

Modeling Uncertainty, Context, and Information Fusion via Lattice-Based Probability

Jun Zhang and Roman Ilin

1 Introduction

This chapter reviews and explores mathematical foundations for probabilistic inference, uncertainty representation, and fusion of disparate information sources. We will revisit probability measures defined on an event space that is modeled as a bounded distributive lattice—this includes as a special case Boolean lattice where each element has unique complementation and upon which standard probability theory has been axiomatized. Following the recent work of Narens (2009, 2011), we will invoke the relative pseudo-complementation operator on a distributive lattice, leading to Heyting algebra (as an extension of Boolean algebra) of event space that supports intuitionistic logic. We then consider basic probability assignment (b.p.a.) on finite distributive lattices, which are linked to lower probability (belief function) and upper probability (plausibility function) on such lattices. Making use of the fact that any topology on a set, that is, a system of subsets satisfying some requirements, forms a distributive lattice, pseudo-complementation can be addressed through the closure operation under such topology prescribed to an event space. Topology provides a rich semantics in terms of both the way subsets are categorized (open, closed, clopen) and the operations that characterize their properties (neighborhood, separation, etc.) and transformation (closure, interior, boundary, etc.). We therefore model contextual information for uncertainty as specifying a topology on the event space. The totality of all topologies (i.e., all contexts) themselves on an event space form

Draft: March 1, 2015.

J. Zhang (✉)

Department of Psychology and Department of Mathematics, University of Michigan,
Ann Arbor, MI, USA
e-mail: junz@umich.edu

R. Ilin

AFRL Sensors Directorate, Wright-Patterson Air Force Base, OH, USA
e-mail: roman.ilin.1@us.af.mil

a bounded and hence complete lattice, ordered by coarse-grading, with the discrete topology (where each elementary event is treated as “clopen”) as the top element, i.e., the finest/largest topology, and the indiscrete topology (consisting of only two events, the null-set and the full set) as the bottom element, i.e., the coarsest/smallest topology). This provides a setting for combining different b.p.a.’s, whose focal (i.e., with non-zero weight) assignments are stipulated to be only on open sets of a topology. Our lattice probability approach, identifying topology with context, deepens the upper/lower probability framework for dealing with uncertainty in two aspects: it provides a principled way for (i) defining “focal elements” (on open sets of the topology) while restraining b.p.a.’s in a given context to satisfy the lattice probability condition, and (ii) combining b.p.a.’s across different contexts through the lattice of topologies. Hence our framework provides a more fundamental mathematical framework compared with current theories (e.g., Dempster-Shafer belief function and Zadeh (1965) fuzzy probability).

1.1 Upper-Lower Probability Theory

A now-popular approach to uncertainty is through upper-lower probability theory, in which probabilistic assessment are given within an interval, meant to reflect tolerance to uncertainty. Starting from a basic probability assignment function $m()$, which assigns non-negative probability mass to (potentially all) subsets of the sample space Ω . The total probability is still required to be normalized to 1.0, but the assignment is not restricted just to its atomic elements (singleton subsets). The lower probability P_* (belief) and upper probability P^* (plausibility) are then defined as

$$P^*(A) = \sum_{X \cap A \neq \emptyset} m(X), \tag{1}$$

$$P_*(A) = \sum_{X \subseteq A} m(X), \tag{2}$$

with $0 \leq P_* \leq P^* \leq 1$. It can be shown that the lower probability P_* becomes a probability measure (and hence equals upper probability P^*) if and only if the basic probability assignment $m()$ is atomic (i.e., only to singletons). This is the case of Bayesian belief function. In general, a belief function is merely monotonic and does not satisfy the additive axiom of a probability measure (see below for more details). The belief function and the basic probability assignment are dual to each other, linked through the so-called Möbius transform:

$$m(A) = \sum_{X \subseteq A} (-1)^{|A|-|X|} P_*(X) \tag{3}$$

$$P_*(A) = \sum_{X \subseteq A} m(X). \tag{4}$$

Dempster-Shafer theory (see Yager and Liu 2008) provided a rule for evidence combination (more fashionably called “information fusion”), as well as a formula for conditioning. Though it has been extensively investigated in the past, its application to uncertainty reasoning in real systems has been limited due to (i) the need to hand-craft the basic probability assignment which is application-dependent; (ii) the lack of efficient computational algorithms to handle combinatoric explosion in the number of variables when computing belief functions defined on a power set.

The upper-lower probability theory provided an interval (with upper and lower bounds) representation of probability measure. This opens the door for representing ignorance and incomplete information. Researchers in Dempster-Shafer theory have been focused on basic probability assignment, and the evidence combination rules. They have rarely, if ever, invoked the theory of submodular functions (and Lovasz extension), which has been well developed in mathematics and recently begun to see wider applications in combinatoric optimization, machine learning, etc. It should be noted that in the past, submodular functions (variously called, capacity, Choquet integral, etc.) have been applied in economics and decision science, for instance, the so-called rank-dependent utility theory in particular. Through the technique of Lovasz extension, combinatoric optimization problems (in discrete variables) can be bypassed through applying convex programming to continuous variables in a vector space. This computation advance opens the door of applying upper-lower probability theory to uncertain reasoning and integrating disparate information in real systems.

The theory of belief functions has an alternative, equivalent approach owed to Fagin and Halpern (1991) who invoked inner and outer probability measures to deal with uncertainty. A non-measurable event, to which no probability measures can be assigned, is meant to be one that the agent does not have sufficient information to assign probability. Non-measurable events nevertheless are provided with an inner (outer) probability measure, which is the probability of the largest (smallest) measurable event contained in (containing) it and hence gives the lower (upper) bound of the degree of belief. The interval assigned by inner and outer measures thus characterizes the degree of uncertainty, akin to the interval provided by the belief-plausibility dichotomy. In fact, belief function and inner probability measure are equivalent. A further theoretical grounding of the upper-lower probability is the idea of rough sets (Pawlak 1982), who formally introduced the ideas of upper and lower sets based on a prescribed equivalence relationship on the set. However, this will not be pursued in the current chapter.

1.2 Non-Boolean Algebra with Pseudo-complementation

Standard probability theory is built upon Boolean algebra of an event space. Recall that given a ground set Ω , a probability measure $\Pr()$ is a function from the power-set $2^\Omega \rightarrow [0, 1]$ that satisfies normalization condition:

$$\Pr(\emptyset) = 0, \quad \Pr(\Omega) = 1,$$

monotonicity condition

$$\Pr(A) \leq \Pr(B) \text{ if } A \subseteq B,$$

and additivity condition

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B),$$

where A, B are any subsets of Ω . Traditionally, a probability measure is based on Boolean algebra over the event space, where an “event” is modeled as any subset of Ω . A collection of subsets (of a set) forms an algebra (of sets) if the unions and intersections of any finite members of the collection remain in that collection (in mathematical jargon, one says that the set-operations of union and intersection are “closed”). When the “union” and “intersection” operations in an algebra are replaced by the “sup” and “inf” operations with respect to the set-containment ordering, it becomes a lattice of sets. In a Boolean algebra, the collection is also required to include the (set-theoretic) complement (negation \neg) of each subset it contains. This leads to the Law of Excluded Middle, namely, the event (subset) A and its negation A^c are not only mutually exclusive but complementary (in the sense that there is no third alternative). In intuitionistic logic, however, the Law of Excluded Middle is not enforced; this is done through relaxing the “unique complementation” requirement on the collection, but instead introducing a pseudo-complementation operation, defined in a way that the output is unique if it exists with respect to any collection. Shifting focus from unique complementation to pseudo-complementation turns the Boolean algebra into a non-Boolean one, thus providing a more general setting for studying propositional logic and for handling probabilistic inference. (There are other possible relaxations to Boolean algebra and Boolean logic, including quantum logic, which will not be discussed in this chapter.)

The study of non-Boolean algebra is closely associated with *lattice theory* the foundation of which, though traceable back to George Boole, was laid by Dedekind in a series of papers in the early 1900s. Lattice encodes algebraic behavior of entailment relation (“if-then” implications) and basic logical connectives (conjunction “and” and disjunction “or”), so it provides the appropriate framework of semantics of inference. However, it was not until 1930–40s when Birkhoff, von Neumann, Stone, Tarski, etc. fully brought out the power of lattice theory with algebraic rigor. McKinsey and Tarski (1944, 1946), in two ground-breaking papers, connected topology with modal logic. They linked the topological properties of the collection of open sets to the pseudo-complementation operation on distributive lattices of sets. By doing so, they showed that topological space provided rich semantics for intuitionistic logic. Their theory motivated modern extension of the so-called Heyting lattice for contemporary modal logic, which will not be further discussed here. Instead, we investigate the structure of probability measures, including belief functions (submodular functions) on distributive lattices and their implications for novel applications to information fusion and uncertainty management.

