

To appear in the *International Journal of Intelligent Systems*
Date of this draft: June 2004

PROBABLE EQUIVALENCE, SUPERPOWER SETS, AND SUPERCONDITIONALS

Bart Kosko
Electrical Engineering Department
University of Southern California
Los Angeles, California 90089-2564

“There is no set corresponding to the inclusion relation between sets.”

Patrick Suppes
Axiomatic Set Theory

Abstract

A natural measure of probabilistic equality between sets leads to two measures of probabilistic conditioning that form the endpoints of a conditioning interval. The interval's lower bound is the standard conditional probability or “subconditional” that describes the probability of a subset relation. The upper bound is a new “superconditional” that describes the probability of the corresponding superset relation. These dual conditioning operators correspond to dual set collections and enjoy optimality relations with respect to these set collections. Fuzzy cubes illustrate these set-collection relations in the two-dimensional case. The subconditional operator corresponds to the usual “power set” of a given set. The dual superconditional operator corresponds to what we call the “superpower set” or the set of all supersets of the given set. The two dual conditioning operators can eliminate each other through simple equalities. They obey dual Bayes theorems but differ in how they respond to statistical independence.

1. Introduction: Similarity Drives the Second-Order Uncertainty of Conditioning

This paper presents the new probabilistic conditioning interval $(P(H | E), Q(H | E))$ from the perspective of similarity or partial equivalence: To what extent does uncertain evidence E resemble an uncertain hypothesis H ? The new interval and its components depend on a natural measure or index $P(E = H)$ of how similar or equivalent E is to H . The next section proposes the ratio definition of uncertain equivalence $P(H = E) = \frac{P(H \cap E)}{P(H \cup E)}$ and explores its conceptual difficulties and its mathematical consequences. The equivalence index $P(E = H)$ gives back both standard conditional probability $P(H | E)$ and the new conditioning operator $Q(H | E)$ as special cases of the partial equivalence that holds between some set events. This introductory section illustrates how such similarity affects standard or first-order conditioning and how it drives the second-order uncertainty of the conditioning process. The simple numerical example below shows that the probabilistic equality $P(E = H)$ and the conditioning interval $(P(H | E), Q(H | E))$ can change to reflect changes in the uncertainty relationship between the evidence E and the hypothesis H while the standard conditional probability $P(H | E)$ alone can ignore these changes.

The length of the interval $(P(H | E), Q(H | E))$ gives a rough measure of the *second-order* uncertainty of the conditioning process. The lower bound $P(H | E)$

describes only first-order uncertainty: It gives the user just one number that describes how the given hypothesis H conditions on the given piece of evidence E . But $P(H | E)$ itself gives the user no information whatsoever about the confidence of any one of these two endpoint numbers because it is just one of these numbers. The new conditioning operator $Q(H | E)$ does give the user such information if the user applies Q jointly with P . The operator Q by itself gives no second-order information for the same reason that P does not. So Q does not offer an alternative to P . It also lacks some of P 's measure-theoretic properties even though it is a formal dual to P . But P and Q together offer both a set of duality relations and the new conditioning interval $(P(H | E), Q(H | E))$ that extend the framework of probabilistic conditioning from first-order uncertainty descriptions to second-order descriptions.

Consider first the components of the conditioning interval $(P(H | E), Q(H | E))$. Let the set H denote a hypothesis with positive prior probability $P(H) > 0$. Let the set E in the same sample space denote some evidence with probability $P(E) > 0$. Then the lower-bound operator $P(H | E)$ is the usual conditional probability $P(H | E) = \frac{P(H \cap E)}{P(E)}$. But the interval's upper bound Q is a new conditioning operator (not itself a probability measure) that we call a *superconditional*. It has the dual ratio form

$Q(H | E) = \frac{P(H)}{P(H \cup E)}$. The next section shows that both P and Q follow as special cases

from the probabilistic similarity or equality measure $P(H = E) = \frac{P(H \cap E)}{P(H \cup E)}$. Each

operator can in turn eliminate the other through the dual "whole-in-the-part" identities

$P(H | E) = Q(H \cap E | E)$ and $Q(H | E) = P(H | H \cup E)$. Theorems 2 and 3 reveal the

dual nature of P and Q based on their respective relationship to collections of event subsets and supersets. A unit square or two-dimensional fuzzy square gives an effective way to visualize (and generalize) these theorems and related propositions.

The “interval theorem” below states that $P(H | E) \leq Q(H | E)$ and so justifies the interval form $(P(H | E), Q(H | E))$. This suggests using the conditioning gap

$Q(H | E) - P(H | E)$ as a rough measure of the second-order uncertainty in the interval

because our confidence in any point value in the interval decreases as the interval length increases. The upper bound Q gives a natural normalizer of this gap or difference. Then

the term $1 - \frac{P(H | E)}{Q(H | E)}$ measures the uncertainty in the interval. So its negation measures

the confidence $c(H | E)$ in the conditioning interval itself: $c(H | E) = \frac{P(H | E)}{Q(H | E)}$. There

is no confidence in the conditioning or $c(H | E) = 0$ in the extreme case when H and E

are disjoint because then $P(H | E) = 0$. There is total confidence or $c(H | E) = 1$ in the

other extreme case when $E \subset H$ or $H \subset E$ because then $P(H | E) = Q(H | E)$.

Now consider two cases of related probabilistic descriptions. The first case assigns the evidence E and the hypothesis H the probabilities $P(E) = .7$ and $P(H) = .6$.

It further assigns the joint events $E \cap H$ and $E \cup H$ the probabilities $P(E \cap H) = .4$

and $P(E \cup H) = .9$. Note that these values obey the inequality constraint

$P(E \cap H) \leq \min(P(E), P(H)) \leq \max(P(E), P(H)) \leq P(E \cup H)$. Then the conditional

probability of the hypothesis H given the evidence E is $P(H | E) = \frac{4}{7}$. The dual

superconditional of H given E is $Q(H | E) = \frac{2}{3}$. The probabilistic equality between H and

E is $P(H = E) = \frac{4}{9}$ or about 44%. This gives a conditioning interval of

$(P(H|E), Q(H|E)) = (\frac{4}{7}, \frac{2}{3})$ and a confidence value of $c(H|E) = \frac{P(H|E)}{Q(H|E)} = \frac{6}{7}$ or

about 86%.

Suppose next that the hypothesis H falls in probability from 60% to 50%. So suppose in this second case that $P(H) = .5$ and further that the joint event $E \cup H$ suffers a like fall in probability from .9 to $P(E \cup H) = .8$. The other two probability values stay the same: $P(E) = .7$ and $P(E \cap H) = .4$. These four probability values still obey the inequality constraint and further obey the equality constraint of modularity:

$P(E) + P(H) = P(E \cap H) + P(E \cup H)$. Then the probabilistic equality increases to

$P(H = E) = \frac{1}{2}$ because now the events E and H have more mass in common. The

superconditional value decreases from $\frac{2}{3}$ to $Q(H|E) = \frac{5}{8}$ or to about 63% while the

conditional probability remains *unchanged* at $P(H|E) = \frac{4}{7}$. These two conditioning

values shrink the conditioning interval to $(P(H|E), Q(H|E)) = (\frac{4}{7}, \frac{5}{8})$ and thus

increase the confidence value to $c(H|E) = \frac{P(H|E)}{Q(H|E)} = \frac{32}{35}$ or about 91%. This simple

example shows that an increase in probabilistic similarity or equality can increase the

second-order confidence of conditioning. Focusing solely on the conditional probability

$P(H|E)$ ignores this change in the probabilistic relationship between E and H .

2. Probable Equality and Conditioning

What is the probability that two events are equal?

This is a simple but problematic question. Consider the probability space (X, \mathcal{A}, P) where the set events A and B belong to the sigma-algebra \mathcal{A} on the sample space X with probability measure P . Then either $A = B$ holds or $A \neq B$ holds with binary certainty. We can assign some probability p in the first case and thus assign $1 - p$ in the second. But just what defines p and how does p relate to the measure P ? The axiom of extensionality [10],[26] states that A and B are equal if they have the same elements. But in general the measure of a lone element or a finite set of elements is zero. And the uncountable nature of most event sigma-algebras precludes defining p in terms of set frequencies.

