11. M. Tribus, *Rational Descriptions, Decisions and Designs*, p. 130. Oxford: Pergamon, 1969.
12. R. Kashyap, Prior Probability and Uncertainty, *IEEE Trans. Information Theory* **IT-17**, 641–650, 1971.
13. H. Chernoff, Rational Selection of Decision Functions, *Econometrica* **22**, 422–443, 1954.
14. J. Milnor, Games Against Nature, in *Decision Processes* (R. Thrall, C. Coombs, and R. Davis, eds.), pp. 49–59. New York: Wiley, 1954.
15. C. Stein, Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution, *Proc. Symp. Math. Statist. and Probability 3rd, Berkeley* **1**, pp. 197–206. Berkeley: Univ. of California Press, 1956.

VII

Logical (Conditional) Probability

VIIA. Introduction

In the preceding chapter we encountered the view that the probability of an event might be assessed from a balance of evidence, or lack of evidence, in favor of the occurrences of each of an exhaustive set of mutually exclusive alternatives, some subset of which comprises the event. However, the classical theory lacked guidelines for the identification of a balance of evidence and was open to inconsistent application. Furthermore, the domain of the classical theory, although not the domains of its reformulations, covered only those situations where the alternatives were of equal weight; only equiprobability assignments were possible. Spurred by the goals of classical probability and its failure to achieve them, there has been a development, by Keynes [1], Jeffreys [2], Koopman [3], Carnap [4], and others, of more flexible theories of logical probability. These theories of inductive logic or nondemonstrative reasoning aim at providing rational or reasonable assessments of the degree to which an hypothesis H is supported by evidence E or, more loosely, the probability that H is true given E.

The formal domain of a theory of logical probability is generally a set of inferences between statements or propositions in a language, rather than the set of statements themselves, and it is distinct from the domain of an empirical theory (set of events or experimental outcomes) and the domain of a subjective theory (set of beliefs of an individual). Formally the distinction between an inference-domain theory and one with a statement-domain is parallel to that between conditional and absolute probability. In general, we can generate a statement-domain theory from an inference-domain theory by restricting the latter to the case of a fixed, reasonably selected, evidence proposition (e.g., a tautology or logical truth). However, the generation of an inference-domain theory from a statement-domain theory is likely to be more difficult (see Subsection IIE3).

Logical probability attempts to explicate induction by defining a logical relation between an evidence statement and an hypothesis statement that is a generalization of the relations of implication and contradiction available from deductive logic. There are several views, however, as to the formal and interpretive nature of the evidence–hypothesis relation.

As was mentioned in Chapter I, there is a weak formal concept of classificatory probability. While ignored by other approaches to probability, the classificatory concept has some interest when examined from the viewpoints of modal logic and logical probability. We then examine Koopman's somewhat stronger theory of comparative logical probability. While Koopman discusses an agreeing quantitative logical probability theory, his subjective interpretation of probability gives us few guidelines for the actual determination of logical probability. Finally, we consider Carnap's more ambitious program for quantitative logical probability. Carnap originally sought a unique, quantitative logical relation, called degree of confirmation (d.c.), to measure the support one statement lends to another. This support is seen as being of an analytical, necessary, or logical nature rather than of a synthetic, contingent, or empirical nature. Just as implication is determinable from the meaning of statements irrespective of their truth, so should a degree of confirmation be determined without regard to either the truth of the statements involved or the contingent, factual aspects of the world. While it appears that Carnap's theory is as yet incomplete, it is sufficiently developed for us to give examples of its application.

The various theories of logical probability have in common, beyond agreement as to the domain of probability properly being that of inferences, the derivation of quantitative probability by guarded use of the principle of indifference and the desire to justify themselves by demonstrating some form of agreement with a relative-frequency outlook. In Carnap's and Koopman's theories, logical probability is either an estimate of or converges to the relative-frequency of the number of individuals for which the hypothesis statement proved to be correct (truth frequency). Going further, Carnap insists that logical probability also be pragmatically justifiable by appeal to a theory of rational decision making. Correct reasoning is to find its *raison d'être* in correct or rational decisions and judgments.

VIIB. Classificatory Probability and Modal Logic

While the classificatory concept of probability ("A is probable") is a very weak characterization of randomness or uncertainty in many interpretations of probability and has therefore been ignored, it finds a place in logical probability through modal logic [5, 6]. Paralleling common usage and notwithstanding our earlier remarks, we will treat a version of classificatory logical probability in which the domain is the set of propositions rather than the set of inferences between propositions. Although we will not do so, the ensuing discussion can be formally applied to inferences.

In brief, a modal probabilistic logic for propositions adjoins to a propositional logic the operator "$\mathscr{P}$" which when prefixed to a proposition p, "$\mathscr{P}p$," is read either as "probably p" or "p is probably true." The operator $\mathscr{P}$ is a mapping from the space of propositions into the space of propositions. The relation between $\mathscr{P}$ and the usual propositional logic symbols of conjunction ($\wedge$), disjunction ($\vee$), negation ($\sim$), implication ($\Rightarrow$), and parenthesis [(,)], can be axiomatized in several, inequivalent, ways. A reasonable set of axioms for $\mathscr{P}$ might include the following:

M1. $(\forall p)\ (\mathscr{P}(p \vee \sim p))$.

M2. $(\forall p) \sim (\mathscr{P}p \wedge \mathscr{P}(\sim p))$.

M3. $(\forall p, q)\ ((p \Rightarrow q) \wedge \mathscr{P}p \Rightarrow \mathscr{P}q)$.

Axiom M1 asserts that a certain kind of tautology is probable. In an event language model in which the propositions describe the occurrences of events, the counterpart of M1 is that the occurrence of the certain event (Ω) is probable. Axioms M1 and M3 together assert the probability of any tautology. Axiom M2 denies the existence of a proposition p for which p and $\sim p$ are both probable. In terms of the usual quantitative probability P, M2 could be modeled for a proposition p_A describing an event A by

$$(\exists t \geqslant \tfrac{1}{2})\ \text{“}\mathscr{P}p_A\text{”} \quad \text{iff} \quad \text{“}P(A) > t.\text{”}$$

Axiom M3 is an axiom of detachment. In the event model M3 corresponds to "B is probable and $A \supseteq B$ implies A is probable." While Hempel [7] has challenged the acceptability of detachment rules in probability, we will accept M3.

Another possible axiom is

M4. $(\forall p)\ (\mathscr{P}p \vee \mathscr{P}(\sim p))$.

Axiom M4 asserts that for any proposition, either it or its contradictory is in the domain of the operator $\mathscr{P}$. However, it might be felt that the implications of M2 and M4 are too strong. In particular, M2 and M4 imply that

$$(\forall p)\,(\mathscr{P}p \vee \sim\mathscr{P}p);$$

that is, for every proposition we can determine whether or not it is probable. Some propositions may be neither probable nor improbable.

In modal logic we are also concerned with the formal properties of iterated modalities, a simple example being "$\mathscr{P}\mathscr{P}p$" (read "probably p is probable"). A possible basis for interpreting the iterated modality of probability might be the following analogy with parametric statistical models, wherein we do not assume M4. Assume that there is a family $\{P_\theta, \theta \in \Theta\}$ of probability measures, one of which correctly describes the random experiment, and that there is a prior distribution π on a suitable σ-field of subsets of Θ that includes singleton events. Introduce the correspondences:

Classificatory logical probability	*Statistical*
Proposition "p_A"	Event "A" described by "p_A"
$\mathscr{P}p_A$	$\sum_{\theta\in\Theta} \pi(\theta)\, P_\theta(A) > t \geqslant \frac{1}{2}$
$\mathscr{P}\mathscr{P}p_A$	$\pi(\{\theta : P_\theta(A) > t\}) > t$

With this interpretation, suitable for the commonly encountered problems of parametric estimation and decision-making, we see that the reduction of iterated modalities is contingent upon Θ and π and cannot be axiomatized. Hence we cannot assert, for example,

$$\mathscr{P}p \Rightarrow \mathscr{P}\mathscr{P}p \quad \text{or} \quad \mathscr{P}\mathscr{P}p \Rightarrow \mathscr{P}p;$$

the truth of these implications depends on extralogical considerations.

The preceding formulation of the modal operator "$\mathscr{P}$" is quite different from any of the several formulations of the modalities of possibility "M" or necessity "N." While

$$(\forall p)\,(p \Rightarrow Mp),$$

it is not true that

$$(\forall p)\,(p \Rightarrow \mathscr{P}p);$$

an event can have occurred without its having been probable to occur. Furthermore, a consequence of M1–M4 is

$$(\forall p)\,(\mathscr{P}p \Leftrightarrow \sim\mathscr{P}(\sim p)).$$

This conflicts with the usually assumed relation

$$(\forall p)\,(Mp \Leftrightarrow \sim N(\sim p)),$$

unless we make the strange identification $M \equiv N$. Hence, "probability" and "possibility" are very different modalities.

