Probability Logic

THEODORE HAILPERIN*

Introduction   Among logicians it is well-known that Leibniz was the first to conceive of a mathematical treatment of logic. Much less known, however, was his insistence that there was need for a new kind of logic that would treat of degrees of probability. Although it isn't clear what Leibniz had in mind for such a logic—understandably, since the subject of probability had just begun in his lifetime and the florescence of modern logic was not to begin until the 19th century—he did envision that it would be a means for estimating likelihoods and a way of proof leading not to certainty but only to probability (see his *Nouveaux Essais*, pp. 372–373). Beginning in his day, and extending through the present century, a number of mathematicians and logicians, e.g., Jacob Bernoulli, J. H. Lambert, A. De Morgan, G. Boole, C. S. Peirce, J. M. Keynes, H. Reichenbach, R. Carnap, and, more recently, D. Scott and P. Krauss have made either forays or detailed attacks on establishing such a logic, but with differing conceptions as to its nature. A few brief remarks will give some idea as to what these were like.

To Bernoulli (as also to Leibniz) probability was degree of certainty, differing from it as part to whole. In his *Ars Conjectandi* (Part IV, Chapter III) he considers the various kinds of arguments (i.e., grounds) for a conclusion (opinion or conjecture) and the problem of estimating their weights so as to compute the probability of the conclusion. Situations involving arguments are divided into three types: those in which the argument is necessarily the case but indicates (proves) the conclusion only contingently; those in which the argument is contingent but when present necessarily proves the conclusion; and those in which the presence of the argument and its proving of the conclusion are both contingent. The "proving power" of an argument is determined by the number of cases in which the argument is, or is not, present and also by the number of

---

cases in which it indicates the conclusion, or doesn't indicate it, or indicates its contrary. As in games of chance, Bernoulli assumes these cases to be equally possible. The algebraic formulas he obtains rely on the assumption of equally likely cases, and also, in some instances, on tacit independence assumptions.

In Lambert's *Neues Organon* (Vol. 2, Part V, Section 169, pp. 184–186) we find problems similar to those of Bernoulli's *Ars Conjectandi* (Part IV), which, however, are treated with more sophistication with respect to both logical and probabilistic aspects. Where Bernoulli talks of the probability of "things" and of "the matter", Lambert talks of the probability of propositions. The logical structure of (syllogistic) inference is taken into account, and due attention is paid to whether or not premises are independent. Where Bernoulli obtains the worth of an argument for a conclusion by computing expected value (treating certainty as a prize and counting cases in which the argument exists and proves the conclusion, or doesn't and proves nothing) Lambert uses the "urn model" idea, representing arguments by heaps of tickets marked and unmarked in proportion to the number of cases in which the argument for the conclusion is valid and invalid, and then considers blindfold selections from the heaps to determine the probability of the argument establishing the conclusion. But, unlike Bernoulli, Lambert doesn't consider situations in which a conclusion follows contingently from premises.

In his *Formal Logic*, subtitled "The Calculus of Inference, Necessary and Probable", De Morgan contends that making a distinction between drawing conclusions from premises whose absolute truth one believes in and those in which one has only partial belief, was quite arbitrary. His (rather small) chapter on probable inference contains items of the following nature.

All arguments are supposed to be logically valid, that is their conclusions necessarily follow from the premises. (This is a restriction in generality from Bernoulli who treated cases of contingent entailment.) If one has independent premises with probabilities a, b, c,...then for such an argument the product abc...is the probability that "the argument is in every way good". A major part of the chapter is devoted to questions on combinations of testimonies of witnesses with given probabilities of being correct, and with combinations of testimonies with arguments. We take these to be not purely logical or probabilistic in character. However, some he looks at do have such a character: "Arguments being supposed logically good, and the probabilities of their proving their conclusions (that is, of all their premises being *true*) being called their validities, let there be a conclusion for which a number of arguments are presented, of validities *a, b, c,* etc. Required the probability that the conclusion is proved" [3, p. 201]. This same problem was treated by Bernoulli and Lambert. We discuss it below in Section 5.

Boole, more than anyone before him, realized and exploited the close relationship between the logic of *not, and,* and *or* and the formal properties of probability [1]. He forced this relationship to be closer than it really is by restricting himself to *or* in the exclusive sense (so that probabilities of an *or*-compound added) and by believing that all events were expressible ultimately in terms of independent events (so that probabilities of an *and*-compound multiplied). As he saw it, the central problem of the theory of probabilities was to obtain the probability of an event (the event expressed as a proposition) in terms of the

probabilities of other events logically related to it in any manner. Though Boole was not conscious of it as such we have here the nub of a probability logic: by taking as premises the assertions that the several starting propositions have the assigned probabilities and as conclusion the assertion that the objective proposition has the (sought for) probability value, we have, in effect, an inference in a Leibnizian "logic of degrees of probability". Boole believed, on the basis of his view of probability, that he had solved the central problem, and in a general form which included conditional probabilities (see [7], Chapters 4 and 5).

One of the examples Boole works out ([1], Example 7, p. 284) is that of finding the probability of the conclusion of a hypothetical syllogism given the probabilities of its premises. Unfortunately he confuses the probability of a conditional (probability of: if A, then B) with conditional probability (probability of B, given A) and solves not the stated problem (which nevertheless his method could handle) but one involving conditional probabilities. While De Morgan restricted himself to arguments in which the conclusion was a necessary consequence of the premises (though these premises could have less than certainty), no such restriction is present in Boole's method. Another novel feature of Boole's work was his general investigation of bounds on probabilities when the data were insufficient to determine a value. The principal result on which we base our work (Section 3 below) is a direct outgrowth of these ideas of Boole.

