Notre Dame Journal of Formal Logic
Volume XX, Number 1, January 1979
NDJFAM

ALTERNATIVE AXIOMATIZATIONS OF ELEMENTARY PROBABILITY THEORY

WILLIAM K. GOOSENS

1 *Outline* In this paper I offer some new axiomatizations of elementary (finite) probability theory. The new axiomatizations have the following distinctive features. First, both conditional and unconditional probabilities are simultaneously axiomatized. Second, the axiomatizations are extensions of Kolmogorov probability theory, and have the usual definitions linking conditional and unconditional probabilities as theorems. Third, while both probabilities are simultaneously axiomatized, the resultant axioms are about as simple as the usual axiomatizations of conditional probability alone. Fourth, the axioms adopted have strong and direct justifications independently of the interpretation of the probability function (axiomatization **L** however is only a permissible extension of intuitions). The reader who wishes to review the axiomatizations before their rationale can at this point go to section **5** (the main axiomatization is system **KK**).

2 *What is unsatisfactory about usual axiomatizations?* I offer alternative axiomatizations because I am dissatisfied with the customary ones. But this dissatisfaction has a special nature. In improving existing axiomatizations, one might reduce the number of axioms, or simplify them, or adopt axioms that make proofs easier and more elegant. Although in my own axiomatizations I am concerned with these factors, none of these traditional formal motivations is the source of my dissatisfaction. Instead, I am basically concerned about how the relationship between conditional and unconditional probabilities is developed (for some related misgivings, see de Finetti [2], pp. 81-83). There are two probability functions about which we have intuitions. The basic formal difference is that unconditional probability is a one-place function, and conditional probability is a two-place function. As a result, there are two probabilistic concepts that can be axiomatized, and the two can be axiomatized independently. What makes a function a one-place probability function, and what makes a function a two-place probability function, are separate questions, and the correct answer for each involves a separate characterizing theory.

Received March 16, 1975

However, if one were merely to conjoin the two separate characterizing theories, the result would not be an adequate theory of probability. The reason is that no theorem at all would follow from the composite theory about the relationship between conditional and unconditional probabilities. Yet most axiomatizations of *probability theory* take only one probability function as primitive. How can a theory with only one primitive serve to prove all the truths of probability? The answer is deceptively simply from a formal viewpoint. One simply defines the other probability function in terms of the primitive. My first objection to the usual axiomatizations is that unconditional and conditional probabilities are joined by means of definition.[1] What exception can be taken to a definition? On the level of justification, definition is a conversation stopper. The formal theory does not itself argue that probabilities are so related, instead the relationship is decreed. This approach goes hand in hand with taking one probability as primitive. The opinion is widely held that all probabilities "really" are conditional, or at least that conditional probabilities are basic. If this belief translates into axiomatically taking only conditional probability as primitive, what may well be a justifiable position is then incorporated into the theory as a dogma. Argument for the fundamentalness of a concept is then pre-axiomatic or extra-formal. The suggestion seems to be that fundamentalness is not subject to proof, but is based on some philosophic intuition.

The universal practice of taking only one probability function as primitive has then had the consequence that questions about the fundamentalness of the different functions and their relationship have been raised outside of, rather than within, the formal theory of probability. These considerations indicate shortcomings of the single-primitive-plus-definition approach to axiomatic probability. A simultaneous axiomatization that takes both functions as primitive has the following advantages. The axiomatization does not presuppose that one function is fundamental, or that probabilities are related definitionally. The question of whether or not one primitive is definitionally eliminable then becomes a matter of proof. The question of whether or not the definitional relationships can be justified by derivation from more fundamental intuitions is also then brought into consideration. As we shall see, simultaneous axiomatization sheds light on both questions. We shall show that there are more fundamental intuitions. Moreover, while unconditional probability is eliminable in favor of conditional probability, the reverse is simply not true unless conditional probability is axiomatized in a weak fashion. The problem centers on conditionalizing factors with zero prior probability. Axiomatizations typically include theorems that hold even if the prior probability is zero, versus being undefined. (This point follows from the discussion in sections 4 and 5.) This asymmetry provides a neglected formal justification for the claim that conditional probabilities are more fundamental. Another advantage of simultaneous axiomatization is that the truths of probability are approached in a manner both natural and unified. Any loss of theoretical eloquence from excess primitives is more than compensated by on how coherent a basis both notions can be developed. Moreover the theory is then developed in a way that better corresponds to how probability is applied. In any interpretation, the entities to which probabilities are attached have some initial probability: calling it the prior, unconditional, antecedent, or conditional probability given a fixed background, makes no difference. In application, it is a one-place function. There are also the probabilities the entities have under new information: these are the conditional probabilities. It seems to me not to be a direct intuition that in application the two probabilities are definitionally related. Thus a definitional relationship ought not be an axiomatic presupposition, but a consequence of natural assumptions concerning the properties and relations of the two probabilities.