VIIC. Koopman's Theory of Comparative Logical Probability

1. Structure of Comparative Probability

Koopman's theory of logical probability [3] applies to inferences between what he, informally, designates as experimental propositions (statements about events whose truth or falsity is determinable from the performance of an experiment). He introduces a quaternary conditional comparative probability relation, involving two evidence statements, E_1 and E_2, and two hypotheses, H_1 and H_2, that is written "$H_1/E_1 \precsim H_2/E_2$" and read "H_1 on E_1 is no more probable than H_2 on E_2." It is presumed that $\precsim$ is prescribed by an individual, based on his intuition, subject to rationality properties established in nine axioms. With respect to the origin of the $\precsim$ relation, Koopman has said,

> . . . the authority for the first ($\precsim$) proposition does not reside in any general law of probability, logic, or experimental science. And the notion presents itself that such primary and irreducible assumptions are grounded on a basis as much of the aesthetic as of the logical order† [3, p. 774].

† Reprinted with permission of the publisher, The American Mathematical Society, from *Bulletin of the American Mathematical Society*, Copyright © 1940, Vol. 46, p. 774.

Koopman's axioms for $\lesssim$ are as follows. The axioms connecting $\lesssim$ to logical implication are

K1. $(E_2 \Rightarrow H_2) \Rightarrow (H_1/E_1 \lesssim H_2/E_2)$.

K2. $(E_1 \Rightarrow H_1) \wedge (H_1/E_1 \lesssim H_2/E_2) \Rightarrow (E_2 \Rightarrow H_2)$.

The axioms extending the $\lesssim$ relation, or imposing constraints on permitted assignments, include

K3. (*Reflexivity*) $H/E \lesssim H/E$.

K4. (*Transitivity*) $(H_1/E_1 \lesssim H_2/E_2) \wedge (H_2/E_2 \lesssim H_3/E_3) \Rightarrow (H_1/E_1 \lesssim H_3/E_3)$.

K5. $(H_1/E_1 \lesssim H_2/E_2) \Rightarrow (\sim H_2/E_2 \lesssim \sim H_1/E_1)$.

The next two axioms essentially state the product rule for conditional probability.

K6a. Let S_1 and S_2 be non-self-contradictory statements and $(S_1 \Rightarrow S_1' \Rightarrow S_1'')$, $(S_2 \Rightarrow S_2' \Rightarrow S_2'')$. Then

$$(S_1/S_1' \lesssim S_2/S_2') \wedge (S_1'/S_1'' \lesssim S_2'/S_2'') \Rightarrow (S_1/S_1'' \lesssim S_2/S_2'').$$

K6b. $(S_1/S_1' \lesssim S_2'/S_2'') \wedge (S_1'/S_1'' \lesssim S_2/S_2') \Rightarrow (S_1/S_1'' \lesssim S_2/S_2'')$.

K7. Assume the notation of K6 and $S_1/S_1'' \lesssim S_2/S_2''$. If either symbol in $(S_2/S_2', S_2'/S_2'')$ has the relation $\lesssim$ to either symbol in $(S_1/S_1', S_1'/S_1'')$, then the remaining symbol in the second set has the relation $\lesssim$ to the remaining symbol in the first set.

The remaining two axioms, omitted here, become theorems if the ordering of $\lesssim$ is complete; that is, if for all E_1, E_2, H_1, and H_2 either $H_1/E_1 \lesssim H_2/E_2$ or $H_2/E_2 \lesssim H_1/E_1$. Koopman, of course, does not require that all H/E pairs be comparable by $\lesssim$; even comparative probability is not assumed to be universally applicable.

Koopman's axioms for the conditional comparative probability of hypothesis statements given evidence statements have parallels with the axioms presented in Section IIE for the conditional comparative probability of events. Curiously, K2 implies that the only event at least as probable as the certain event is the certain event or equivalently, that null-equivalent events must be null. We did not make this assumption, nor is it usual in probability theory. Interestingly though in the complexity approach to probability it appears that a zero probability event never occurs [8].

2. Relation to Conditional Quantitative Probability

The transition from a "subjectively" selected comparative relation $\lesssim$ to a quantitative probability assignment for H/E is based on the postulated existence of an infinite sequence of "n-scales."

Definition. A set $\{S_1, ..., S_n\}$ is an n-scale if and only if:

(1) At least one S_i is non-self-contradictory.

(2) The conjunction $S_i \wedge S_j$ (to be read "S_i and S_j") of any two distinct statements is a contradiction.

(3) If E is the logical disjunction of the $\{S_i\}$,

$$E = \bigvee_{i=1}^{n} S_i,$$

where $\bigvee_{i=1}^{n} S_i$ is read as "S_1 or S_2 or ... or S_n," then

$$(\forall i, j)\; S_i/E \lesssim S_j/E.$$

Koopman's axiom is

K8. $(\forall n)\, \exists\{S_i\}$ an n-scale.

Axiom K8 can be recognized as a stronger form of Savage's almost uniform partition hypothesis (Subsection IIC4). As we would expect from the discussion of Savage's axiom, K8 only assures us of the existence of a finitely additive, almost agreeing probability P; that is,

$$(H/E \lesssim H'/E') \Rightarrow P(H/E) \leqslant P(H'/E'),$$

but the reverse implication is not necessarily true. Nor does it follow from K8 that $P(H/E)$ has the usual property of conditional probability that

$$P(H/E \wedge E') = \frac{P(H \wedge E'/E)}{P(E'/E)}.$$

Insight into the existence of a quantitative theory is available from the analyses of Sections IIC and IIE, and we will not reinvestigate this question here.

3. Relation to Relative-Frequency

A relation between Koopman's quantitative logical probability P and relative-frequency can be drawn as follows. Consider a language describing repeated experiments in which $\{a_i\}$ represent the trials, Sa_i is the statement of a "success" on trial i, and 0_n is the integer corresponding to the number of successful trials in the first n repetitions. Koopman [9, p. 185] establishes

Theorem 1. Let the evidence E be such that

(1) $(\forall n)(\forall \{i_j\}, \{k_j\}) \left(\left(\bigvee_{j=1}^{n} Sa_{i_j}/E \right) \lesssim \left(\bigvee_{j=1}^{n} Sa_{k_j}/E \right) \right);$

(2) $\lim_{n\to\infty} (0_n/n) = p.$

Then

$$P(Sa_i/E) = p.$$

The evidence E must be such that the order of the trials is irrelevant (permutation invariance); the logical probability of the events of at least one success in n trials does not depend on which n trials we examine. The theorem also assumes that the limit of the relative-frequency of success actually exists, is known, and stated in E. The permutation invariance hypothesis is just de Finetti's hypothesis of exchangeable events (see Sections IVB and VIIE).

4. Conclusions

The significance of Theorem 1 is not the justification of the use of Koopman's logical probability as the estimate of "true" relative-frequency but rather a verification of the reasonableness of the proposed definition of quantitative logical probability. It would be surprising if under the stringent hypotheses of the theorem we found that $P(Sa_i/E) \neq p$. As for the determination of probability directly from relative-frequency, Koopman is well aware that a conclusion asserting a probability relation must be preceded by an antecedent assumption involving a probability relation.

Koopman's analysis of comparative and quantitative probability leaves to the intuition of the user the selection of the comparative relation and a suitable sequence of n-scales. He does not believe it possible to supply compulsory instructions for either selection and is content to leave matters with an axiomatic system capable of supporting an infinite variety of comparative assignments.

VIID. Carnap's Theory of Logical Probability

1. Introduction

In Carnap's view the theory of logical probability concerns a quantitative relation $C(H/E)$ between an hypothesis H and evidence E, called the degree of confirmation function (d.c.), that expresses the degree to which E implies or supports H. Inductive reasoning or inference is a process of assessing the degrees of confirmation of various hypotheses given an evidence statement rather than a process of choosing the uniquely correct or true hypothesis. The choice of a rational or "best" d.c. is to be made on analytic or logical grounds rather than on synthetic or empirical grounds and is to be independent of contingent facts. In this, induction parallels deduction, wherein determinations of implication or contradiction are made without regard to the contingent truth of the statements. All empirical or synthetic knowledge is presumed included in the evidence E. We reserve to Section VIIG our doubts as to the meaningfulness of this analytic/synthetic dichotomy.

Carnap structures the confirmation function C through three sets of axioms [10].† The first set of axioms relates to strict coherence and arises from a desire to develop C so that it will be useful for the purpose of making rational decisions. The second set of axioms concerns invariance properties for C and is motivated by the desires that there be no *a priori* distinguished predicates or individuals and that the relation between propositions H and E depends only on the subset of predicates and individuals they reference. The third set of axioms is concerned with ensuring the ability to learn from experience; they prescribe the behavior of C in certain "intuitively clear" instances of inductive inference.

The sort of languages $\mathscr{L}$ for which Carnap was able to develop his theory of logical probability most fully have finitely many individuals $\{a_i\}$ and finitely many families $\{\mathscr{P}_i\}$ of one-place "primitive" predicates. By a family $\mathscr{P}_i$ of predicates we mean a finite collection $\{P_{i,j}\}$ of predicates such that each individual is describable by one and only one predicate in each family. Other terms and notations of interest to us are as follows.