1.3 Why Lattice?

Lattice is an algebraic object that can be defined in two equivalent ways, (i) as a set with a non-strict partial order defined on it, and the set is closed with respect to the inf and sup operations induced from such non-strict partial order; (ii) as a set with two algebraic operations (“meet” and “join”) defined on it satisfying basic axioms like commutativity and associativity, while the two operations must also be compatible by satisfying an “absorption” relation. The set is required to be closed with respect to the meet and join operations. The details will be reviewed below—here we emphasize the fact that a lattice has, *simultaneously*, order and algebraic structures. As an example, a collection of subsets of a set, under suitable constraints about the collection, can form a lattice (of sets), which behaves somewhat similarly (but with important differences) to the power-set; it provides the “right” amount of relaxation of Boolean algebra on the power-set. Two important classes of lattices, the distributive lattice and the (non-distributive) orthomodular lattice, turn out to be the mathematical tools in service of intuitionistic logic and quantum logic, respectively; they extend Boolean logic in different directions.

Recall that a *Boolean lattice (algebra)* $(B, \vee, \wedge, \neg, 0, 1)$ is a special kind of lattice with two binary operations join (\vee) and meet (\wedge) operations that distributes over one another, a bottom element 0 and a top element 1 such that $a \vee 0 = a$, $a \wedge 1 = a$ for all $a \in B$, and a unary operation \neg “complementation” with unique output such that $a \vee \neg a = 1$ and $a \wedge \neg a = 0$, for all $a \in B$. A typical example of Boolean algebra is the lattice of power-set of a set, ordered by set inclusion; here, \vee and \wedge are set-theoretic \cup and \cap , respectively. Boolean lattice is where classic probability theory with classic propositional logic is anchored upon. In fact, Stone (1936) proved an important result of Boolean algebra: a lattice is Boolean if and only if it is isomorphic to a field of sets.

One relaxation of Boolean lattices is the so-called “distributive lattice”, namely, a lattice in which the operators \wedge and \vee still preserve their roles, but without the \neg operation nor upper and/or lower bounds. An example of distributive lattice is the so-called *Brouwer lattice* $(B, \vee, \wedge, \prime, 0)$, which is bounded from below (the 0 element) and admits an additional unary operator \prime called pseudo-complementation. More generally, the *Heyting algebra* $(H, \vee, \wedge, \rightarrow)$ is endowed with an additional binary operator \rightarrow , so-called relative pseudo-complementation, defined as follows: the relative pseudo-complement x of a with respect to b , denoted as $a \rightarrow b$, is the largest element x that meets a to b : i.e., $x \leq (a \rightarrow b)$ iff $(a \wedge x) \leq b$. The Brouwerian pseudo-complement \prime is a special case with $a' \equiv a \rightarrow 0$, which satisfies $a \wedge a' = 0$. Yet $a = (a')'$ does not hold in general; nor does $a \vee a' = 1$ hold. So in a Brouwer lattice, \prime stands in place of the complement \neg operation as in Boolean algebra.

Pseudo-complementation operator may sound unnatural, but it embodies the “intuitionistic logic”, which suspends the Law of Excluded Middle; it can be traced back to Brouwer’s philosophy of mathematical foundation. It is a satisfying conclusion that every finite distributive lattice admits a relative pseudo-complementation operator. So distributive lattice provides a concise extension to Boolean lattice when

one relaxes the Law of Excluded Middle. In the 1930s, Kolmogorov used Heyting algebra as a logic for describing mathematical constructions, while Gödel employed them as a basis for modal logics that are useful for understanding proof theory in mathematical logic. A recent interest of intuitionistic logic appeared in cognitive psychology, where a variant was employed as a basis for propositions that are neither verifiable nor refutable, and formed a basis for formulating a concept “incompleteness” or “ambiguity” that people presumably take into account in making probability judgments (Narens 2009, 2011). This chapter will follow this important move to provide an alternative to Bayesian probability theory based on distributive lattices.

2 Mathematical Background

2.1 An Introduction to Lattice Theory

2.1.1 Lattice as Poset

Lattice theory is a mature topic of mathematics, so here we follow standard introduction to this subject (e.g. Birkhoff 1933; Davey and Priestley 2002). Lattice is a kind of ordered set, that is, a set with a prescribed order structure. A *partially ordered set* (poset) (X, \preceq) is a set X equipped with a binary relation \preceq such that the binary relation is (i) reflexive; (ii) transitive; and (iii) antisymmetric. Reflexivity of \preceq means that $x \preceq x$ always holds. Transitivity of \preceq means that if $x \preceq y$, $y \preceq z$ then $x \preceq z$. Antisymmetry of \preceq means that x and y must be the same element whenever $x \preceq y$ and $y \preceq x$ hold at the same time. Note that, strictly speaking, the order \preceq defined above should be called “non-strict partial order”. The reflexivity requirement makes it more like the so-called “pre-order”, which is a binary relation that only obeys (i) and (ii). On the other hand, if (i) is replaced by irreflexivity and (iii) is replaced by asymmetry, then the binary relation is called *strict partial order*, usually denoted by $<$. In this case $\neg(x < x)$ (irreflexivity) holds and that if $x < y$ then it cannot be true that $y < x$ and vice versa (asymmetry). In the lattice theory, we focus on non-strict partial order \preceq .

In a poset X , it can happen that, between two arbitrary elements x, y , neither $x \preceq y$ nor $y \preceq x$ holds—we say x, y are *incomparable*. When all elements of a poset X are pairwise comparable, i.e., either $x \preceq y$ or $y \preceq x$, then the order is total, and the set is linearly ordered.

Let S be a subset of a poset X , $S \subseteq X$. If there is an element $x \in X$ such that $s \preceq x, \forall s \in S$, then x is said to be an *upper bound* of set S . An upper bound x is called a *least upper bound* (or “supremum”) of S (denoted $x = \sup S$), if for any upper bound y of S , $x \preceq y$. Likewise, we define a *lower bound* of a set $S \subseteq X$ as any element x such that $x \preceq s, \forall s \in S$. The *greatest lower bound* (or “infimum”) of S , denoted $\inf S$, is a lower bound x of S such that $y \preceq x$ for any other lower bound y of S . Note that, because of the anti-symmetric nature of \preceq , $\sup S$, if it exists, is unique. Likewise, $\inf S$ is unique if it exists.

Taking S in the above discussions to be a binary set $\{a, b\}$, we denote $a \vee b = \sup\{a, b\}$, $a \wedge b = \inf\{a, b\}$, where \vee and \wedge are called *join* and *meet*, respectively. A *lattice* is defined as a poset (X, \preceq) in which both $\sup\{a, b\}$ and $\inf\{a, b\}$ exist for any pair a, b of elements of X . A *complete lattice* is a poset (X, \preceq) of which every non-empty subset (not just the binary subsets as in a general lattice) has an infimum (greatest lower bound) and a supremum (least upper bound). If only either \sup or \inf is required to exist, then it is called a join or meet *semi-lattice*, respectively; semi-lattices are weaker concepts than lattices, of course.

Bounded lattice, complement, and pseudo-complement. A lattice (X, \preceq) is called *bounded* if there is both a top and a bottom element, respectively denoted as 1 and 0 , that is the upper bound and lower bound for all pairs of elements of X . A bounded lattice is, thus, always complete. In a bounded lattice, for any element a , its *complement* is defined as any element b such that $a \vee b = 1$ and $a \wedge b = 0$. In general, a lattice element may have more than one complement or none—this is very different from set-theoretic complementation, where the complement always exists and is unique. For instance, in a bounded lattice with linear order (i.e., a chain), 0 and 1 are the only elements that have complements. In a bounded lattice, for any two elements a, b , we can define *relative pseudo-complement* of a with respect to b as the largest element x that satisfy $a \wedge x \preceq b$. The above-mentioned chain has relative pseudo-complement for all pairs of its elements. Of course, on an arbitrary lattice, relative pseudo-complement may not exist for all pairs of elements. A bounded lattice in which relative pseudo-complement exists for all pairs of elements is called a Heyting lattice/algebra.

Joint-prime and meet-prime elements. In a lattice, we would like to distinguish certain elements that are more “primitive” than others, in the sense that they are not “generated” by joins and meets of other elements. Let (X, \preceq) be a bounded lattice. We call an element $j \neq 0$ of X *join-prime* if $j \preceq a \vee b$ implies $j \preceq a$ or $j \preceq b$ for all $a, b \in X$. We use $J(X)$ denote the set of join-prime elements of X . Dually, we call an element $m \neq 1$ of X *meet-prime* if $a \wedge b \preceq m$ implies $a \preceq m$ or $b \preceq m$ for all $a, b \in X$. We use $M(X)$ to denote the set of meet-prime elements of X .

We call $U \subset X$ an *upset* of X if $x \in U$ and $x \preceq y$ imply $y \in U$. The set of all upsets of X is denoted $U(X)$, which forms a lattice itself, with set-containment \subseteq as the (non-strict partial) order on $U(X)$. Dually, D is called a *downset* of X if $x \in A$ and $y \preceq x$ imply $y \in D$. The set of all downsets of X is denoted $D(X)$, which forms a lattice as well, with set-containment as induced order on $D(X)$. Furthermore, the mapping $x \mapsto U(x) = \{y \in X : x \preceq y\}$, when viewed as a map from X to $U(X)$, i.e., when viewing $U(x)$ as an element $U(X)$, preserves the order relation of elements in X . The same order-isomorphic property holds for the mapping $x \mapsto D(x) = \{y \in X, y \preceq x\}$.