The more immediate question is whether the probability p is trivial. The binary nature of the equality $A = B$ and the inequality $A \neq B$ suggests that p should itself be binary: either $p = 0$ or $p = 1$ but we do not know which. Then the uncertainty appears subjective.

Random set theory shows how to resolve this problem but at the price of increased abstraction. Consider the closely related problem whether a population mean m lies in a given confidence interval $Z \pm c$ for random variable Z and positive constant c . Then either $m \in (z - c, z + c)$ or $m \notin (z - c, z + c)$ for any realization z of Z . So the probability p that m belongs to any such realized interval is 1 or 0 according as m does or does not belong to the interval. Indeed this problem accounts for the use of the awkward adjective “confidence” rather than the more natural “probabilistic.” A standard solution views the confidence interval as a *random set* [9] or a measurable mapping from the

probability space to some measurable space. The random set $(Z - c, Z + c)$ leaves regular deterministic sets $(z - c, z + c)$ as realizations much as a foot leaves footprints. The mean lies in some of the realized confidence intervals but not in others and does so in each case with binary certainty. The frequency or probability structure of the set realizations depends on the probability distribution of the random variable Z and more generally on the distribution of the random set.

The same reasoning views the events A and B as random sets that leave set realizations for each random “experiment.” Some of the experiments result in equality realizations $A = B$ while the rest result in inequality realizations $A \neq B$. Then the probability p describes the frequency or occurrence probability of the equality realizations for some underlying but unknown random set and its probability distribution.

These conceptual difficulties involving the equality relation $A = B$ stem from the relation’s logical structure: It *is* a logical relation and not itself a set. This is the import of the above Suppes epigraph. The intersection $A \cap B$ and the union $A \cup B$ are sets while the set inclusion $A \subset B$ and the set equality $A = B$ are not sets but logical relations between sets. Defining the equivalence probability p as the term $P(A = B)$ does not technically make sense because the probability measure P applies only to sets. Random sets offer but one way to make formal sense of such a term.

Another closely related research problem is how to make probabilistic sense of the if-then conditional of classical binary logic [1],[21]. But formal efforts to add related abstract “conditional elements” of the form “ $B|A$ ” to *sigma-algebras* have ended in failure [17]. One version [11] of this so-called Lewis triviality theorem states that the probability of a set-theoretic “conditional” $P(A \rightarrow B)$ equals the standard conditional

probability $P(B|A)$ only if events A and B are independent under P :

$P(A \rightarrow B) = P(B|A)$ implies $P(A \cap B) = P(A)P(B)$. So the proposed probabilistic conditional does not condition at all except in a trivial sense because then $P(A \rightarrow B) = P(B)$.

One way around the Lewis triviality theorem is simply to abandon sigma-algebras and work instead with conditional-event algebras [3],[8],[18]-[20],[24] or other Boolean algebraic structures. We present a more conservative approach that works within the set-theoretic confines of sigma-algebras and thus that preserves the standard theory of finite positive measures.

We propose indices or measures of equality and conditioning based on logically equivalent set-theoretic relations. The starting point is the fact that the set relation $A = B$ is logically equivalent to other set relations. Some of these relations allow a definition of probable equality that depends only on P and on events in the sigma-algebra \mathbb{A} . Equation (1) below captures just such a relation in a definition of probable equivalence.

The proposed defining set-theoretic relations are well-known logical equivalents of the definition of set-theoretic equality. Recall that the relation $A = B$ holds if and only if $A \subset B$ and $B \subset A$. This holds iff $A \cap B = A \cup B$ --and this is the only such biconditional for set equivalence that depends just on the sets A and B and the basic dual operations of intersection and union. Other set equivalences can produce very different results. The equivalent relation $A \cap B = A \cup B$ in turn leads to a natural normalized index or measure of probable equivalence $E(A, B)$:

$$E(A, B) = P(A = B) = \frac{P(A \cap B)}{P(A \cup B)} . \quad (1)$$

The definition (1) implies that $P(A \neq B) = 1 - E(A, B) = \frac{P(A \cup B) - P(A \cap B)}{P(A \cup B)}$.

We assume throughout that $P(A) > 0$ and $P(B) > 0$.

The defining ratio of probabilities (1) has the further advantage that researchers in diverse fields have independently arrived at some form of it as a measure of equality or similarity. The ratio resembles the Tanimoto similarity measure in the theory of pattern recognition [6]. Fuzzy theorists have proposed the same ratio but with finite fuzzy cardinalities in place of probabilities [12], [16]. And mathematical psychologists have used a like ratio to measure perceptual similarity [27]. An indirect historical connection comes from the early work of logical positivist Rudolph Carnap. He grounded his observer-based logical world view or “methodological solipsism” on an undefined partial measure of equivalence or similarity coupled with symbolic logic [4].

The equality operator $E(A, B)$ in (1) balances the elements common to the two events A and B against the total elements in the two events. The operator gives at once a new way to view the probability of any event A : $P(A)$ is just the probability that event A equals the entire sample space X since $P(A = X) = P(A)$. This probability equals one of course if $P(A) = P(X)$. The ratio (1) gives $P(A = B) = 0$ in the disjoint case $A \cap B = \emptyset$. It gives $P(A = B) = 1$ when $A \cap B = A \cup B$ and thus when $A = B$. The ratio also satisfies the monotone condition that any reasonable measure of equality should satisfy [12]: $A \subset B \subset C$ implies that $E(A, C) \leq E(B, C)$ since $A \subset B$ iff $A \cap B = A$ iff $A \cup B = B$.

Now consider the related concept of probable inclusion or subsethood [13]-[16].

What is the probability that $A \subset B$? The logical relation $A \subset B$ holds iff $A \cap B = A$.

The equality $A \cap B = A$ allows (1) to measure the probability that A is a subset of B :

$$P(A \subset B) = P(A \cap B = A) = E(A, A \cap B) \quad (2)$$

$$= \frac{P(A \cap B)}{P(A)} \quad (3)$$

$$= P(B|A) \quad (4)$$

Hence probable equality gives back the familiar ratio (3) of conditional probability

$P(B|A)$. Now the relation follows as a theorem and not as the usual *ad hoc* definition.

Note that $P(A) = P(A|X)$ or $P(A) = P(X \subset A)$.

But symmetry raises a like question about probable *superset*hood: What is the probability that $B \supset A$? The dual relation $B \supset A$ holds iff $A \cup B = B$. The equality $A \cup B = B$ also allows (1) to measure the probability that B is a superset of A :

$$P(B \supset A) = P(A \cup B = B) = E(B, A \cup B) \quad (5)$$

$$= \frac{P(B)}{P(A \cup B)} \quad (6)$$

$$= Q(B|A) \quad (7)$$

Note that $P(A) = Q(A|X)$ or $P(A) = P(A \supset X)$. The ratio definition (1) and the

inequality $P(A \cap B) \leq \min(P(A), P(B))$ imply the upper bound

$P(A = B) \leq \min(Q(A|B), Q(B|A))$ just as $P(A \cup B) \geq \max(P(A), P(B))$ implies $P(A = B) \leq \min(P(A|B), P(B|A))$. This reflects that the biconditional nature of equality is a stronger condition on sets A and B than is the extent to which either set contains the other. We note that the operator $Q(B|A)$ has the same ratio form for set events as one of the six conditioning operators that Walker [28] independently proposed in his study of conditional-event algebras of propositions (but with the severe restriction that A and B be mutually exclusive).

