Belief functions on lattices

Michel Grabisch

Belief functions on lattices

Michel GRABISCH
Université Paris I – Panthéon-Sorbonne
email Michel.Grabisch@lip6.fr

Abstract

We extend the notion of belief function to the case where the underlying structure is no more the Boolean lattice of subsets of some universal set, but any lattice, which we will endow with a minimal set of properties according to our needs. We show that all classical constructions and definitions (e.g., mass allocation, commonality function, plausibility functions, necessity measures with nested focal elements, possibility distributions, Dempster rule of combination, decomposition w.r.t. simple support functions, etc.) remain valid in this general setting. Moreover, our proof of decomposition of belief functions into simple support functions is much simpler and general than the original one by Shafer.

Keywords: belief function, lattice, plausibility, possibility, necessity

1 Introduction

The theory of evidence, as established by Shafer [16] after the work of Dempster [4], and brought into a practically usable form by the works of Smets in particular [17, 18], has become a popular tool in artificial intelligence for the representation of knowledge and making decision. In particular, many applications in classification have been done [5, 6]. The main advantage over more traditional models based on probability is that the model of Shafer allows for a proper representation of ignorance.

On a mathematical point of view, belief functions, which are at the core of the theory of evidence, possess remarkable properties, in particular their links with the Möbius transform [15] and the co-Möbius transform [9, 10], called *commonality* by Shafer. Remarking that belief functions are non negative isotone functions defined on the Boolean lattice of subsets, one may ask if all these properties remain valid when more general lattices are considered. The aim of this paper is precisely to investigate this question, and we will show that amazingly they all remain valid. A first investigation of this question was done by Barthélemy [1], and our work will complete his results. We are not aware of other similar works, except the one of Kramosil [12], where belief functions are defined on Boolean lattices but take value in a partially ordered set, and the notion of bi-belief proposed by Grabisch and Labreuche [11], where the underlying lattice is 3^n.

On an application point of view, one may ask about the usefulness of such a generalization, apart from its mathematical beauty. A general answer to this is that the objects we manipulate (events, logical propositions, etc.) may not form a Boolean lattice, i.e., distributive and complemented. Thus, a study on a weaker yet rich structure has its interest. Let us give some examples.

- Case where the universal set Ω is the set of possible outcomes, states of nature, etc. In the classical case, all subsets of Ω (called events) are considered, but it may happen that some events are not observable or realizable, meaningful, etc. Then, the structure of the events is no more the Boolean lattice 2^Ω.

- Case where the universal set Ω is the set of propositional variables, either true or false. As argued by Barthélemy [1], in non-classical logics, the set of propositions need not be 2^Ω, and as we will see later, probability theory applies as far as the lattice induced by propositional calculus is distributive, and this covers intuitionistic logic and paraconsistent logic. If distributivity does not hold, then belief functions appear as a natural candidate, since as it will be shown, belief functions can live on any lattice.

- Case where the universal set is the set of players/agents in some cooperative game or multiagent situation. Subsets of Ω are called coalitions, and most of the time, it happens that some coalitions are infeasible, .i.e., they cannot form, due to some inherent impossibility depending on the context. For example, in voting situations, clearly all coalitions of political parties cannot form. The same holds for agents or players in general where some incompatibilities exist between them.

- Knowledge extraction and modeling: objects under study are often structured as lattices. For example, the popular Formal Concept Analysis of Ganter and Wille [8] build lattices of concepts, from a matrix of objects described by qualitative attributes.

- Finally, in some cases, objects of interest are not subsets of some universal set. This is the case for example when one is interested into the collection of partitions of some set (again, this happens in game theory under the name "game in partition function form" [20], and also in knowledge extraction where the fundamental problem is to partition attributes), or when objects of interest are "bi-coalitions" like for bi-belief functions. A bi-coalition is a pair of subsets with empty intersection, and it may represent the set of criteria which are satisfied and the one which are not satisfied.

The paper is organized as follows. Section 2 recalls necessary material on lattices and classical belief functions. Section 3 gives the main results on belief defined over lattices, while the last one examine the case of necessity measures.

Throughout the paper, we will deal with finite lattices.

