1}^{n} q_i$$

$$= \sum_{i=1}^{n} [\, q_i \ln \frac{q_i}{p_i} + \sum_{j=1}^{p} \lambda_j T_j(i)\, q_i + \mu\, q_i].$$

Then,

$$\frac{\partial F}{\partial q_r} = \ln \frac{q_r}{p_r} + 1 + \sum_{j=1}^{p} \lambda_j T_j(r) + \mu, \quad r = 1, 2, \ldots, n. \quad (4.2.1)$$

Setting the equations (4.2.1) equal to zero yields,

$$\ln \frac{q_r}{p_r} = - \sum_{j=1}^{p} \lambda_j T_j(r) - \mu - 1$$

so that

$$\frac{q_r}{p_r} = e^{-\sum_{j=1}^{p} \lambda_j T_j(r) - \mu - 1}$$

and

$$q_r = p_r e^{-\sum_{j=1}^{p} \lambda_j T_j(r) - \mu - 1}, \qquad r = 1, 2, \ldots, n. \qquad (4.2.2)$$

Using the restraining equations, the equations (4.2.2) may be solved for λ_j and μ to yield the critical point(s). Note, that regardless of the values of the λ_j and μ, q_r is greater than zero since p_r is greater than zero for all r. The second partials are

$$\frac{\partial^2 F}{\partial q_r^2} = \frac{1}{q_r} > 0$$

and

$$\frac{\partial^2 F}{\partial q_s \partial q_r} = 0, \quad s \neq r.$$

Since all of the mixed partials are zero and the second partials with respect to the same variable are all positive, from the second directional derivative test (cf. Kaplan [16] pp. 128-129) the critical point(s) is a minimum. Since all critical points are minima there is a unique minimum and the theorem is proved.

Applications

The situation where n = 4 which arises when two propositions are considered will be studied in great detail. This is an application that occurs frequently in practice. The situation where n = 8 which arises when three propositions are considered will be discussed briefly. The applications considered will be presented in a series of cases. Before considering these cases, the notational scheme that will be employed for the situation when two propositions are considered will be presented. Diagrammatically the situation is as follows:

Stage 1

	b	b'	
a	$p_1(a.b)$	$p_1(a.b')$	$p_1(a)$
a'	$p_2(a'.b)$	$p_1(a'.b')$	$p_1(a')$
	$p_1(b)$	$p_1(b')$	1

Change due to outside information →

Stage 2

	b	b'	
a	$p_2(a.b)$	$p_2(a.b')$	$p_2(a)$
a'	$p_2(a'.b)$	$p_2(a'.b')$	$p_2(a')$
	$p_2(b)$	$p_2(b')$	1

The notation will be abbreviated to

$p_1(a.b) = p_1$	$p_2(a.b) = q_1$
$p_1(a.b') = p_2$	$p_2(a.b') = q_2$
$p_1(a'.b) = p_3$	$p_2(a'.b) = q_3$
$p_1(a'.b') = p_4$	$p_2(a'.b') = q_4$
$p_1(a) = u$	$p_2(a) = v$
$p_1(a') = 1 - u$	$p_2(a') = 1 - v$
$p_1(b) = r$	$p_2(b) = s$
$p_1(b') = 1 - r$	$p_2(b') = 1 - s$

The outside information will consist of changes in the various marginals; that is, for example, it will be given that r changes to s.

Case I. In this case two propositions are considered and no outside information is given; that is, neither v nor s are known. There are then no linear restrictions on the q_i except that they sum to one. Thus, from Theorem 4.1 it follows that $q_i = p_i$, $i = 1, 2, 3, 4$. This is a highly reasonable result. There is no outside information and, thus, there should be no change in any of the plausibilities considered.

Case II. Again, two propositions are considered. In this case, though, one of the marginals will be assumed to change due to outside information. The choice is arbitrary and it will be assumed that the plausibility $p_1(b) = r$ changes to $p_2(b) = s$. The restraining equations are then

$$q_1 + q_3 = s, \text{ and} \tag{4.3.1}$$

$$q_1 + q_2 + q_3 + q_4 = 1. \tag{4.3.2}$$

In the notation of Theorem 4.2, $i = 1, 2, 3, 4$ and $j = 1$ with $k_1 = s$ and $T_1(1) = 1$, $T_1(2) = 0$, $T_1(3) = 1$, and $T_1(4) = 0$. The $T_1(i)$ are obtained from (4.3.1). Thus, the minimum is given by (let $\lambda_1 = \lambda$)

$$q_1 = p_1 e^{-\lambda - \mu - 1} \tag{4.4.1}$$

$$q_2 = p_2 e^{-\mu - 1} \tag{4.4.2}$$

$$q_3 = p_3 e^{-\lambda - \mu - 1} \tag{4.4.3}$$

$$q_4 = p_4 e^{-\mu - 1} \tag{4.4.4}$$

From (4.4.1) and (4.4.3) it follows that $q_1/q_3 = p_1/p_3$ and from (4.4.2) and (4.4.4) it follows that $q_2/q_4 = p_2/p_4$. But $q_1 + q_3 = s$ so that

$$s = q_1 + q_3 = \frac{p_1}{p_3} q_3 + q_3 = (\frac{p_1}{p_3} + 1)\, q_3 = \frac{(p_1 + p_3)}{p_3} q_3 = \frac{r}{p_3} q_3$$

and

$$q_3 = \frac{s}{r} p_3.$$

Then,

$$q_1 = s - q_3 = s - \frac{s}{r} p_3 = s \frac{(r - p_3)}{r} = \frac{s}{r} p_1.$$