The new conditioning operator $Q(B|A)$ is dual to the standard conditioning measure $P(B|A)$. Both the relationship of $Q(B|A)$ to supersets and the results in the next section suggest that we call $Q(B|A)$ the *superconditional* probability. This further suggests that we call $P(B|A)$ the *subconditional* probability. The later ratio uses the intersection of the two events while the former uses the union even though both conditioning operators arise from a logical equivalence between set relations. A like duality holds for the “whole in the part” eliminations $P(B|A) = Q(A \cap B|A)$ and $Q(B|A) = P(B|A \cup B)$. Theorem 1 below shows that the two operators are not equal in general. And the subconditional is a proper probability measure while the superconditional is not because $Q(B|A) + Q(B^c|A) > 1$. This occurs when $A \subset B$.

All three operators behave similarly with respect to the contrapositive biconditional that states that $A \subset B$ iff $B^c \subset A^c$. The inequality $P(B|A) \neq P(A^c|B^c)$ holds in general as do the inequalities $Q(B|A) \neq Q(A^c|B^c)$ and $E(A, B) \neq E(A^c, B^c)$. But equality does hold in all three cases in the special case when $P(A) + P(B) = 1$.

Several equalities directly connect the three operators. Either one of the two conditioning operators can eliminate the probable equivalence that gives rise to them:

$P(A \cap B | A \cup B) = E(A, B) = Q(A \cap B | A \cup B)$. The familiar *modular equality*

$$P(A) + P(B) = P(A \cap B) + P(A \cup B) \quad (8)$$

also implies that probable equivalence (1) reduces to a function of the subconditional (4):

$$E(A, B) = \frac{P(A|B) P(B|A)}{P(A|B) + P(B|A) - P(A|B) P(B|A)} . \quad (9)$$

The modular equality also reduces probable equivalence to the superconditional (7):

$$E(A, B) = Q(A|B) + Q(B|A) - 1 . \quad (10)$$

This gives a type of conservation law between subconditionals and superconditionals:

$$\frac{1}{Q(A|B) + Q(B|A) - 1} = \frac{1}{P(A|B)} + \frac{1}{P(B|A)} - 1 . \quad (11)$$

The ratio definition (1) leads to the more intuitive relation:

$$E(A, B) = P(A|B)Q(B|A) = P(B|A)Q(A|B) . \quad (12)$$

These identities reflect the conjunctive nature of probable equality. It measures how much A is a subset of B and how much B is a superset of A . The modular equality further implies that the superconditional probability bounds the subconditional probability. This gives the following “interval theorem.”

Theorem 1. $P(B|A) \leq Q(B|A)$ with equality if $A \subset B$. (13)

Proof. The deterministic inclusion $A \subset B$ gives $P(B|A) = 1 = Q(B|A)$. And (8) and the inequality $P(A) \leq P(A \cup B)$ prove the result:

$$P(B|A) = \frac{P(A) + P(B) - P(A \cup B)}{P(A)} \quad (14)$$

$$= 1 - \frac{P(A \cup B) - P(B)}{P(A)} \quad (15)$$

$$\leq 1 - \frac{P(A \cup B) - P(B)}{P(A \cup B)} \quad (16)$$

$$= Q(B|A) \quad . \quad \text{Q.E.D.} \quad (17)$$

The converse conditioning operators also equal each other if the certain inclusion

$$A \subset B \text{ holds: } P(A|B) = \frac{P(A)}{P(B)} = Q(A|B) = E(A,B). \text{ The inequality in Theorem 1}$$

justifies calling the interval $(P(B|A), Q(B|A))$ a *conditioning interval*. This interval is always a deterministic subset of the unit interval $[0, 1]$.

So defining the measure $P(A = B)$ as the probable equivalence $E(A, B)$ in (1) gives back the standard definition of conditional probability and leads to a dual conditioning operator. The last section shows that the new superconditional operator $Q(B|A)$ also obeys a Bayes theorem but differs from the subconditional operator $P(B|A)$ in how it responds to independent events. The two dual conditioning operators define upper and lower bounds on conditioning and thus give rise to a conditioning interval. The next two sections show that the two conditioning operators reflect the dual structure of their underlying set collections.

3. Dual Set Collections: Subpower and Superpower Sets

The standard term “power set” reflects the combinatorial fact that if a set X has n elements then its set of all subsets has 2^n elements (because of the binomial theorem). This has given rise to the superscripted notation 2^X to denote this set collection. The same symbol scheme applies to any subset $A \subset X$. If A has $k \leq n$ elements then its power set 2^A contains 2^k elements.

But what about all *supersets* of A in X ? This set collection contains 2^{n-k} elements because each set in it is the union of A and one of the 2^{n-k} subsets of its complement A^c .

Thus each pair A and A^c forms a type of set-theoretic basis for the space X . This is the idea behind the proof of the next two propositions:

Proposition 1. Set $A \subset X$ has 2^{n-k} supersets if X contains n elements and if A contains $k \leq n$ elements.

Proof. A^c contains $n - k$ elements because A contains k elements. So the power set of A^c contains 2^{n-k} subsets $C \subset A^c$. Now suppose $D \subset X$ is a superset of A : $A \subset D$. Then D has the form $D = A \cup C$ where C is some subset of A^c . So the power set of A^c enumerates the supersets of A . Thus A has 2^{n-k} supersets. Q.E.D.

Proposition 1 shows that the set collection of A 's supersets equally warrants the adjective "power" in its description. But what symbol denotes this set of all supersets? There seems to be no standard symbol or even name for this set collection even though mathematical concepts from convex hulls to generated sigma-algebras depend on a set's collection of supersets. Lattice theory uses symbols to denote various classes of supersets but it does not use a dedicated symbol to denote the class of all supersets [2]. Yet this set collection is the dual collection to the "power set" of subsets. The sample-space structure of probabilistic conditioning lets us sidestep the delicate issue of when such a set collection exists in general because specifying a sample space X ensures that there is a superset "ceiling" for all set events A . This in turn ensures the existence of a collection of supersets for any A . So we ignore whether there is a "set of all sets" or any other sets "above" the ceiling X [26].

We will use the term *superpower set* to refer to the collection of all supersets of a given set $A \subset X$. We defy common practice but maintain conceptual consistency and denote the superpower set with the superscripted symbol 2^A . Then symmetry requires that the term *subpower set* refer to the collection of all subsets and that the subscripted symbol 2_A denote this set collection.

The new notation simplifies the statement of many dual relations. The standard subset relation of contraposition that $A \subset B$ iff $B^c \subset A^c$ now has the form $A \in 2_B$ iff $A^c \in 2^{B^c}$. This underlies the set-theoretic isomorphism between subpower set 2_A and the complement superpower set 2^{A^c} :

Proposition 2. $2_A \sim 2^{A^c}$. (18)

Proof. Consider the map $f: 2_A \rightarrow 2^{A^c}$ where $f(A') = A^c \cup A'$ for some subset $A' \subset A$.

The result follows if f is a bijection. We first show that f is one-to-one. Suppose

$f(A') = f(A'')$ for any two subsets $A' \in 2_A$ and $A'' \in 2_A$. Then $A^c \cup A' = A^c \cup A''$

by the definition of f . So $A \cap (A^c \cup A') = A \cap (A^c \cup A'')$ iff $A \cap A' = A \cap A''$ iff

$A' = A''$ since A' and A'' are subsets of A . So f is one-to-one. We next show that f is

onto. Pick any $D \in 2^{A^c}$. So $A^c \subset D$. Then $D = A^c \cup A'$ for some subset $A' \in 2_A$. So

$D = f(A')$ by the definition of f . A like argument also shows that the dual map

$g: 2^{A^c} \rightarrow 2_A$ is a bijection if $g(A') = A \cap A'$ for superset $A' \in 2^{A^c}$. Q.E.D.

The new notation states that now the null set's superpower set 2^\emptyset is the same as the old "power set" of X : $2^\emptyset = 2_X$. The subpower set 2_\emptyset contains only the null set \emptyset while the superpower set 2^X contains only the whole space X . Note also that now the dual to $\bigcap_{C \in 2_A} C = \emptyset$ is $\bigcup_{C \in 2^A} C = X$ and that the reflexive relation $A \subset A$ implies $\bigcup_{C \in 2_A} C = \bigcap_{C \in 2^A} C = A$. And note that the relation $2^k + 2^{n-k} < 2^n$ holds for $n \geq 3$. The product relation $2^k 2^{n-k} = 2^n$ reflects the fact that the global subpower set 2_X is isomorphic to the product set $2_A \times 2^A$ for any set A .