C. S. Peirce was one of the great formal logicians of the 19th century who also thought intensively about the nature of probability. It is thus surprising that his long essay "A Theory of Probable Inference" of 1883 [11] has no systematic consideration of the formal grounds for justifying probable inferences. Indeed, one gets the impression that he thinks probable inference is not entirely formal. For example, in citing the wide differences, despite analogies, between "syllogistic" and probable inference he writes:

> (3) A cardinal distinction between the two kinds of inference is, that in demonstrative reasoning the conclusion follows from the existence of the objective facts laid down in the premises; while in probable reasoning these facts themselves do not render the conclusion probable, but account has to be taken of various subjective circumstances—of the manner in which the premises have been obtained, of there being no countervailing considerations; etc.; in short, good faith and honesty are essential to good logic in probable reasoning. ([11], p. 435)

By contrast, in Keynes [10], the explicit aim is to formalize the theory of probability. There is a philosophical analysis of probability as a relation between propositions—considered by Keynes to be a logical relation—and axioms for it are presented. This relation behaves similarly to conditional probability which, like the notions of randomness and statistical independence, will not be part of our study. We are thus construing 'logic' in a narrow sense which excludes investigations such as that of Keynes and, similarly where the interest is in analysing, and obtaining a definition of, 'degree of confirmation'.

After the advent of many-valued logics in the 1920s, with both finite and infinitely many values, it was natural to associate probability with a logic of infinitely many values. The period 1932–1936 saw a spate of papers on the subject. The most zealous proponent of treating probability as a many-valued

logic was Reichenbach whose first paper on the idea appeared in 1932. A good account of his work appears in his 1949 book [12]. In opposition to Reichenbach, Tarski [16], reflecting the ideas of various Polish philosophers, declared that a revamping of the present system of logic in order to obtain a methodologically correct foundation for probability (which was Reichenbach's aim) was neither necessary nor desirable—a chief criticism being the "nonextensionality" of probability as an operator on propositions. For instance, in many-valued logics the "truth"-value of a logical sum (alternation) is determined by those of its components, whereas the probability of a logical sum is not a (single-valued) function of the probabilities of its components. Reichenbach counters this criticism of nonextensionality by what is in essence a hidden variable argument. Based on the probability formula

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$
$$= P(A) + P(B) - P(A)P(B|A) \ ,$$

he writes the table

$P(A)$	$P(B)$	$P(B|A)$	$P(A \vee B)$	$P(A \wedge B)$
p	q	u	$p + q - pu$	pu

restrictive conditions: $\dfrac{p + q - 1}{p} \leq u \leq \dfrac{q}{p}$,

thus making $P(A \vee B)$ depend on three arguments one of which involves a conditional probability. Derivable formulas of his probability logic are either formulas stating that a logical expression has probability 1 (e.g., $P(A \vee \neg A) = 1$) or formulas coming from the well-known calculus of probabilities expressing relations between probabilities (e.g., $P(B) = P(A)P(B|A) + P(\neg A)P(B|\neg A)$). Tarski [16] declared that Reichenbach's probability logic is not a many-valued logic but rather a chapter of the usual two-valued logic.

In Carnap [2], as in Keynes, the central notion is that of a (confirmation) relation between pairs of sentences and with a function assigning numerical values to such pairs. Unlike Keynes, there is a deeper analysis of the formal language.

More directly related to our topic are considerations which assign probability values to simple sentences and which study its transmission through to more complicated sentences (as one does for truth in Tarski semantics). In Gaifman [5] we have a generalization of the semantic notion of a model for a first-order language in which probability values replace truth-values. Gaifman's idea is extended to infinitary languages and intensively investigated in Scott and Krauss [14], where a probability logic is developed on this basis.

At the other extreme, in contrast to the Scott and Krauss work, here we remain entirely on the propositional level. We develop a probability logic appropriate to this level from a consideration of the truth-table idea, slightly generalized, along with properties of probability. The notion of probability is presupposed as part of the semantics, and questions as to its nature will play no more role than does the nature of truth in usual logic. A small set of properties, shared by essentially all theories of probability, is all that we shall appeal

to. What we come up with could conceivably be viewed as a many-valued logic in that propositions take on more than two values—viz., the real numbers in the interval [0,1]—but it differs from what has been known and studied as many-valued logic in a number of important details.

For one thing, although the *formal* language of our probability logic has the same symbols as that of two-valued logic there is an uncoupling of the connection between value assigning functions ("truth"-functions) and connectives—values associated with a formula are not strictly determined by its linguistic structure but by its probability-related structure, namely by its representation in terms of basic conjunctions on its variables (corresponding to 'elementary event', 'state-description', and the like, in probability treatments). Moreover, the logic is not given syntactically, i.e., by specifying some set of starting formulas and rules of inference for deriving others, but semantically, by defining 'logical consequence' using the notion of a 'probability model'. Although the definition of 'model' for probability logic is pretty much what one would expect, the definition of 'logical consequence' has an unusual feature in that its assertions concern formulas having a value in a set of values (rather than some designated value (or values)), and in such assertions all probability values are on a par. When the sets of values are subintervals of [0,1] there is a decision procedure for determining whether or not a logical consequence relation holds.