2 Background

2.1 Lattices

We begin by recalling necessary material on lattices (a good introduction on lattices can be found in [3] and [14]), in a finite setting. A *poset* is a set P endowed with a partial order $\leq$ (reflexive, antisymmetric, transitive). A *lattice* L is a poset such that for any

$x, y \in L$ their least upper bound $x \vee y$ and greatest lower bound $x \wedge y$ always exist. For finite lattices, the greatest element of L (denoted $\top$) and least element $\bot$ always exist. x *covers* y (denoted $x \succ y$) if $x > y$ and there is no z such that $x > z > y$. Let P be a poset, $Q \subseteq P$ is a *downset* if for any $y \in P$ such that $y \leq x$, $x \in Q$, then $y \in Q$. The set of all downsets of P is denoted by $\mathcal{O}(P)$.

A *linear lattice*, or *chain*, is such that $\leq$ is a total order. A chain C in L is *maximal* if no element $x \in L \setminus C$ can be added so that $C \cup \{x\}$ is still a chain.

Lattices can be represented by their *Hasse diagram*, where nodes are elements of the lattice, and there is an edge between x and y, with x above y, if and only if $x \succ y$. Fig. 1 shows three lattices. The middle and right ones are two different diagrams of the lattice of subsets of $\{1, 2, 3\}$ ordered by inclusion.

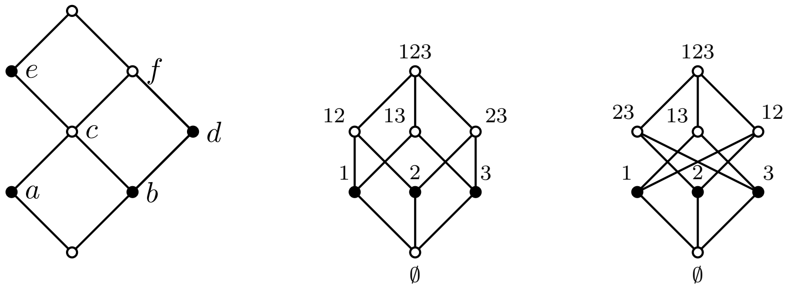

Figure 1: Examples of lattices

Let P, Q be two posets, and consider $f : P \to Q$. f is *isotone* (resp. *antitone*) if $x \leq y$ implies $f(x) \leq f(y)$ (resp. $f(x) \geq f(y)$). P and Q are *isomorphic* (resp. *anti-isomorphic*), denoted by $P \cong Q$ (resp. $P \cong Q^\partial$), if it exists a bijection f from P to Q such that $x \leq y \Leftrightarrow f(x) \leq f(y)$ (resp. $f(x) \geq f(y)$). Isomorphic posets have same Hasse diagrams, up to the labelling of elements.

For any poset $(P, \leq)$, one can consider its *dual* by inverting the order relation, which is denoted by $(P, \leq^\partial)$ (or simply P^∂ if the order relation is not mentionned), i.e., $x \leq y$ if and only if $y \leq^\partial x$. *Autodual* posets are such that $P \cong P^\partial$ (i.e., they have the same Hasse diagram). The lattices of Fig. 1 are all autodual, and Fig. 2 shows their dual.

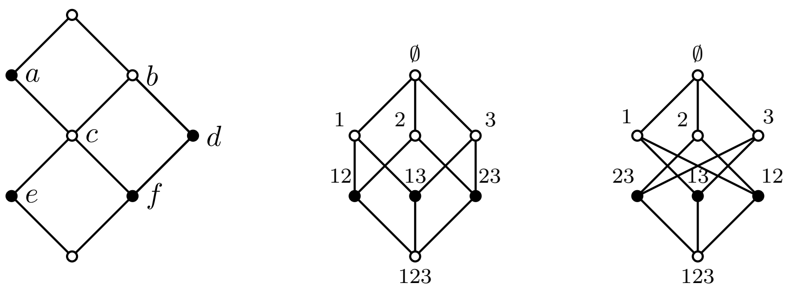

Figure 2: Dual of the lattices of Fig. 1

A lattice L is *lower semimodular* (resp. *upper semimodular*) if for all $x, y \in L$, $x \vee y \succ x$ and $x \vee y \succ y$ imply $x \succ x \wedge y$ and $y \succ x \wedge y$ (resp. $x \succ x \wedge y$ and $y \succ x \wedge y$ imply $x \vee y \succ x$ and $x \vee y \succ y$). A lattice being upper and lower semimodular is called *modular*. The lattice is *distributive* if $(x \vee y) \wedge z = (x \wedge z) \vee (y \wedge z)$ holds for all $x, y, z \in L$.