Definition. A complete description of an individual is a predicate Q of the form

$$Q = \bigwedge_i P_{i,j_i}$$

where the conjunction is over all families.

† Carnap's final remarks on logical probability appear to be those in [22].

The set of all complete descriptions of an individual will be denoted by $\mathcal{Q}$, and q will denote the number of elements in $\mathcal{Q}$. If there are K families of predicates and $\mathcal{P}_i$ has n_i elements, then

$$q = \prod_{i=1}^{K} n_i .$$

Definition. A state description S is a proposition of the form

$$S = \bigwedge_j Q_{i_j} a_j ,$$

where the conjunction is over all individuals.

A state description is a complete description of the world consisting of all the individuals $\{a_i\}$. The set of all state descriptions will be denoted by $\mathcal{S}$.

Definition. The range $\mathcal{R}(H)$ of a proposition H is given by

$$\mathcal{R}(H) = \{S : S \in \mathcal{S}, S \Rightarrow H\}.$$

For example, we may take N individuals $\{a_i\}$ to refer to the N repetitions of an experiment such as the toss of a die, and assume one family $\mathcal{P} = \{P_i\}$ of predicates, where "$P_i a_j$" means "i spots appeared on trial j." When there is only one family of predicates each P_i is itself a complete description. A state description S would be of the form "$\bigwedge_{j=1}^{N} P_{i_j} a_j$" meaning "$i_1$ spots appeared on trial 1 and ... and i_N spots appeared on trial N". The range of, say, $P_j a_k$ is given by

$$\mathcal{R}(P_j a_k) = \Big\{S : (\exists j_1, ..., j_{k-1}, j_{k+1}, ..., j_N)\Big(S \doteq P_j a_k \wedge \bigwedge_{n \neq k} P_{j_n} a_n\Big)$$

and contains 6^{N-1} state descriptions.

After presenting and examining Carnap's axioms for logical probability or d.c., we consider the relation of d.c. to relative-frequency phenomena, illustrate possible applications for d.c., and proceed to a brief critique of d.c.

2. Compatibility with Rational Decision-Making

In his later writings Carnap grew more insistent on the point that the formal analytic theory of d.c. be compatible with applications to rational decision-making. To relate d.c. to decision-making we introduce

the notion of a gamble g on a set $\{H_i\}$ of mutually contradictory statements $((\forall i \neq j) \sim (H_i \wedge H_j))$; g assigns a numerical payoff $g(H_i)$ to each H_i. The d.c. $C(H/E)$ is to have the role of evaluating gambles through the "expected value"

$$v(g) = \sum_i C(H_i/E)\, g(H_i \wedge E)$$

when

$$E \Rightarrow \bigvee_i H_i , \qquad (\forall i \neq j)\, E \Rightarrow \sim(H_i \wedge H_j).$$

Gamble g_1 is to be as good as g_2 if and only if $v(g_1) \geqslant v(g_2)$. This evaluation procedure would be more compelling if g is a utility function [11]; but it is not. The *status quo* is taken as a gamble with zero payoffs. Hence, g is as good as the *status quo* if and only if $v(g) \geqslant 0$.

The requirement of strict coherency arises from the rationality of rejecting gambles such that given E you can never achieve a positive payoff but can sustain a negative payoff. Formally we require of the d.c. that to be useful in decision-making it satisfy

L1. (*Strict coherence*) $\Big(E \Rightarrow \bigvee_i H_i , (\forall i \neq j) E \Rightarrow \sim(H_i \wedge H_j),$

$$(\forall i) g(H_i \wedge E) \leqslant 0, (\exists j)\, g(H_j \wedge E) < 0\Big) \Rightarrow v(g) < 0.$$

A characterization of the class of strictly coherent d.c. is given by

Theorem 2 (de Finetti–Kemeny [12]). C is strictly coherent iff (throughout E, E', and E'' are not self-contradictory):

(1) $0 \leqslant C(H/E) < \infty$.
(2) $(H \Leftrightarrow H', E \Leftrightarrow E') \Rightarrow C(H/E) = C(H'/E')$.
(3) $((E \Rightarrow H), (E' \not\Rightarrow H')) \Rightarrow (C(H/E) > C(H'/E'))$.
(4) $(E \Rightarrow \sim(H \wedge H')) \Rightarrow (C(H \vee H'/E) = C(H/E) + C(H'/E))$.
(5) $C(H \wedge E'/E) = C(H/E \wedge E')\, C(E'/E)$.

Hence $C(H/E)$ has the basic formal properties of a finitely additive conditional probability providing that we adopt the convention that

$$(E \Rightarrow H) \Rightarrow C(H/E) = 1,$$

and remember that

$$C(H/E) = 1 \Rightarrow (E \Rightarrow H).$$

Furthermore, if we select a tautology (logical truth) T and define

$$m(H) = C(H/T),$$

then we can represent

$$C(H/E) = \frac{m(H \wedge E)}{m(E)}.$$

The function m has the formal properties of a finitely additive probability measure with the added constraint that

$$m(H) = 1 \Rightarrow H \text{ a tautology.}$$

In terms of the language $\mathscr{L}$ described in Subsection VIID1, we see that the confirmation C can be represented by

$$C(H/E) = \frac{\sum_{S \in \mathscr{R}(H \wedge E)} m(S)}{\sum_{S \in \mathscr{R}(E)} m(S)}.$$

The strictly coherent confirmation functions, called regular by Carnap, satisfy the constraint

$$(\forall S \in \mathscr{S})\,(m(S) > 0).$$

If we wish to extend $\mathscr{L}$ to include countably many individuals, then we have to relax the requirement of strict coherence to that of coherence:

$$(\forall S \in \mathscr{S})\,(m(S) \geqslant 0), \qquad C(H/E) = \frac{\sum_{S \in \mathscr{R}(H \wedge E)} m(S)}{\sum_{S \in \mathscr{R}(E)} m(S)}.$$

3. Axioms of Invariance

In his attempts to narrow the class of confirmation functions to a unique "best" confirmation function, Carnap [10] invoked several axioms of invariance much as is done in classical probability. In terms of the language $\mathscr{L}$ we may state Carnap's axioms as

L2. $C(H/E)$ is invariant under any permutation of the individuals $\{a_i\}$.

L3. $C(H/E)$ is invariant if the set of individuals $\{a_i\}$ is augmented provided that no quantifiers $(\forall, \exists)$ appear in either H or E.

L4. $C(H/E)$ is invariant under any permutation of the predicates $\{P_{ij}\}$ within a family $\mathscr{P}_i$.

L5. $C(H/E)$ is invariant under any permutation of two families $\mathscr{P}_i$ and $\mathscr{P}_j$ provided that $n_i = n_j$.

A strengthened version of L4 and L5 is

L6. $C(H/E)$ is invariant under any permutation of the complete descriptions in $\mathscr{Q}$.

Finally, Carnap suggests

L7. $C(H/E)$ is invariant under augmentation of the set $\{\mathscr{P}_i\}$ of families of predicates.

The axioms apply as well of course to $m(H) = C(H/T)$.

To understand the implications of these invariance axioms it suffices, since C is determined by m and m is finitely additive, to characterize $m(\bigwedge_{j=1}^{n} Q_{i_j} a_j)$ where the conjunction $\bigwedge_{j=1}^{n} Q_{i_j} a_j$ would be a state description if $\mathscr{L}$ had only n individuals. A first result is given by

Lemma 1. If C satisfies L1–L3, then there is a unique measure F such that

$$R_q = \left\{(p_1, ..., p_q) : p_i \geqslant 0, \sum_{i=1}^{q} p_i = 1\right\},$$

$$F(R_q) = 1,$$

$$m\left(\bigwedge_{j=1}^{n} Q_{i_j} a_j\right) = \int \cdots \int \left(\prod_{j=1}^{n} p_{i_j}\right) F(dp_1, ..., dp_q),$$

and q is the number of complete descriptions.