1 The true-false logic of propositions   Although the true-false logic of propositions is obviously well-known, we nevertheless wish to sketch this elementary portion of standard logic in a form that will enable us to see how probability logic generalizes it.

One can either start out by assuming a (potentially infinite) list of un-analyzed "atomic" sentences of some arbitrary, but fixed, language or, as we shall be doing, assuming a list of formal propositional variables taking on such propositions as values. Then, using

(i) propositional variables: $A_1, \ldots, A_n, \ldots$
(ii) logical connectives: $\neg, \wedge, \vee$
(iii) parentheses: (,)

we may construct in customary fashion the *formulas* of the (two-valued) propositional logic.

The fundamental relation governing inference, the relation of logical consequence, is defined in terms of models. We use a slightly variant definition of model so as to bring out that the validity of an inference depends only on a finite number of propositional variables. A *model M* (adequate for a set of formulas $\chi_1, \ldots, \chi_N$) is an assignment of values 0 or 1 to propositional variables $A_1, \ldots, A_n$, where this list is long enough to include all variables present in $\chi_1, \ldots, \chi_N$. (The use of a consecutive initial string of variables is merely for convenience.) The values 0 and 1, rather than atomic propositions, are assigned to the variables since for the purposes of logical inference there is no need to distinguish among false, and among true, atomic propositions. We write $v_M(A_i)$ for the value assigned to A_i by M. In customary fashion we extend the value-assigning function v_M to all formulas ϕ on $A_1, \ldots, A_n$ by setting $v_M(\phi) = Tf_\phi(v_M(A_1), \ldots, v_M(A_n))$, where Tf_ϕ is the truth-function associated

with ϕ, definable by composition of functions from the three truth-functions normally associated with $\neg$, $\wedge$, $\vee$. If $v_M(\phi) = 1$, then we say M is a model of ϕ, or that ϕ is true in M.

A formula ψ (abstractly or schematically standing for a proposition) is a (*valid*) *logical consequence* of the set of formulas $\phi_1, \ldots, \phi_m$, if:

(1)   For all models M (adequate for $\phi_1, \ldots, \phi_m$, ψ): if $v_M(\phi_1) = 1, \ldots,$ $v_M(\phi_m) = 1$, then $v_M(\psi) = 1$.

This consequence relation is customarily symbolized by $\phi_1, \ldots, \phi_m \vDash \psi$.

Property (1) can be trivially generalized by allowing 0 in place of any or all of the 1's (trivially, because $v_M(\phi) = 0$ is equivalent to $v_M(\neg\phi) = 1$), so obtaining

(2)   For all models M (adequate for $\phi_1, \ldots, \phi_m$, ψ): if $v_M(\phi_1) = v_1, \ldots,$ $v_M(\phi_m) = v_m$, then $v_M(\psi) = v_0$, where $v_1, \ldots, v_m, v_0$ is any fixed (i.e., independent of M) selection of 0's and 1's.

We symbolize (2) by writing

(3) $V(\phi_1) = v_1, \ldots, V(\phi_m) = v_m \vDash V(\psi) = v_0,$

the symbol 'V' being chosen as suggestive of 'verity'.

If we had a truth-table entered from $A_1, \ldots, A_n$ and computing values for $\phi_1, \ldots, \phi_n$, ψ, then we can assert (3) if in every row in which the values under $\phi_1, \ldots, \phi_m$ are, respectively, $v_1, \ldots, v_m$, the value under ψ in each of these rows is v_o. In general, suppose we delete the portion of the table headed by $A_1, \ldots, A_n$ and in the remaining portion collapse together, or identify, all rows having the same distribution of truth-values under $\phi_1, \ldots, \phi_m$. From this collapsed table we can read off a consequence relation of the form (3) when each of the identified rows has the same value for ψ, though not if both values 0 and 1 occur. Yet this latter case is information, and we can represent it by writing

(4) $V(\phi_1) = v_1, \ldots, V(\phi_m) = v_m \vDash V(\psi) \in \{0,1\}$.

More generally, we can introduce a consequence relation of the form,

(5) $V(\phi_1) \in \alpha_1, \ldots, V(\phi_m) \in \alpha_m \vDash V(\psi) \in \beta,$

where the symbols $\alpha_1, \ldots, \alpha_m$, β represent any of the sets $\{0\}$, $\{1\}$, $\{0,1\}$. As an example, the eight rows of a truth-table computing values for $\phi_1 = A_1 \rightarrow A_3$, $\phi_2 = A_2 \vee A_3$, $\psi = (A_1 \rightarrow A_2) \wedge (A_1 \vee A_2)$ is, in collapsed form,

	$A_1 \rightarrow A_3$	$A_2 \vee A_3$	$(A_1 \rightarrow A_2) \wedge (A_1 \vee A_2)$
(5 rows)	1	1	$\{0,1\}$
(1 row)	0	1	1
(2 rows)	$\{0,1\}$	0	0

leading to three statements of the form (5):

$$V(\phi_1) \in \{1\}, \ V(\phi_2) \in \{1\} \vDash V(\psi) \in \{0,1\}$$
$$V(\phi_1) \in \{0\}, \ V(\phi_2) \in \{1\} \vDash V(\psi) \in \{1\}$$
$$V(\phi_1) \in \{0,1\}, \ V(\phi_2) \in \{0\} \vDash V(\psi) \in \{0\}.$$

This generalized form of logical consequence, while of little interest for two-valued logic, will serve as a prototype for that of probability logic.