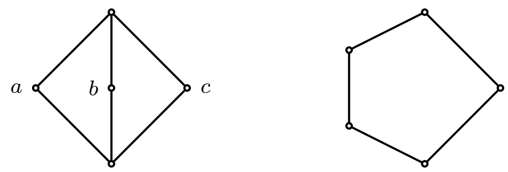

Figure 3: The lattices M_3 (left) and N_5 (right)

$(L, \leq)$ is said to be *lower (upper) locally distributive* if it is lower (upper) semimodular, and it does not contain a sublattice isomorphic to M_3. These are weaker conditions than distributivity, and if L is both lower and upper locally distributive, then it is distributive.

An element $j \in L$ is *join-irreducible* if $j = x \vee y$ implies either $j = x$ or $j = y$, i.e., it cannot be expressed as a supremum of other elements. Equivalently j is join-irreducible if it covers only one element. Join-irreducible elements covering $\bot$ are called *atoms*, and the lattice is *atomistic* if all join-irreducible elements are atoms. The set of all join-irreducible elements of L is denoted $\mathcal{J}(L)$. On Fig. 1 and 2, they are figured as black nodes.

Similarly, *meet-irreducible elements* cannot be written as an infimum of other elements, and are such that they are covered by a single element. We denote by $\mathcal{M}(L)$ the set of meet-irreducible elements of L. *Co-atoms* are meet-irreducible elements covered by $\top$.

For any $x \in L$, we say that x *has a complement in* L if there exists $x' \in L$ such that $x \wedge x' = \bot$ and $x \vee x' = \top$. The complement is unique if the lattice is distributive. L is said to be *complemented* if any element has a complement. On Fig. 1 (left), no element has a complement, except top and bottom, while the two others are complemented lattices.

Boolean lattices are distributive and complemented lattices, and in a finite setting, they are of the type 2^N for some set N, i.e. they are isomorphic to the lattice of subsets of some set, ordered by inclusion (see Fig. 1 (middle,right)). Boolean lattices are atomistic, and atoms correspond to singletons, while co-atoms are of the form $N \setminus \{i\}$ for some $i \in N$.

An important property is that in a lower locally distributive lattice, any element x can be written as an irredundant supremum of join-irreducible elements in a unique way (this is called the *minimal decomposition* of x). We denote by $\eta^*(x)$ the set of join-irreducible elements in the minimal decomposition of x, and we denote by $\eta(x)$ the *normal decomposition* of x, defined as the set of join-irreducible elements smaller or equal to x, i.e., $\eta(x) := \{j \in \mathcal{J}(L) \mid j \leq x\}$. Hence $\eta^*(x) \subseteq \eta(x)$, and

$$x = \bigvee_{j \in \eta^*(x)} j = \bigvee_{j \in \eta(x)} j.$$

Put differently, the mapping η is an isomorphism of L onto $\mathcal{O}(\mathcal{J}(L))$ (Birkhoff's theorem).

Likewise, any element in a upper locally distributive lattice can be written as a unique irredundant infimum of meet-irreducible elements. The decomposition are denoted by μ and μ^*. Specifically, $\mu(x) := \{m \in \mathcal{M}(L) \mid m \geq x\}$, and $x = \bigwedge_{m \in \mu(x)} m$.

The *height function* h on L gives the length of a longest chain from $\bot$ to any element in L. A lattice is *ranked* if $x \succ y$ implies $h(x) = h(y) + 1$. A lattice is lower locally distributive if and only if it is ranked and the length of any maximal chain is $|\mathcal{J}(L)|$.

2.2 The Möbius and co-Möbius transforms

We follow the general definition of Rota [15] (see also [2, p. 102]). Let $(L, \leq)$ be a poset which is locally finite (i.e., any interval is finite) having a bottom element. For any function f on $(L, \leq)$, the *Möbius transform* of f is the function $m : L \longrightarrow \mathbb{R}$ solution of the equation:

$$f(x) = \sum_{y \leq x} m(y). \tag{1}$$

This equation has always a unique solution, and the expression of m is obtained through the Möbius function $\mu : L^2 \to \mathbb{R}$ by:

$$m(x) = \sum_{y \leq x} \mu(y, x) f(y) \tag{2}$$

where μ is defined inductively by

$$\mu(x, y) = \begin{cases} 1, & \text{if } x = y \\ -\sum_{x \leq t < y} \mu(x, t), & \text{if } x < y \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Note that μ depends solely on L.