Proof. All proofs of results can be found in the Appendix to this chapter. ∎

Equivalently, we may write that under L1–L3

$$m\left(\bigwedge_{i=1}^{n} Q_{i_j} a_j\right) = \int \cdots \int \left(\prod_{i=1}^{q} p_i^{0_i}\right) F(dp_1, ..., dp_q), \qquad (*)$$

where 0_i is the number of occurrences of Q_i in $\{Q_{i_j}\}$;

$$0_i \geqslant 0, \qquad \sum_{i=1}^{q} 0_i = n.$$

Given Lemma 1, we can directly, but not too fruitfully, explore the consequences of adjoining L4 and L5 to L1–L3. If there are K families of predicates with the ith family containing n_i predicates, then define an array

$$X = [X_{i,j}]$$

having K rows with the ith row being a permutation of $(1, ..., n_i)$. The set of all such arrays will be denoted $\mathscr{X}$. The array $X \in \mathscr{X}$ applied to the families of predicates $\{\mathscr{P}_i\}$ with $\mathscr{P}_i = \{P_{i,1}, ..., P_{i,n_i}\}$ induces a new family $\{\mathscr{P}_i'\}$ with

$$P'_{i,j} = P_{i,x_{i,j}}.$$

Furthermore, X induces a permutation of the complete description $\mathscr{Q}$ with

$$Q_i = \bigwedge_j P_{ij} \rightarrow \bigwedge_j P_{i,x_{i,j}} = Q_{\pi_i}.$$

The set of all such permutations π of $\mathscr{Q}$ induced by some $X \in \mathscr{X}$ will be denoted $\pi_{\mathscr{Q}}$. We can now state

Lemma 2. If C satisfies L1–L4, then m is characterized by (*) with the proviso that the measure F has the invariance property

$$(\forall \pi \in \pi_{\mathscr{Q}}) F(dp_1, ..., dp_q) = F(dp_{\pi_1}, ..., dp_{\pi_q}).$$

Lemma 3. If C satisfies L1–L5, then m is characterized as in Lemma 2 with the additional constraint on F that

$$(\forall r, s)\, n_r = n_s, \qquad X = [X_{i,j}], \qquad X' = [X'_{i,j}],$$

$$X'_{i,j} = \begin{cases} X_{i,j} & \text{if } i \neq r, s \\ X_{r,j} & \text{if } i = s \\ X_{s,j} & \text{if } i = r, \end{cases}$$

π and π' the permutations induced on $\mathscr{Q}$ by X and X' (respectively),

$$F(dp_{\pi_1}, ..., dp_{\pi_q}) = F(dp_{\pi_1'}, ..., dp_{\pi_q'}).$$

The role of L6 is more readily indicated than were the roles of L4 and L5.

Lemma 4. If C satisfies L1–L3 and L6, then m satisfies (*) with the proviso that F is symmetric, that is,

$$(\forall \{\pi_i\} \text{ permutation of } \{1, ..., q\}) F(dp_1, ..., dp_q) = F(dp_{\pi_1}, ..., dp_{\pi_q}).$$

Let $\{0^i\}$ denote the rank ordered $(0^i \geqslant 0^{i+1})$ set of frequencies $\{0_i\}$.

Corollary. If C satisfies L1–L3 and L6, then there is a unique, symmetric measure F such that

$$m\left(\bigwedge_{j=1}^{n} Q_{i_j} a_j\right) = \int \cdots \int \left(\prod_{i=1}^{q} p_i^{0^i}\right) F(dp_1, ..., dp_q).$$

Unfortunately, while L6 and L7 are individually attractive they are inconsistent.

Lemma 5. There is no C satisfying L1–L3, L6, and L7.

The set (L1–L5 and L7) is consistent but not categoric. Further axioms are required to define a uniquely "best" confirmation function. Before turning to such axioms, we observe that in the impossibility of substituting L6 for L4 and L5, when L7 is desired, lies an inadequacy of logical probability that cannot be corrected by subsequent axioms. Whatever d.c. we are led to will unavoidably assign degrees of confirmation in a manner that depends on features of $\mathscr{L}$ which may be irrelevant to the meaning of the hypothesis and evidence statements. That is to say, changing the language $\mathscr{L}$ without apparently changing the meaning of H and E will lead to a change in $C(H/E)$.

4. Learning from Experience

Carnap supplemented the characterization of d.c. by postulating properties of C which govern its behavior in certain instances of inductive reasoning. Unfortunately, of the three axioms he proposed, the first is implied by the second, the second does little to narrow the class of acceptable d.c., and the third is merely Johnson's sufficiency postulate (Subsection IVJ2, axiom A2) previously discussed in connection with relative-frequency.

The first axiom concerns the property of instantial relevance, and in a weak form is given by

L8a. $C\left(Q_k a_{n+2}/Q_k a_{n+1} \wedge \left(\bigwedge_{j=1}^{n} Q_{i_j} a_j\right)\right) \geqslant C\left(Q_k a_{n+2}/\bigwedge_{j=1}^{n} Q_{i_j} a_j\right).$

This axiom asserts that confirmation does not decrease when we add confirming instances to the evidence.

Lemma 6. If C satisfies L1–L3, then C satisfies L8a.

A strict form of instantial relevance would be

L8b. $$C\left(Q_k a_{n+2} / Q_k a_{n+1} \wedge \left(\bigwedge_{j=1}^{n} Q_{i_j} a_j\right)\right) > C\left(Q_k a_{n+2} \Big/ \bigwedge_{j=1}^{n} Q_{i_j} a_j\right).$$

Lemma 7. If C satisfies L1–L3, then C satisfies L8b if and only if the measure F in the characterization (*) of m has nondegenerate marginal distributions.

Hence, even strict instantial relevance does little to restrict the set of possible confirmation functions. The reader should also consult [22, pp. 228–251] for related analyses of which we were unaware.

A second axiom concerns the asymptotic behavior of C as the number n of individuals $\{a_i\}$ increases. Let $0_k(n)$ be the number of occurrences of Q_k in $(Q_{i_1}, \ldots, Q_{i_n})$.

L9. $$(\forall\{Q_{i_j}\}) \lim_{n\to\infty} \left[C\left(Q_k a_1 \Big/ \bigwedge_{j=2}^{n} Q_{i_j} a_j\right) - \frac{0_k(n)}{n}\right] = 0.$$

The significance of L9 is immediate from

Lemma 8. If C satisfies L1–L3, then C satisfies L9 if and only if the measure F in (*) has as support the set

$$R_q = \left\{(p_1, \ldots, p_q) : p_i \geqslant 0, \sum_{i=1}^{q} p_i = 1\right\}.$$

We note that L9 does not considerably narrow the class of possible confirmation functions. Furthermore L6 is now seen to imply L8b, and therefore the pair of axioms is no more useful than is L9 alone.

The final axiom is the Johnson sufficiency postulate [14].

L10. $$(\exists g)\, C\left(Q_k a_{n+1} \Big/ \bigwedge_{j=1}^{n} Q_{i_j} a_j\right) = g(0_k, n).$$

The exact distribution of occurrences of predicates other than Q_k is irrelevant with respect to the confirmation of $Q_k a_{n+1}$. Carnap asserts that when there is only one family $\mathscr{P}_1$ of n_1 predicates in $\mathscr{L}$, then the d.c. satisfying L1–L4 and L10 can be characterized as follows:

$$(\exists 0 < \lambda < \infty)\, C_\lambda\left(Q_k a_{n+1} \Big/ \bigwedge_{j=1}^{n} Q_{i_j} a_j\right) = \frac{0_k + \lambda/n_1}{n + \lambda}.$$

This one-parameter family of d.c. $\{C_\lambda\}$ comprises what Carnap grandly called the continuum of inductive methods [15].

To derive the measure m_λ corresponding to C_λ, note that if T is any tautology, then

$$m_\lambda\left(\bigwedge_{j=1}^{n} Q_{i_j} a_j\right) = C_\lambda\left(\bigwedge^{n} Q_{i_j} a_j / T\right) = C_\lambda(Q_{i_1} a_1 / T) \prod_{r=1}^{n-1} C\left(Q_{i_{r+1}} a_{r+1} \Big/ \bigwedge_{j=1}^{r} Q_{i_j} a_j\right).$$

From the definition of C_λ we have that

$$C_\lambda(Q_{i_1} a_1 / T) = \frac{1}{q},$$

$$C_\lambda\left(Q_{i_{r+1}} a_{r+1} / \bigwedge^{r} Q_{i_j} a_j\right) = \frac{0_{i_{r+1}}(r) + \lambda/q}{r + \lambda},$$

where $0_{i_{r+1}}(r)$ is the number of individuals in $(a_1, \ldots, a_r)$ that are described by $Q_{i_{r+1}}$. It is now immediate that

$$m_\lambda\left(\bigwedge^{n} Q_{i_j} a_j\right) = \frac{1}{q}\left(\prod_{r=1}^{n-1} [r + \lambda]^{-1}\right) \prod_{r=1}^{n-1} \left[0_{i_{r+1}}(r) + \frac{\lambda}{q}\right].$$

The properties of m_λ become more apparent after rewriting. Let $\{0^i\}$ be the rank ordered $(0^i \geqslant 0^{i+1})$ set of frequencies of occurrence $\{0_i\}$. With the understanding that $\prod_{i=1}^{0} f(i) = 1$, we find that

$$m_\lambda\left(\bigwedge_{j=1}^{n} Q_{i_j} a_j\right) = \frac{1}{q}\left(\prod_{r=1}^{n-1} [r + \lambda]^{-1}\right) \prod_{j=1}^{q} \left(\prod_{i=1}^{0^j} \left[i + \frac{\lambda}{q}\right]\right).$$

5. Selection of a Unique Confirmation Function

If we have only a single family of predicates, and we follow Carnap, then we can conclude with a representation for confirmation functions that involves one free parameter $\lambda \in (0, \infty)$. Selection of a unique C_λ confirmation function requires us to choose λ. Carnap [15, pp. 56–79] has indicated several arguments that might form the basis for a selection of λ, although these arguments do not all lead to the same choice of λ. One interesting argument is that λ reflects our *a priori* anticipations as to the balance between the occurrences of individuals satisfying each of the n_1 predicates in the single family $\mathscr{P}_1$. Small values of λ correspond to anticipating that the predicates $\{P_{1i}\}$ do not occur equally often in descriptions of $\{a_j\}$, whereas large values of λ correspond to anticipating a balance in the frequencies of occurrence of the $\{P_{1i}\}$ in descriptions of the individuals $\{a_j\}$. Such a vague argument, however, leaves much to be desired in an *a priori* theory of rational induction and rational decision-making.