2 Probability-valued logic   Purely abstract forms of many-valued logics, introduced by "truth"-tables with either finite or infinitely many values assigned to propositional arguments, have been known since the 1920s. The subject has had fairly extensive development: current bibliographies (as in [13]) list hundreds of items, mainly along lines of a mathematico-structural nature. Application as usable logic, i.e., for validating inferences, is not very prominent. What hinders successful application is the semantic problem, crucial to which is the meaning of the abstract values when there are more than two. Without a clear idea of what the values stand for there is no firm guidance as to how to define 'model' and 'logical consequence', notions which, in our opinion, are necessary for something to be called logic.

In view of difficulties stemming from lack of clarity in this connection, the possibility of using probabilities as values seems especially attractive; for although the nature of probability may be in philosophical dispute, we do have a clear idea of its properties, particularly in relation to *not*, *and*, and *or*. A comparison of these properties in the two cases, i.e., for truth and for probability, will suggest an appropriate definition of 'model' for probability logic. We assume in both cases a common formal language, that described in Section 1.

For *truth*-value assignments to formulas the properties needed are that each formula (as representative of a proposition) has either 0 or 1 as a value, and the logical connectives behave truth-functionally in the well-known manner exemplified by the conventional truth-tables. But it is clear that we cannot, in general, extend probability values for variables $A_1, \ldots, A_n$ to formulas (on these variables). However, if we restrict ourselves to formulas constructed from some initial set by means of only negation and alternation-with-disjoint-components, then we can extend probabilities from this initial set to all such formulas. But, as all formulas (on $A_1, \ldots, A_n$) are expressible as a disjunction (or negation of a disjunction) of basic conjunctions (on $A_1, \ldots, A_n$), we can, by taking basic conjunctions as the initial set, extend probabilities to all formulas. We are thus led to the following definition.

Let k_j $(j = 1, \ldots, 2^n)$ be real numbers such that for each j, $k_j \in [0,1]$, and such that the sum of all is 1. A *probability model* M (adequate for a set of formulas $\chi_1, \ldots, \chi_N$) is an assignment of the numbers k_j to the basic conjunctions K_j on variables $A_1, \ldots, A_n$, where this list of variables is long enough to include all those occurring in $\chi_1, \ldots, \chi_N$.

We write $P_M(K_j)$ for the value assignment by M to K_j (i.e., $P_M(K_j) = k_j$) and extend P_M to all formulas on $A_1, \ldots, A_n$ by setting $P_M(\phi) = \Sigma^{(\phi)} k_j$, where the summation is over all j for which K_j implies ϕ or, lacking such, we set $P_M(\phi) = 0$. By 'K_j implies ϕ' we mean with K_j and ϕ taken as formulas in the ordinary two-valued sense. Thus we are including two-valued logic as part of our semantic apparatus. The so-defined function P_M satisfies the following general properties of probability:

P1.   (i) $P(\phi) = 0$, if ϕ implies $A_1 \wedge \neg A_1$
         (ii) $P(\phi) \le P(\psi)$, if ϕ implies ψ

P2. $P(\neg\phi) = 1 - P(\phi)$
P3. $P(\phi \vee \psi) = P(\phi) + P(\psi)$, if $\phi \wedge \psi$ implies $A_1 \wedge \neg A_1$.

Note that if in a probability model M some $k_{i_0} = 1$ (and hence all other k_j's are 0) then for $i = 1, \ldots, n$, the formula A_i has P_M value 0, or 1, according as A_i appears negated, or unnegated, in K_{j_0}. Such models then coincide with two-valued models.

We now define logical consequence for probability logic. For this we use as our prototype the more general form of two-valued logic introduced in Section 1.

Let $\alpha_1, \ldots, \alpha_m, \beta$ be subsets of $[0,1]$. The (*probability*) *logical consequence* relation, denoted by

(1) $P(\phi_1) \in \alpha_1, \ldots, P(\phi_m) \in \alpha_m \models P(\psi) \in \beta$,

is defined as: For all probability models M (adequate for $\phi_1, \ldots, \phi_m, \psi$): if $P_M(\phi_1) \in \alpha_1, \ldots, P_M(\phi_m) \in \alpha_m$, then $P_M(\psi) \in \beta$.

The intuitive picture here is that of a "truth"-table entered from basic conjunctions $K_1, \ldots, K_{2^n}$ with additional columns headed $\phi_1, \ldots, \phi_m, \psi$. The $K_1, \ldots, K_{2^n}$ are assigned all possible sets of 2^n real numbers from $[0,1]$, the sum of such numbers in each set being 1. Each assignment (row) determines a probability model and corresponding values for $\phi_1, \ldots, \phi_m, \psi$. The premise conditions in (1) select out of the $2^{\aleph_0}$ rows of the table those in which the probabilities of the ϕ_i are, respectively, in the sets α_i, the relation (1) then holding if for each of these rows ψ has a probability in β.

When a subset α of $[0,1]$ is a singleton, say, $\alpha = \{a\}$, then we shall write $P(\phi) = a$ in place of $P(\phi) \in \alpha$.

Theorem 2.1 $\models P(\phi) = 1$ *iff ϕ is a (two-valued) tautology.*

Proof: (a) If ϕ is a tautology then, for any M, $P_M(\phi) = 1$. (b) If ϕ is not a tautology then its expansion as an alternation of basic conjunctions (we assume it has one, otherwise it is equivalent to $A_1 \wedge \neg A_1$ and hence $P_M(\phi) = 0$) is missing at least one conjunction, say K_{j_1}. There are models in which $k_{j_1} \neq 0$; in any such $P_M(\phi) \neq 1$.