The reflexive relation $A \subset A$ leads at once to a characterization of any set $A \subset X$ in terms of its subpower and superpower sets:

Proposition 3. $A = 2_A \cap 2^A$. (19)

Proposition 3 has a simple geometry even in the case of $n = 2$ —if we view its generalization to multivalued or "fuzzy" sets. Standard binary sets admit no easy visualization in low (or high) dimensions. Let $a: X \rightarrow \{0,1\}$ be the indicator function of binary set A : $a(x) = 1$ if $x \in A$ and $a(x) = 0$ if $x \notin A$. Hence the 2^n sets in the subpower set 2_X define the 2^n vertices of the Boolean n -cube: $2_X = \{0,1\}^n$. The Boolean n -cube is in turn a subset of the n -dimensional unit hypercube: $\{0,1\}^n \subset [0,1]^n$. The latter relation shows how finite fuzzy set theory formally subsumes binary set theory: A is a vague or fuzzy subset of X iff A has a multivalued indicator function $a: X \rightarrow [0,1]$ [29]. Then we

can view fuzzy set A as a point in the “fuzzy cube” $[0,1]^n$ [14]-[16]:

$A = (a_1, \dots, a_n) = (a(x_1), \dots, a(x_n)) \in [0,1]^n$. This gives geometric content to the fact that $A \subset B$ iff $a(x) \leq b(x)$ for all $x \in X$ because this biconditional holds for multivalued as well binary indicator functions. So both all fuzzy subsets of A and all fuzzy supersets of A define a hyper-rectangle in $[0,1]^n$ or I^n . Proposition 3 states that these two hyper-rectangles touch at exactly the point $A \in [0,1]^n$. Figure 1 shows these two touching rectangles in the fuzzy 2-square for fuzzy set $A = (\frac{1}{3}, \frac{3}{4})$ in which element x_1 belongs to A only partially (but deterministically) to degree $\frac{1}{3}$ and element x_2 belongs to A to the higher degree $\frac{3}{4}$. We denote the fuzzy subpower set of A as I_A and denote the fuzzy superpower set of A as I^A where $I = [0, 1]$. Then the reflexive fact that $A \subset A$ now gives the far more general result that $A = I_A \cap I^A$. The inclusion $A \subset A$ still holds for any fuzzy set because $a(x) \leq a(x)$ holds trivially for all x and for all possible set functions.

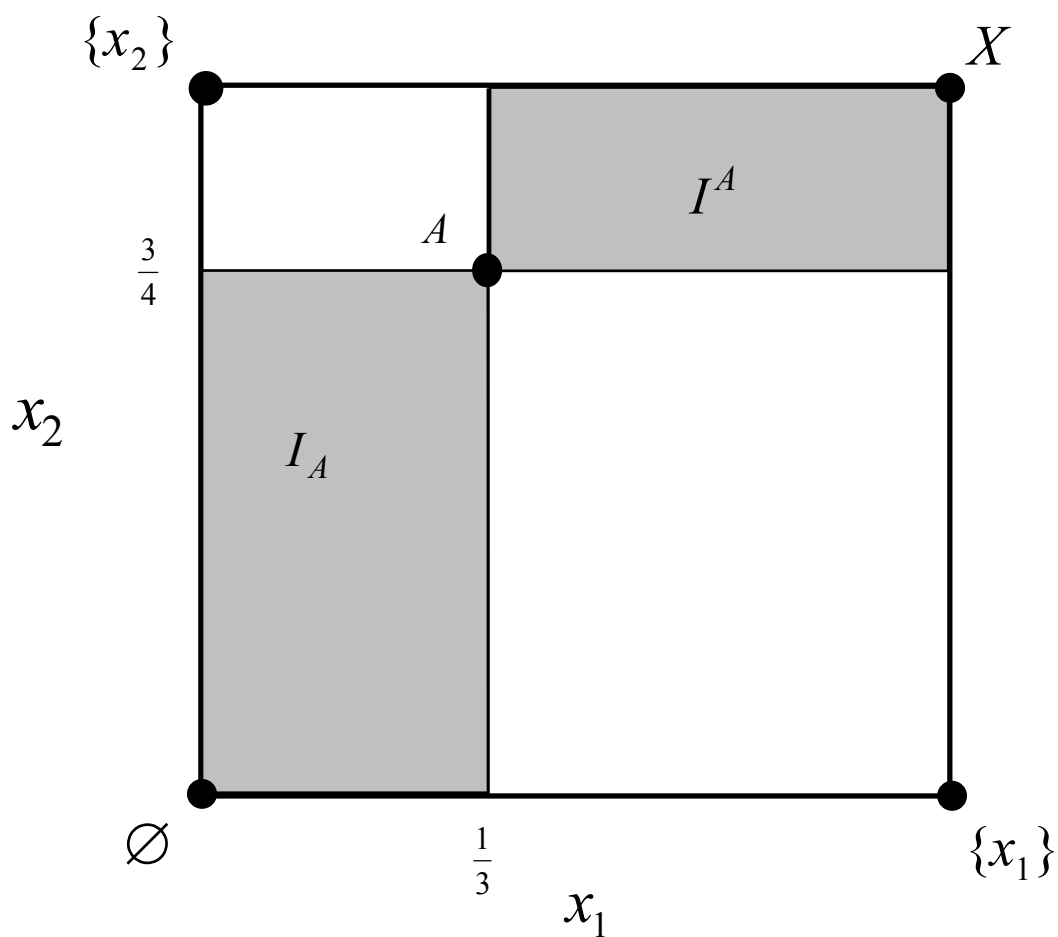


Figure 1: Geometric content of $A = I_A \cap I^A$ if A is a fuzzy subset of the binary set $X = (x_1, x_2) = (1, 1)$. Fuzzy set $A = (\frac{1}{3}, \frac{3}{4})$ has fuzzy subpower set I_A and fuzzy superpower set I^A . The simpler result in Proposition 3 that $A = 2_A \cap 2^A$ for binary sets A describes set behavior only at the cube vertices.

A like unit-square figure shows the geometric content of Proposition 2 for fuzzy sets: Rotating the square by 180 degrees swaps the subpower-set rectangle I_A with the complement superpower-set rectangle I^{A^c} . This leads to a natural strengthening of Proposition 2 for *all* fuzzy or multivalued sets:

Proposition 4. $I_A \sim I^{A^c}$. (20)

Proof. $D \in I_A$ iff $d(x) \leq a(x) \quad \forall x$ (21)

iff $1 - a(x) \leq 1 - d(x) \quad \forall x$ (22)

iff $A^c \subset D^c$ (23)

iff $D^c \in I^{A^c}$. Q.E.D. (24)

Figure 2 shows this symmetry between the subpower set of A and the superpower set of A^c as the congruence of two solid rectangles.