In his earlier work Carnap [4, pp. 562–577] had proposed, albeit with little enthusiasm, a confirmation function C^* as the most appropriate choice; C^* is defined through m^* by

$$m^*\left(\bigwedge_{j=1}^{n} Q_{i_j}a_j\right) = \frac{(q-1)!\prod_{j=1}^{q}(0_j!)}{(n+q-1)!}.$$

C^* is easily seen to satisfy L1–L6. We may interpret m^* as arising from a desire to assign equal weight to all state descriptions having the same set of ranked frequencies $\{0^i\}$ and equal weight to all statements (called structure descriptions) of the form

$$\{0^i\} = \{m_i\}.$$

The measures m^* and m_λ agree in assigning equal weight to all statements having the same structure descriptions $\{0^i\}$ but disagree in their assignments of weights to different structure descriptions.

The grounds for preferring one confirmation function to another as the explication of correct or rational inductive inference or as the basis for rational decision-making include the intuitive or *a priori* acceptability of the axioms used to characterize the "best" confirmation function as well as our willingness to forgo those properties of a desirable confirmation function that the "best" function does not possess. As we have seen in the discussion of L6 and L7 there does not exist a function having all of what might be thought to be the desirable properties of confirmation. With respect to the choice of axioms for inductive logic, Carnap† has commented:

> (a) The reasons are based upon our intuitive judgments concerning inductive validity, i.e., concerning inductive rationality of practical decisions (e.g., about bets).
> Therefore:
> (b) It is impossible to give a purely deductive justification of induction.
> (c) The reasons are *a priori* [10, p. 978].

The interested reader can find arguments for the acceptability of the axioms we have presented in the referenced writings of Carnap and Kemeny.

† P. A. Schilpp, ed., *The Philosophy of Rudolf Carnap*, p. 978. La Salle, Illinois: The Open Court Publishing Co., 1963.

VIIE. Logical Probability and Relative-Frequency

Carnap has suggested that while logical probability attaches only to inferences, it can also serve as a rational estimate of empirical probability, especially in the relative-frequency interpretation. A language $\mathscr{L}$ to describe the model of unlinked, indefinitely repeated experiments would contain individuals $\{a_i\}$, where a_i is the outcome of the ith experiment, and a single family $\mathscr{P}$ of predicates $\{Q_i\}$ that completely describe the possible outcomes of the experiment. An evidence statement might be of the form $\bigwedge_{j=1}^{n} Q_{i_j}a_j$, and we might be interested in assessing the probability of the hypothesis $Q_k a_{n+1}$. The hypotheses of unlinkedness and indefinite repeatability seem to justify L2 and L3 for C. It is then immediate from Lemma 8 that if C satisfies L1–L3 and the measure F in (*) has R_q as support, then

$$\lim_{n\to\infty}\left[C\left(Q_k a_{n+1}\Big/\bigwedge_{j=1}^{n} Q_{i_j}a_j\right) - \frac{0_k(n)}{n}\right] = 0.$$

Hence, if the frequency $0_k(n)/n$ of outcomes Q_k converges to ρ_k, then so will C; that is, logical probability will approximate relative-frequency-based empirical probability when the latter is "appropriate."

If we do not wish to assume that F has R_q as support and/or are wisely not content with asymptotic arguments, then Carnap has proposed a less direct link between logical and relative-frequency-based probability. We first define a logical probability version of an estimator.

Definition. If a function f of the states of the world takes on the possible values $\{f_i\}$ and H_i is the hypothesis that $f = f_i$, then the estimate of f relative to evidence E based on a confirmation function C is

$$e(f, E, C) = \sum_i f_i C(H_i/E).$$

These estimates are similar to expected values, and their merits are argued by Carnap [4, Chapter IX]. If, for example, we have a set of statements $L_1, ..., L_r$ and wish to estimate the proportion f of true statements in the set, given evidence E and a confirmation function C satisfying L1, then letting H_m be the hypothesis that exactly m of $\{L_i\}$ are true we have

$$e(f, E, C) = \sum_{m=1}^{r} \frac{m}{r} C(H_m/E).$$

Carnap [4, p. 543] proves the following.

Theorem 3. Under the preceding definition of terms,

$$\sum_{m=1}^{r} mC(H_m/E) = \sum_{i=1}^{r} C(L_i/E).$$

The estimate of the proportion of true statements in a set can be determined directly from the degrees of confirmation of the statements.

We can now relate degree of confirmation to the estimation of relative-frequency-based empirical probability. Let there be r trials; let L_i be the statement of success on trial i; and let E be an evidence statement implying that the outcomes of the trials are uninfluenced by their order, that is, $(\forall i)\ C(L_i/E) = C(L_1/E)$. It is now apparent from Theorem 3 that $C(L_1/E)$ is the estimate of the relative-frequency of success in the r trials. Furthermore, since this conclusion is independent of r, we see that for suitable E, $C(L_1/E)$ is also an estimate of the limit of relative-frequency as r goes to infinity. Note that this estimate does not require the assumption that the actual relative-frequency of success has a limit. Thus $C(L_1/E)$ is an estimate of relative-frequency-based empirical probability in the same sense that $\sum_{m=1}^{r} mC(H_m/E)$ is an estimate of the number of true $\{L_i\}$.

Clearly, since any C satisfying $L1$ can serve to calculate an estimate of relative-frequency-based probability, the "estimate" is in nowise bound to be empirically correct. Theorem 3 is only a mathematical tautology concerning a suggestively labeled quantity and cannot, despite appearances, be expected to yield empirically significant conclusions.

VIIF. Applications of C^* and C_λ

As a first illustration of the use of quantitative logical probability consider an experiment such as repeated die tossing in which there is one family $\{Q_i\}$ of predicates to describe the outcome of each of the repeated trials $\{a_i\}$. Let a success Sa_i at trial i be defined by

$$Sa_i = \bigvee_{k=1}^{w} Q_{j_k} a_i;$$

a success occurs if any of w outcomes described by $Q_{j_1}, ..., Q_{j_w}$ occurs. Given evidence E of the form

$$E = \bigwedge_{j=1}^{n} Q_{i_j} a_j ,$$

we may be interested in an hypothesis H of the form Sa_{n+1}. If there were s successes in the first n trials, then recourse to C^* yields,

$$C^*(H/E) = (s + w)/(n + q).$$

The form of this conditional probability for a future success is identical to that of the Bayes estimator $\hat{p}$ of the probability p of success calculated using a quadratic loss function $(\hat{p} - p)^2$ and a Beta prior density for p; that is,

$$\frac{(q-1)!}{(w-1)!(q-w-1)!} p^{w-1}(1-p)^{q-w-1} \qquad \text{for} \quad p([0, 1]).$$

From the representation for m_λ given at the end of Subsection VIID4, we can easily calculate that

$$C_\lambda(H/E) = (s + w\lambda/q)/(n + \lambda).$$

Clearly, for this problem choosing $\lambda = q$ would produce agreement between C^* and C_λ as to the degree of confirmation of the hypothesis of a success at a future trial given the record of past outcomes. Furthermore, C_λ can also be viewed as the Bayes estimator $\hat{p}$ of the probability p of success calculated using a quadratic loss and Beta prior density for p. Hence both C^* and C_λ are not unreasonable confirmation functions in that they are statistically admissible for some problem.

As an additional illustration of the use of C^* and C_λ we consider the following communications problem. A source or transmitter, about which we have little prior knowledge, generates messages that are binary sequences of length N. These messages are communicated through a poorly understood channel and received as possibly different binary sequences of length N. We collect data in the form of M pairs of transmitted and received sequences; the pair may be thought of as a binary sequence of length $2N$ with the initial segment the transmitted sequence. We are now informed that a given $(M + 1)$th sequence has been received and wish to infer what has been transmitted. For example we may wish to know the degree of confirmation for the hypothesis that more 0's were transmitted than 1's. A formulation of this problem in the terms of logical probability is as follows.