The *co-Möbius transform* of f, denoted by q, is defined by [9, 10]:

$$q(x) := \sum_{y \geq x} m(y), \quad x \in L. \tag{4}$$

2.3 Belief functions and related concepts

We recall only necessary definitions. For details, the reader is referred to, e.g., [17, 18], or the monograph [13].

Let Ω be a finite space. A function $m : 2^\Omega \to [0, 1]$ is said to be a *mass allocation function* (or simply a *mass*) if $m(\emptyset) = 0$ and $\sum_{A \subseteq \Omega} m(A) = 1$. A subset $A \subseteq N$ is said to be a *focal element* if $m(A) > 0$.

A *belief function* on Ω is a function $\mathrm{bel} : 2^\Omega \to [0, 1]$ generated by a mass allocation function as follows:

$$\mathrm{bel}(A) := \sum_{B \subseteq A} m(B), \quad A \subseteq \Omega. \tag{5}$$

Note that $\mathrm{bel}(\emptyset) = 0$ and $\mathrm{bel}(\Omega) = 1$. One recognizes m as being the Möbius transform of bel (apply Eq. (1) to $(L, \leq) := (2^\Omega, \subseteq)$). The inverse formula, obtained by using (2) and (3), is:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mathrm{bel}(B). \tag{6}$$

Given a mass allocation m, the *plausibility function* is defined by:

$$\mathrm{pl}(A) := \sum_{B | A \cap B \neq \emptyset} m(B) = 1 - \mathrm{bel}(A^c), \quad A \subseteq \Omega. \tag{7}$$

Similarly, the *commonality function* is defined by:

$$q(A) := \sum_{B \supseteq A} m(B), \quad A \subseteq \Omega. \tag{8}$$

It is the co-Möbius transform of bel (see (4)). Remark that $q(\emptyset) = 1$.

A *capacity* on Ω is a set function $v : 2^\Omega \to [0,1]$ such that $v(\emptyset) = 0$, $v(\Omega) = 1$, and $A \subseteq B$ implies $v(A) \leq v(B)$ (*monotonicity*). Plausibility and belief functions are capacities. For any capacity v, its *conjugate* is defined by $\overline{v}(A) := 1 - v(A^c)$. Hence, plausibility functions are conjugate of belief functions (and vice versa). A capacity is *k-monotone* ($k \geq 2$) if for any family of k subsets $A_1, \dots, A_k$ of Ω, it holds:

$$v(\bigcup_{i \in K} A_i) \geq \sum_{I \subseteq K, I \neq \emptyset} (-1)^{|I|+1} v(\bigcap_{i \in I} A_i), \tag{9}$$

with $K := \{1, \dots, k\}$. A capacity is *totally monotone* if it is k-monotone for every $k \geq 2$.

Shafer [16] has shown that a capacity is totally monotone if and only if it is a belief function, hence there exists some mass allocation generating it.

Given two mass allocations m_1, m_2, the *Dempster's rule of combination* computes a combination of both masses into a single one:

$$m(A) =: (m_1 \oplus m_2)(A) := \sum_{B_1 \cap B_2 = A} m_1(B_1) m_2(B_2), \quad \forall A \subseteq \Omega, A \neq \emptyset, \tag{10}$$

and $m(\emptyset) := 0$. Note that m is no more a mass allocation in general, unless some normalization is carried out. It is well known that the Dempster rule of combination can be computed through the commonality functions much more easily. Specifically, calling q, q_1, q_2 the commonality functions associated to m, m_1, m_2, one has:

$$q(A) = q_1(A) q_2(A), \quad \forall A \subseteq \Omega. \tag{11}$$

A *simple support function focused on A* is a particular belief function bel_A whose mass allocation is:

$$m_A(B) := \begin{cases} 1 - w_A, & \text{if } B = A \\ w_A, & \text{if } B = \Omega \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

with $0 < w_A < 1$. Smets [19], using results of Shafer, has shown that any belief function such that $m(\Omega) \neq 0$ can be decomposed using only simple support functions as follows:

$$\mathrm{bel} = \bigoplus_{A \subseteq \Omega} \mathrm{bel}_A \tag{13}$$

with

$$w_A = \prod_{B \supseteq A} q(B)^{(-1)^{|B \setminus A|+1}}, \quad \forall A \subseteq \Omega. \tag{14}$$

In the above decomposition, coefficients w_A may be greater than 1. If this happens, the corresponding bel_A is no more a belief function.