Since properties P1–P3 hold for any P_M we can list any of the simple identities derivable from these properties as probability consequence relations. For example, $P(\phi \vee \psi) = P(\phi) + P(\psi) - P(\phi \wedge \psi)$ can be rephrased as

$$P(\phi) = a, P(\psi) = b, P(\phi \wedge \psi) = c \models P(\phi \vee \psi) = a + b - c .$$

However for the more substantial results we are going to arrive at we need, not particular identities, but some general theorems. Since these theorems are not well-known and, also, provide an effective procedure for obtaining the strongest conclusion from a set of premises, we devote a section to a short exposition.

3 Some theorems about probability algebras   In the preceding section we considered probability measures on sets of formulas built up from the propositional variables $A_1, \ldots, A_n$ and the connectives $\neg, \wedge, \vee$. But since (by P1(ii)) logically equivalent formulas have the same probability values, we can simplify the structure carrying the probability measure by "identifying" logically equivalent formulas, namely by going over to the Lindenbaum-Tarski algebra,

i.e., the Boolean algebra of equivalence classes of logical formulas modulo logical equivalence. However we shall not bother to change notation, using $\neg$, $\wedge$, $\vee$ in place of $^-$, $\cap$, $\cup$, (and with $\varnothing$, the zero element of the algebra being the equivalence class determined by $A_1 \wedge \neg A_1$). The general notion here is that of a *probability algebra*: an ordered pair $\langle B, P \rangle$, where B is a Boolean algebra and P a real-valued function, a probability function, defined on elements of B, with the properties

(P1)   (i) $P(A) \geq 0$
       (ii) $P(A) = 0$ if $A = \varnothing$
(P2) $P(\neg A) = 1 - P(A)$
(P3) $P(A \vee B) = P(A) + P(B)$ if $A \wedge B = \varnothing$.

(In strictness the 'P' should carry an indication, e.g., a subscript, of the Boolean algebra it is associated with but we shall usually omit it, introducing it only when necessary.)

Given (P1)–(P3) it is readily shown that[1]

$$\max\{0, P(A) + P(B) - 1\} \leq P(AB) \leq \min\{P(A), P(B)\} \ ,$$

a result giving upper and lower bounds for $P(AB)$ in terms of $P(A)$ and $P(B)$ which are good for any probability algebra. (For a "primitive" treatment of this inequality relation see [1], pp. 298–299.)

Are these "best possible", i.e., could there be another *function of $P(A)$ and $P(B)$* giving bounds for $P(AB)$, good for all probability algebras, which would be narrower in some cases? As Fréchet shows (in [4]), these bound functions are indeed best possible since, for example, in the case of the upper bound, no matter what values $P(A)$ and $P(B)$ may have, there is a probability algebra and elements A and B in it such that $P(AB) = \min\{P(A), P(B)\}$.

This result of Fréchet's was generalized in [6], where the following was established:

Theorem 3.1   *Given any Boolean polynomial expression* $\phi(A_1, \ldots, A_n)$: (i) *there are numerical-valued n-ary functions L_ϕ and U_ϕ depending only on the (Boolean) structure of ϕ, such that the inequalities*

(*) $L_\phi(a_1, \ldots, a_n) \leq P(\phi(A_1, \ldots, A_n)) \leq U_\phi(a_1, \ldots, a_n)$

hold in any probability algebra for which $P(A_i) = a_i$ ($i = 1, \ldots, n$); (ii) the bounds in (*) are best possible; and (iii) the functions L_ϕ and U_ϕ are effectively determinable (by solving a linear programming problem) from the structure of ϕ.

Here are a couple of examples of U_ϕ's. If $\phi(A_1, A_2) = A_1 \vee A_2$, then

$$U_\phi(a_1, a_2) = \min\{1, a_1 + a_2\} \ ;$$

and if $\phi(A_1, A_2, A_3) = \neg A_1 A_2 A_3 \vee \neg A_2 A_1 A_3 \vee \neg A_3 A_1 A_2$, then

$$U_\phi(a_1, a_2, a_3) = \min\left\{ \frac{a_1 + a_2 + a_3}{2}, a_1 + a_2, a_1 + a_3, a_2 + a_3, \right.$$

$$\left. 1 - a_1 + 1 - a_2 + 1 - a_3 \right\} \ .$$

We give a brief indication of the ideas behind the proof of Theorem 3.1 (full details are in [6]). As an illustration we take the simple example of finding U_ϕ for $\phi(A_1, A_2) = \neg A_1 \vee A_2 = A_1 A_2 \vee \neg A_1 A_2 \vee \neg A_1 \neg A_2$. In any probability algebra in which A_1 and A_2 are elements we have the quantities $k_1 = P(A_1 A_2)$, $k_2 = P(A_1 \neg A_2)$, $k_3 = P(\neg A_1 A_2)$, and $k_4 = P(\neg A_1 \neg A_2)$, where $A_1 A_2$, $A_1 \neg A_2$, $\neg A_1 A_2$, and $\neg A_1 \neg A_2$ are the four *basic conjunctions* on A_1, A_2. Then, as shown in [6], the least upper bound of the values $P(\phi(A_1, A_2))$, as one varies the probability algebra and the elements A_1, A_2 of the algebra (subject to the condition that $P(A_i) = a_i$ $(i = 1, 2)$) is the same as finding the maximum of $g = k_1 + k_3 + k_4 [= P(\phi(A_1, A_2))]$, subject to the constraints