Another source of my dissatisfaction with customary axiomatizations concerns the exact content of the ways in which conditional and unconditional probabilities are related. Starting with unconditional probability, the conditional probability is defined to be a certain ratio of unconditional probabilities, where the denominator—the probability of the conditionalizing factor—is not zero. First of all, the restriction to conditionalizing factors with non-zero probability is not reflected in any direct intuition about conditional probabilities, but this objection will be postponed until section **4**. I have no doubt that the claimed relationship, as far as it goes, is true. But the definition is not direct and obvious in the way that the axioms are. By adopting it as a definition it is trivially provable without being correspondingly trivial. More importantly, the definition prematurely closes the effort to try to find more basic and fundamental relationships between conditional and unconditional probabilities from which the ratio definition can be proved.

That the ratio definition is not intuitively fundamental is evidenced by the widespread effort to justify it by derivation from other intuitions. Interestingly, most justifications involve a particular interpretation of probability. If probabilities are relative frequencies, betting ratios, or the ratio of favorable to total cases, or if probabilities are interpreted as areas, more or less ingenious and clever justifications of the ratio definition follow. Such justifications, however, are piecemeal, and at most justify the definition for this or that application of probability theory. In this paper I seek a more general and more purely formal basis for the ratio definition—and therefore a justification not tied to any particular interpretation of probability.

In axiomatizations where conditional probabilities are taken as primitive, the unconditional probability of A is taken to be the conditional probability of A on some logical truth. This itself is intuitive and unproblematic, since a logical truth amounts to no additional information whatsoever. What is less satisfying is the way in which one then proves the ratio definition. The widely adopted axiom is

$$\mathrm{P}((A \,\&\, B)|E) = \mathrm{P}(A|(B \,\&\, E))\mathrm{P}(B|E)$$

Substituting a logical truth for E immediately yields the ratio relationship as a special case. Misgivings about the fundamentalness of the ratio relationship then carry over to this axiom. There are various mathematical considerations that have been shown to lead to the ratio definition. Such approaches, though, usually involve mathematical apparatus that goes beyond that used in axiomatizing probability—e.g. properties of entropy and information. What I seek is an approach which is not only more intuitive, but also capable of replacing the usual definitions in an axiomatic development, and therefore does not go beyond the usual apparatus of probability theory. To the alternative justification I now turn.

3 *Probability and the logic of conditionals* If the interface between conditional and unconditional probabilities is not to depend on the interpretation of the probability function, what intuitions are left? The answer is that we can use the intuition that conditional probabilities are conditionals. The basic form of inference using conditionals governs how they can be de-conditionalized, or as we shall say, conditionalized out. From the conditional "if A then B" one can obtain B by adding A, i.e. "A and if A then B" entails "B". Notice that this inference shows how the conditional assertion ("if A then B") relates to unconditional assertions ("A" and "B"). Likewise, the key relationship between conditional and unconditional probabilities concerns how to pass from conditional to unconditional probability. Where $\mathrm{P}(A|B)$ is to represent the probability of A conditional on B, what does one need to add to generate unconditional probabilities? A basic form of inference is that $\mathrm{P}(A)$ equals $\mathrm{P}(A|B)$ if $\mathrm{P}(B) = 1$. But this relationship can easily be generalized. Where B has prior probability 1, $\mathrm{P}(A|B)$ would be $\mathrm{P}(A)$. But where B has prior probability less than 1, $\sim B$ needs to be considered. The probability that A has can be viewed as deriving from two possible sources: B and $\sim B$. Since B and $\sim B$ are exclusive, the probability that A derives from them will be additive. And since B and $\sim B$ are exhaustive, all of the probability of A will derive from these sources. Now $\mathrm{P}(A|B)$ gives $\mathrm{P}(A)$ if $\mathrm{P}(B) = 1$. But B may be uncertain, in which case $\mathrm{P}(A|B)$ contributes to $\mathrm{P}(A)$ in *direct proportion* to $\mathrm{P}(B)$. Since the same is true for $\sim B$, we therefore derive

$$\mathrm{P}(A) = \mathrm{P}(A|B)\mathrm{P}(B) + \mathrm{P}(A|\sim B)\mathrm{P}(\sim B)$$