A *necessity function* or *necessity measure* is a belief function whose focal elements form a chain in $(2^\Omega, \subseteq)$, i.e., $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n$ (Dubois and Prade, [7]). The characteristic property of necessity functions is that for any subsets A, B, $\mathrm{N}(A \cap B) = \min(\mathrm{N}(A), \mathrm{N}(B))$, where N denotes a necessity function.

Conjugates of necessity functions are called *possibility functions*, denoted by Π, and are particular plausibility functions. It is easy to see that their characteristic property is that for any subsets A, B, $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$. This characteristic property implies that Π is entirely determined by its value on singletons, i.e., $\Pi(A) = \max_{\omega\in A} \Pi(\{\omega\})$ for any $A \subseteq \Omega$. For this reason, $\pi(\omega) := \Pi(\{\omega\})$ is called the *possibility distribution* associated to Π. Note that necessarily there exists $\omega_0 \in \Omega$ such that $\pi(\omega_0) = 1$. Although this is generally not considered, one may define as well a *necessity distribution* $\nu(\omega) := \mathrm{N}(\Omega \setminus \omega)$, with the property that $\mathrm{N}(A) = \min_{\omega\in A^c} \nu(\omega)$.

Let π be a possibility distribution on $\Omega := \{\omega_1, \ldots, \omega_n\}$, and assume that for some permutation σ on $\{1 \ldots, n\}$, it holds $\pi(\omega_{\sigma(1)}) \leq \pi(\omega_{\sigma(2)}) \leq \cdots \leq \pi(\omega_{\sigma(n)}) = 1$. Then it can be shown that the focal elements of the mass allocation associated to Π are of the form $A_{\sigma(i)} := \{\omega_{\sigma(i)}, \ldots, \omega_{\sigma(n)}\}$, $i = 1, \ldots, n$, and $m(A_{\sigma(i)}) = \pi(\omega_{\sigma(i)}) - \pi(\omega_{\sigma(i-1)})$, with the convention $\pi(\omega_{\sigma(0)}) = 0$.

3 Belief functions and capacities on lattices

Let $(L, \leq)$ be a finite lattice. A *capacity* on L is a function $v : L \to [0, 1]$ such that $v(\bot) = 0$, $v(\top) = 1$, and $x \leq y$ implies $v(x) \leq v(y)$ (isotonicity).

To define the conjugate of a capacity, a natural way would be to write $\overline{v}(x) := 1 - v(x')$, where x' is the complement of x. But this would impose that L is complemented, which is very restrictive. For example, the lattice 3^n underlying bi-belief functions is not complemented. Moreover, if distributivity is imposed in addition, then only Boolean lattices are allowed, and we are back to the classical definition. We adopt a more general definition.

Definition 1 *A lattice L is of* De Morgan type *if it exists a bijective mapping $n : L \to L$ such that for any $x, y \in L$ it holds $n(x \vee y) = n(x) \wedge n(y)$, and $n(\top) = \bot$. We call such a mapping a* $\vee$-negation.

The following is immediate.

Lemma 1 *Let L be a De Morgan lattice, with n a $\vee$-negation. Then:*

(i) $n(\bot) = \top$.

(ii) $n^{-1}(x \wedge y) = n^{-1}(x) \vee n^{-1}(y)$, *for all $x, y \in L$ (n^{-1} is called a* $\wedge$-negation*).*

(iii) If j is join-irreducible, then $n(j)$ is meet-irreducible, and if m is meet-irreducible, then $n^{-1}(m)$ is join-irreducible.

Proof: (i) $n(x \vee \bot) = n(x) = n(x) \wedge n(\bot)$, for all $x \in L$, which implies $n(\bot) = \top$ because n is a bijection.

(ii) Putting $x' := n(x)$ and $y' := n(y)$, we have $n^{-1}(x' \wedge y') = n^{-1}(n(x \vee y)) = x \vee y = n^{-1}(x') \vee n^{-1}(y')$.