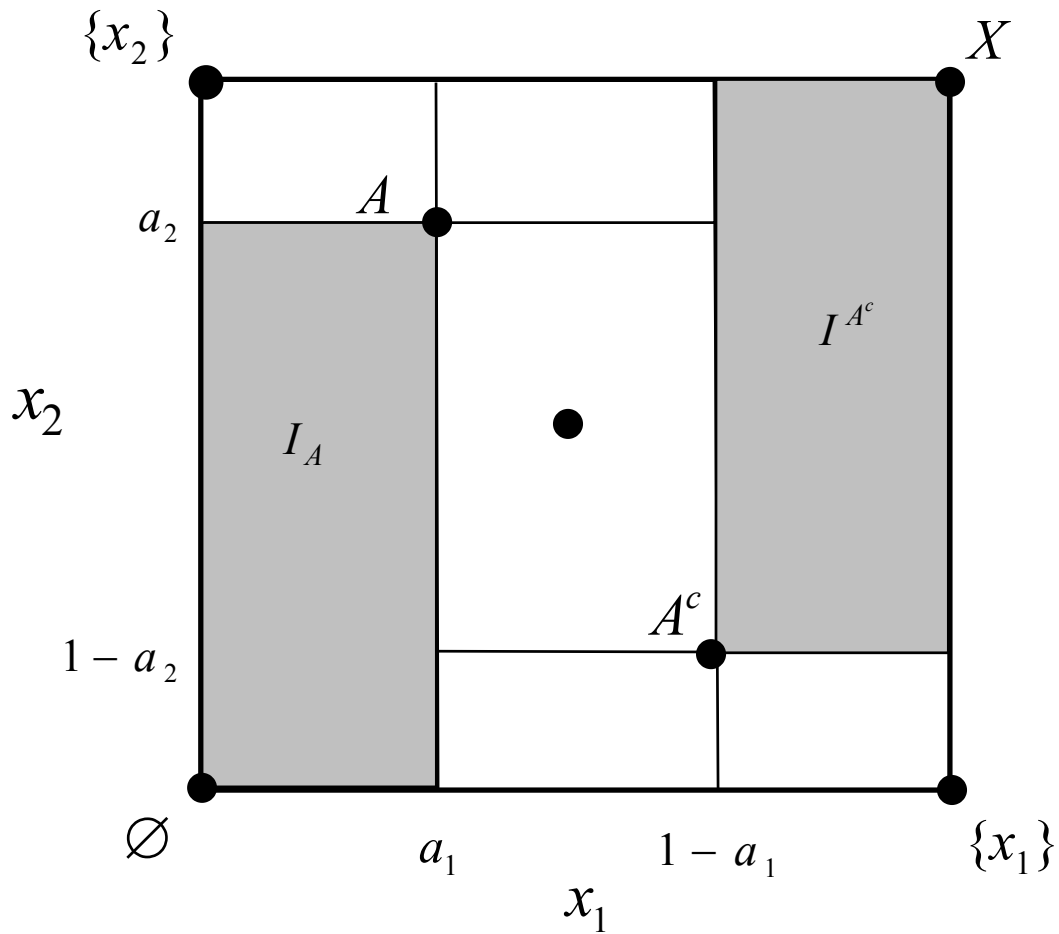


Figure 2: Complement-based symmetry of the subpower set and superpower set in Proposition 4. The fuzzy subpower set I_A of A is isomorphic to the fuzzy superpower set I^{A^c} of A^c for any set A .

4. Optimality of Subconditionals and Superconditionals

The next two theorems are duality theorems. They show that the subconditional $P(A|B)$ bears an optimal relationship to the subpower set 2_A and that the superconditional $Q(B|A)$ bears a dual optimal relationship to the superpower set 2^A . The first theorem states these optimality relations in terms of probable equality. The second theorem states them in terms of the pseudo-metric $d(A, B) = P(A \cup B) - P(A \cap B)$.

Theorem 2 shows how an arbitrary measurable set B stands in relation to the subpower set 2_A and the superpower set 2^A of some other arbitrary measurable set A . Part (a) says that the subconditional probability $P(A|B)$ maximizes $P(C|B)$ over the subpower set 2_A because no subset $C \subset A$ is more like (or is more likely to equal) the arbitrary set B than is the intersection subset $A \cap B$. Part (b) likewise says that the superconditional probability $Q(B|A)$ maximizes $Q(B|C)$ over the superpower set 2^A because no superset $C \supset A$ is more like (or is more likely to equal) the arbitrary set B than is the union superset $A \cup B$.

Theorem 2. (a) $P(A|B) = E(B, A \cap B) \geq P(C|B) \geq E(B, C)$ for all $C \in 2_A$ (25)

(b) $Q(B|A) = E(B, A \cup B) \geq Q(B|C) \geq E(B, C)$ for all $C \in 2^A$. (26)

Proof. (a) Pick $C \in 2_A$ and thus $C \subset A$. Then $P(C) \leq P(A)$ and $B \cap C \subset A \cap B$ for any (measurable) set B . Then $P(A \cap B) \geq P(B \cap C)$ leads to

$$E(B, A \cap B) = P(A|B) = \frac{P(A \cap B)}{P(B)} \geq \frac{P(B \cap C)}{P(B)} \geq \frac{P(B \cap C)}{P(B \cup C)} = E(B, C) \quad (27)$$

since $B \subset B \cup C$.

(b) Pick $C \in 2^A$ and thus $A \subset C$. Then $A \cup B \subset B \cup C$ for any (measurable) set B . So $P(A \cup B) \leq P(B \cup C)$ and this gives

$$E(B, A \cup B) = Q(B|A) = \frac{P(B)}{P(A \cup B)} \geq \frac{P(B)}{P(B \cup C)} \geq \frac{P(B \cap C)}{P(B \cup C)} = E(B, C) \quad (28)$$

since $B \cap C \subset B$. Q.E.D.

The next theorem looks more closely at the metrical structure of the subconditional and superconditional operators relative to the respective subpower and superpower set collections. This requires applying a distance measure to these set collections. So define the usual pseudo-metric d on the probabilistic symmetric difference:

$$d(A, B) = P(A \Delta B) \quad (29)$$

$$= P(A \cup B) - P(A \cap B) \quad (30)$$

The two-place set operation d is a pseudo-metric rather than a metric because the zero case $d(A, B) = 0$ does not imply that $A = B$ whereas a metric would [23]. But the pseudo-metric does define a proper metric on the equivalence classes on the subsets of X

if we identify sets A and B when $d(A, B) = 0$. The pseudo-metric also obeys the triangle inequality:

$$d(A, B) \leq d(A, C) + d(C, B) \quad \text{for any set } C. \quad (31)$$

Figure 3 shows that equality holds in (31) for the optimality conditions in Theorem 3.

A simple argument motivates the use of the pseudo-metric d for probabilistic comparisons among sets in subpower and superpower sets. Suppose $C \in 2_{A \cap B}$ and thus $C \subset A \cap B$. Then $P(C) \leq P(A \cap B)$ by the monotonicity of P . Then

$$d(B, A \cap B) = P(B) - P(A \cap B) \quad (32)$$

$$\leq P(B) - P(C) \quad (33)$$

$$= P(B \cup C) - P(B \cap C) \quad (34)$$

$$= d(B, C) \quad (35)$$

because $C \subset A \cap B$ implies that $B = B \cup C$ and $C = B \cap C$ for any such C . So B is closer to the intersection $A \cap B$ in the subpower set $2_{A \cap B} \subset 2_A$ than to any other subset C in the subpowerset $2_{A \cap B}$. A symmetric argument shows that $d(B, A \cup B) \leq d(B, C)$ and thus that the same B is closer to the union $A \cup B$ in the superpower set $2^{A \cup B} \subset 2^A$ than to any other superset C in the superpower set $2^{A \cup B}$. Theorem 3 extends these dual results to the entire subpower set 2_A and to the entire superpower set 2^A .

The first pseudo-metrical part of Theorem 3 below extends the result $d(B, A \cap B) \leq d(B, C)$ from the subcollection $2_{A \cap B}$ to the entire subpower set 2_A . Thus it shows the subpower-set optimal status that the subconditional $P(A|B)$ enjoys because of the defining intersection term $A \cap B$. The second pseudo-metrical part of Theorem 3 extends $d(B, A \cup B) \leq d(B, C)$ from the subcollection $2^{A \cup B}$ to the entire superpower set 2^A . Thus it shows the dual superpower-set optimal status that the superconditional $Q(B|A)$ enjoys because of its dual defining union term $A \cup B$.

We first present two propositions that reveal the Pythagorean-type structure of the pseudo-metrical relations involving the subpower sets and superpower sets. Proposition 5 states two dual Pythagorean-like identities that follow from the modular equality. The first relates an arbitrary set B to the subpower set 2_A . The second relates B to the superpower set 2^A .