The importance of $U(X)$ and $D(X)$ is that they provide a “good” model of the original set X —they are order-isomorphic with respect to \preceq , the order prescribed on X and used to construct $U(X)$, $D(X)$ in the first place. The discussions in the last paragraph can be summarized as the statement:

Lemma 2.1 *Let X be a set endowed with pre-order \preceq . For each $x, y \in X$, the following three conditions are equivalent:*

- (i) $x \preceq y$;
- (ii) $U(y) \subseteq U(x)$;
- (iii) $D(x) \subseteq D(y)$.

More interestingly, while not all $x \in X$ are join-prime (or meet-prime) elements in X , $U(x)$ (or $D(x)$) is a join-prime (or meet-prime) element of $U(X)$ (or $D(X)$).

2.1.2 Lattice as Algebra

Note that \wedge and \vee are binary operations on a lattice: they map $X \times X \rightarrow X$. Both operations satisfy

- (i) associativity: $a \wedge (b \wedge c) = (a \wedge b) \wedge c$ and $a \vee (b \vee c) = (a \vee b) \vee c$;
- (ii) commutativity: $a \wedge b = b \wedge a$ and $a \vee b = b \vee a$;
- (iii) absorption: $a \wedge (a \vee b) = a = a \vee (a \wedge b)$.

A special case of absorption is idempotency: $a \wedge a = a = a \vee a$, which can be obtained by replacing, in (iii), b with $a \wedge b$ or $a \vee b$. Viewed in another way, a lattice is a set X endowed with two binary operations \wedge, \vee that satisfy (i)–(iii). In this case, letting $a \preceq b$ iff $a \wedge b = a$, or equivalently iff $a \vee b = b$, we turn a lattice as an algebra into a lattice as an ordered set. We use $L = (X, \preceq)$ to denote the lattice as a poset and $L = (X, \wedge, \vee)$ to denote the lattice as an algebra, but the reader should keep in mind this dualistic model of any lattice L .

Various forms of complementation Viewing lattice as an algebra allows the introduction of a variety of complementation operation on a lattice. Below, we investigate at least four such notions of “complementation” of a bounded lattice L , all as a unary map: $L \rightarrow L$.

- (i) (Regular) Complement (\neg): $\neg a$ is any element x in L that satisfies $a \wedge x = 0$; $a \vee x = 1$. There maybe multiple such elements.
- (ii) Orthocomplement (\perp): a special kind of complement, requiring *additionally* $(a^\perp)^\perp = a$ (involutive), and that $a \preceq b \rightarrow b^\perp \preceq a^\perp$ (order-reversing). Hence, \perp is an order-isomorphism: $a \preceq b$ if and only if $b^\perp \preceq a^\perp$.
- (iii) De Morgan complement (\sharp): a bijective mapping $\sharp : L \rightarrow L$ such that for any $a, b \in L$, $\sharp(a \vee b) = (\sharp a) \wedge (\sharp b)$, $\sharp(\sharp a) = a$, and $\sharp(1) = 0$. In other words, \sharp is “ \vee -negation”. Denote its inverse operation $(\sharp)^{-1} \equiv \flat$, the “ \wedge -negation”. Then it follows that $\flat(a \wedge b) = (\flat a) \vee (\flat b)$, and $\flat(0) = 1$.
- (iv) Pseudo-complement (\prime): weaker than regular complement, a' is the largest x (uniquely given) such that $a \wedge x = 0$ (without imposing the requirement of $a \vee x = 1$). Pseudo-complement is a special form of relative pseudo-complement, i.e., with respect to the element 0 .

Note that while \perp and the two De Morgan complements \sharp and \flat are involutive by definition (meaning that applying twice leads to identity mapping), \neg and \prime are not. De Morgan complement was introduced by Moisil (1935) and investigated by Monteiro (1980); a further requirement of $a \wedge \sharp(a) \leq b \vee \sharp(b)$ for all a, b leads to the so-called Kleene algebra. These various “complementations” affect the existence and rules of probability measure defined on the corresponding lattices.

Compared with regular complement \neg , orthocomplement \perp of a given element a selects out a special element out of many complements of a , with the additional property of being “orthogonal” to a —a binary relation \perp (not necessarily symmetric) is said to describe “ a orthogonal to b ”, $a \perp b$, iff $a \leq b^\perp$. However, there can be more than one orthocomplementation operations definable on a complemented lattice. A lattice equipped with an orthocomplement operation is called ortholattice. If an ortholattice is uniquely complemented (i.e., if \neg and hence \perp is unique), then it is a Boolean lattice (algebra).

On the other hand, pseudo-complement \prime , if it exists, is always unique. A complete distributive lattice always admits a pseudo-complement for each element. If a distributive lattice is uniquely complemented (i.e., if \neg is unique), then its complement \neg must be the same as its pseudo-complement \prime —in this case the lattice is Boolean, which models the event space underlying classic probability measure.

As Narens (2014) pointed out, there are two ways to relax/generalize Boolean algebra/lattices to non-Boolean ones admitting appropriate notions of complementation. The first generalization is through distributive lattice which, when bounded, always admits pseudo-complementation. A distributive lattice endowed with pseudo-complementation is called a Brouwerian lattice (a subclass of Heyting algebra) which provides the setting for intuitionistic logic. The second generalization is through orthomodular lattice, a special kind of ortholattice (i.e., admitting “orthocomplement” operation) upon which the so-called “modularity law” is enforced on orthocomplement pairs. Modular lattice is a relaxation to distributive lattice (since all distributive lattices are modular lattices, but not vice versa) with imposing the modularity condition as instead of the more restrictive distributivity condition on all of its pairs. Ortholattice is, in general, non-modular, so orthomodular lattice is in general non-distributive, and provides the setting to model quantum logic. A lattice that is simultaneously orthocomplemented and distributive is a Boolean algebra. So in this sense, the above two approaches, namely, intuitionistic logic and quantum logic, provide two “independent and complementary” ways of relaxing Boolean lattice/algebra. Below, we focus on the first route of generalizing Boolean lattices, through distributive lattices.

2.2 Distributive Lattice

Viewing lattice as an algebra allows further classification of lattices. First, it can be shown that in any lattice $L = (X, \wedge, \vee)$,

$$(a \wedge b) \vee (a \wedge c) \preceq a \wedge (b \vee c), \quad (5)$$

$$a \vee (b \wedge c) \preceq (a \vee b) \wedge (a \vee c). \quad (6)$$

When equality in (5) holds, then we say in L “meet distributes over join”; when equality in (6) holds, then we say in L “join distributes over meet”. It can be proven that these two equalities imply each other for any lattice, so if any one of them is satisfied, we call it *distributive lattice*. Another equivalent characterization of distributive lattice is that the following holds for any three elements a, b, c :

$$(a \wedge b) \vee (b \wedge c) \vee (c \wedge a) = (a \vee b) \wedge (b \vee c) \wedge (c \vee a).$$

Weaker than the notion of distributivity is the notion of modularity. First, in analogous to (5) and (6), the following holds in any lattices:

$$a \vee (b \wedge c) \preceq a \wedge (b \vee c), \quad \forall a, c \text{ such that } c \preceq a.$$

If the converse inequality holds in a certain lattice L , that is

$$a \wedge (b \vee c) \preceq a \vee (b \wedge c), \quad \forall a, c \text{ such that } c \preceq a,$$

then such a lattice L is called a *modular lattice*. Equivalently, if the equality

$$a \wedge (b \vee c) = a \vee (b \wedge c)$$

holds for all elements in L satisfying $c \preceq a$, then L is modular. A distributive lattice is always a modular lattice, but not vice versa (i.e., there exist non-distributive modular lattices).

Distributive lattices are also characterized by the absence of “pentagon” N_5 (two chains, with one and two elements each) and “crown” M_3 (three chains, each with one element) configurations as sublattices. As examples of distributive lattices, given any ground set X with pre-order on it, the set of all its upsets $(U(X), \cup, \cap)$ and the set of all its downsets $(D(X), \cup, \cap)$ both form distributive lattices (ordered by set-containment).

In a bounded distributive lattice, for each element a , one can define its *relative pseudo-complement* with respect to any other element b , denoted as $a \rightarrow b$ (or a^b):

$$x \preceq (a \rightarrow b) \text{ iff } (a \wedge x) \preceq b. \quad (7)$$

In a general lattice, an element a is said to be relative pseudo-complemented if $a \rightarrow b$ exists for all b . When $b = 0$, the relative pseudo-complement becomes pseudo-complement, and denoted $'$ as discussed before. A pseudo-complemented lattice that satisfies the relation $a' \vee (a')' = 1$ is called a Stone lattice. A pseudo-complemented lattice becomes a Boolean lattice iff $a \vee a' = 1, \forall a \in X$, that is, iff $a = (a)'$, $\forall a \in X$.