There are $M + 1$ individuals $\{a_i\}$ representing the pairs of transmitted and received binary sequences of length N. There are 2^{2N} predicates $\{Q_i\}$, each one specifying one of the possible binary sequences of length $2N$. Let $\{Q_{j_k}\}$ denote the 2^N predicates describing each possible sequence of length $2N$ which terminates in the observed received sequence and $\{Q_{i_j}\}$

describe the M known pairs of transmitted–received sequences. Then the range $\mathscr{R}(E)$ of the evidence or data is

$$\mathscr{R}(E) = \left\{ S : (\exists Q_{j_k} \in \{Q_{j_r}\})\ S = Q_{j_k} a_{M+1} \wedge \left(\bigwedge_{j=1}^{M} Q_{i_j} a_j \right) \right\}.$$

The range $\mathscr{R}(H \wedge E)$ of the conjunction of hypothesis H [more 0's sent than 1's in the $(M+1)$th message] and evidence E is the subset of $\mathscr{R}(E)$ for which Q_{j_k} described only those sequences where there are more 0's than 1's in the first N symbols and the observed $(M+1)$th received sequence for the remaining N symbols. If n_1 is the number of the M observed transmitted–received sequences for which the first N symbols contained more 0's than 1's and the last N symbols agreed with the last N symbols of the $(M+1)$th sequence (whose initial segment we are inferring), and n_2 is the number of observed sequences agreeing in the last N symbols with the $(M+1)$th sequence, then use of m^* yields

$$C^*(H/E) = \begin{cases} \dfrac{n_1 + 2^{N-1}}{n_2 + 2^N} & \text{if } N \text{ is odd,} \\[2ex] \dfrac{n_1 + \frac{1}{2}\left(2^N - \binom{N}{N/2}\right)}{n_2 + 2^N} & \text{if } N \text{ is even.} \end{cases}$$

To calculate $C_\lambda(H/E)$ through m_λ as represented at the end of Subsection VIID4 we need to introduce the following.

n_3 number of distinct Q_j in E such that Q_j describes a sequence with more 0's sent than 1's and a received sequence that agrees with that of a_{M+1}.

n_4 number of distinct Q_j in E such that Q_j describes a sequence in which the received sequence agrees with that of a_{M+1}.

Recourse to m_λ yields

$$C_\lambda(H/E) = \frac{n_1 + \lambda 2^{-2N} n_3}{n_2 + \lambda 2^{-2N} n_4}.$$

Hence we have found logical probabilities for the $(M+1)$th transmitted sequence to contain more 0's than 1's given the received sequence and M transmitted–received pairs of sequences. Both C^* and C_λ yield probabilities that differ from a straight relative-frequency assessment of n_1/n_2.

VIIG. Critique of Logical Probability

1. Roles for Logical Probability

Possible roles for logical probability include

(1) formalization of inductive reasoning via a measure of inferential support,
(2) source of rational estimates of empirical probability,
(3) explication of classical probability,
(4) basis for rational decision-making.

Too little is known about the modal and comparative concepts of logical probability to discuss them here. Extensive discussions as to whether present concepts of quantitative logical probability can, or do, fulfill any of these roles are available in Hempel [7], Kyburg [16], and Schilpp [10], and are not reproduced here. While we will indicate a few objections to the way present concepts of logical probability fulfill the above-mentioned roles, it is more difficult to determine whether these objections must bear against any concept of logical probability.

A serious objection to Carnap's concept of logical probability is that it is not applicable to the full range of uses of inductive reasoning. The limited languages that are treated by Carnap are unequal to the description of most of the usual scientific observations such as those involving numerical measurements. Hence we cannot apply Carnap's logical probability to discuss the degree to which laboratory observations support an hypothesis or theory. Furthermore, in Carnap's formulation all "laws" (propositions involving universal quantifiers) always have a zero degree of confirmation given any finite number of supporting instances. Albeit, universal statements being of no importance in the formulation of real problems, we are not inclined to take this latter defect of the Carnapian concept of logical probability seriously.

Reformulations of logical probability by Scott and Krauss and by Hintikka promise to avoid the aforementioned difficulties. Scott and Krauss [17] and Krauss [18] avail themselves of model theory in an attempt to define logical probability for richer and more realistic languages. Hintikka [19] has developed a version of logical probability wherein universal statements can have positive degrees of confirmation.

While we feel that logical probability may serve to explicate an objective version of classical probability, through an algorithmic assessment of the support provided by statements of prior knowledge for the various alternative outcomes, we are less inclined to agree that logical probability is the proper basis for estimating empirical (say relative-frequency-based) probability. There is a significant informal and subjective component in the selection of good estimates that may be uneliminatable; an experimenter can rarely write out in some simple language all he knows about an experiment.

There may be a role for logical probability in rational decision-making, although the form it will take is as yet unclear. The Anscombe–Aumann development of subjective probability (Section VIIIC) suggests one way of linking logical probability with decision-making. However, this link has yet to be detailed and the resulting relationship justified.

2. Formulation of Logical Probability

Questions of formulation concern the axiomatic structure of logical probability and the manner in which a choice is made, consistent with the axioms, of a logical probability relation to resolve a specific problem of inductive inference. Our discussion of modal logical probability completely ignored the problem of selecting a specific modal quantifier $\mathscr{P}$ and thus represents but a fragment of a useful theory. Koopman supplied us not only with a weak axiomatization of comparative logical probability but also with the directive that the specific choice be made subjectively. However, in contrast to the development of subjective probability in Chapter VIII, where the choice of subjective measure is guided by the desire for good decisions, Koopman in no way guides our subjective selection of order relations. At best Koopman has motivated and supplied a starting point for a useful theory of comparative logical probability.

Carnap and others attempted axiomatic specifications of quantitative logical probability or degree of confirmation (d.c.). Carnap's original objective of a categoric axiomatization to uncover the unique logically necessary d.c. would have resolved the problem of a specific choice of relation. However, the goal of a logically necessary d.c. seems to have been abandoned in the face of criticisms that were leveled at all of the candidates. Furthermore, results such as Lemma 5 suggest that there cannot exist a "best" d.c. function. The many desirable properties of rational inductive inference procedures, as reflected in the axioms for logical probability, are inconsistent. Failing a "best" choice of d.c., we need to know more about the bases for choosing between those d.c. functions satisfying the agreed upon axioms so as to resolve specific problems of inductive inference properly. There has been some discussion of this question, particularly as it affects C_λ [15, pp. 56–79]. However,

we expect that it will prove to be very difficult to illuminate this subject.

Turning to the axiomatic formulations themselves, we have previously noted the existence of systems of logical probability intended to overcome certain problems with Carnap's construction [17, 18]. A generic difficulty with these various formulations is the dependence of the resulting d.c. on the choice of language in which the evidence and hypothesis are expressed. Syntactical considerations intrude on questions that seem to be purely semantical. (This difficulty is analogous to the irksome dependence of complexity evaluations on the arbitrary choice of AUTM.) Possibly though this conflict with Salmon's seemingly reasonable criterion of linguistic invariance (Subsection IVJ2) is inevitable and its direct confrontation a virtue of the approaches to logical probability.

3. Justifying Logical Probability

A widely espoused criticism of logical probability is, in the words of Black [20],

> The most difficult question that any "logical" theory has to answer is how *a priori* truths can be expected to have any bearing upon the practical problem of anticipating the unknown on the basis of nondemonstrative reasons.

Carnap responded to this objection by pointing out that the degree of confirmation of a statement by an evidence statement E is not *a priori* in that it contains the synthetic content of our asserting E to be true. We might also note the relevance, to a defense of logical probability, of the difficulties encountered in selecting a unique d.c. While Carnap desires a definition of a d.c. that can be stated in purely logical terms without reference to contingent matters of fact, the choice of a best d.c. seems to involve extralogical, intuitive considerations. These considerations are at least in part to be distilled from our awareness of the purposes of a theory of probability as well as our extensive experience with inductive reasoning and are not *a priori*; *vide* the arguments for choosing λ in C_λ [15, pp. 56–79]. Of course, this defense is also an attack against the conception that the best d.c. could be selected on logical or analytic grounds alone. Perhaps part of the difficulty here stems from the dubious assumption of a dichotomy between the analytic and the synthetic [21].

It is essential that the relations between logical probability, empirical probability, and practical concerns be clarified and strengthened, if possible. We noted in Section VIIE that logical probability is compatible with relative-frequency-based empirical probability. Albeit our studies in Chapter IV and elsewhere lead us to take little comfort from this. More significant has been Carnap's interest in pragmatically justifying logical probability as leading to a fair betting quotient, or better yet, as appropriate for calculating the expected utilities of decisions. However, the justification of logical probability through a role in rational decision-making does not seem to have been fully argued, the several invocations of such a justification notwithstanding. Strict coherence, when possible, and it is not always possible, may be necessary, but it is hardly a sufficient guide to decision-making. Overall, Carnap's conception of logical probability seems most nearly to be a clarification and refinement of classical probability.

The attempt to base a theory of probability on its role in decision-making is at the root of the theories of subjective probability. As we see in Chapter VIII, subjective theories typically admit the use of individual judgment, as did Koopman, and do not yield the uniquely "correct" probability distribution that was Carnap's original goal.

VII. Appendix: Proofs of Results

Lemma 1. If C satisfies L1–L3, then there is a unique measure F such that

$$R_q = \left\{(p_1, ..., p_q) : p_i \geqslant 0, \sum_{i=1}^{q} p_i = 1\right\},$$

$$F(R_q) = 1,$$

$$m\left(\bigwedge_{j=1}^{n} Q_{i_j} a_j\right) = \int \cdots \int \left(\prod_{j=1}^{n} p_{i_j}\right) F(dp_1, ..., dp_q),$$

and q is the number of complete descriptions.