Also,

$$1 - s = q_2 + q_4 = \frac{p_2}{p_4} q_4 + q_4 = (\frac{p_2}{p_4} + 1)\, q_4 = \frac{(p_2 + p_4)}{p_4}\, q_4 = \frac{(1 - r)}{p_4}\, q_4$$

and

$$q_4 = \frac{(1-s)}{(1-r)}\, p_4.$$

Then,

$$q_2 = 1 - s - p_4 = 1 - s\, \frac{(1-s)}{(1-r)}\, p_4 = (1-s)\, \frac{(1-s-p_4)}{(1-r)} = \frac{(1-s)}{(1-r)}\, p_2.$$

Now, $p_2(a) = v$ is unknown. However,

$$v = q_1 + q_2 = \frac{s}{r}\, p_1 + \frac{(1-s)}{(1-r)}\, p_2.$$

Similarly, it is found that

$$1 - v = \frac{s}{r}\, p_3 + \frac{(1-s)}{(1-r)}\, p_4.$$

In terms of the original notation which is, perhaps, more informative the solution is

$$p_2(a.b) = \frac{p_2(b)}{p_1(b)}\, p_1(a.b) \qquad (4.5.1)$$

$$p_2(a.b') = \frac{p_2(b')}{p_1(b')}\, p_1(a.b') \qquad (4.5.2)$$

$$p_2(a'.b) = \frac{p_2(b)}{p_1(b)}\, p_1(a'.b) \qquad (4.5.3)$$

$$p_2(a'.b') = \frac{p_2(b')}{p_1(b')}\, p_1(a'.b') \qquad (4.5.4)$$

$$p_2(a) = \frac{p_2(b)}{p_1(b)}\, p_1(a.b) + \frac{p_2(b')}{p_1(b')}\, p_1(a.b') \qquad (4.5.5)$$

$$p_2(a') = \frac{p_2(b)}{p_1(b)}\, p_1(a'.b) + \frac{p_2(b')}{p_1(b')}\, p_1(a'.b') \qquad (4.5.6)$$

If $r = s$, that is, $p_1(b) = p_2(b)$, then inspection of equations (4.5.1) to (4.5.6) reveals that the solution reduces to the solution obtained in Case I. This is reasonable since if $r = s$, then the outside information has produced no change and, therefore, no plausibilities should change.

If a and b are independent, then the solution becomes

$$p_2(a.b) = p_1(a)\ p_2(b) \tag{4.6.1}$$

$$p_2(a.b') = p_1(a)\ p_2(b') \tag{4.6.2}$$

$$p_2(a'.b) = p_1(a')\ p_2(b) \tag{4.6.3}$$

$$p_2(a'.b') = p_1(a')\ p_2(b) \tag{4.6.4}$$

$$p_2(a) = p_1(a)\ p_2(b) + p_1(a)\ p_2(b') = p_1(a) \tag{4.6.5}$$

$$p_2(a') = p_1(a')\ p_2(b) + p_1(a')\ p_2(b') = p_1(a') \tag{4.6.6}$$

The results (4.6.5) and (4.6.6) are not wholly unexpected in view of Axiom 3.4. They imply that if a and b are independent, then any change in the plausibility of b does not produce a change in the plausibility of a. This is a most pleasing result.

Recall that the information function $i(q, p)$ was not defined when either p or q (or both) is zero. However, it is possible to make a definition for the total information function that is consistent in a certain sense when p is zero. Suppose that $a \leq b$; that is, a implies b. Then $p(a.b') = 0$ always. If this result is substituted into the equations (4.5.1) through (4.5.6) the following equations are obtained.

$$p_2(a.b) = p_2(a) = \frac{p_2(b)}{p_1(b)}\ p_1(a.b) = \frac{p_2(b)}{p_1(b)}\ p_1(a) \tag{4.7.1}$$

$$p_2(a.b') = 0 \tag{4.7.2}$$

$$p_2(a'.b) = \frac{p_2(b)}{p_1(b)}\,[p_1(b) - p_1(a)] \tag{4.7.3}$$

$$p_2(a'.b') = \frac{p_2(b')}{p_1(b')}\,p_1(a'.b') = p_2(b') \tag{4.7.4}$$

$$p_2(a') = p_2(b) - \frac{p_2(b)}{p_1(b)}\,p_1(a) + p_2(b') = 1 - \frac{p_2(b)}{p_1(b)}\,p_1(a) \tag{4.7.5}$$

Now, in the total information if the denominator of any ln term is zero set the whole term equal to zero. For the example being considered this results in

$$I = q_1 \ln \frac{q_1}{p_1} + q_3 \ln \frac{q_3}{p_3} + q_4 \ln \frac{q_4}{p_4}.$$

The restraints are

$$q_1 + q_3 = s$$

$$q_4 = 1 - s$$

so that the last term of I is not necessary. The resulting problem is to minimize

$$I = q_1 \ln \frac{q_1}{p_1} + q_3 \ln \frac{q_3}{p_3} \tag{4.8.1}$$

subject to

$$q_1 + q_3 = s. \tag{4.8.2}$$

Diagrammatically the problem is this.