When a distributive lattice is finite, it is relative pseudo-complemented for all of its elements; in particular, it is pseudo-complemented. On a Boolean lattice, pseudo-complement operation is identical to regular complement operation, and relative pseudo-complementation $a \rightarrow b$ is given by $\neg a \vee b$. Huntington's Theorem says the opposite is true as well: a lattice is Boolean iff it is pseudo-complemented and that the pseudo-complementation is also a complementation. Any element of a bounded distributive lattice can have at most one complement. So if a distributive lattice is complemented, then it must be uniquely complemented, and hence Boolean.

2.2.1 Brouwer and Heyting Algebra

We can define *Brouwer complementation operation* $'$ axiomatically as a unary operator satisfying the following properties (the use of the same symbol $'$ as we used for pseudo-complementation operation defined in terms of order is intentional, see below):

- (i) $a \wedge a' = 0$;
- (ii) $(a \vee b)' = a' \wedge b'$;
- (iii) $a \leq (a)'$, or equivalently, $a = a \wedge (a)'$;
- (iv) $1' = 0$.

It can be deduced that $a' = ((a)')'$ and $0' = 1$. A lattice $L = (X, \wedge, \vee)$ with lower bound 0 and equipped with a unary operation $'$ is called a *Brouwer algebra* if it is closed with respect to the Brouwer complement) defined above. It turns out that a Brouwer algebra is necessarily a distributive lattice, and the Brouwer complementation defined above (when the lattice is viewed as a poset) is precisely the pseudo-complementation operation defined earlier when the (distributive lattice) is viewed as a poset. So Brouwer complement and pseudo-complement turns out to be equivalent, merely reflecting a difference in viewing the lattice as an algebra (former) versus a poset (latter).

De Morgan's laws under Brouwer algebra manifest as follows:

$$\begin{aligned} a' \wedge b' &= (a \vee b)', \\ a' \vee b' &\leq (a \wedge b)', \\ (a \wedge b)' &= ((a)' \wedge (b)')' = ((a' \vee b')')'. \end{aligned}$$

Hence, it is useful to introduce the binary operator \sqcup in a Brouwer algebra B :

$$a \sqcup b \equiv ((a \vee b)')$$

Consider the subset S of elements of B which satisfy $a = (a)'$. Then S is closed with respect to the operations \wedge, \sqcup , so (S, \wedge, \sqcup) forms a Boolean algebra with respect to those two operations.

Recall that relative pseudo-complementation operation is defined in (7), when a lattice is viewed $L = (X, \preceq)$. There, \wedge should be read as “greatest lower bound” using the language of ordered set, so relative pseudo-complement of a with respect to b is the largest element c such that the greatest lower bound, of this element c and a , should be less than b . When the lattice is viewed as an algebra $L = (X, \wedge, \vee)$, relative pseudo-complementation has been axiomatized by Monteiro (1980) as a binary operator satisfying:

- (i) $a \rightarrow a = b \rightarrow b$;
- (ii) $(a \rightarrow b) \wedge b = b$;
- (iii) $(a \rightarrow b) \wedge a = a \wedge b$;
- (iv) $a \rightarrow (b \wedge c) = (a \rightarrow b) \wedge (a \rightarrow c)$;
- (v) $(a \vee b) \rightarrow c = (a \rightarrow b) \wedge (a \rightarrow c)$;
- (vi) $0 \rightarrow a = 0$.

A distributive lattice $L = (X, \wedge, \vee)$ that admits the above operation can be called a Heyting algebra. The binary operation \rightarrow is, unlike \vee and \wedge , neither commutative nor associative. Brouwer complement $'$ is simply $a' \equiv a \rightarrow 0$.

2.2.2 Representation of Distributive Lattices

Representation of a lattice means to find a lattice isomorphism, typically using the lattice of sets as the target. For distributive lattices, the ring of set (closed under union and intersection operations) or field of set (closed under an additional set-complement operation) provide good candidates. It is well-known that:

- (i) A lattice is distributive iff it is isomorphic to a ring of sets (Birkhoff 1933; Stone 1936);
- (ii) A lattice is Boolean iff is isomorphic to a field of set (Stone 1936).

For sets with finite elements, the above results are intuitively understood. It is easy to envision that, in finite lattice, being Boolean is being isomorphic to the power-set of some finite set. For Boolean lattices with uncountable elements, much subtleties arise. Stone’s characterization, for instance, involved topological considerations of compactness. The same complications apply to characterizing distributive lattices. In finite case, each distributive lattice $L = (X, \preceq)$ can be represented as the lattice of upsets (downsets) of some poset; that “some poset” is the dual poset $L^d = (X, \succeq)$ of the join-prime elements of the original lattice L . (Recall that join-prime elements means that they only have a single down-link in the Hasse diagram.) For distributive lattices possibly with infinitely many elements but without infinitely descending chains, each element a is the join of join-prime elements of X underneath a , that is: $a = \bigvee \{j \in J(X) : j \preceq a\}$. Such lattices X can be represented as the sublattice of upsets $U(\mathcal{F}(X))$ of the set of prime filters $\mathcal{F}(X)$ of X . So these characterization results become very technical. Priestley in 1990s found a characterization of bounded distributive lattices in terms of Priestley spaces, or equivalently pairwise

Stone spaces. A related characterization of Heyting algebra by the so-called Esakia space is also obtained using the framework of category theory.

2.2.3 Topology as Distributive Lattice

A topological space is a pair (X, \mathcal{T}) , where X is a set and \mathcal{T} is a collection of subsets of X , called *open sets*, containing \emptyset, X and closed under finite intersections and arbitrary unions. Immediately, one sees that this definition (as Hausdorff introduced) implies that any topology is a complete, distributive lattice of sets, in which set-containment is the order and \vee and \wedge are just \cup and \cap ; \mathcal{T} is a sublattice of $P(X)$, the Boolean lattice of power-set of X . This is essentially Birkhoff (1933) characterization of distributive lattices.

Recall that a “closed set” of a topology is any set-theoretic complement of an open set of the topology. The collection \mathcal{C} of the closed sets, containing \emptyset, X and closed under arbitrary intersection and finite union, also forms a complete distributive lattice and is a sublattice of $P(X)$.

In a topological space X , *open neighborhood* of a point x in X is defined as any open set containing x . We say x is in the interior of a subset $A \subseteq X$ if there is an open neighborhood U of x that is contained in A . The set $\text{Int}(A)$ contains all interior points of A . We say that x belongs to the closure of a subset $A \subseteq X$ if each open neighborhood U of x has nonempty intersection with A . The set $\text{Cl}(A)$ denotes the closure of A . It is easy to verify that the interior operator Int satisfies:

- (i) $\text{Int}(X) = X$,
- (ii) $\text{Int}(A) \subseteq A$,
- (iii) $\text{Int}(A) = \text{Int}(\text{Int}(A))$,
- (iv) $\text{Int}(A \cap B) = \text{Int}(A) \cap \text{Int}(B)$;

and the closure operator Cl satisfies

- (i) $\text{Cl}(\emptyset) = \emptyset$,
- (ii) $A \subseteq \text{Cl}(A)$,
- (iii) $\text{Cl}(\text{Cl}(A)) = \text{Cl}(A)$,
- (iv) $\text{Cl}(A \cup B) = \text{Cl}(A) \cup \text{Cl}(B)$.

The interior and closure operators are dual to each other: $\text{Int}(A) = X - \text{Cl}(X - A)$; $\text{Cl}(A) = X - \text{Int}(X - A)$. In fact, let any operator satisfy the four conditions of Cl above and call a subset $A \subseteq X$ “closed” if $A = \text{Cl}(A)$. Then $\mathcal{T} = \{A : X - A \text{ is closed}\}$ is a topology on X , and every topology on X arises this way.

A natural question arises: Is there a connection between the pseudo-complementation operator and closure operator (or the dually defined interior operator)? The answer was affirmatively provided by McKinsey and Tarski (1944, 1946): pseudo-complementation operation in a distributive lattice and interior operation in a topology are in one-to-one correspondence: $a \rightarrow b = \text{Int}((\neg a) \vee b)$.

2.3 Probability and Belief Functions on Lattice

We next review the known facts about the feasibility of introducing probability functions or belief functions on a lattice. The mathematical tool that plays a key role is Möbius transform on sets with partial order (Rota 1964).