Proof. By L1, C can be represented by a measure m. By L2 and L3 true for all n, m is an exchangeable measure. The representation of m is that stated for exchangeable measures by de Finetti and Hewitt and Savage [13].

Lemma 2. If C satisfies L1–L4, then m is characterized by (*) with the proviso that the measure F has the invariance property

$$(\forall \pi \in \pi_{\mathscr{Q}}) F(dp_1, ..., dp_q) = F(dp_{\pi_1}, ..., dp_{\pi_q}).$$

Proof. From L4

$$(\forall \pi \in \pi_{\mathscr{Q}})\, m\left(\bigwedge_{j=1}^{n} Q_{i_j} a_j\right) = m\left(\bigwedge_{j=1}^{n} Q_{\pi_{i_j}} a_j\right).$$

Hence by (*)

$$\int \cdots \int \left(\prod_{i=1}^{q} p_i^{0_i}\right) F(dp_1, ..., dp_q) = \int \cdots \int \left(\prod_{i=1}^{q} p_{\pi_i}^{0_{\pi_i}}\right) F(dp_1, ..., dp_q).$$

Note that each permutation $\pi \in \pi_{\mathscr{Q}}$ has an inverse $\pi^{-1} \in \pi_{\mathscr{Q}}$. Changing variables in the right-hand side of the preceding equation yields

$$\int \cdots \int \left(\prod_{i=1}^{q} p_i^{0_i}\right) F(dp_1, ..., dp_q) = \int \cdots \int \left(\prod_{i=1}^{q} p_i^{0_i}\right) F(dp_{\pi_1^{-1}}, ..., dp_{\pi_q^{-1}}).$$

Since the measure F is unique, the lemma follows with π^{-1} in place of π. ∎

Lemma 3. If C satisfies L1–L5, then m is characterized as in Lemma 2 with the additional constraint on F that

$$(\forall r, s)\, n_r = n_s, \qquad X = [X_{i,j}], \qquad X' = [X'_{i,j}],$$

$$X'_{i,j} = \begin{cases} X_{i,j} & \text{if } i \neq r, s \\ X_{r,j} & \text{if } i = s \\ X_{s,j} & \text{if } i = r, \end{cases}$$

π and π' the permutations induced on $\mathscr{Q}$ by X and X' (respectively),

$$F(dp_{\pi_1}, ..., dp_{\pi_q}) = F(dp_{\pi_1'}, ..., dp_{\pi_q'}).$$

Proof. Parallels that of Lemma 2 and is omitted. ∎

Lemma 4. If C satisfies L1–L3 and L6, then m satisfies (*) with the proviso that F is symmetric, that is,

$$(\forall \{\pi_i\} \text{ permutation of } \{1, ..., q\}) F(dp_1, ..., dp_q) = F(dp_{\pi_1}, ..., dp_{\pi_q}).$$

Proof. This lemma is a corollary to Lemma 2 restricted to the case of only one family of predicates. ∎

Lemma 5. There is no C satisfying L1–L3, L6, and L7.

Proof. The assumption that we can adjoin arbitrarily many families of predicates to $\mathscr{L}$ and still leave $m(\bigwedge Q_{i_j} a_j)$ invariant corresponds, through Lemma 4, to assuming the existence of an infinite family of consistent, symmetric measures $\{F_q(dp_1, ..., dp_q)\}$. Hence from the de Finetti and Hewitt and Savage [13] characterization of symmetric (exchangeable) families of measures we have the following.

There exist a probability measure H and a conditional probability measure G such that

$$(\forall q)(\forall A)\, F_q(A) = \int H(d\alpha) \int_A \cdots \int \prod_{i=1}^{q} G(dp_i \mid \alpha).$$

However, we assert that this characterization of F_q is incompatible with the requirement

$$(\forall q)\, F_q(R_q) = 1.$$

To verify the asserted incompatibility, consider the following mutually exclusive and exhaustive possibilities:

(1) $(\exists A)\,(H(A) > 0)\,(\forall \alpha \in A) \int_{1/q}^{1} G(dp \mid \alpha) > 0;$

(2) $(\exists A)\,(H(A) > 0)\,(\forall \alpha \in A) \int_{0}^{1/q} G(dp \mid \alpha) > 0;$

(3) $(\exists A)\,(H(A) = 1)\,(\forall \alpha \in A)\, G\left(p = \frac{1}{q} \,\middle|\, \alpha\right) = 1.$

Under (1),

$$F_q\left(\left\{(p_1, \ldots, p_q) : \sum_{i=1}^{q} p_i > 1\right\}\right) > \int_A H(d\alpha) \left\{\int_{1/q}^{\infty} G(dp \mid \alpha)\right\}^q > 0.$$

Similarly, under (2),

$$F_q\left(\left\{(p_1, \ldots, p_q) : \sum_{i=1}^{q} p_i < 1\right\}\right) > 0.$$

Both of these conclusions conflict with $F_q(R_q) = 1$, leaving only (3). However, (3) is not invariant with respect to the number q of complete descriptions and therefore violates L7. ∎

Lemma 6. If C satisfies L1–L3, then C satisfies L8a.

Proof. From the characterization (∗) of m and introducing the notation

$$X = \prod_{i \neq k} p_i^{0_i}, \qquad Y = p_k,$$

$$EXY^n = \int \cdots \int \left(\prod_{i \neq k} p_i^{0_i}\right) p_k{}^n F(dp_1, \ldots, dp_q),$$

we have that

$$C\left(Q_k a_{n+2} / Q_k a_{n+1} \wedge \left(\bigwedge_{j=1}^{n} Q_{i_j} a_j\right)\right) = EXY^{n+2} / EXY^{n+1},$$

$$C\left(Q_k a_{n+2} \Big/ \bigwedge_{j=1}^{n} Q_{i_j} a_j\right) = EXY^{n+1} / EXY^{n}.$$

Note that the random variables X and Y are nonnegative. Hence, by Schwarz's inequality,

$$E^2 XY^{n+1} = E^2(\sqrt{XY^{n+2}})(\sqrt{XY^n}) \leqslant E(XY^{n+2})\, E(XY^n).$$

It is immediate that

$$C\left(Q_k a_{n+2} / Q_k a_{n+1} \wedge \left(\bigwedge_{j=1}^{n} Q_{i_j} a_j\right)\right) \geqslant C\left(Q_k a_{n+2} \Big/ \bigwedge_{j=1}^{n} Q_{i_j} a_j\right). \quad ∎$$

Lemma 7. If C satisfies L1–L3, then C satisfies L8b if and only if the measure F in the characterization (∗) of m has nondegenerate marginal distributions.

Proof. Examination of the proof of Lemma 6 shows that equality between confirmations occurs only when there is equality in the Schwarz inequality

$$E^2(\sqrt{XY^{n+2}})(\sqrt{XY^n}) \leqslant E(XY^{n+2})\, E(XY^n).$$

As is well known, there is equality in the Schwarz inequality only if there is linear dependence; that is,

$$(\exists \alpha)\, \sqrt{XY^{n+2}} = \alpha \sqrt{XY^n} \qquad \text{(a.e. } F).$$

Hence the condition for equality between the confirmations is that

$$(\forall k)\, p_k \text{ is a constant} \qquad \text{(a.e. } F).$$

The lemma follows immediately. ∎

Lemma 8. If C satisfies L1–L3, then C satisfies L9 if and only if the measure F in (*) has as support the set

$$R_q = \left\{(p_1, \ldots, p_q) : p_i \geqslant 0, \sum_{i=1}^{q} p_i = 1\right\}.$$

Proof. We first show that if F as R_q as support and m satisfies L1–L3, then m satisfies L9. Define

$$f(\mathbf{p}) = \sum_{i=1}^{q} \rho_i \log p_i ,$$

where $\rho_i = 0_i/n$ and $\mathbf{p}$ denotes the vector $(p_1, \ldots, p_q)$. It follows from the continuity and concavity of $\log x$ that f is continuous and concave in the interior of R_q. It is easily shown that f has a unique maximum in R_q at $\mathbf{p} = \boldsymbol{\rho}$, where $\boldsymbol{\rho}$ denotes the vector $(\rho_1, \ldots, \rho_q)$. Furthermore, for all $\mathbf{q}$ and $0 \leqslant \lambda \leqslant 1$, $f(\lambda\mathbf{q} + (1-\lambda)\boldsymbol{\rho})$ is strictly decreasing in λ. To verify these statements, note that

$$\frac{\partial f(\lambda\mathbf{q} + (1-\lambda)\boldsymbol{\rho})}{\partial\lambda} = \sum_{i=1}^{q} \frac{\rho_i(q_i - \rho_i)}{\lambda q_i + (1-\lambda)\rho_i},$$

and write

$$\rho_i = -\lambda(q_i - \rho_i) + (1-\lambda)\rho_i + \lambda q_i$$

to find

$$\frac{\partial f}{\partial\lambda} = \sum_{i=1}^{q} (q_i - \rho_i) - \lambda \sum_{i=1}^{q} \frac{(q_i - \rho_i)^2}{\lambda q_i + (1-\lambda)\rho_i}.$$

Noting that as $\mathbf{q}$ and $\boldsymbol{\rho}$ are both in R_q,

$$\sum (q_i - \rho_i) = 0,$$

we see that $\partial f/\partial\lambda$ is strictly negative unless $\mathbf{q} = \boldsymbol{\rho}$.