Stage 1

	b	b'	
a	p_1	0	u
a'	p_3	p_4	1-u
	r	1-r	1

r changes to s $\longrightarrow$

Stage 2

	b	b'	
a	q_1	0	v
a'	q_3	q_4	1-v
	s	1-s	1

It is apparent that $v = q_1$, $q_4 = 1 - s$, and that the only undetermined quantities are q_1 and q_3. Since r changes to s the restraining equation (4.8.2) obtains.

Applying Theorem 4.2 to (4.8.1) and (4.8.2) results in

$$q_1 = p_1 e^{-\lambda - \mu - 1}$$

$$q_3 = p_3 e^{-\lambda - \mu - 1}$$

so that

$$\frac{q_1}{q_3} = \frac{p_1}{p_3}.$$

Therefore,

$$q_1 = \frac{p_1}{p_3} q_3 = \frac{p_1(s - q_1)}{p_3}$$

and

$$\frac{q_1}{p_1} + \frac{q_1}{p_3} = \frac{s}{p_3}$$

so that

$$q_1 = \frac{p_1}{r} s.$$

In terms of the original notation the solution is

$$p_2(a.b) = p_2(a) = \frac{p_2(b)}{p_1(b)} p_1(a).$$

Thus, the two solutions are equivalent. Inspection of the example shows that this is so since $p(a, b') = 0$ always and the effect of dropping $q_2 \ln \frac{q_2}{p_2}$ from the total information is the same as if p_2 were not zero and q_2 were known. Thus, any time a p_i is zero that term will be dropped from the total information if p_i must be zero always. If a p_i is zero and the preceding condition is not met, then I is undefined.

Two examples will now be considered that will indicate how the theory developed in this paper can relate to present theories.

Example II.1. In this example proposition a implies proposition b so that $p_2(a) = [p_2(b)/p_1(b)]\, p_1(a)$. The example is typical of problems treated in statistics and subjective inference. A given physical phenomenon is to be observed. It is assumed that a probability law which describes the phenomenon is known except for the value of a constant called the parameter. This law will be denoted by $P_\omega(x)$ where ω is the unknown parameter. Then, it is assumed that $\omega = \omega_0$ and a sample of n x's is observed. If $\omega = \omega_0$, then the probability law is completely specified so that the probability of all possible samples can be computed. Note that this probability exists even if the probability law is (absolutely) continuous. This is so since an exact real number can never be recorded by any measuring device. A given measuring device always rounds off after some given decimal. This accounts for the fact that in practice samples are observed that have the same observation repeated. If the probability law is continuous, then the probability of observing any given number is zero and yet samples are drawn that have repeats. The obvious way to avoid this dilemma is to recognize the fact that an exact real number can never be measured. It is typical in statistical practice to ignore this fact and deal with the sample likelihood or the product of the probability density functions governing each observation. The simplifying assumption is generally made that the probability density is the same for each observation. The sample likelihood is employed because to do so greatly simplifies the mathematical computations. However, in this example likelihood will not be used. The probability law governing the observation of n x's will be denoted by $P_\omega(x_1, \ldots, x_n)$. The problem is then this. If $\omega = \omega_0$, then $P_\omega(x_1, \ldots, x_n) = P_{\omega_0}(x_1, \ldots, x_n)$. Now, $(x_1', \ldots, x_n')$ is observed. Does this support the assumption that $\omega = \omega_0$? In statistics this question is answered by performing a test of hypothesis. The subjectivist answers the question by assuming a subjective probability law on ω and then uses Bayes Theorem to compute

$$P(\omega_0|x') = \frac{P_{\omega_0}(x_1', \ldots, x_n')}{P(x')} P(\omega_0) \tag{4.9}$$

where $P(x') = \sum_\omega P_\omega(x_1', \ldots, x_n')P(\omega)$ and it is assumed that the distribution on ω is discrete. The value of $P(\omega_0|x')$ is used by the subjectivist to answer the question do the observed data support the assumption $\omega = \omega_0$. It is possible to give a solution to the problem in terms of concepts developed in this paper. Let

$a \equiv$ The true value of ω is ω_0.

$b \equiv$ The probability of observing any sample is given by $P_{\omega_0}(x_1, \ldots, x_n)$.