Proposition 5. (a) $d(A, B) = d(A, A \cap B) + d(A \cap B, B)$ (36)

(b) $d(A, B) = d(A, A \cup B) + d(A \cup B, B)$. (37)

Proof. $d(A, B) = P(A \cup B) - P(A \cap B)$ (38)

$= P(A) + P(B) - P(A \cap B) - P(A \cap B)$ (39)

$= [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)]$ (40)

$$= d(A, A \cap B) + d(A \cap B, B) \quad \text{in (a)} \quad (41)$$

$$= P(A \cup B) - P(A) - P(B) + P(A \cap B) \quad (42)$$

$$= [P(A \cup B) - P(A)] + [P(A \cup B) - P(B)] \quad (43)$$

$$= d(A, A \cup B) + d(A \cup B, B) \quad \text{in (b). Q.E.D.} \quad (44)$$

The next proposition extends Proposition 5 to arbitrary sets in the subpower set 2_A and not just the intersection $A \cap B$ on its “border.” It likewise extends Proposition 5(b) to sets in the superpower set 2^A and not just the union $A \cup B$ on its “border.” These two sets in fact lie on the borders of the respective convex and compact fuzzy subpower set I_A and fuzzy superpower set I^A as Figure 3 shows.

Propositon 6:

$$(a) \quad d(B, B \cap C) = d(B, A \cap B) + d(A \cap B, B \cap C) \quad \text{if } C \in 2_A. \quad (45)$$

$$(b) \quad d(B, B \cup C) = d(B, A \cup B) + d(A \cup B, B \cup C) \quad \text{if } C \in 2^A. \quad (46)$$

Proof. (a) Suppose $C \in 2_A$. Then $C \subset A$ and so $C = A \cap C$. Then

$B \cap C = A \cap B \cap C$. So

$$d(A \cap B, B \cap C) = d(A \cap B, A \cap B \cap C) \quad (47)$$

$$= P(A \cap B) - P(A \cap B \cap C) \quad (48)$$

because $A \cap B \cap C \subset A \cap B$. This implies that

$$d(B, A \cap B) + d(A \cap B, B \cap C) = P(B) - P(A \cap B) + P(A \cap B) - P(A \cap B \cap C) \quad (49)$$

$$= P(B) - P(A \cap B \cap C) \quad (50)$$

$$= d(B, A \cap B \cap C) \quad (51)$$

$$= d(B, B \cap C) \quad (52)$$

because $B \cap C = A \cap B \cap C$. This proves part (a).

(b) The proof is dual to the proof of part (a): Suppose $C \in 2^A$. Then $A \subset C$ and so $C = A \cup C$. Then $B \cup C = A \cup B \cup C$. So

$$d(A \cup B, B \cup C) = d(A \cup B, A \cup B \cup C) \quad (53)$$

$$= P(A \cup B \cup C) - P(A \cup B) \quad (54)$$

because $A \cup B \subset A \cup B \cup C$. This implies that

$$d(B, A \cup B) + d(A \cup B, B \cup C) = P(A \cup B) - P(B) + P(A \cup B \cup C) - P(A \cup B) \quad (55)$$

$$= P(A \cup B \cup C) - P(B) \quad (56)$$

$$= d(B, A \cup B \cup C) \quad (57)$$

$$= d(B, B \cup C) \quad (58)$$

because $B \cup C = A \cup B \cup C$. Q.E.D.

Theorem 3 reflects the straight-line geometry of the two-dimensional fuzzy cube in Figure 3. The defining intersection term $A \cap B$ of the subconditional $P(A|B)$ is closer to B than is any other subset $C \subset A$ for any arbitrary set B that is not a subset of A . The dual defining union term $A \cup B$ of the superconditional $Q(B|A)$ is closer to B than is any other superset $C \supset A$. Note also that part (a) of Theorem 3 is trivial if B is a subset of A just as part (b) is trivial if B is a superset of A . The proof of Theorem 3 follows the same dual lines as the does the proof of Theorem 2.

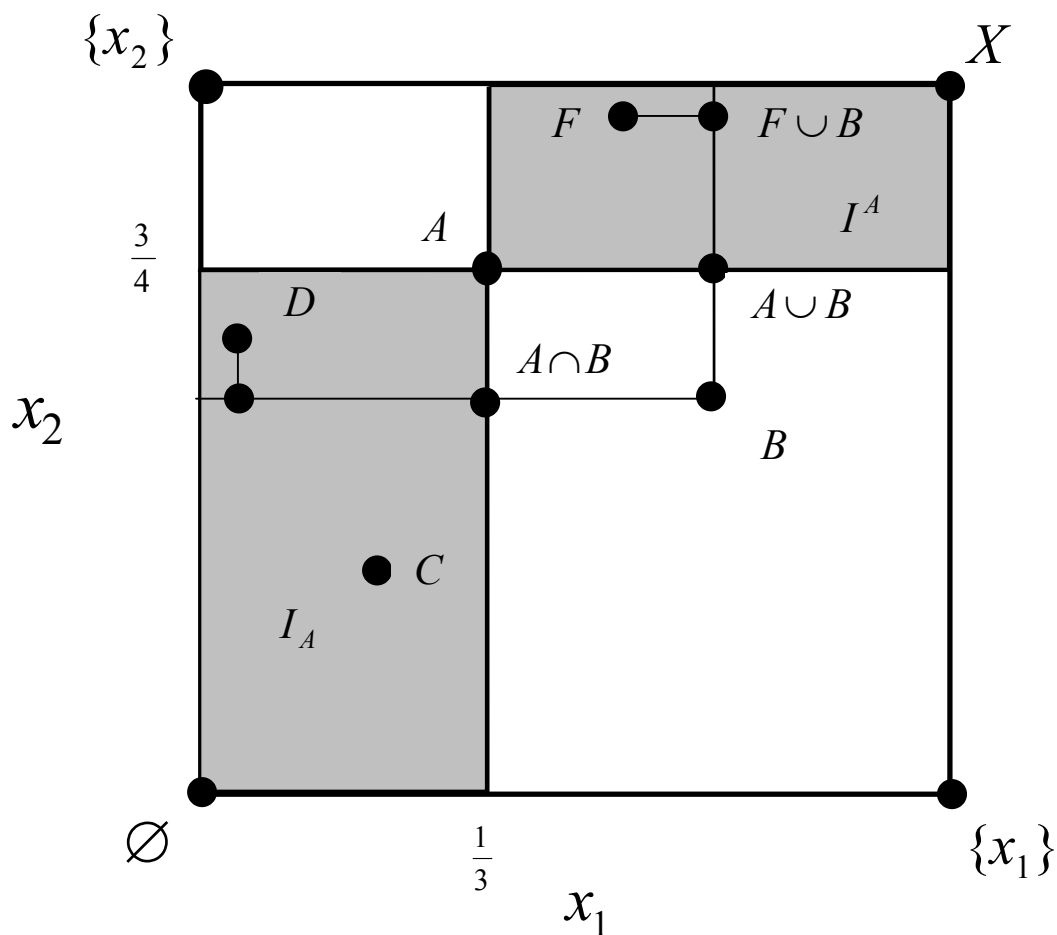


Figure 3: Geometric content of Theorem 3—metrical structure of how the subconditional operator $P(B|A)$ relates to the fuzzy subpower set I_A and how the superconditional operator $Q(B|A)$ relates to the fuzzy superpower set I^A . Suppose fuzzy set B is not a proper subset of fuzzy set A and thus $B \notin I_A$. Suppose fuzzy set C is a subset of A and thus $C \in I_A$. Then the defining intersection term $A \cap B$ of the subconditional operator $P(A|B)$ is closer to B than is any other subset C or D . A dual relation holds if B is not a proper superset of A and thus if $B \notin I^A$. Then the defining

union term $A \cup B$ of the superconditional operator $Q(B|A)$ is closer to B than is any other superset C or D .