An inequality that will be useful to us is derived as follows. Note that

$$(\forall \mathbf{q}, \boldsymbol{\rho} \in R_q)\ \lambda q_i + (1-\lambda)\rho_i \leqslant 1,$$

to conclude that

$$\frac{\partial f}{\partial\lambda} \leqslant -\lambda \sum_{i=1}^{c} (q_i - \rho_i)^2.$$

Hence,

$$f(\mathbf{q}) = f(\boldsymbol{\rho}) + \int_0^1 \frac{\partial f}{\partial\lambda}\, d\lambda \leqslant f(\boldsymbol{\rho}) - 1/2 \sum_{i=1}^{q} (q_i - \rho_i)^2.$$

Define the open ball $B(\boldsymbol{\rho}, \epsilon)$ in R_q with center $\boldsymbol{\rho}$ and radius ϵ by

$$B(\boldsymbol{\rho}, \epsilon) = \left\{(p_1, \ldots, p_q) : p_i \geqslant 0, \sum_{i=1}^{q} p_i = 1, \sum_{i=1}^{q} (p_i - \rho_i)^2 < \epsilon^2\right\}.$$

From the preceding inequality we conclude that

$$(\forall \boldsymbol{\rho}, \epsilon) \sup_{p \notin B(\boldsymbol{\rho},\epsilon)} f(\mathbf{p}) \leqslant f(\boldsymbol{\rho}) - \epsilon^2/2.$$

Define

$$R^\delta = \{(p_1, \ldots, p_q) : (\forall_i)\, p_i \geqslant \delta\} \cap R_q .$$

From the continuity of f on R^δ and the compactness of R^δ we have

$$(\forall\epsilon > \delta > 0)(\exists\gamma_0 > 0)(\forall\boldsymbol{\rho} \in R_q)(\forall\mathbf{p} \in B(\rho, \gamma_0) \cap R^\delta)\, f(\mathbf{p}) > f(\boldsymbol{\rho}) - \epsilon^2/4.$$

Furthermore, from the hypothesis that F has R_q as support we can choose γ_0 such that

$$(\forall\epsilon > \delta > 0)(\exists c_{\gamma_0,\delta})(\forall\boldsymbol{\rho})\, F(B(\boldsymbol{\rho}, \gamma_0) \cap R^\delta) \geqslant c_{\gamma_0,\delta} > 0.$$

Noting that

$$\prod_{i=1}^{q} p_i^{0_i} = \left[\prod_{i=1}^{q} p_i^{\rho_i}\right]^n = e^{nf(\mathbf{p})},$$

we see that for $0 < \gamma \leqslant \min(\epsilon, \gamma_0)$,

$$\int_{B(\boldsymbol{\rho},\epsilon)} \cdots \int e^{nf(\mathbf{p})} F(dp_1, \ldots, dp_q) \geqslant \int_{B(\boldsymbol{\rho},\gamma)} \cdots \int e^{nf} F(dp_1, \ldots, dp_q)$$
$$\geqslant c_{\gamma,\delta} \exp(n[f(\boldsymbol{\rho}) - \epsilon^2/4]),$$

$$\int_{\overline{B(\boldsymbol{\rho},\epsilon)}} \cdots \int e^{nf(\mathbf{p})} F(dp_1, \ldots, dp_q) \leqslant \exp(n[f(\boldsymbol{\rho}) - \epsilon^2/2]) F(\bar{B}) \leqslant \exp(n[f(\boldsymbol{\rho}) - \epsilon^2/2]).$$

Hence,

$$(\forall\{\boldsymbol{\rho}_n\}) \lim_{n\to\infty} \frac{\int \cdots \int_{\overline{B(\boldsymbol{\rho}_n,\epsilon)}} \exp[nf_n(\mathbf{p})]\, F(dp_1, \ldots, dp_q)}{\int \cdots \int_{B(\boldsymbol{\rho}_n,\epsilon)} \exp[nf_n(\mathbf{p})]\, F(dp_1, \ldots, dp_q)} \leqslant \lim_{n\to\infty} \frac{\exp(-n\epsilon^2/4)}{c_{\gamma_0,\delta}} = 0,$$

where

$$\boldsymbol{\rho}_n = (\rho_1(n), \ldots, \rho_q(n)), \qquad f_n(\mathbf{p}) = \sum_{i=1}^{q} \rho_i(n) \log p_i .$$

Applying this result to the d.c. C we find that

$$(\forall\epsilon > 0)(\forall\{\boldsymbol{\rho}_n\}) \lim_{n\to\infty} \left[C\left(Q_k a_1 \Big/ \bigwedge_{j=2}^{n} Q_{i_j} a_j\right) - \frac{\int \cdots \int_{B(\boldsymbol{\rho}_n,\epsilon)} p_k \exp(nf_n)\, F(dp_1, \ldots, dp_q)}{\int \cdots \int_{B(\boldsymbol{\rho}_n,\epsilon)} \exp(nf_n)\, F(dp_1, \ldots, dp_q)}\right] = 0.$$

It follows upon taking ϵ arbitrarily small that

$$\lim_{n\to\infty}\left[C\left(Q_k a_1 / \bigwedge_{j=2}^{n} Q_{i_j} a_j\right) - \rho_k(n)\right] = 0,$$

and the sufficiency of the hypothesis is now apparent.

To verify the necessity that F have R_q as support we need only assume to the contrary that

$$(\exists \rho, \epsilon) F(B(\rho, \epsilon)) = 0.$$

It can now be easily shown by examining a sequence $\{Q_{i_j}\}$ such that $\rho_i(n) \to \rho_i$ that

$$C - \rho_k(n) \nrightarrow 0.$$

We omit the details. ▮

The proof of the preceding lemma could also have been recast in probabilistic terms involving the convergence of $E(p_k \mid \{0_i(n)\})$.

References

1. J. M. Keynes, *A Treatise on Probability*. New York: Harper, 1962.
2. H. Jeffreys, *Theory of Probability*. London and New York: Oxford Univ. Press (Clarendon), 1948.
3. B. Koopman, The Bases of Probability, *Bull. Amer. Math. Soc.* **46**, 763–774, 1940.
4. R. Carnap, *Logical Foundations of Probability*, 2nd ed. Chicago, Illinois: Univ. of Chicago Press, 1962.
5. G. H. von Wright, *An Essay in Modal Logic*. Amsterdam: North-Holland Publ., 1951.
6. G. E. Hughes and M. J. Cresswell, *An Introduction to Modal Logic*. London: Methuen, 1968.
7. C. G. Hempel, *Aspects of Scientific Explanation*. Glencoe, Illinois: Free Press, 1965.
8. P. Martin-Löf, The Definition of Random Sequences, *Information and Control* **9**, 619, 1966.
9. B. Koopman, Intuitive Probability and Sequences, *Ann. of Math.* **42**, 185, 1941.
10. P. A. Schilpp, ed., *The Philosophy of Rudolf Carnap*, pp. 969–979. LaSalle, Illinois: Open Court, 1963.
11. P. C. Fishburn, *Utility Theory for Decision Making*. New York: Wiley, 1970.
12. J. G. Kemeny, Fair Bets and Degree of Confirmation, *J. Symbolic Logic* **20**, 263–273, 1955.
13. E. Hewitt and L. J. Savage, Symmetric Measures on Cartesian Products, *Trans. Amer. Math. Soc.* **80**, 484, 489, 1955.
14. I. J. Good, *The Estimation of Probabilities*, p. 26. Cambridge, Massachusetts: MIT Press, 1965.

15. R. Carnap, *The Continuum of Inductive Methods*. Chicago, Illinois: Univ. of Chicago Press, 1952.
16. H. E. Kyburg, Jr., I. Recent Work in Inductive Logic, *Amer. Philos. Quart.* **1**, 249–287, 1964.
17. D. Scott and P. Krauss, Assigning Probabilities to Logical Formulas, in *Aspects of Inductive Logic* (J. Hintikka and P. Suppes, eds.), pp. 219–264. Amsterdam: North-Holland Publ., 1966.
18. P. Krauss, Representation of Symmetric Probability Models, *J. Symbolic Logic* **34**, 183–193, 1969.
19. J. Hintikka, A Two-Dimensional Continuum of Inductive Methods, in *Aspects of Inductive Logic* (J. Hintikka and P. Suppes, eds.), pp. 113–132. Amsterdam: North-Holland Publ., 1966.
20. M. Black, *Margins of Precision*, p. 127. Ithaca, New York: Cornell Univ. Press, 1970.
21. W. V. O. Quine, Two Dogmas of Empiricism, reprinted in *From a Logical Point of View*. Cambridge, Massachusetts: Harvard Univ. Press, 1953.
22. R. Carnap and R. C. Jeffrey, eds. *Studies in Inductive Logic and Probability*, I. Berkeley and Los Angeles: Univ. of California Press, 1971.