Now, before performing the experiment the beliefs in a and b are given by the initial plausibilities $p_1(a)$ and $p_1(b)$. Observe that a implies

b. That is, if $\omega = \omega_0$, then $P_\omega(x_1, \ldots, x_n) = P_{\omega_0}(x_1, \ldots, x_n)$. Thus,

$$p_2(a) = \frac{p_2(b)}{p_1(b)} p_1(a) \qquad (4.10)$$

where $p_2(a)$ is the belief in a after having observed $(x_1', \ldots, x_n')$. The observing of $(x_1', \ldots, x_n')$ results in $p_1(b)$ changing to $p_2(b)$. It is possible to compare (4.9) and (4.10). To begin with, both formulas are similar in appearance. In fact, $p_1(a)$ corresponds to $P(\omega_0)$ and $p_2(a)$ corresponds to $P(\omega_0|x')$. However, the ratio $p_2(b)/p_1(b)$ does not correspond to the ratio $P_\omega(x_1', \ldots, x_n')/P(x')$ and it is this fact that would make the solutions differ in general. To the writer of this paper there is more intuitive appeal in using the ratio $p_2(b)/p_1(b)$. The value of $p_1(b)$ expresses the belief in the probability law before sampling and the value of $p_2(b)$ expresses the belief in the probability law after sampling $(x_1', \ldots, x_n')$. It seems reasonable for $p_2(a)$ to be related to $p_1(a)$ by the ratio $p_2(b)/p_1(b)$. Unfortunately, in order to use the theory developed in this paper $p_1(a)$ and $p_1(b)$ must be somehow obtained and the more vexing problem of how to change $p_1(b)$ to $p_2(b)$ must be solved.

Example II. 2. In this example it will be shown that with certain assumptions Bayes Theorem holds for plausibilities. Consider (4.5.5) and suppose that b is found to be true. While the problem is not treated in this paper, in any reasonable theory that determines how outside information changes plausibilities if it is found that b is true, then $p_2(b)$ should be set equal to one. If $p_2(b) = 1$, then $p_2(b') = 0$ and (4.5.5) becomes

$$p_2(a) = \frac{p_1(a.b)}{p_1(b)}. \qquad (4.11)$$

Now, (4.11) is the form of the definition of conditional probability which is used to derive Bayes Theorem.

Case III. Two propositions are again considered. However, in this case it is assumed that both marginals change; that is, it will be assumed that $p_1(a) = u$ changes to $p_2(a) = v$ and that $p_1(b) = r$ changes to $p_2(b) = s$. The restraints are then

$$q_1 + q_2 = v \qquad (4.12.1)$$

$$q_1 + q_3 = s \qquad (4.12.2)$$

$$q_1 + q_2 + q_3 + q_4 = 1 \qquad (4.12.3)$$

In the notation of Theorem 4.2, $i = 1, 2, 3, 4$ and $j = 1, 2$ with $k_1 = v$ and $k_2 = s$. The $T_j(i)$ are obtained from (4.12.1) and (4.12.2) and are $T_1(1) = T_1(2) = 1$, $T_1(3) = T_1(4) = 0$, $T_2(1) = T_2(3) = 1$ and $T_2(2) = T_2(4) = 0$. Thus, the minimum is given by

$$q_1 = p_1 e^{-\lambda_1 - \lambda_2 - \mu - 1} \qquad (4.13.1)$$

$$q_2 = p_2 e^{-\lambda_1 - \mu - 1} \qquad (4.13.2)$$

$$q_3 = p_3 e^{-\lambda_2 - \mu - 1} \qquad (4.13.3)$$

$$q_4 = p_4 e^{-\mu - 1} \qquad (4.13.4)$$

From (4.13.1) and (4.13.2) it follows that

$$\frac{q_1}{q_2} = \frac{p_1}{p_2} e^{-\lambda_2}$$

and from (4.12.1) it then follows that

$$v = \frac{p_1}{p_2} q_2 e^{-\lambda_2} + q_2 = q_2 \frac{e^{-\lambda_2}}{p_2} (p_1 + p_2 e^{\lambda_2})$$

Therefore,

$$q_2 = \frac{p_2 v e^{\lambda_2}}{(p_1 + p_2 e^{\lambda_2})} \tag{4.14.1}$$

and

$$q_1 = \frac{p_1 v}{(p_1 + p_2 e^{\lambda_2})} . \tag{4.14.2}$$

Then, from (4.13.3) and (4.13.4) it follows that

$$\frac{q_4}{q_3} = \frac{p_4}{p_3} e^{\lambda_2}$$

Now, $q_3 + q_4 = 1 - v$ so that

$$1 - v = q_3 + q_3 \frac{p_4}{p_3} e^{\lambda_2} = \frac{q_3}{p_3} (p_3 + p_4 e^{\lambda_2})$$

and

$$q_3 = \frac{p_3(1 - v)}{(p_3 + p_4 e^{\lambda_2})} . \tag{4.14.3}$$

By employing (4.12.2) it follows that from (4.14.2) and (4.14.3) that

$$s = q_1 + q_3 = \frac{p_1 v}{(p_1 + p_2 e^{\lambda_2})} + \frac{p_3(1 - v)}{(p_3 + p_4 e^{\lambda_2})}$$

or $s(p_1 + p_2 e^{\lambda_2})(p_3 + p_4 e^{\lambda_2}) = p_1 v (p_3 + p_4 e^{\lambda_2}) + p_3(1 - v)(p_1 + p_2 e^{\lambda_2})$