Theorem 3. (a) $d(B, A \cap B) \leq d(B, C)$ for any subset $C \in 2_A$. (59)

(b) $d(B, A \cup B) \leq d(B, C)$ for any superset $C \in 2^A$. (60)

Proof. (a) Suppose first that C is a subset of A : $C \in 2_A$ or $C \subset A$. Then

$B \cap C \subset A \cap B$ and so $P(B \cap C) \leq P(A \cap B)$. Then

$$d(B, A \cap B) = P(B) - P(A \cap B) \quad (61)$$

$$\leq P(B) - P(B \cap C) \quad (62)$$

since $P(B \cap C) \leq P(A \cap B)$

$$\leq P(B \cup C) - P(B \cap C) \quad (63)$$

since $B \subset B \cup C$

$$= d(B, C). \quad (62)$$

(b) Suppose next that C is a superset of A : $C \in 2^A$ or $A \subset C$. Then

$A \cup B \subset B \cup C$ and so $P(A \cup B) \leq P(B \cup C)$. Then

$$d(B, A \cup B) = P(A \cup B) - P(B) \quad (63)$$

$$\leq P(B \cup C) - P(B) \quad (64)$$

since $P(A \cup B) \leq P(B \cup C)$

$$\leq P(B \cup C) - P(B \cap C) \quad (65)$$

since $B \cap C \subset B$

$$= d(B, C) \quad . \quad \text{Q.E.D.} \quad (66)$$

A like result holds pointwise in the more general finite fuzzy-set case for any l^p metric based on summed fit values or membership degrees. The compact and convex structure of fuzzy hypercubes and their hyper-rectangles leads to the result

$$l^p(B, I_A) = \min l^p(B, C) = l^p(B, A \cap B) \quad (67)$$

where the minimum is over all subsets $C \subset A$. This optimality result reflects the l^p version of the n -dimensional Pythagorean theorem that holds on finite fuzzy cubes and that holds pointwise in more general fuzzy spaces [14]-[16]. The dual fuzzy result is

$$l^p(B, I^A) = \min l^p(B, C) = l^p(B, A \cup B) \quad (68)$$

where the minimum is over all supersets $C \supset A$.

The pseudo-metric d suggests an alternative way to define probable equivalence—simply as the distance between two sets:

$P(A \neq B) = d(A, B) = P(A \cup B) - P(A \cap B)$. This differs from (1) by the normalization term $P(A \cup B)$. It is also the set-level analogue of the equality measure that Gaines [7] proposed at the statement level for probability logics. But this nonnormalized equality measure fails to give back the standard conditional probability measure

$P(B | A) = \frac{P(A \cap B)}{P(A)}$ because of the modular equality:

$$P(B = A \cup B) = 1 - P(A \cup B) + P(B) \quad (69)$$

$$= 1 - P(A) + P(A \cap B) \quad (70)$$

$$= 1 - P(A \neq A \cap B) \quad (71)$$

$$= P(A = A \cap B) \quad (72)$$

So the nonnormalized equality measure equates the subconditional with the superconditional. But then $P(A = A \cap B) \neq P(B | A)$ because the term $P(A \neq A \cap B)$ lacks the normalizer $P(A)$. The normalized equality measure (1) avoids this problem but at the “expense” of creating the inequality condition $P(B|A) \leq Q(B|A)$ of Theorem 1.

5. A Bayes Conditioning Interval and the Problem of Independence

Bayes Theorem lies at the heart of the theory of classical conditional probability. It gives a simple way to find the conditional probability $P(B | A)$ as a scaled version of the converse conditional probability $P(A | B)$. This in turn endows probabilistic learning with a powerful way to find the conditional probability of a set-based hypothesis H given set-based evidence E in terms of its converse-- $P(H | E)$ in terms of $P(E | H)$. Theorem 4 shows that the superconditional operator $Q(B | A)$ enjoys the same converse relation and with the *same* update or “learning” factor (the ratio of the prior probability of the hypothesis to the probability of the evidence).

Theorem 4. (a)
$$P(H | E) = P(E | H) \left[\frac{P(H)}{P(E)} \right] \quad (73)$$

(b)
$$Q(H | E) = Q(E | H) \left[\frac{P(H)}{P(E)} \right] . \quad (74)$$

Proof. Part (a) follows from the identity $P(B | A) = \frac{P(A \cap B)}{P(A)}$ either as the standard definition of conditional probability or as the subconditional result in (2). The dual result in part (b) follows likewise from the superconditional ratio in (6) because

$$Q(B | A) = \frac{P(B)}{P(A \cup B)} \frac{P(A)}{P(A)} \quad (75)$$

$$= \frac{P(A)}{P(A \cup B)} \left[\frac{P(B)}{P(A)} \right] \quad (76)$$

$$= Q(A|B) \left[\frac{P(B)}{P(A)} \right] . \text{ Q.E.D.} \quad (77)$$

The so-called theorem on total probabilities of elementary probability theory expands the denominator term $P(E)$ in the Bayes ratios of Theorem 4 into

$P(E) = \sum_k P(H_k)P(E|H_k)$ if a denumerable collection of pairwise-disjoint and

exhaustive hypotheses $\{H_k\}$ partitions the sample space X : if $X = \bigcup_k H_k$ and

$H_k \cap H_j = \emptyset$ if $k \neq j$. Then Theorem 4 and the whole-in-part eliminations

$P(B|A) = Q(A \cap B|A)$ and $Q(B|A) = P(B|A \cup B)$ imply an immediate but complex-looking corollary:

Corollary. $Q(H_j|E) = \frac{P(H_j)Q(E|H_j)}{\sum_k P(H_k)Q(H_k \cap E|H_k)}$ if $\{H_k\}$ partitions X . (78)

Two conceptual points follow at once from the dual Bayes Theorems in Theorem 4. The first is that Bayes Theorem is not unique to the standard conditional probability $P(B|A)$. The same Bayes properties of this subconditional operator apply to the superconditional operator and perhaps to other probabilistic operators. The corollary does show that they differ in how they expand the evidence probability over a partition of alternatives.

The second point is again that probabilistic conditioning on evidence E produces a conditioning *interval* $(P(H|E), Q(H|E))$. The standard conditional probability $P(B|A)$ is merely the lower bound of this conditioning interval. The functional form of Bayes Theorem itself does not select this lower bound as the unique conduit of conditional information. The interval structure suggests instead that if there is a “real” or “true” conditional probability (or probability of a conditional) then this quantity lies between the endpoint extremes. Or the interval structure could mean that there is no unique point estimate of conditioning in general. There is only the uncertainty interval itself.

The interval structure of $(P(H|E), Q(H|E))$ also resembles the lower and upper probabilities of Dempster [5] and the closely related belief and plausibility bounds of Shafer’s probabilistic theory of evidence [25]. The belief of event B sums a nonnegative mass function over all subsets A of B while the more generous plausibility sums it over all A that have nonempty intersection with B . But neither this belief nor plausibility sum resembles the formal structure of the subconditional $P(B|A)$ or the superconditional $Q(B|A)$.

A closer match lies in the related upper and lower uncertainty bounds of Pawlak's rough set theory [22]. This uncertainty theory uses the subset bound B_* and the superset bound B^* in $B_* \subset B \subset B^*$. The subset bound B_* is the union of all subsets $A \subset B$ in the set algebra. The superset bound B^* is the intersection of all supersets $B \subset A$ in the algebra. Two sets B and C are equivalent if $B_* = C_*$ and $B^* = C^*$. Then rough sets are the equivalence classes of sets that arise from this equivalence relation. But this relation still assumes that the set equivalence is certain. And the inclusion $B_* \subset B \subset B^*$ relates sets whereas $P(B | A)$ and $Q(B | A)$ involve only ratios of probabilities of sets.