As long as $B_1, \ldots, B_n$ include all of the ways A can be possibly true, and the B's are inconsistent with one another, the same intuitions apply. The obvious generalization is

(*) where A entails the disjunction of $B_1, \ldots, B_n$ and $(B_i \;\&\; B_j)$ is inconsistent for $i \neq j$, $\mathrm{P}(A) = \sum \mathrm{P}(A|B_i)\mathrm{P}(B_i)$.

The summation sign derives from the fact that exclusive cases are considered. The equality derives from the fact that the B's form exhaustive cases of A. The conditional probabilities represent the contribution to $\mathrm{P}(A)$ of the factor conditionalized upon, assuming the factor. The multiplication by the unconditional probability represents the proportion of the conditional probability that is contributed to $\mathrm{P}(A)$. Where B is certain to be false,

B contributes nothing to $P(A)$, and where B is certain to be true, $P(A) = P(A|B)$. But B is known only with some probability.

Thesis (*) then is directly intuitive independently of the interpretation of the probability function. Corresponding, I suggest that (*) is appropriate to take as axiomatic, and that (*) can be used as a fundamental building block for joining conditional and unconditional probabilities. Notice too that (*) does not have the form that would allow it to serve as a definition of either conditional or unconditional probabilities, and therefore (*) could not be used to bridge the two probabilities in an axiomatic development that took only one probability function as primitive. The effort to minimize the number of primitives in probability theory has therefore tended to obscure the fundamentalness of (*), forcing the bridge between conditional and unconditional probabilities to be made using only relationships that could serve as definitions.

The way in which (*) can be used to derive the usual links is instructive. Using B and $\sim B$ as inconsistent cases, and substituting $(A \& B)$ for A, we get

(**) $P(A \& B) = P((A \& B)|B)P(B) + P((A \& B)|\sim B)P(\sim B)$

By arguing that the second term equals 0 and the first reduces to $P(A|B)P(B)$, we immediately derive the ratio relationship that if $P(B) > 0$, then $P(A|B) = P(A \& B)/P(B)$. Where $\mathbf{T}$ is a logical truth, adding that $P(\mathbf{T}) = 1$ then gives $P(A) = P(A|\mathbf{T})$ by substitution of $\mathbf{T}$ for B in the ratio definition. The different axiomatizations offered herein take their cue from how (**) can be simplified. We shall explore how the inference from (**) to the ratio definition can be axiomatically derived. We then seek axiomatizations that employ (*) so as to derive the usual links between conditional and unconditional probabilities, and to result in systems which include the usual axiomatizations. The answer to this purely mathematical question constitutes the final section of this paper. One final point, however, must be developed before the alternative axiomatizations can be fully appreciated.

4 *Some differences between axiomatizations concerning conditional probabilities* Whereas the axiomatization of unconditional probability is highly standardized, there is variation in the properties attributed to conditional probability. These differences are reflected in the following five theses. (Note that *T2* entails *T1*, and *T5* plus *T3* entails *T2*).

T1 $P(A|A) = 1$
T2 *If A entails B, then* $P(B|A) = 1$.
T3 $P(B|A) + P(\sim B|A) = 1$.
T4 *If $(B \& E)$ is inconsistent, then* $P((B \vee E)|A) = P(B|A) + P(E|A)$
T5 *If A entails $\sim B$, then* $P(B|A) = 0$.

I take the standard version of probability theory to be the Kolmogorov [4] axiomatization plus the ratio definition of conditional probability. In the standard version, none of these theses are theorems. The reason is that

$\mathrm{P}(B|A)$ is determined only where $\mathrm{P}(A) \neq 0$. What is provable is each of the five under the condition that $\mathrm{P}(A) \neq 0$. That the theses are true for this condition I regard as unproblematic. However, are any of the theses true under less restrictive conditions, i.e. where $\mathrm{P}(A) = 0$? First there is the case in which A is a contradiction. Notice that for this case it is impossible to satisfy all of the theses.[2] Since a contradiction entails everything, acceptance of *T2* forces *T3, T4,* and *T5* to be rejected. The sum of the probabilities in *T3* is 2, not 1. In *T4,* the left probability is 1, the right 2. And in *T5* $\mathrm{P}(B|A)$ is 1, not 0. Moreover, *T1* and *T4* are inconsistent. Likewise, acceptance of *T5* necessitates the rejection of *T1, T2,* and *T3.* An inconsistent A entails its negation, so by *T5* $\mathrm{P}(A|A) = 0$, not 1 as in *T1* and in *T2,* and the sum of the probabilities in *T3* would be 0, not 1.