By simple algebraic manipulation a quadratic in e^{λ_2} is obtained. It is

$$e^{2\lambda_2} + \left[\frac{(s - v)p_1 p_4 + (s + v - 1)p_2 p_3}{p_2 p_4 s} \right] e^{\lambda_2} - \frac{p_1 p_3}{p_2 p_4} \frac{(1 - s)}{s} = 0. \tag{4.14.4}$$

The constant term in (4.14.4) is always greater than zero so that (4.14.4) has a solution with one positive and one negative root. The positive root is the only permissible one since $e^{\lambda_2} > 0$ always. Therefore, there is a unique solution to (4.14.4) for e^{λ_2}.

It was previously shown that

$$\frac{q_1}{q_2} = \frac{p_1}{p_2} e^{-\lambda_2}.$$

Now, from (4.13.1) and (4.13.3) it follows that

$$\frac{q_1}{q_3} = \frac{p_1}{p_3} e^{-\lambda_1}.$$

Thus,

$$\frac{p_1}{p_3} q_3 e^{-\lambda_1} = \frac{p_1}{p_2} q_2 e^{-\lambda_2}$$

or

$$e^{-\lambda_1} = \frac{p_3}{p_2} \frac{q_2}{q_3} e^{-\lambda_2}$$

and employing (4.14.1) and (4.14.3) yields

$$e^{-\lambda_1} = \frac{v}{(1 - v)} \frac{(p_3 + p_4 e^{\lambda_2})}{(p_1 + p_2 e^{\lambda_2})}. \qquad (4.14.5)$$

Now, by employing (4.12.3) and using the solutions for $e^{-\lambda_1}$ and $e^{-\lambda_2}$ it follows that

$$e^{-\mu - 1} = [p_1 e^{-\lambda_1 - \lambda_2} + p_2 e^{-\lambda_1} + p_3 e^{-\lambda_2} + p_4]^{-1}. \qquad (4.14.6)$$

It is now possible to solve for the q_i by using the solutions for $e^{-\lambda_1}$,

$e^{-\lambda_2}$, and $e^{-\mu - 1}$ and equations (4.13.1) through (4.13.4).

Now let r = s and v = v, that is, $p_1(b) = p_2(b)$ and $p_1(a) = p_2(a)$. Then (4.14.4) becomes

$$e^{2\lambda_2} + \left[\frac{p_1p_3}{p_2p_4}\frac{(1 - r)}{r} - 1\right]e^{\lambda_2} - \frac{p_1p_3}{p_2p_4}\frac{(1 - r)}{r} = 0. \qquad (4.15.1)$$

Set $\frac{p_1p_3}{p_2p_4}\frac{(1 - r)}{r} = w$ so that (4.15.1) may be written as

$$e^{2\lambda_2} + (w - 1)e^{\lambda_2} - w = 0 \qquad (4.15.2)$$

where $w > 0$. The roots of (4.15.2) are $-w$ and 1 and since $e^{\lambda_2} > 0$ always, the root 1 is the only acceptable one. Thus, $e^{\lambda_2} = 1$ and therefore, $\lambda_2 = 0$. With $\lambda_2 = 0$ (4.15.5) becomes

$$e^{-\lambda_1} = \frac{v}{(1 - v)}\frac{(p_3 + p_4)}{(p_1 + p_2)} = \frac{u}{(1 - u)}\frac{(1 - u)}{u} = 1$$

so that $\lambda_1 = 0$. Finally (4.15.6) becomes

$$e^{-\mu - 1} = [p_1 + p_2 + p_3 + p_4]^{-1} = 1$$

or $\mu = -1$. It is now possible to solve (4.13.1) through (4.13.4) for the q_i. The solution is, obviously, $q_i = p_i$, i = 1, 2, 3, 4. Thus, again, if there is no change in the marginals, then there is no change in any of the plausibilities and the solution is the same as in Case I.

Suppose now that a and b are independent. Since no explicit solution for the q_i has been obtained it will be necessary to return to equations (4.13.1) through (4.13.4) to obtain a solution when a and b are independent. The equations with independence are

$$q_1 = ure^{-\lambda_1 - \lambda_2 - \mu - 1} \tag{4.16.1}$$

$$q_2 = u(1 - r)e^{-\lambda_1 - \mu - 1} \tag{4.16.2}$$

$$q_3 = (1 - u)re^{-\lambda_2 - \mu - 1} \tag{4.16.3}$$

$$q_4 = (1 - u)(1 - r)e^{-\mu - 1} \tag{4.16.4}$$

From (4.16.1) and (4.16.2) it follows that

$$\frac{q_1}{q_2} = \frac{r}{1 - r}e^{-\lambda_2}$$

and from (4.16.3) and (4.16.4) it follows that

$$\frac{q_3}{q_4} = \frac{r}{1 - r}e^{-\lambda_2}.$$

Thus,

$$\frac{q_1}{q_2} = \frac{q_3}{q_4}.$$

Now,

$$v = q_1 + q_2 = \frac{q_2 q_3}{q_4} + q_2 = \frac{q_2}{q_4}(q_3 + q_4) = \frac{q_2}{q_4}(1 - v)$$