$$
\begin{aligned}
k_1 + k_2 & = a_1 \\
(1) \quad k_1 + k_3 & = a_2 \\
k_1 + k_2 + k_3 + k_4 &= 1, \text{ and } k_i \geq 0 \ (i = 1, \ldots, 4),
\end{aligned}
$$

or, in matrix notation to bring out the pattern of 0's and 1's,

$$
(2) \quad
\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}
\begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix}
=
\begin{bmatrix} a_1 \\ a_2 \\ 1 \end{bmatrix}, \quad
\begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} \geq 0 \ ;
$$

that is, finding the maximum of a linear form $g = k_1 + k_3 + k_4$ subject to the linear inequality constraints (2). By the duality theorem of linear programming this problem is equivalent to finding the minimum of $r = a_1 x_1 + a_2 x_2 + 1 \cdot x_3$ subject to the constraints

$$
(3) \quad
\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}
\geq
\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \ .
$$

Note that in (3) the initial matrix is the transpose of that of (2), and that the quantities a_1, a_2 no longer appear in the constraint conditions but are coefficients in the form $a_1 x_1 + a_2 x_2 + 1 \cdot x_3$; also that the matrix on the right-hand side in (3) reproduces in order the k's associated with $P(\phi(A_1, A_2))[= 1 \cdot k_1 + 0 \cdot k_2 + 1 \cdot k_3 + 1 \cdot k_4]$, and is uniquely determined by the structure of ϕ. From linear programming theory we know that if conditions (3) are consistent then the minimum of the linear form r occurs at a corner point (x_1, x_2, x_3) of the polytope defined by (3). In our particular case these points are found to be $(-1, 1, 1)$ and $(0, 0, 1)$ giving then

$$(4) \quad \min r = U_\phi(a_1, a_2) = \min\{1 - a_1 + a_2, 1\}.$$

Best lower bounds are found by noting that

$$L_\phi(a_1, \ldots, a_n) = 1 - U_{\neg\phi}(a_1, \ldots, a_n) \ .$$

Consistency conditions for the linear system (1) are easily seen by inspection to be $0 \leq a_1 \leq 1$, $0 \leq a_2 \leq 1$. In general, there is an effective procedure for determining consistency of any system of linear inequations which may also include equations (see, for example, [15], Section 1.1).

The result of Theorem 3.1 can be considerably extended:

Theorem 3.2   *Given Boolean polynomial expressions* ψ, $\phi_1, \ldots, \phi_m$ *in variables* $A_1, \ldots, A_n$: (i) *there are 2m-ary numerical functions* $L_\psi^{(\phi)}$ *and* $U_\psi^{(\phi)}$ *depending only on the structures of* ψ, $\phi_1, \ldots, \phi_m$, *such that the inequalities*

$$(**) \quad L_\psi^{(\phi)}(a_1, \ldots, a_m, b_1, \ldots, b_m) \leq P(\psi(A_1, \ldots, A_n))$$
$$\leq U_\psi^{(\phi)}(a_1, \ldots, a_m, b_1, \ldots, b_m)$$

hold in any probability algebra for which

$$a_i \leq P(\phi_i(A_1, \ldots, A_n)) \leq b_i \quad (i = 1, \ldots, m) \ ;$$

(ii) *the bounds in (**) are best possible; and* (iii) *the function* $L_\psi^{(\phi)}$ *and* $U_\psi^{(\phi)}$ *are effectively determinable (by solving a linear programming problem) from the structures of* ψ, $\phi_1, \ldots, \phi_m$.

In this case, unlike for Theorem 3.1 where the consistency conditions are the simple $0 \leq a_i \leq 1$, the requirements here imposed by the constraints $a_i \leq P(\phi_i) \leq b_i$ demand considerably more. Nevertheless, the system of linear inequalities in the $k_j (j = 1, \ldots, 2^n)$ which expresses the constraints is effectively decidable as to consistency.

A special case of Theorem 3.2 which will be of interest to us is that for which $\phi_i(A_1, \ldots, A_n) = A_i$. Here the constraints become

$$a_i \leq P(A_i) \leq b_i \quad (i = 1, \ldots, n)$$

and we are thus considering bounds on $P(\psi(A_1, \ldots, A_n))$ not in terms of values a_i of the $P(A_i)$, but in terms of intervals $[a_i, b_i]$ within which the $P(A_i)$ lie.

The guiding ideas for the proof of Theorem 3.2 are similar to those for Theorem 3.1 (see [6], Section 6).

We point out that, as far as the determination of best lower and upper bounds for a $\psi(A_1, \ldots, A_n)$ is concerned, appeal to probability algebras is replaceable by assignments of values to the k_j taken as variables—with, of course, the requirements

$$k_j \in [0,1] \quad \text{and} \quad \sum_{j=1}^{2^n} k_j = 1 \ .$$

The only way the probability algebras enter into the problem is via the values k_j. Hence varying all possible such values is equivalent to varying all possible probability algebras. In other words, the full generality of probability algebras is not needed, as we are concerned only with (finite) Boolean algebras which are generated by n generators and (call these generators $A_1, \ldots, A_n$) which have the property that $\bigvee' K_j = \varnothing$, where $\bigvee'$ indicates an alternation over a proper subset of the basic conjunctions on $A_1, \ldots, A_n$, namely those having $k_j = 0$. Such probability algebras correspond precisely to our probability models.