The conditioning interval further suggests that the conditioning gap or difference $Q(H | E) - P(H | E)$ gives a rough measure of confidence in the conditioning process. Confidence in the conditioning accuracy should fall as this gap grows. The inequality $P(B | A) \leq Q(B | A)$ of Theorem 1 gives a normalized confidence measure c :

$$c(H | E) = \frac{P(H | E)}{Q(H | E)} . \quad (79)$$

Conditioning confidence should grow directly with $c(H | E)$. The ratio gives maximal confidence in the rare equality case when $P(H | E) = Q(H | E)$. The ratio also gives veto power to the standard conditional probability $P(H | E)$ when it equals zero. This holds approximately when the prior hypothesis $P(H)$ is nearly zero or when the evidence E is nearly the set complement of the hypothesis H .

There is no dual result for the concept of statistical independence. Sets A and B are independent iff the joint probability factors into a product of marginals--

$P(A \cap B) = P(A)P(B)$. This occurs iff a more intuitive conditioning property holds:

$P(B | A) = P(B)$. A related conditional characterization of statistical independence is

that A and B are independent iff $P(B | A) = P(B | A^c)$. Note that $c(B | A) = P(A \cup B)$ if

the events A and B are independent. This reflects how the superconditional $Q(B | A)$

responds to factoring the joint probability into a product of marginals.

The superconditional $Q(B | A)$ has its own independence-like conditioning relation: $Q(B | A) = P(B)$ iff $P(A \cup B) = 1$. This result is sure to occur if A is the complement of event B : $Q(B | B^c) = P(B)$.

But the independence outcome $Q(B | B^c) = P(B)$ presents a conceptual problem. How can B occur with nonzero probability when its complement B^c occurs? How can this happen when $P(B | B^c) = 0$ under the same conditions? Independence compounds the problem because $Q(B | B^c) = P(B)$ has the form of an independence relation when the set B functionally depends on B^c through their indicator functions. One response to this problem is simply that Q is not a probability measure. So this counter-intuitive behavior may just be a non-measure artifact.

A more subtle response is that superconditionals may require a new intuition about superset behavior. That intuition rests on the simple identity $P(B) + P(B^c) = 1$:

Why can't B occur with some probability if B^c occurs only with some probability? This

differs from the certain occurrence of the two sets and the resulting non-contradiction impossibility $B \cap B^c = \emptyset$ or excluded-middle necessity $B \cup B^c = X$.

The two certain outcomes $Q(B | B^c) = 0$ and $Q(B | B^c) = 1$ arguably make sense as limiting cases of the probability magnitude $P(B)$ because they occur respectively when $P(B) = 0$ and when $P(B) = 1$. The certain outcome $Q(B | B^c) = 0$ occurs iff B is a measure-zero event and thus B is sure not to occur. So its complement B^c is a sure event and is sure to occur. We don't expect to observe an event if we expect to observe its opposite. This intuition varies directly with the occurrence probability $P(B)$ of event B . The other extreme outcome $Q(B | B^c) = 1$ corresponds to the symmetric intuition: We do expect to observe an event if we don't expect to observe its opposite.

This argument for the superconditional result $Q(B | B^c) = P(B)$ finds the subconditional outcome $P(B | B^c) = 0$ problematic. Should the conditional probability $P(B | B^c)$ be *completely* insensitive to the magnitude of the occurrence probability $P(B)$? That insensitivity appears to be more an artifact of using intersection in the definition of $P(B | A) = \frac{P(A \cap B)}{P(A)}$ than the result of prior intuitions about probabilistic conditioning.

Note further that $P(B | B^c) = 0$ does not hold in general for vague events because

$P(A \cap A^c) > 0$ holds in the general case of the probability of a measurable fuzzy event A where $P(A) = E[a]$ for measurable set function a [30].

A third response may be the most practical. The outcome $Q(B | B^c) = P(B)$ occurs iff $c(B | B^c) = 0$. So the veto power of the subconditional $P(B | A)$ in the confidence measure (79) renders the issue moot.

Acknowledgements

The author thanks the reviewers as well as Vladik Kreinovich, Rod Taber, Fred Watkins, and Ronald Yager for helpful comments.

References

- [1] Adams, E. W., "On the Logic of High Probability," *Journal of Philosophical Logic*, vol. 15, 255 – 279, 1986.
- [2] Birkhoff, G., *Lattice Theory*, Providence, RI: American Mathematical Society, Providence, RI, 1967.
- [3] Calabrese, P. G., "A Theory of Conditional Information with Applications," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 12, 1676 – 1684, December 1994.
- [4] Carnap, R., *The Logical Structure of the World*, University of California Press, 1967.
- [5] Dempster, A., "Upper and Lower Probabilities Induced by a Multivalued Mapping," *Annals of Statistics*, vol. 38, 325 – 341, 1967.
- [6] Duda, R. O., Hart, P. E., and Stork, D. G., *Pattern Classification*, second edition, John Wiley & Sons, 2001.
- [7] Gaines, B. R., "Fuzzy and Probability Uncertainty Logics," *Information and Control*, vol. 38, no. 2, 154 – 169, 1978.
- [8] Goodman, I. R., "Toward a Comprehensive Theory of Linguistic and Probabilistic Evidence: Two New Approaches to Conditional Event Algebra," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 12, 1685 – 1689, December 1994.

- [9] Goutsias, J., Mahler, R. P. S., and Nguyen, H. T., editors, *Random Sets: Theory and Applications*, Springer-Verlag, 1997.
- [10] Halmos, P. R., *Naïve Set Theory*, Springer-Verlag, 1974.
- [11] Jeffrey, R., *Formal Logic: Its Scope and Limits*, second edition, McGraw-Hill, 1981.
- [12] Klir, G. J., and Yuan, B., *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, 1995.
- [13] Kosko, B., “Fuzzy Entropy and Conditioning,” *Information Sciences*, vol. 40, no. 2, 165 – 174, 1986.
- [14] Kosko, B., “Fuzziness vs. Probability,” *International Journal of General Systems*, vol. 17, no. 2, 211 – 240, 1990.
- [15] Kosko, B., *Neural Networks and Fuzzy Systems*, Prentice Hall, 1991.
- [16] Kosko, B., *Fuzzy Engineering*, Prentice Hall, 1996.
- [17] Lewis, D., “Probabilities of Conditionals and Conditional Probabilities,” *The Philosophical Review*, vol. 85, no. 3, 297 – 315, July 1976.
- [18] McGee, V., “Conditional Probabilities and Compounds of Conditionals,” *The Philosophical Review*, vol. 98, no. 4, 485 – 541, 1989.
- [19] Milne, P., “Bruno de Finetti and the Logic of Conditional Events,” *British Journal of Philosophy of Science*, vol. 48, 195 – 232, 1997.
- [20] Nguyen, H. T., “Some Mathematical Structures for Computational Information,” *Information Sciences*, vol. 128, 67 – 89, 2000.
- [21] Nilsson, N. J., “Probabilistic Logic,” *Artificial Intelligence*, vol. 28, 71 – 87, 1986.
- [22] Pawlak, Z., *Rough Sets: Theoretical Aspects of Reasoning About Data*, Kluwer, 1991.
- [23] Rudin, W., *Principles of Mathematical Analysis*, third edition, McGraw-Hill, 1976.
- [24] Schay, G., “An Algebra of Conditional Events,” *Journal of Mathematical Analysis and Applications*, vol. 24, 334 – 344, 1968.
- [25] Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

- [26] Suppes, P., *Axiomatic Set Theory*, Dover, 1972.
- [27] Tversky, A., “Features of Similarity,” *Psychological Review*, vol. 84, no. 4, 327 – 352, July 1977.
- [28] Walker, E. A., “Stone Algebras, Conditional Events, and Three-Valued Logic,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 12, 1699 – 1707, December 1994.
- [29] Zadeh, L. A., “Fuzzy Sets,” *Information and Control*, vol. 8, no. 3, 338 – 353, 1965.
- [30] Zadeh, L. A., “Probability Measures of Fuzzy Events,” *Journal of Mathematical Analysis and Applications*, vol. 23, no. 2, 421 – 427, 1968.