so that

$$\frac{q_2}{q_4} = \frac{v}{1 - v}$$

and, also,

$$\frac{q_1}{q_3} = \frac{v}{1 - v}.$$

But

$$s = q_1 + q_3 = q_3 \frac{v}{(1 - v)} + q_3 = q_3 \frac{1}{(1 - v)}$$

so that

$$q_3 = s(1 - v). \qquad (4.17.1)$$

Then,

$$q_1 = q_3 \frac{v}{(1 - v)} = sv. \qquad (4.17.2)$$

Now,

$$v = q_1 + q_2$$

so that

$$q_2 = v - q_1 = v - sv = (1 - s)v. \qquad (4.17.3)$$

Finally,

$$q_4 = \frac{(1 - v)}{v} q_2 = (1 - s)(1 - v). \qquad (4.17.4)$$

In terms of the original notation the solution is

$$p_2(a.b) = p_2(a)p_2(b) \qquad (4.18.1)$$

$$p_2(a.b') = p_2(a)p_2(b') \qquad (4.18.2)$$

$$p_2(a'.b) = p_2(a')p_2(b) \qquad (4.18.3)$$

$$p_2(a'.b') = p_2(a')p_2(b) \qquad (4.18.4)$$

Thus, it turns out that if a and b are independent, then the plausibilities of the various conjunctions of a and b are the products of the separate plausibilities. Therefore, minimizing I comes up with the correct result. For if a and b are independent, then $p(a.b) = p(a)p(b)$ always and similarly for the other conjunctions.

<u>Case IV</u>. In this the concluding case, three propositions will be considered briefly. With three propositions it is possible to impose three restraints that express the plausibilities of the individual propositions in addition to the restraint that the sum of all q's is one. If either two or three restraints are imposed the solutions are difficult to obtain and are not explicit. Therefore, only the case of one additional restraint will be considered. Assume that the plausibility of a changes due to outside information and denote $p_2(a)$ by s, $p_1(a)$ by r, and let the following correspondence of propositions and plausibilities hold:

abc: p_1 and q_1

abc': p_2 and q_2

ab'c: p_3 and q_3

ab'c': p_4 and q_4

a'bc: p_5 and q_5

a'bc': p_6 and q_6

a'b'c: p_7 and q_7

a'b'c': p_8 and q_8

The linear restraint is given by

$$q_1 + q_2 + q_3 + q_4 = s \tag{4.19}$$

and from Theorem 4.2 it follows that

$$q_1 = p_1 e^{-\lambda - \mu - 1} \tag{4.20.1}$$

$$q_2 = p_2 e^{-\lambda - \mu - 1} \tag{4.20.2}$$

$$q_3 = p_3 e^{-\lambda - \mu - 1} \tag{4.20.3}$$

$$q_4 = p_4 e^{-\lambda - \mu - 1} \tag{4.20.4}$$

$$q_5 = p_5 e^{-\mu - 1} \tag{4.20.5}$$

$$q_6 = p_6 e^{-\mu - 1} \tag{4.20.6}$$

$$q_7 = p_7 e^{-\mu - 1} \tag{4.20.7}$$

$$q_8 = p_8 e^{-\mu - 1} \tag{4.20.8}$$

Summing (4.20.1) through (4.20.4) yields

$$s = re^{-\lambda - \mu - 1}$$

so that

$$e^{-\lambda - \mu - 1} = \frac{s}{r}. \tag{4.21.1}$$

Similarly, working with (4.20.5) through (4.20.8) yields

$$e^{-\mu - 1} = \frac{(1 - s)}{(1 - r)}. \tag{4.21.2}$$

Now, solving for $p_2(b)$ results in

$$\begin{aligned} p_2(b) &= q_1 + q_2 + q_5 + q_6 \\ &= p_1 \frac{s}{r} + p_2 \frac{s}{r} + p_5 \frac{(1 - s)}{(1 - r)} + p_6 \frac{(1 - s)}{(1 - r)} \\ &= (p_1 + p_2) \frac{s}{r} + (p_5 + p_6) \frac{(1 - s)}{(1 - r)}. \end{aligned}$$

In terms of the original notation the solution for $p_2(b)$ is

$$p_2(b) = p_1(a.b)\,\frac{p_2(a)}{p_1(a)} + p_1(a'.b)\,\frac{p_2(a')}{p_1(a')}. \qquad (4.22)$$

Obviously, similar equations hold for $p_2(b')$, $p_2(c)$, and $p_2(c')$. Thus, with three propositions if only one individual proposition changes plausibilities from outside information, then the solution for the remaining two is identical to the solution with two propositions when one proposition changes due to outside information. Again, minimizing I works quite well.