4 Probability logic for intervals   By use of the results stated in the preceding section we can completely characterize the class of valid inference forms for

probability logic when the sets involved are subintervals of [0,1]. First a special case.

Theorem 4.1   *Let L_ψ and U_ψ be the functions furnishing lower and upper bounds as in Theorem 3.1. Define an interval-valued function Pf_ψ by setting*

$$Pf_\psi(a_1, \ldots, a_n) = [L_\psi(a_1, \ldots, a_n), U_\psi(a_1, \ldots, a_n)] \ .$$

Then for $a_1, \ldots, a_n \in [0,1]$,

$$P(A_1) = a_1, \ldots, P(A_n) = a_n \vDash P(\psi) \in Pf_\psi(a_1, \ldots, a_n) \ .$$

Proof: Let M be a probability model for which $P_M(A_i) = a_i$ $(i = 1, \ldots, n)$. By virtue of the discussion at the end of the preceding section we may consider P_M as a probability function for a probability algebra and hence conclude that $P_M(\psi) \in Pf_\psi(a_1, \ldots, a_n)$, since the interval $Pf_\psi(a_1, \ldots, a_n)$ contains $P_M(\psi)$ for any probability algebra in which $P_M(A_i) = a_i$ $(i = 1, \ldots, n)$.

The consistency conditions for the premises in Theorem 4.1 are simply that the a_i be probability values, i.e., numbers in the interval [0,1]. This can be proved formally, but is quite evident if one thinks of the A_i as regions of the unit square, each with area equal respectively to the assigned probability.

For the general theorem we first introduce an interval-valued function $Pf_\psi^{(\phi)}$ of $2m$ arguments $a_1, \ldots, a_m, b_1, \ldots, b_m$ but, with an abuse of notation, write its arguments as $\alpha_1, \ldots, \alpha_m$, where $\alpha_i = [a_i, b_i]$. Using notation from Theorem 3.2, we set

$$Pf_\psi^{(\phi)}(\alpha_1, \ldots, \alpha_m) = [L_\psi^{(\phi)}(a_1, \ldots, a_m, b_1, \ldots, b_m),$$
$$U_\psi^{(\phi)}(a_1, \ldots, a_m, b_1, \ldots, b_m)] \ .$$

Theorem 4.2   *Let $\alpha_1, \ldots, \alpha_m, \beta$ be probability intervals. Then*

(i) $P(\phi_1) \in \alpha_1, \ldots, P(\phi_m) \in \alpha_m \vDash P(\psi) \in \beta$

if and only if

(ii) $Pf_\psi^{(\phi)}(\alpha_1, \ldots, \alpha_m) \subseteq \beta$, *where $Pf_\psi^{(\phi)}$ is the probability interval-valued function defined above.*

Proof: Immediate by Theorem 3.2.

5 Examples   In the following examples of inference forms for probability logic we restrict ourselves to premises in which the probability intervals contain a single value and the conclusion interval is the strongest possible, i.e., as given by $Pf_\psi^{(\phi)}$. Although unsophisticated techniques would suffice to establish these simple inferences, we shall nevertheless use the full-blown linear algebra technique in Example 1 in order to illustrate the general method. In the remaining examples we merely state results.

Example 1 (Generalizing modus ponens):

$$P(A_1) = p, \ P(A_1 \to A_2) = q \vDash P(A_2) \in [p + q - 1, q] \ ,$$

with consistency conditions on the parameters p and q: $p \leq 1$, $q \leq 1$, $p + q \geq 1$.

To show this let k_1, k_2, k_3, k_4 be the values associated with the basic con-

junctions on A_1, A_2. On expressing the conditions in the premises in terms of the k's, and adjoining the probability requirements on the k's, we obtain

(1)
$$\begin{aligned}
k_1 + k_2 \qquad\qquad &= p \\
k_1 \qquad + k_3 + k_4 &= q \\
k_1 + k_2 + k_3 + k_4 &= 1 \\
k_1,\, k_2,\, k_3,\, k_4 &\geq 0.
\end{aligned}$$

It is readily seen that these imply the stated consistency conditions on p and q. (In general one can always use elimination techniques, e.g., Fourier elimination, for determining consistency of systems of linear inequations.) To find the best upper bound for $P(A_2)$ subject to conditions (1) one needs to find max $k_1 + k_3$, subject to

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} = \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} \geq 0 \ .$$

The dual form for this linear programming problem is to find min $px_1 + qx_2 + x_3$, subject to

(2)
$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \geq \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \ .$$

By simple algebraic means one finds the one corner point of the polytope specified by (2), namely, $(0,1,0)$. Hence $p \cdot 0 + q \cdot 1 + 0 = q$ is the minimum value. The best lower bound is found by use of complementary techniques, giving the result as stated.

Example 2 (The "hypothetical syllogism"):

$$P(A_1 \to A_2) = p, \ P(A_2 \to A_3) = q \models P(A_1 \to A_3) \in [p + q - 1, 1] \ ,$$

with consistency conditions $p \leq 1$, $q \leq 1$, $p + q \geq 1$.

The next example illustrates the effect of additional information. In Example 1 the premises show no relation between A_1 and A_2. If we replace A_1 by $A_1 \lor A_2$ the result remains true, but a sharper result can be given, namely (dropping the redundant A_2 from $A_1 \lor A_2 \to A_2$):

Example 3:

$$P(A_1 \lor A_2) = p, \ P(A_1 \to A_2) = q \models P(A_2) = p + q - 1 \ ,$$

with the same consistency conditions as in Example 1.