Popper [5], p. 357, has clearly seen the perils of inconsistency embedded in *T1* through *T5* when taking conditional probabilities as primitive. Popper showed earlier editions of Harold Jeffreys' *Theory of Probability* to be inconsistent on just this point [5], p. 357. Yet similar errors occur in later axiomatizations. Salmon [9], p. 59, adopts *T2* and *T4.* No less a careful worker than Renyi [7], p. 70, once adopted *T1* and a generalized version of *T4.*

Second, consider the case where $\mathrm{P}(A) = 0$ even though A is not a contradiction. Unlike the case of a contradiction, all five theses are consistent for such A's. Indeed, all five theses seem eminently reasonable, and given that they are consistent for such A's, there seems to be little reason for not adopting them. Correspondingly, the restriction of these theses to the case where $\mathrm{P}(A) \neq 0$ seems artificial. The fact that a consistent A has prior probability 0 does not seem at all to prevent the things it entails from having conditional probability 1, and the things whose negation it entails to have conditional probability 0.

One desirable extension of standard probability theory, then, is to hold all five theses on the condition that A is consistent. Carnap [1], p. 38, Renyi [7], p. 38, and Jeffreys [3] all provide consistent axiomatizations which accept all five theses for consistent conditionalizations, and no more, as does the main axiomatization we offer. However, in these axiomatizations unconditional probability is essentially introduced definitionality as conditional probability on a logical truth. In all of them, the ratio definition then follows as a special case of an axiom, thus bridging conditional and unconditional probabilities in the way we wish to avoid.

With respect to the five theses, two other extensions of standard probability are possible. One is to accept as universally true *T1* and *T2.* To my knowledge the only consistent axiomatization that has adopted *T1* universally is Popper's [5]. The other alternative is to accept as universally true *T5.* To my knowledge, no one has before proposed such a system. We show how *T5* can be combined with the idea in (*) to result in surprisingly simple extensions of standard probability theory. These last two extensions are inconsistent with one another. Although I find them interesting, in so far as both go beyond holding *T1* through *T5* conditional on consistent A's, I see no philosophical reason to prefer one axiomatization over the other. These different extensions do matter, though, as other axioms can then be weakened or simplified. To the long promised alternative axiomatizations we now turn.

5 *Alternative axiomatizations* We will assume that probability theory is developed as an adjunct to the propositional calculus, rather than the more standard set-theoretic framework. We therefore use 'A', 'B', 'C', 'D', 'E', and '**T**' for propositional assertions, and '&' for conjunction, 'v' for disjunction, and '~' for negation. Logically equivalent assertions are assumed to have identical probabilities. Standard elementary probability theory, as based on unconditional probability, consists of the following four assertions, which we call system **KU** (for Kolmogorov theory of unconditional probability):

KU

(1) $P(A) \geq 0$.
(2) $P(\mathbf{T}) = 1$ *if* **T** *is a logical truth.*
(3) $P(A \vee B) = P(A) + P(B)$ *if* $(A \,\&\, B)$ *is inconsistent.*
(4) *If* $P(A) > 0$, *then* $P(B|A) = P(B \,\&\, A)/P(A)$.

There is no standard formulation of probability theory based on conditional probability. However, let us call any theory based on conditional probabilities a *Kolmogorav* theory if and only if it is equivalent to **KU**. To my knowledge, no one who has proposed axiomatizations of conditional probability has investigated if the resultant system is equivalent to **KU**. The following system **KC** is provably equivalent to **KU**.

KC

(5) $P(\mathbf{T}|\mathbf{T}) > 0$, *where* **T** *is a logical truth.*
If $P(B|\mathbf{T}) > 0$ *and* **T** *is a logical truth, then* (6)-(9):
(6) $P(A|B) \geq 0$.
(7) *if* B *entails* A, *then* $P(A|B) = 1$.
(8) $P(A|B) + P(\sim A|B) = 1$.
(9) *if* $P((B \,\&\, A)|E) > 0$, *then* $P((E \,\&\, A)|B) = P(A|B)P(E|(B \,\&\, A))$.
(10) $P(A) = P(A|\mathbf{T})$, *where* **T** *is a logical truth.*

This system is a modified version of von Wright [12], p. 176. We will not here prove that **KU** is logically equivalent to **KC**. The proof itself is easy, with the exception of the proof of (3) from **KC**. Here the reader will find von Wright useful [12] especially pp. 187-188.