CHAPTER V

SUMMARY

In this paper a study of the concept of a subjective inferential theory of probability is made. The term plausibility is used to name this form of subjective probability. Intuitively, plausibilities are to be thought of as expressing an "ideal" person's degree of belief in the truth or falsity of some proposition. The set of propositions under study is assumed to form a Boolean algebra and truth or falsity is expressed by a homomorphism from this Boolean algebra onto the Boolean algebra $\{O, I\}$. If a proposition maps into I it is true and if it maps into O it is false. The difficulty in practice is that seldom is the homomorphism known; thus, there is a need for some method of expressing degrees of belief about propostions whose truth or falsity is not known. This concept of truth or falsity is also employed in defining independent propostions.

The primary problem investigated is this. If initially the plausibilities of certain of the propositions are given and then some of these are known to change by a known amount, then how should the plausibilities of the remaining propositions change. In general, in any scientific theory the propositions are related. Thus, if the plausibility of one proposition has changed from one value to another, then if a person is acting "rationally," the plausibilities of propositions related to this proposition should change in some calculable manner.

This problem is a special case of the general inference problem.

A solution to this special case of the general inference problem is obtained in this paper. The solution is accomplished in two steps. First, axiomatically, a theory of plausibilities and a total information function for a special set of propositions are derived. The theory of plausibilities is similar to elementary probability theory with the exception that no theory concerning conditional plausibilities is derived. This is so since no definition or use is made of the concept of a conditional plausibility. The total information function derived corresponds to the mean information function extensively investigated by Kullback [21]. The total information is a function of the plausibilities of the propositions in the special set for which total information is defined. The second step is to use plausibilities and information in some manner to obtain the desired solution. The technique employed is to minimize the total information subject to the restrictions imposed by knowing the values of the plausibilities of those propositions whose plausibilities have changed.

The use of the minimization principle is investigated for various applications. It is found that minimizing total information (subject to restraints) gives certain desirable results:

(1) If two propositions are considered and the plausibility of neither is known to change, then the plausibility of any proposition made up from these two by any of the operations of disjunction, conjunction, or negation does not change.

(2) If two propositions are considered and the plausibility of, say, the first changes value, then the plausibility of the

second is related to the first by a function that is a generalization of Bayes Theorem. This function has the property that if the change in the plausibility of the first is zero, then there is no change in the plausibility of the second. If the two propositions are independent, then the plausibility of the second does not change. If the first implies the second, then an alternative solution is obtained for the Bayesian inference problem.

(3) If two propositions are considered and the plausibility of both changes, then it is possible to obtain a solution for the value of the plausibility of any proposition obtained from these two by any of the operations of disjunction, conjunction, or negation. If the change in plausibility of both is zero, then the plausibility of any proposition made up from these two does not change. If the two propositions are independent, then the plausibility of the conjunction of the first or its negation and the second or its negation is the product of their respective plausibilities.

(4) If three propositions are considered and the plausibility of, say, the first changes value, then the solution for the plausibility of the second and the third is the same as the solution obtained in part (2). Because of the complexity of the solution the case where either two or three of the propositions change plausibilities was not considered.

It is hoped that the theories of plausibilities and information developed in this paper can be employed in a meaningful manner to obtain solutions to various special problems of the general inference problem. The technique of minimizing the total information gives reasonable results in all of the cases investigated. Two difficulties are, however, encountered. First, in considering three or more but at most a finite number of propositions the solution for the cases when two or more of the propositions change plausibilities is extremely complex. Second, there appears to be no reasonable way to extend the technique of minimizing the total information to the case when the Boolean algebra considered has an infinity of elements. Before this problem could even be investigated the algebra of plausibilities would somehow have to be extended so that plausibilities are countably additive.

BIBLIOGRAPHY

1. Anderson, T. W. An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, 1958.

2. Anscombe, F. J. and R. J. Aumann. "A Definition of Subjective Probability." Annals of Mathematical Statistics, 34, (1963), 199-205.

3. Birkhoff, G. and S. MacLane. A Survey of Modern Algebra, 3rd edition. New York: Macmillan, 1965.

4. Blackwell, D. and M. A. Girshick. Theory of Games and Statistical Decisions. New York: John Wiley and Sons, 1954.

5. Carnap, R. Logical Foundations of Probability, 2nd edition. Chicago: The University of Chicago Press, 1962.

6. Chernoff, H. "Rational Selection of Decision Functions." Econometrica, 22, (1954), 422-443.

7. Cox, R. T. The Algebra of Probable Inference. Baltimore: The Johns Hopkins Press, 1961.

8. Cramér, H. Mathematical Methods of Statistics. Princeton: Princeton University Press, 1946.

9. De Finetti, B. "Foresight: Its Logical Laws, Its Subjective Sources." Studies in Subjective Probability (ed. Kyburg, Smokler). New York: John Wiley and Sons, 1964.

10. Fisher, R. A. Statistical Methods and Scientific Inference. New York: Hafner Publishing Company, 1956.

11. Good, I. J. Probability and the Weighing of Evidence. New York: Hafner Publishing Company, 1950.

12. Halmos, P. R. "The Basic Concepts of Algebraic Logic." Algebraic Logic (ed. Halmos). New York: Chelsea Publishing Company, 1962.