In the preceding three examples the formula in the conclusion was a necessary consequence (in the two-valued sense) of the formulas in the premises. Here is a simple example where this is not the case:

Example 4:

$$P(A_1 \to A_2) = p \models P(A_2 \to A_1) = [1 - p, 1] \ .$$

We return to the problem, referred to in our Introduction, of combining separate arguments for a conclusion. Using De Morgan's language we translate 'A is an argument for C' by '$A \to C$'; 'A is logically good (for C)' by '$P(A \to C) = 1$'; 'A proves C' by '$A \wedge (A \to C)$'; and 'the validity of A' by '$P(A)$'. Note that the validity of A is equal to the probability that A proves C if the argument A is logically good. The problem considered is: given arguments $A_1, \ldots, A_n$ (for C), all being logically good, and with respective validities $a_1, \ldots, a_n$, what is the probability that the conclusion is "proved"? Since $(A_1 \to C) \wedge \ldots \wedge (A_n \to C)$ is logically equivalent to $(A_1 \vee \ldots \vee A_n) \to C$ the question is equivalent to asking for $P(A_1 \vee \ldots \vee A_n)$ given that $P(A_i) = a_i$ $(i = 1, \ldots, n)$. De Morgan gives the result $1 - (1 - a_1)(1 - a_2) \ldots (1 - a_n)$, which is correct if $P(\neg A_1 \neg A_2 \ldots \neg A_n) = P(\neg A_1) P(\neg A_2) \ldots P(\neg A_n)$, i.e., if the $A_1, \ldots, A_n$ are stochastically independent. Without this assumption we have

Example 5:

$$P(A_1) = a_1, \ldots, P(A_n) = a_n \vDash P(A_1 \vee A_2 \vee \ldots \vee A_n)$$
$$\in [\max\{a_1, \ldots, a_n\}, \min\{1, a_1 + a_2 + \ldots + a_n\}] \ .$$

Our methods readily extend to cases in which the arguments for C can be contingent ($P(A_i \to C) = b_i$), and also to the inclusion in the premises of interrelations among the A_i when these can be logically expressed. In contrast to Riechenbach's having to use an additional variable to express $P(A_1 \vee A_2)$ in terms of $P(A_1)$ and $P(A_2)$, note how Example 5 expresses the *interval* for $P(A_1 \vee A_2)$ solely in terms of $P(A_1)$ and $P(A_2)$.

One might observe that the upper and lower bounding functions occurring in Examples 1–5 are obtainable by functional compositions from the two functions

$$f(x,y) = \min\{1, 1 - x + y\}$$
$$g(x) = 1 - x \ ,$$

which are the conditional and negation of Łukasiewicz many-valued logic on the unit interval. However, this is not a general feature. For example, as observed in Section 3, $\neg A_1 A_2 A_3 \vee \neg A_2 A_1 A_3 \vee \neg A_3 A_1 A_2$ has for the best upper bound

$$\min\left\{ \frac{a_1 + a_2 + a_3}{2}, a_1 + a_2, a_1 + a_3, a_2 + a_3, 1 - a_1 + 1 - a_2 + 1 - a_3 \right\} \ ,$$

which cannot be obtained by composition from the Łukasiewicz functions.

NOTE

1. Henceforth we shall use juxtaposition for conjunction.

REFERENCES

[1] Boole, G., *The Laws of Thought*, Macmillan, London, 1854.

[2] Carnap, R., *The Logical Foundations of Probability*, 2nd Ed., University of Chicago Press, Chicago, Illinois, 1962.

[3] De Morgan, A., *Formal Logic*, Taylor and Walton, London, 1847.

[4] Fréchet M., "Generalizations du théorème des probabilities totales," *Fundamenta Mathematicae*, vol. 25 (1935), pp. 379–387.

[5] Gaifman, H., "Concerning measures in first-order calculi," *Israel Journal of Mathematics*, vol. 2 (1964), pp. 1–17.

[6] Hailperin, T., "Best possible inequalities for the probability of a logical function of events," *American Mathematical Monthly*, vol. 72 (1965), pp. 343–359.

[7] Hailperin, T., *Boole's Logic and Probability*, North-Holland, Amsterdam, 1976.

[8] Kappos, D. A., *Strukturtheorie der Wahrscheinlichkeitsfelder und -räume*, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1960.

[9] Kappos, D. A., *Probability Algebras and Stochastic Spaces*, Academic Press, New York and London, 1969.

[10] Keynes, J. M., *A Treatise on Probability*, Macmillan, London, 1921.

[11] Peirce, C. S., "A theory of probable inference," pp. 126–181 in *Studies in Logic*, Little, Brown & Co., Boston, Massachusetts, 1883.

[12] Reichenbach, H., *The Theory of Probability*, University of California Press, Berkeley, 1949.

[13] Rescher, N., *Many-Valued Logic*, McGraw-Hill, New York, 1969.

[14] Scott, D. and P. Krauss, "Assigning probabilities to logical formulas," in *Aspects of Inductive Logic*, ed. J. Hintikka and P. Suppes, North-Holland, Amsterdam, 1966.

[15] Stoer, J. and C. Witzgall, *Convexity and Optimization in Finite Dimensions I*, Springer-Verlag, New York-Heidelberg-Berlin, 1970.

[16] Tarski, A., "Wahrscheinlichkeitslehre und mehrwertige Logik," *Erkenntnis*, vol. 5 (1935–36), pp. 174–175.

Department of Mathematics
Lehigh University
Bethlehem, Pennsylvania 18015