2.3.1 Möbius Transform and Monotone Functions

Rota (1964) considered a poset (X, \preceq) with a bottom element 0. For any function f on (X, \preceq) , the Möbius transform of f is a function $m : X \rightarrow R$ that is the solution of the equation:

$$f(x) = \sum_{y \preceq x} m(y).$$

The above equation always has a unique solution, given through the Möbius function $\mu : X \times X \rightarrow R$ by:

$$m(x) = \sum_{y \preceq x} \mu(y, x) f(y)$$

where μ is defined inductively by

$$\mu(x, y) = \begin{cases} 1, & \text{if } x = y \\ -\sum_{x \preceq t \prec y} \mu(x, t), & \text{if } x < y \\ 0, & \text{otherwise} \end{cases}.$$

Note that μ depends solely on X . One can also define the co-Möbius transform of f as:

$$g(x) = \sum_{y \succeq x} m(y).$$

A *capacity* on a set X is a function $f : 2^X \rightarrow [0, 1]$ such that (i) $f(\emptyset) = 0$, $f(X) = 1$ (normalization); and (ii) $A \subseteq B \subseteq X$ implies $f(A) \leq f(B)$ (monotonicity). Functions satisfying (ii) is called a *1-monotone* (or *strict monotone*) condition.

A function is said to be *k-monotone*¹ ($k \geq 2$) if for *any* family of k subsets A_1, \dots, A_k of X , there holds:

$$f\left(\bigcup_{i=1}^k A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1, 2, \dots, k\}} (-1)^{|I|+1} f\left(\bigcap_{i \in I} A_i\right). \quad (8)$$

¹Barthélemy (2000) used the term “weakly monotone”.

In particular, the case for a 2-monotone function f is explicitly written as (condition of “convexity” or “supermodularity”):

$$f(A_1) + f(A_2) \leq f(A_1 \cap A_2) + f(A_1 \cup A_2) \tag{9}$$

for all subsets A_1, A_2 of X . Obviously, when f is k -monotone, it is k' -monotone for all $2 \leq k' \leq k$. If f is a k -monotone function and also satisfies $f(\emptyset) = 0$ and $f(\{x\}) \geq 0$ for all $x \in X$, then f is 1-monotone—in this case, f then becomes a k -monotone capacity.

A function is said to be *totally monotone* if it is k -monotone for every $k \geq 2$. It can be proved that when $|X| = n$, total monotonicity is equivalent to $(n - 2)$ -monotonicity for a capacity function.

For $k \geq 2$, when equality in (8) holds, we say that the function f is a k -valuation. A *probability function* is both a capacity and a total valuation (i.e., k -valuation for every k). In fact, on a distributive lattice, the condition of 2-valuation (13) is sufficient for f to be a k -valuation for any k ; the proof invokes the identity $0 = (1 - 1)^n$ with binomial expansion

$$\binom{m}{0} = \sum_{k=1}^m (-1)^{k-1} \binom{m}{k},$$

which expresses the inclusion-exclusion principle.

Characterization of capacity and probability functions. The characterization of a capacity function is through its Möbius transform. The following statement is known:

Lemma 2.2 (Chateauneuf and Jaffray 1989) *A set function f is a capacity if and only if its Möbius transform m satisfies $m(\emptyset) = 0$; $\sum_{A \subseteq X} m(A) = 1$ and that for all $A \subseteq X$,*

$$\sum_{\{x\} \subseteq A \subseteq X} m(A) \geq 0$$

for all $x \in X$. In particular, $m(\{x\}) \geq 0$ for all $x \in A$.

In their paper, Chateauneuf and Jaffray (1989) also discussed probability functions that dominate a given capacity, in the context of trans shipment problem.

Shafer (1976) showed that f is a k -monotone ($k \geq 2$) capacity function if and only if its Möbius transform m satisfies

$$\sum_{C \subseteq B \subseteq X} m(B) \geq 0$$

for all subsets $C \subseteq X$ with $2 \leq |C| \leq k$. (For $k = 2$, the subsets with $|C| = k$ cannot be included.) Equivalently, the condition can be written as

$$\sum_{B \subseteq \bigcup_{i=1}^k A_i; B \not\subseteq \{A_1, \dots, A_k\}} m(B) \geq 0$$

for any $A_1, \dots, A_k \subseteq X$. In particular, 2-monotone functions are characterized by their Möbius transform satisfying

$$\sum_{B \subseteq (A_1 \cup A_2); B \not\subseteq A_1; B \not\subseteq A_2} m(B) \geq 0$$

for all subsets A_1, A_2 of X ; or by

$$\sum_{\{x_1, x_2\} \subseteq B \subseteq A} m(B) \geq 0$$

for all subset A of X and all $x_1, x_2 \in X, x_1 \neq x_2$. Shafer (1976) showed that a totally monotone capacity is equivalent to its Möbius transform m being non-negative.

If any 1-monotone function is 2-monotone on a lattice L , then L must be linearly ordered. When the inequality sign of (9) is reversed:

$$f(A_1) + f(A_2) \geq f(A_1 \cap A_2) + f(A_1 \cup A_2)$$

the function f is called a *submodular* function. Submodular functions have the following so-called “diminishing marginal return” properties:

$$f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B)$$

for all $A \subseteq B \subseteq X$ and $x \in X \setminus B$; and

$$f(A \cup \{x\}) + f(A \cup \{y\}) \geq f(A \cup \{x, y\}) + f(A)$$

for all $A \subseteq X$ and $x, y \in X \setminus A$. A 3-monotone function f satisfies:

$$f(A) + f(B) + f(C) + f(A \cap B \cap C) \leq f(A \cup B \cup C) + f(A \cap B) + f(B \cap C) + f(A \cap C).$$

Note that the concept of k -monotone function uses \cap -operation. As a counterpart, using \cup -operation instead leads to the so-called k -alternating function:

$$f\left(\bigcap_{i \in K} A_i\right) \leq \sum_{I \subseteq K, I \neq \emptyset} (-1)^{|I|+1} f\left(\bigcup_{i \in I} A_i\right). \tag{10}$$

2.3.2 Valuation on a Lattice

Note that while the results in the last section are mostly dealing with real-valued functions on the power-set, now we study real-valued functions on a lattice. Valuation of a lattice is the assignment of a real-valued function on it. A valuation function f is called *strictly monotone* or simply *monotone* when $f(a) \leq f(b)$ iff $a \leq b$.

Given any real-valued function f on $L = (X, \wedge, \vee)$, let us construct, for arbitrary $a, b \in X$,

$$\delta(a, b) \equiv f(a) + f(b) - 2f(a \wedge b).$$

It is easily seen that $\delta(a, a) = 0$; $\delta(a, b) = \delta(b, a)$. The triangular inequality

$$\delta(a, c) \leq \delta(a, b) + \delta(b, c)$$

amounts to the condition

$$f(a \wedge b) + f(b \wedge c) \leq f(b) + f(a \wedge c) \quad (11)$$

which, after taking $b = a \vee c$, leads to

$$f(a) + f(c) \leq f(a \vee c) + f(a \wedge c). \quad (12)$$

So $\delta(a, b)$ is a metric on L when f satisfies (12). Note that this is precisely the *2-monotone* condition except that f is defined on a lattice as opposed to be on the power-set as in (12).

When equality in (12) holds, that is,

$$f(a \wedge b) + f(a \vee b) = f(a) + f(b) \quad (13)$$

for $a, b \in X$, then f is called a *2-valuation*. In this case,

$$\delta(a, b) = f(a \vee b) - f(a \wedge b). \quad (14)$$

A lattice with monotone 2-valuation is called a *metric lattice*; in fact, one can show that the metric given by (14) also satisfies triangular inequality.

In analogous to 2-valuation, we call a function f on lattice L a *3-valuation* if

$$f(a) + f(b) + f(c) + f(a \wedge b \wedge c) = f(a \vee b \vee c) + f(a \wedge b) + f(b \wedge c) + f(a \wedge c)$$

for all lattice elements a, b, c . Clearly, 3-valuation implies 2-valuation, but not vice versa.

It is interesting to note that a metric lattice is always a modular lattice, which is weaker than a distributive lattice. A modular lattice of finite length is always metric.

The following is known Birkhoff (1967)—they link the properties of the lattice (modular or distributive) to the existence of strictly monotone valuations:

- (i) L is modular if and only if it admits a strictly monotone 2-valuation;
- (ii) L is distributive if and only if it is modular and every strictly monotone 2-valuation on L is a 3-valuation.

- (iii) L is distributive if and only if it admits a strictly monotone 3-valuation;
- (iv) L is distributive if and only if it is modular and every strictly monotone 2-valuation on L is a k -valuation for any $k > 2$.

In other words, the existence of a strict monotone 2-valuation characterizes modularity, while the existence of a strictly monotone k -valuation (any $k > 2$) characterizes distributivity.

2.3.3 Belief Function on a Lattice

A belief function has two equivalent definitions:

- (i) A 1-monotone function whose Möbius transform m is non-negative;
- (ii) A totally monotone capacity function.