A second interesting logical question is how one could axiomatize exactly those theorems of the Kolmogorov theory that involve only conditional probabilities. We can likewise answer this formal question. An assertion involving only conditional probabilities is provable from **KU** if and only if it is provable from (5)-(9). In this sense, (5)-(9) provides an axiomatization of the theory of conditional probability embedded in **KU**. The proof this claim is simple. Clearly, the form of (10) makes it a definitional extension of (5)-(9), and therefore **KC** is a conservative extension of (5)-(9) (see Shoenfield [10], pp. 57-61, for the logic of this claim). Thus, any assertion formulable in (5)-(9) is provable from (5)-(9) if and only if it is provable from (5)-(10). But since **KU** is equivalent to (5)-(10), the theorems of (5)-(9) are exactly those assertions about only conditional probabilities provable from **KU**.

As indicated earlier, my dissatisfaction with the standard axiomatizations centers on the fundamentalness of the consequents of (4) and (9), and the restriction of conditional probabilities in the antecedent of (4) and (6)-(9). First, we offer an alternative axiomatization in which all of the following can be proved: (*), the customary links between conditional and unconditional probabilities, standard theories of unconditional probability, plus all five theses mentioned in the previous section relative to consistent conditionalizations. We call this system **KK**.

KK

A1 $\mathrm{P}(B|A) \geq 0$.
A2 *Where* **T** *is a logical truth and A is consistent,* $\mathrm{P}(\mathbf{T}) = \mathrm{P}(\mathbf{T}|A) = 1$.

Where A entails $(B \vee D)$ *and* $(B \,\&\, D)$ *is inconsistent, both*

A3 *If A is consistent,* $\mathrm{P}(E|A) = \mathrm{P}((E \,\&\, B)|A) + \mathrm{P}((E \,\&\, D)|A)$.
A4 $\mathrm{P}(A) = \mathrm{P}(A|B)\mathrm{P}(B) + \mathrm{P}(A|D)\mathrm{P}(D)$.

First note that **KK** allows the second argument of the probability function in general to have 0 probability, and in the case of *A1* and *A4*, even to be inconsistent. The first two axioms I call normalizing axioms, because they represent convenient choices of numbers for extrema. Note, too, that *A3* and *A4* are respectively equivalent to the following.

Where A entails that at least one of $B_1, \ldots, B_n$ *is true, and* $(B_i \,\&\, B_j)$ *is inconsistent for* $i \neq j$, *both*

Theorem 1 *If A is consistent, then* $\mathrm{P}(E|A) = \sum \mathrm{P}((E \,\&\, B_i)|A)$.

Theorem 2 $\mathrm{P}(A) = \sum \mathrm{P}(A|B_i)\mathrm{P}(B_i)$.

Although we adopt the binary case as axiomatic, the general case is no less intuitive. Theorem 2 is (*), discussed in section **3**. For A to entail at least one of $B_1, \ldots, B_n$ amounts to the condition that every way A can be true is included among the B's. The Theorems 1 and 2 assert that the probability of events on A derives from their overlap with the B's, and that A's probability can be derived from the B's. The summation derives from the exclusiveness of the B's, the identity from their inclusiveness of A. As Theorem 1 claims, where A is true, the B's include all of the ways A is true. Whatever probability E then has can then be divided up among the B's, and since the B's are exclusive, by summing over the probability of E's overlap with the B's, we get exactly E's probability. The outstanding feature of **KK**, in my opinion, is that the axioms have a direct and intuitive rationale.

As a special case of *A4* we have

Theorem 3 $P(A) = P(A|B)P(B) + P(A|{\sim}B)P({\sim}B)$.

Where C is a contradiction, both B, ${\sim}B$ and B, ${\sim}B$, C satisfy the antecedent of Theorem 2. Hence

Theorem 4 *Where C is a contradiction,* $P(B|C)P(C) = 0$.

As an interesting consequence of *A3*, we have

Theorem 5 *If A is consistent and A entails* ${\sim}B$, *then* $P(B|A) = 0$.