13. Halmos, P. R. Lectures on Boolean Algebras. Princeton: D. Van Nostrand Company, 1963.

14. Jaynes, E. T. "Probability Theory in Science and Engineering." Colloquim Lectures in Pure and Applied Science No. 4. Dallas: Socony Mobil Oil Company, Inc., 1958.

15. Jeffreys, H. Theory of Probability, 3rd edition. Oxford: Oxford University Press, 1961.

16. Kaplan, W. Advanced Calculus. Reading: Addison-Wesley, 1952.

17. Kemeny, J. G., A. Schleifer, J. L. Snell, and G. L. Thompson. Finite Mathematics with Business Applications. Englewood Cliffs: Prentice-Hall, 1962.

18. Keynes, J. M. A Treatise on Probability. London: Macmillan, 1921 (Reprinted: Harper Torchbooks, New York, 1962).

19. Kolmogorov, A. Foundations of the Theory of Probability, 2nd English edition. New York: Chelsea Publishing Company, 1956.

20. Koopman, B. O. "The Bases of Probability." Studies in Subjective Probability. (ed. Kyburg, Smokler). New York: John Wiley and Sons, 1964.

21. Kullback, S. Information Theory and Statistics. New York: John Wiley and Sons, 1959.

22. Luce, R. D. and H. Raiffa. Games and Decisions. New York: John Wiley and Sons, 1957.

23. Neyman, J. and E. S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." Phil. Trans. Roy. Soc., A231, (1932), 289-337.

24. Neyman, J. "L'estimation statistique, traitée comme un problème classique de probabilité." Actualités scientifiques et industrielles No. 739, pp. 25-27. Paris: Hermann et Cie, 1938.

25. Neyman, J. Lectures and Conferences on Mathematical Statistics and Probability, 2nd revised edition. Washington, D. C.: Department of Agriculture, 1952.

26. Parzen, E. Stochastic Processes. San Francisco: Holden-Day, 1962.

27. Polya, G. Mathematics and Plausible Reasoning, Volume II: Patterns of Plausible Inference. Princeton: Princeton University Press, 1954.

28. Pratt, J. W., H. Raiffa, and R. Schlaifer. "The Foundations of Decision Under Uncertainty: An Elementary Exposition." Journal of the American Statistical Association, 59, (1964), 353-375.

29. Raiffa, H. and R. Schlaifer. Applied Statisitical Decision Theory. Boston: Harvard University Press, 1961.

30. Ramsey, F. The Foundations of Mathematics and Other Logical Essays. (ed. Braithwaite). New York: Humanities Press, 1950.

31. Rényi, A. "On Measures of Entropy and Information." Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, I, (1961), 547-562.

32. Savage, L. J. The Foundations of Statistics. New York: John Wiley and Sons, 1954.

33. Savage, L. J. "The Foundations of Statistics Reconsidered." Studies in Subjective Probability (ed. Kyburg, Smokler). New York: John Wiley and Sons, 1964.

34. Stewart, C. C. "The Statistical Decision Problem." (unpub. Masters report). Stillwater: Oklahoma State University, 1966.

35. Von Mises, R. Mathematical Theory of Probability and Statistics (ed. and complemented by H. Geringer). New York: Academic Press, 1964.

36. Von Neumann, J. and O. Morgenstern. Theory of Games and Economic Behavior, 3rd edition. Princeton: Princeton University Press, 1953.

37. Wald, A. "Contributions to the Theory of Statistical Estimation and Testing Hypothesis." Annals of Mathematical Statistics, 10, (1939), 299-326.

38. Wald, A. Statistical Decision Functions. New York: John Wiley and Sons, 1950.

VITA

William Talbert Tucker, Jr.

Candidate for the Degree of

Doctor of Philosophy

Thesis: A THEORY OF SUBJECTIVE INFERENTIAL PROBABILITY

Major Field: Mathematics and Statistics

Biographical:

Personal Data: Born in Memphis, Tennessee, March 11, 1934, the son of Mr. and Mrs. William T. Tucker, Sr.

Education: Graduated from Messick High School, Memphis, Tennessee, in June, 1952; received the Bachelor of Mechanical Engineering degree from Georgia Institute of Technology, Atlanta, Georgia, in June, 1957; received the Master of Science degree from Southern Methodist University, Dallas, Texas, with a major in Mechanical Engineering, in August, 1961; received the Master of Science degree from Southern Methodist University, Dallas, Texas, with a major in Statistics, in May, 1963; completed requirements for the Doctor of Philosophy degree in July, 1966.

Professional Experience: Technical Engineer, General Dynamics/Fort Worth, Fort Worth, Texas, 1957-1961; Graduate Assistant in the Department of Statistics, Southern Methodist University, Dallas, Texas, 1961-1963; Quality Control Engineer, Texas Instruments, Dallas, Texas, during the summer of 1962; Statistician in the Geophysics Laboratory, Southern Methodist University, Dallas, Texas during the summers of 1963, 1964 and 1965; Graduate Teaching Assistant in the Department of Mathematics and Statistics, Oklahoma State University, Stillwater, Oklahoma, 1963-1966; registered Professional Engineer in the state of Texas.