The function m was called *basic probability assignment* (b.p.a.) in Dempster-Shafer Theory, and those subsets with non-zero probability assignment are called “focal” elements. The Möbius function on the elements of power-set is given as

$$\mu(A, B) = \begin{cases} (-1)^{B \setminus A} & \text{if } A \subseteq B \\ 0 & \text{otherwise} \end{cases} .$$

The theory of belief functions on general lattices was recently investigated by Barthélemy (2000), Grabisch (2008). An important conclusion is that any lattice admits a total monotone function:

Lemma 2.3 (Barthélemy 2000) *On any lattice L and for any function $m : L \rightarrow [0, 1]$ such that $m(0) = 0$; $\sum_{x \in L} m(x) = 1$, then the function $f(x) = \sum_{y \leq x} m(y)$ is a totally monotone function and satisfies $f(0) = 0$; $f(1) = 1$.*

That is, for any mass assignment (b.p.a.), the corresponding inverse Möbius transform is a belief function. If two totally monotone functions on a lattice are identical, then their inducing b.p.a.’s must also be identical. Zhou (2013) showed that the converse is also true: the Möbius transform of any totally monotone capacity function on a lattice must be non-negative. In other words, for any capacity f on L , total monotonicity of f and non-negativity of its Möbius transform m are equivalent. Considering the smallest Boolean algebra (lattice) of which a given finite distributive lattice L is a sublattice, Zhou showed that a belief function on L is a probability function iff all focal elements (i.e., those with positive assignments of b.p.a.’s) are join-irreducible in L . So join-irreducible elements are akin to singletons in Boolean algebra.

To conclude, while a belief function can be defined on any lattice, a probability function (as a total-valuation, normalized and strictly monotone function) can only be defined on a distributive lattice.

3 Upper-Lower Probability Anchored on Topology

3.1 Topologizing Dempster-Shafer Theory

The Dempster-Shafer theory is constructed on the standard Boolean algebra of event space, with basic probability assignment (b.p.a.) on non-atomic element in general. Zadeh (1965) fuzzy set theory is also built upon Boolean algebra, with b.p.a. assigned to an ascending sequence of subsets. As a natural extension of event space structure for probability measures, belief functions on lattices are an interesting and natural topic of investigation. Indeed, recent studies (Barthélemy 2000; Grabisch 2008; Zhou 2013) show that any lattice admits a belief function (and hence an associated non-negative probability mass assignment), and any distributive lattice further admits a probability measure (that is, a 2-valuation that is a capacity). However, none of these research look at the role of pseudo-complementation operator in place of complementation operator of a Boolean lattice. Nor have they investigated the probability theory in a hierarchical setting, which is crucial for modeling multiple contexts and context change. As discussed earlier, pseudo-complementation amounts to specifying a topology on the ground set, which is important to provide semantics to a probability theory. Below, we initiate a new approach to information fusion by (i) investigating probability measures defined on a particular distributive lattice, namely, the topology of a set; and (ii) investigating the belief function defined on the full lattice of all topologies on the given set.

3.2 Lattice of Topologies

Given a set X , the set of all topologies on X form a bounded (and hence complete) lattice \mathcal{L}_X , ordered by “refinement”, or relative coarseness (i.e., inclusion of collection of open sets), of the two topologies under comparison, see survey by Larson and Andima (1975). For two topologies τ and σ , their meet is set-wise intersection $\tau \cap \sigma$, $t \in (\tau \cap \sigma)$ if $t \in \tau$ and $t \in \sigma$; $\tau \cap \sigma$ contains the open sets common to the two topologies. Their join $\tau \cup \sigma$ is the topology *generated* by the intersections of all open sets $\{t_1 \cap t_2 | t_1 \in \tau; t_2 \in \sigma\}$. The top element of this lattice \mathcal{L}_X is the discrete topology where each element is treated as a clopen set; this is the finest (largest) topology on X . The bottom element of \mathcal{L}_X is the indiscrete topology, consisting of only \emptyset, X ; this is the coarsest (smallest) topology. According to Larson and Andima (1975), there is no known formula for the number of topologies on a finite set, only that the number is between 2^n and $2^{n(n-1)}$ where $n = |X|$. For $n = 3, 4, 5, 6, 7$, the number of elements in the lattice is 29, 355, 6942, 209527, 9535241. Figure 1 gives the case of 29 topologies, organized as a lattice, for a three-element set $X = \{a, b, c\}$.

In general, the compact topologies (a topology that every cover has a finite cover) are closed downward in this lattice, since if a topology τ has fewer open sets than σ and σ is compact, then τ is compact. Similarly, the Hausdorff topologies (any

Lattice of topologies of $\{a,b,c\}$

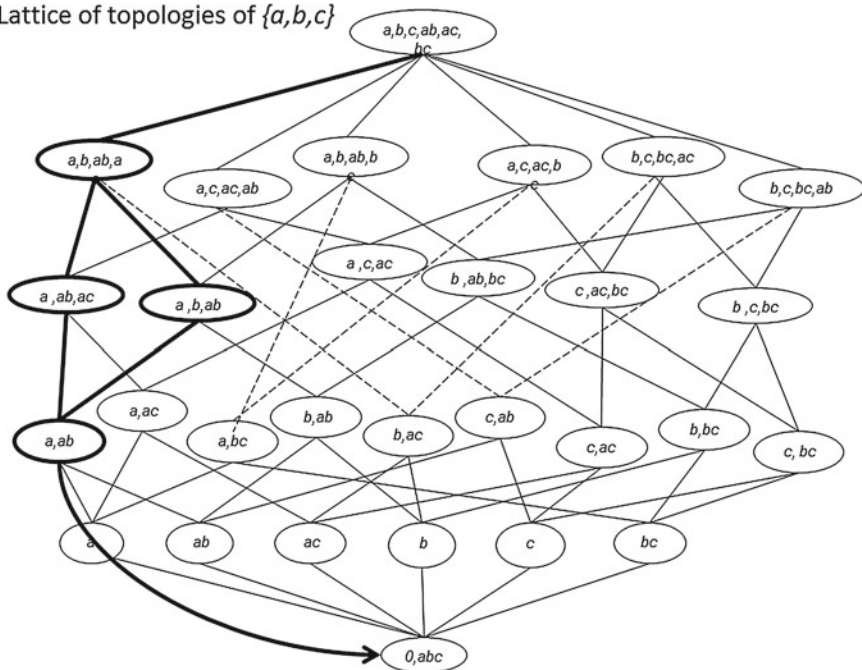


Fig. 1 The lattice of topologies for a 3-element set. Each node shows the elements of the corresponding topology with the empty set and the full set X omitted for best readability. The letter-string, say “bc”, stands for the set $\{b, c\}$. The lattice from Fig. 3 is embedded in the complete lattice and shown by *thick lines*

two points are separated by disjoint open sets) are closed upward, since if τ is Hausdorff and contained in σ , then σ is Hausdorff. Thus, in this lattice of topology, the compact topologies inhabit the bottom of the lattice (where indiscrete topology lies as the extreme) and the Hausdorff topologies the top (where discrete topology lies as the extreme). These two types of topologies run into each other in the middle, known as compact Hausdorff topologies. They form an anti-chain in the lattice: no two compact Hausdorff topologies are comparable and they are all distinct.

The lattice of topologies \mathcal{L}_X , when the ground set X is denumerable, is also known to be complemented. Embedded in \mathcal{L}_X is a sublattice of all T_1 topologies—a T_1 topology on a ground set X is one where each singleton subset $\{x\} \subset X$ is closed. The T_1 topology sublattice is, however, not complemented, not modular (and hence not distributive), but it is both upper and lower semi-modular.

Larson and Andima (1975) noted that, even for a finite set X , the lattice \mathcal{L}_X is non-distributive, non-modular, neither upper nor lower semi-modular, and not self-dual. It has only trivial lattice homomorphisms. For any lattice l , there exists a ground set X such that l can be embedded into the lattice \mathcal{L}_X of topologies on X . Moreover, Valent and Larson (1972) and Rosický (1975) showed that a finite lattice l is distributive if and only if it can be realized as an interval of T_1 topologies on a set X , that is,

there are two T_1 -topologies τ and σ on X such that the subinterval $[\tau, \sigma]$ of \mathcal{L}_X is isomorphic to l . Knight et al. (1997) further strengthened its realizability from T_1 to Hausdorff.

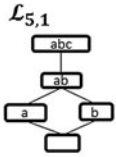
Investigating the lattice of topologies along with each individual topologies allows us to construct a hierarchical scheme of upper-lower probabilities. This is done as follows. Referring to Fig. 1, which depicts all possible topologies on $X = \{a, b, c\}$. Each topology represents a distinct “context”, and each will be assigned a probability measure. Figure 2 gives a few examples of the topologies, with graphic coding for open sets, closed sets, clopen sets, and sets that are neither open nor closed. On the lattice of topologies, we will prescribe a belief function (rather than a probability measure), from which we will construct the probability mass. This is the top-level of the hierarchical scheme, which models different contexts. Probability mass (b.p.a.) will, in general, not be assigned to singletons, i.e., any single topology, reflecting the fact that contexts are not “independent”. In our hierarchical scheme, switching contexts amounts to switching topologies.

Though the lattice of topologies depicts all possible contexts, sometime we may restrict ourselves to a distributive sublattice for convenience. Figure 3 gives such a case. In this case, we may assign probability measure to such sublattice of topologies.