For suppose A is consistent and A entails ${\sim}B$. Now ${\sim}B$ is equivalent to $(({\sim}B \,\&\, D) \vee ({\sim}B \,\&\, {\sim}D))$. Hence $({\sim}B \,\&\, D)$, $({\sim}B \,\&\, {\sim}D)$ satisfies the antecedent to Theorem 1. But likewise $({\sim}B \,\&\, D)$, $({\sim}B \,\&\, {\sim}D)$, B satisfies the antecedent. Letting E be a logical truth T, the two applications of Theorem 1 then entail Theorem 5.

Where C is a contradiction, putting in C for both A and B in Theorem 3, we obtain $P(C) = P(C|C)P(C) + P(C|{\sim}C)P({\sim}C)$. By Theorem 4 the first term is 0, and by Theorem 5, $P(C|{\sim}C) = 0$, yielding

Theorem 6 *Where C is a contradiction,* $P(C) = 0$.

As another basic consequence of *A3* we get

Theorem 7 *If A is consistent,* $P(B|A) + P({\sim}B|A) = 1$.

Letting a logical truth T be E, we can apply *A3* to B, ${\sim}B$ to get $P(\mathsf{T}|A) = P(B|A) + P({\sim}B|A)$, which by *A2*, is 1. We then also have

Theorem 8 *If A is consistent and A entails B, then* $P((E \,\&\, B)|A) = P(E|A)$.

Since A is consistent, by *A3* we get $P(E|A) = P((E \,\&\, B)|A) + P((E \,\&\, {\sim}B)|A)$. Since A entails B, A entails ${\sim}(E \,\&\, {\sim}B)$, so by Theorem 5 $P((E \,\&\, {\sim}B)|A) = 0$, yielding Theorem 8.

Now we are in a position to prove a central theorem.

Theorem 9 $P(A \,\&\, B) = P(A|B)/P(B)$.

By Theorem 3 $P(A \,\&\, B) = P((A \,\&\, B)|B)P(B) + P((A \,\&\, B)|{\sim}B)P({\sim}B)$. At least one of B, ${\sim}B$ is consistent. We show the theorem for each case. (This proof can be simplified by noting that B entails $(B \vee C)$, where C is a contradiction, C then replacing ${\sim}B$).

Case 1: B is consistent. By Theorem 9 $P((A \,\&\, B)|B) = P(A|B)$. Either ${\sim}B$ is consistent or not. If ${\sim}B$ is consistent, by Theorem 5 $P((A \,\&\, B)|{\sim}B) = 0$. If ${\sim}B$ is not consistent, by Theorem 6 $P({\sim}B) = 0$. In either case the second factor in $P(A \,\&\, B)$ is 0, yielding the theorem.

Case 2: ${\sim}B$ is consistent. By Theorem 5 then, $P(A \,\&\, B) = P((A \,\&\, B)|B)P(B)$. Either B is consistent or not. If B is consistent, by Theorem 8 $P(A \,\&\, B) = P(A|B)P(B)$. If B is not consistent, then $(A \,\&\, B)$ is not consistent. By Theorem 6, $P(A \,\&\, B) = P(A|B)P(B) = 0$.

As corollaries of Theorem 9 we get

Theorem 10 *If* $P(B) > 0$, *then* $P(A|B) = P(A \& B)/P(B)$.

Theorem 11 *If* $P(B) = 0$, *then* $P(A \& B) = 0$.

Applying Theorem 9 to Theorem 3, we get

Theorem 12 $P(A) = P(A \& B) + P(A \& {\sim}B)$.

Since $P(\mathbf{T}) = 1$ by *A2*, substitution of **T** into Theorem 10 yields

Theorem 13 $P(A) = P(A|\mathbf{T})$ *if* **T** *is a logical truth.*

To show that the unconditional probability axioms are satisfied, we must show (1)-(4). (1) trivially follows from *A1* and Theorem 13, (2) is a part of *A2*, and (4) is Theorem 10. To show (3), we first need to prove

Theorem 14 $P(B) + P({\sim}B) = 1$.

Theorem 14 follows directly from Theorem 7 and Theorem 13.