3.3 A Hierarchical Scheme for Upper-Lower Probability

As Zhou (2013) has recently shown, on a distributive lattice, basic probability assignment (b.p.a.’s) would only need to be given to joint-irreducible elements of a lattice for it to be consistent with the Bayesian framework. Any topology on a set fulfills the requirement of a distributive lattice L , where each open set is just an element of L , with set-inclusion \subseteq identified as the order $<$ on L , and set-union \cup and set-intersection \cap are \vee and \wedge operations on L . The closure operation that comes with the given topology is related to the pseudo-complementation operation L . Following Narens (2009, 2011), we use this operation to model intuitionistic “negation” in propositional logic. Moreover, we simultaneously consider two or more topologies defined on the same ground set, each with its own b.p.a.’s. As the set of topologies form a lattice itself, and any lattice admits a belief function see Barthélemy (2000), we will endow a belief function on this “lattice of topology” \mathcal{L} as specifying a higher-level (in a hierarchical Bayesian model) upper and lower probability of different contexts. This allows us to accomplish fusion of evidence (different b.p.a.’s) that is soundly rooted in topological semantics, and achieve a hierarchical inference structure that surpasses most current hierarchical schemes of probability inference.

At the bottom level of our hierarchical scheme, we will treat all open sets in a given topology of a ground set as focal elements for basic probability assignments. This is feasible because, for finite set at least, a topology amounts to a distributive lattice, and from Zhou (2013), Bayesian probability measure is possible as long as the b.p.a.’s are assigned to joint-irreducible elements of the lattice; these will be the “elementary events” in the topological event space.



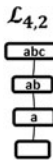
Set-theoretic complements of *Open Sets* (wave patterns) are the *Closed Sets* (dot patterns), listed below. Diamond fill are *Clopen sets*.

Complements

- $\{\}' = \{a, b, c\}$
- $\{a\}' = \{b, c\}$
- $\{b\}' = \{a, c\}$
- $\{a, b\}' = \{c\}$
- $\{a, b, c\}' = \{\}$

Topological Closures

- $Cl \{\} = \{\}$
- $Cl \{a\} = \{a, c\}$
- $Cl \{b\} = \{b, c\}$
- $Cl \{c\} = \{c\}$
- $Cl \{a, b\} = \{a, b, c\}$
- $Cl \{a, c\} = \{a, c\}$
- $Cl \{b, c\} = \{b, c\}$



Set-theoretic complements of *Open Sets* (wave patterns) are the *Closed Sets* (dot patterns), listed below. Diamond fill are *Clopen sets*.

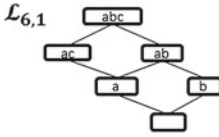
Complements

- $\{\}' = \{abc\}$
- $\{a\}' = \{bc\}$
- $\{ab\}' = \{c\}$
- $\{abc\}' = \{\}$

Topological Closures

- $Cl \{\} = \{\}$
- $Cl \{a\} = \{a, b, c\}$
- $Cl \{b\} = \{b, c\}$
- $Cl \{c\} = \{c\}$
- $Cl \{a, b\} = \{a, b, c\}$
- $Cl \{a, c\} = \{a, b, c\}$
- $Cl \{b, c\} = \{b, c\}$

Fig. 2 Some examples of topologies, where open sets (wave fill), closed sets (dot fill), clopen sets (diamond fill), and sets that are neither open nor closed (blank) are all graphically coded. Also indicated is the closure operation on each node



Set-theoretic complements of *Open Sets* (wave patterns) are the *Closed Sets* (dot patterns), listed below. Diamond fill are *Clopen sets*.

Complements

- $\{\}' = \{abc\}$
- $\{a\}' = \{bc\}$
- $\{b\}' = \{ac\}$
- $\{ab\}' = \{c\}$
- $\{ac\}' = \{b\}$
- $\{abc\}' = \{\}$

Topological Closures

- $Cl \{\} = \{\}$
- $Cl \{a\} = \{a, c\}$
- $Cl \{b\} = \{b\}$
- $Cl \{c\} = \{c\}$
- $Cl \{a, b\} = \{a, b, c\}$
- $Cl \{a, c\} = \{a, c\}$
- $Cl \{b, c\} = \{b, c\}$

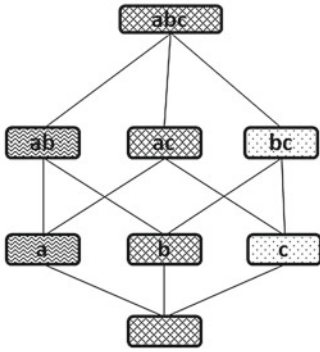


Fig. 2 (continued)

A preliminary implementation of our hierarchical scheme of uncertainty reasoning and probabilistic inference based on topological event space is reported in Ilin and Zhang (2014). The general setting is sensor networks, where bodies of evidence for each sensor network needs to be “fused”. There, we devised a flow-down algorithm for basic probability assignment only on join-irreducible elements of a distributive lattice (topological event space). We further invoke the lattice of topologies for representing different sensor networks, which are treated as different contexts for uncertainty reasoning.

To summarize, our scheme draws its source from two principled approaches to probabilistic inference and uncertainty management: the Dempster-Shafer theory for upper-lower probability constructed from basic probability assignments, and Narens’ (2009, 2011), Narens and Saari (2015) approach to topological event space for lattice-based probability. Our idea, which is fueled by recent mathematical results on the existence of belief functions on a general lattice and probability measure on a distributive lattice, is to construct the upper-lower probability on topological event spaces by (i) stipulating a principled way for basic probability assignments to elements of a topology; and (ii) stipulating (even a sublattice of) the lattice of topologies on the same sample space for modeling switching between and integrating across contexts for b.p.a. assignments. Combining basic probability assignments in a topological event space to obtain upper-lower probabilities with the lattice of topologies to model

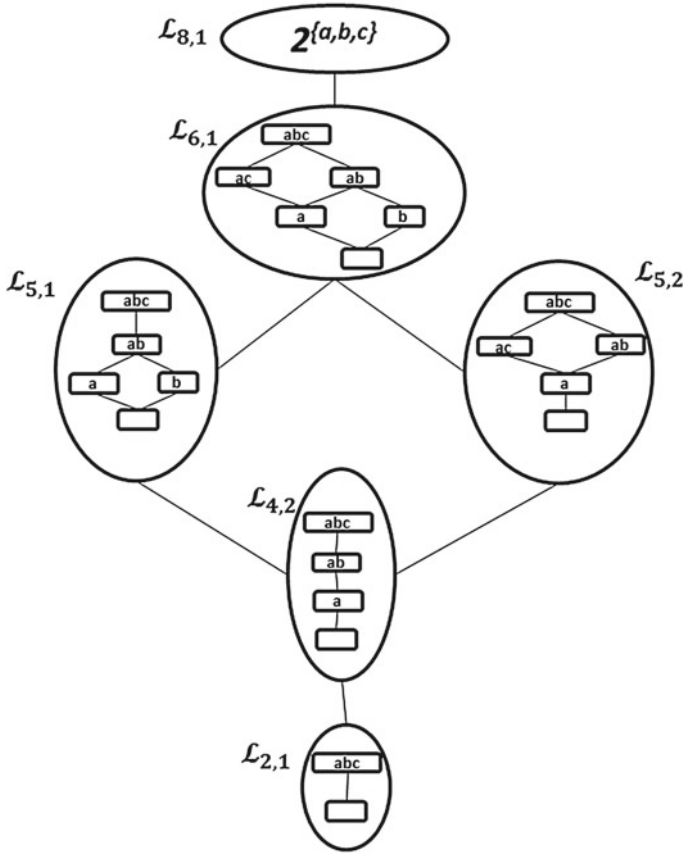


Fig. 3 A distributive sublattice of the lattice of topologies, which supports the assignment of probability measure to “singletons”

hierarchical inference structure has never, to our knowledge, been attempted. Ours can be viewed as a non-Bayesian hierarchical model. The advantage is its ability to model rich sets of contextual information using different topologies for the same underlying sample space and at the same time to reduce the size of the event space by not having to consider the full Boolean (combinatorial) structure.

4 Discussions

Our proposed scheme of upper-lower probability theory based on topological event space and lattice theory is complementary to several other recent developments in extending classical probability calculus.

4.1 Relation to Topological Characterization of “Rare Events”

Chichilnisky (2010) investigated a probability theory that is capable of dealing with unexpected contingencies (“black swan”). Specifically, she proposed a topological framework (Chichilnisky 2010) to deal with measure-zero (“rare”) events which yet have catastrophic consequences. Chichilnisky’s analysis centered on Villegas and Arrows “Axiom of Monotone Continuity”, which invoked a topology that neglects rare events. By replacing it with sup-norm topology of the L_∞ space, she obtained a probability theory which contains a countably additive term as well as a purely finitely additive term.