Consider, then, that $P(A \vee B) = P({\sim}({\sim}A \& {\sim}B))$, which by Theorem 14 is $1 - P({\sim}A \& {\sim}B)$. By Theorem 14, $1 = P(A) + P({\sim}A)$. By Theorem 12 applied to ${\sim}A$, we get $1 = P(A) + P({\sim}A \& B) + P({\sim}A \& {\sim}B)$. Hence $1 - P({\sim}A \& {\sim}B)$, which is $P(A \vee B)$, equals $P(A) + P({\sim}A \& B)$. By Theorem 12 applied to B, we get $P({\sim}A \& B) = P(B) - P(A \& B)$. Substituting, we get

Theorem 15 $P(A \vee B) = P(A) + P(B) - P(A \& B)$.

Applying Theorem 6 to Theorem 15 we immediately obtain (3):

Theorem 16 $P(A \vee B) = P(A) + P(B)$ *if* $(A \& B)$ *is inconsistent.*

As noted earlier, the axioms of **KU** entail those of **KC**. So, trivially since **KK** entails **KU**, **KK** entails **KC**. The proofs are all easy, so we here omit them. They center on Theorems 10 and 13, and involve simple algebra.

The final theorems we want to show for **KK** are the five theses in section **4**, qualified for consistent factors. *T3* is Theorem 7, and *T5* is just Theorem 5. To show *T4*, suppose *A* is consistent. Using Theorem 7 and *A3*, the same proof for Theorem 11 can be utilized to show

Theorem 17 *If A is consistent and* $(B \& E)$ *is inconsistent, then*

$$P((B \vee E)|A) = P(B|A) + P(E|A).$$

T2 is easily shown to follow from Theorems 5 and 7. Suppose *A* is consistent and *A* entails *B*. Then *A* entails ${\sim\sim}B$. By Theorem 5 then, $P({\sim}B|A) = 0$. By Theorem 7 we have $P(B|A) = 1$. Hence

Theorem 18 *If A is consistent and A entails B, then* $P(B|A) = 1$.

As a special case of Theorem 18 we have *T1*:

Theorem 19 *If A is consistent, then* $P(A|A) = 1$.

This completes my development of system **KK**, which is a simple, powerful, and yet intuitive extension of Kolmogorov probability theory. However, if we either are willing to go beyond theses *T1* through *T5*, for consistent factors, or else are not concerned with extending Kolmogorov probability theory, interesting variants of **KK** are possible. By strengthening *A3*, for example, by simply dropping the qualification that A be consistent, we immediately extend **KK** to accept universally *T5*: all probabilities conditional on a contradiction will have to be 0. This augmented system is simpler and stronger than **KK**, but goes beyond intuitions in my opinion.

If we keep the dissatisfaction with the way conditional and unconditional probabilities are linked, weaker axiomatizations are possible which are still extensions of the standard theory. One can substitute special cases and consequences of *A3* and *A4* for them, but the resultant systems are not simpler. For example, in system **L**, which universally accepts *T5*, the following are provable: part of (*), the usual links, and Kolmogorov probability theory.

L

L1 $\mathsf{P}(A) \geqslant 0$.
L2 $1 = \mathsf{P}(A) + \mathsf{P}(\sim A)$.
L3 *If* A *entails* $\sim B$, *then* $\mathsf{P}(B|A) = 0$.
L4 *If* A *entails* B, *then* $\mathsf{P}((E \,\&\, B)|A) = \mathsf{P}(E|A)$.
L5 $\mathsf{P}(A) = \mathsf{P}(A|B)\mathsf{P}(B) + \mathsf{P}(A|\sim B)\mathsf{P}(\sim B)$.

One can easily show that the theorems needed to prove standard probability theory for **KK** hold for **L**. The key to the weakenings is that in **L** one can very simply and directly prove $\mathsf{P}(A \,\&\, B) = \mathsf{P}(A|B)\mathsf{P}(B)$ and $\mathsf{P}(C) = 0$ if C is a contradiction. Substituting $(A \,\&\, B)$ for A in *L5* and applying *L3* and *L4* immediately yields the first assertion. Since everything entails the negation of a contradiction C, substituting C for A in *L5* and applying *L3* immediately yields the second assertion. The remainder of the Theorems through 16 follow routinely for **L** as for **KK**, showing that **L** too is an extension of Kolmogorov probability theory.

Finally, how does **KK** compare in power to the axiomatizations of conditional probability that accept theses *T1* through *T5* exactly for consistent conditionalizations? The answer is that **KK** is usually slightly weaker. For example, consider the following axiom system, which is a slight modification of Carnap [1], p. 38.