Her careful analysis of measure-zero events is tightly related to the Ascending/Descending Chain Condition (ACC/DCC) in formulating probability measure. Recall that a partially ordered set (poset) is said to satisfy the ascending chain condition (ACC) if every strictly ascending sequence of elements $\{a_i, i \in \mathbb{N}\}$ eventually terminates. That is, given any sequence $a_1 \leq a_2 \leq a_3 \leq \dots$, there exists a positive integer n such that $a_n = a_{n+1} = a_{n+2} = \dots$. The descending chain condition (DCC) can be analogously defined. ACC and DCC are essentially finiteness properties satisfied by some algebraic structures, e.g., ideals in certain commutative rings.

It is interesting that an analysis of ascending and descending sequence also underlies the characterization of the properties of belief functions. Shafer (1979) defined the notion of “continuity” and “condensation” as regularity conditions for belief (lower-probability) functions to be defined on an infinite set. A belief function f is called continuous if

$$f\left(\bigcap_i A_i\right) = \lim_{i \rightarrow \infty} f(A_i)$$

for every decreasing sequence $A_1 \supseteq A_2 \supseteq \dots$ of subsets of X . A similar definition for plausibility (upper-probability) functions g to be continuous is

$$g\left(\bigcup_i A_i\right) = \lim_{i \rightarrow \infty} g(A_i)$$

for every increasing sequence $A_1 \subseteq A_2 \subseteq \dots$ of subsets of X . A belief function f is called condensable if

$$f\left(\bigcap \mathcal{A}\right) = \inf_{A \in \mathcal{A}} f(A)$$

where \mathcal{A} is a down-net of X , that is, a collection of subsets such that if $A_1, A_2 \in \mathcal{A}$, then there exists $A_3 \in \mathcal{A}$ such that $A_3 \subseteq (A_1 \cap A_2)$. Similarly, a plausibility function g is called condensable if

$$g\left(\bigcup \mathcal{A}\right) = \sup_{A \in \mathcal{A}} g(A)$$

where \mathcal{A} is a up-net of X , that is, a collection of subsets such that if $A_1, A_2 \in \mathcal{A}$, then there exists $A_3 \in \mathcal{A}$ such that $A_3 \supseteq (A_1 \cup A_2)$. The continuity and condensible conditions turn out to be sufficient and necessary conditions for representing a belief function by a \cap -homomorphism into the algebra of measure space (i.e., basic probability assignment on a multiplicative subclass of power-set), see Shafer (1979).

4.2 Relation to Quantum Logic and Quantum Probability

In recent years, quantum probability theory has been invoked in explaining certain phenomena in cognitive psychology, such as conjunction fallacy and order effect in human decision-making literature see Busemeyer and Bruza (2012). The computation model is, however, based on Hilbert space formulation of probability amplitude and quantum physical interpretation of probability measure. Narens (2014) proposed to interpret the cognitive phenomena using quantum logic rather than quantum physics. It is now accepted that the logic underlying quantum physical phenomenology is associated with orthomodular lattices, where orthocomplement is singled out as the “maximal” complement element that each element possesses, and where modularity requirement is imposed upon orthocomplemented pairs only.

Historically, von Neumann first attempted lattice-theoretic characterization of quantum measurements by resorting to orthocomplemented modular lattices. Modular lattice provides a good model for projective geometry. As discussed in Sect. 2.3.2, valuation of a lattice provides the tools for introducing the so-called dimension function on a lattice. In 1930s, von Neumann successfully gave lattice-theoretic treatment of dimension in complete complemented modular lattices. There, dimension is determined, up to a positive linear transformation, by the following two properties: It is conserved by perspective mappings (“perspectivities”) and it is ordered by inclusion. According to Birkhoff, the deepest part of von Neumann’s theory is the equivalence of perspectivity with “projectivity by decomposition”, which gives rise to the transitivity of perspectivity as a corollary. The “dimension function” of the von Neumann algebra takes value not only in a discrete set $\{0, 1, \dots, n\}$, but can also be a value in the unit interval $[0, 1]$. This is the “continuous geometry” setting. Kaplansky later (1955) showed that any orthocomplemented complete modular lattice is a continuous geometry.

It is rather curious that the existence of a strict monotone 2-valuation characterizes modularity, while the existence of a strictly monotone k -valuation (any $k > 2$) characterizes distributivity (see Birkhoff 1933). So for non-distributive modular lattices, a strict monotone 2-valuation does not satisfy 3-valuation condition, though any lattice admits totally monotone functions (which satisfy 2-monotone and 3-monotone conditions). So one wonders what prevents a 2-valuation from becoming a 3-valuation in a non-distributive modular lattice. Understanding this obstruction can lead to insights about distinct probability calculus in intuitionistic and quantum cases.

4.3 Closing Remarks

This chapter reviews some well-established mathematical theory of lattice and its connection to topology, as well as recent results about belief functions and probability measures defined on lattices. We then put forth the idea of a hierarchical scheme for modeling fusion of evidence based on constructing the lattice of topologies over a given sample space, where each topology encodes context for sensor measurement as specified by the basic probability assignment function. This approach provides a rigorous mathematical grounding for modeling uncertainty and information fusion based on upper and lower probabilities originally put forth by the Dempster-Shafer model.

References

- Barthélemy, J.-P. (2000). Monotone functions on finite lattices: An ordinal approach to capacities, belief and necessity functions. *Preferences and decisions under incomplete knowledge* (pp. 195–208). Physica-Verlag HD.
- Birkhoff, G. (1933). On the combination of subalgebras. *Proceedings of the Cambridge Philosophical Society*, 29, 441–464.
- Birkhoff, G. (1967). *Lattice theory*. Providence, RI: American Mathematical Society.
- Busemeyer, J., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge University Press.
- Chateauneuf, A., & Jaffray, J.-Y. (1989). Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17, 263–283.
- Chichilnisky, G. (2010). The foundations of statistics with Black Swans. *Mathematical Social Sciences*, 59, 184–192.
- Davey, B., & Priestley, H. (2002). *Introduction to lattices and order*. Cambridge: University Press.
- Fagin, R., & Halpern, J. Y. (1991). Uncertainty, belief, and probability. *Computational Intelligence*, 7, 160–173.
- Grabisch, M. (2008). *Belief functions on lattice*. arXiv preprint [arXiv:0811.3373v1](https://arxiv.org/abs/0811.3373v1).
- Ilin, R., & Zhang, J. (2014). 2014. Information fusion with uncertainty modeled on topological event spaces. In *Proceedings of IEEE Symposium on Foundation of Computation Intelligence (FOCI 2014)*
- Kaplansky, I. (1955). Any orthocomplemented complete modular lattice is a continuous geometry. *Annals Mathematics*, 61, 524–541.
- Knight, R. W., Gartside, P., & McIntyre, D. W. (1997). All finite distributive lattices occur as intervals between Hausdorff topologies. *Order*, 14, 259–265.
- Larson, R. E., & Andima, S. J. (1975). The lattice of topologies: A survey. *Journal of Mathematics*, 5(2).
- McKinsey, J., & Tarski, A. (1944). The algebra of topology. *Annals of Mathematics*, 45, 141–191.
- McKinsey, J., & Tarski, A. (1946). On closed elements in closure algebras. *Annals of Mathematics*, 47, 122–162.
- Moisil, G. C. (1935). Recherches sur l’algebre de la logique. *Annales scientifiques de l’université de Jassy*, 22, 1–117.
- Monteiro, A. (1980). Sur les algebres de Heyting symtriques. *Portugaliae Mathematica*, 39, 1–237.
- Narens, L. (2011). Probabilistic lattices: Theory with an application to decision theory. In Dzhafarov et al. (Eds.) *Descriptive and normative approaches to human behavior* (pp. 161–202).

- Narens, L., & Saari, D. (2015). Modeling decisions involving ambiguous, vague, and rare events. This volume.
- Narens, L. (2009). A foundation for support theory based on a non-Boolean event space. *Journal of Mathematical Psychology*, 53, 399–407.
- Narens, L. (2014). Alternative probability theories for cognitive psychology. *Topics in Cognitive Science*, 6, 114–120.
- Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences*, 11, 341–356.
- Rosický, J. (1975). Modular, distributive and simple intervals of the lattice of topologies. *Archivum Mathematicum*, 11, 105–114.
- Rota, G.-C. (1964). On the foundation of combinatorial theory. I. Theory of Möbius functions. *Probability Theory and Related Fields*, 2, 340–368.
- Shafer, G. (1976). *A Mathematical Theory of Evidence* (Vol. 1). Princeton: Princeton University Press.
- Shafer, G. (1979). Allocations of probability. *The Annals of Probability*, 7, 827–839.
- Stone, M. H. (1936). The theory of representations for Boolean algebras. *Transactions of the American Mathematical Society*, 40, 37–111.
- Valent, R., & Larson, R. E. (1972). Basic intervals in the lattice of topologies. *Duke Mathematical Journal*, 39, 401–411.
- Yager, R. R., & Liu, L. (Eds.) (2008). *Classic works of the Dempster-Shafer theory of belief functions*. Springer.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338–353.
- Zhou, C. (2013). Belief functions on distributive lattices. *Artificial Intelligence*, 201, 1–31.