R

R1 $\mathsf{P}(B|A) \geqslant 0$.
R2 *If* A *is consistent, then* $\mathsf{P}(A|A) = 1$.
R3 *If* A *is consistent, then* $\mathsf{P}(B|A) + \mathsf{P}(\sim B|A) = 1$.
R4 *If* $(A \,\&\, E)$ *is consistent, then* $\mathsf{P}((E \,\&\, B)|A) = \mathsf{P}(E|A)\mathsf{P}(B|(A \,\&\, E))$.
R5 $\mathsf{P}(A) = \mathsf{P}(A|\mathbf{T})$ *where* $\mathbf{T}$ *is a logical truth.*

It is not too difficult to show that **R** entails **KK**, and that **KK** entails all of **R** but *R4*. And *R4* (the assertion we wish to avoid as axiomatic because the ratio definition is a special case of it) is nearly provable. If $P(A \,\&\, E) > 0$, then $P(A) > 0$, and *C4* is easily proved using Theorem 10. And if $P(E|A) = 0$, then $P((E \,\&\, B)|A) = 0$, making both sides 0. Thus we can show

Theorem 20 *If* $P(A \,\&\, E) > 0$ or $P(E|A) = 0$, then $P((E \,\&\, B|A) = P(E|A) \times P(B|(A \,\&\, E))$.

Note that Theorem 20 covers all cases except where $P(A) = 0$. For if $P(A) \neq 0$ and $P(A \,\&\, E) = 0$, since $P(A \,\&\, E) = P(E|A)P(A)$, we must have $P(E|A) = 0$. Note also that putting in a logical truth for *A* in Theorem 20 yields the ratio definition. How **KK** could be extended to be equivalent to **R**, and whether or not it should be so extended, are questions I do not currently have answers for.

NOTES

1. By definition I do not mean that there is in the theory an assertion labeled "definition". I mean the purely formal point that one probability function is introduced into the theory via an assertion which has the form of a definition, whether it is called a definition or axiom or whatever (*cf.* Suppes [11], chapter 8).

2. In saying that these theses are inconsistent, I assume that P is a function. Reichenbach [6] universally accepts all five theses, but can do so consistently only because he takes as primitive not a two-place function but a three-place relation–roughly, "*B* is probable to degree *r* where *A* is true." Then the notation "$P(B|A) = r$" is appropriate only where *r* is unique. In the one case where *A* is inconsistent, *B* is held to be probable to degree *r* on *A*, for *all* real numbers *r* [6], p. 6. Then *T1* shows merely that *B* is probable to degree 1 on a contradictory *C*, and *T5* that *B* is probable to degree 0 on *C*, and *T4* that the sum of two degrees of probability on *C* is a degree of probability on *C*. Reichenbach's move is certainly a permissible and eloquent way to preserve the universality of these theses. However, accepting universally what he calls the univocality of probability [6], p. 55–that for *A* and *B* there is one and only one thing which is the degree that *A* is probable on *B*–other assertions must be restricted to at least consistent conditionalizations. Note that probabilistic entailment can still be a generalization of logical entailment if *T2* is universally accepted.

REFERENCES

[1] Carnap, R., "A basic system of inductive logic, Part I," pp. 33-165 in *Studies in Inductive Logic and Probability*, vol. I, R. Carnap and R. C. Jeffrey (eds.), University of California Press (1971).

[2] de Finetti, B., *Probability, Induction, and Statistics*, John Wiley & Sons, London (1972).

[3] Jeffreys, H., *Theory of Probability*, third edition, Oxford (1961).

[4] Kolmogorov, A., *Foundations of the Theory of Probability*, Chelsea, New York (1956).

[5] Popper, K. R., *The Logic of Scientific Discovery*, Harper & Row, New York (1965). New appendices *iv and *v.

[6] Reichenbach, H., *The Theory of Probability*, University of California (1949).

[7] Renyi, A., *Foundations of Probability*, Holden-Day, San Francisco (1970).

[8] Renyi, A., *Probability Theory*, North-Holland, Amsterdam (1970).

[9] Salmon, W. C., *The Foundations of Scientific Inference*, University of Pittsburgh, Pittsburgh (1967).

[10] Shoenfield, J. R., *Mathematical Logic*, Addison-Wesley, New York (1967).

[11] Suppes, P., *Introduction to Logic*, Van Nostrand (1957).

[12] von Wright, G., *A Treatise on Induction and Probability*, Littlefield, Adams, & Co., New Jersey (1960).

Lawrence University
Appleton, Wisconsin