(iii) If j is join-irreducible, $j = x \vee y$ implies that $j = x$ or $j = y$. Hence, $n(j) = n(x \vee y) = n(x) \wedge n(y)$ is either $n(x)$ or $n(y)$, which means that $n(j)$ is meet-irreducible. ■

A complemented lattice with unique complement is of De Morgan type with $n(x) := x'$. If L is isomorphic to its dual L^∂, i.e. it is autodual, then it is of De Morgan type since it suffices to take for $n(x)$ the element in the Hasse diagram of L^∂ which takes the place of x in the Hasse diagram of L. In this case, we call n a *horizontal symmetry*. In general, n is not unique since there is no unique way to draw Hasse diagrams. Taking lattices of Fig. 1 as examples, for the left one, we would have $n(a) = e$, for the middle one $n(12) = 1$, and for the right one $n(12) = 3$ (see Fig. 2). Since middle and right lattices are the same, this shows that several n exist in general. Note that n for the right lattice is nothing else than the usual complement.

The following result shows that in fact the only De Morgan type lattices are those which are autodual.

Proposition 1 *A lattice L is of De Morgan type if and only if it is autodual.*

Proof: We already know that if L is autodual, then it is of De Morgan type. Conversely, assuming it is of De Morgan type, it suffices to show that n is an anti-isomorphism. We already know that n is a bijection. Taking $x \leq y$ implies that $x \vee y = y$, hence $n(x \vee y) = n(y) = n(x) \wedge n(y)$, which implies $n(y) \leq n(x)$. Conversely, $n(y) \leq n(x)$ implies $n(y) \wedge n(x) = n(y) = n(x \vee y)$, hence $x \vee y = y$ since n is a bijection, so that $x \leq y$. ■

In general, n and n^{-1} differ, that is, n is not always *involutive*. Take for example the lattice M_3 of Fig. 3, and n defined by $n(\top) = \bot$, $n(\bot) = \top$, $n(a) = b$, $n(b) = c$ and $n(c) = a$. Clearly, n is a $\vee$-negation, but $n(n(a)) = c \neq a$. The $\vee$-negation is involutive whenever n is a horizontal symmetry on the Hasse diagram. If n is involutive, it is simply called a *negation*.

Definition 2 *Let L be an autodual lattice, and n a $\vee$-negation on L. For any capacity v, its* $\vee$-conjugate *and* $\wedge$-conjugate *(w.r.t. n) are defined respectively by*

$$
\begin{aligned}
{}^{\vee}\overline{v}(x) &:= 1 - v(n(x)) \\
{}^{\wedge}\overline{v}(x) &:= 1 - v(n^{-1}(x)),
\end{aligned}
$$

for any $x \in L$. If n is a negation, then $\overline{v}(x) := 1 - v(n(x))$ is the conjugate *of v.*

The following is immediate.

Lemma 2 *Let L be an autodual lattice, and n a $\vee$-negation on L. For any capacity v, it holds*

(i) ${}^{\vee}\overline{v}$ *and* ${}^{\wedge}\overline{v}$ *are capacities on L.*

(ii) ${}^{\vee}\overline{{}^{\wedge}\overline{v}} = {}^{\wedge}\overline{{}^{\vee}\overline{v}} = v$.

Proof: (i) ${}^{\vee}\overline{v}(\top) = 1 - v(n(\top)) = 1$, similarly for $\bot$. Isotonicity of ${}^{\vee}\overline{v}$ follows from antitonicity of n and isotonicity of v.

(ii) ${}^{\vee}\overline{{}^{\wedge}\overline{v}}(x) = 1 - {}^{\wedge}\overline{v}(n(x)) = 1 - (1 - v(x)) = v(x)$. ■

The following definition of belief functions is in the spirit of the original one by Shafer. We used it also for defining bi-belief functions [11].

Definition 3 *A function* bel : $L \to [0,1]$ *is called a* belief function *if* $\text{bel}(\top) = 1$, $\text{bel}(\bot) = 0$, *and its Möbius transform is non negative.*

Referring to (1), we recall that

$$\text{bel}(x) = \sum_{y \leq x} m(y), \quad \forall x \in L. \tag{15}$$

Note that $\text{bel}(\top) = 1$ is equivalent to $\sum_{x \in L} m(x) = 1$, and $\text{bel}(\bot) = 0$ is equivalent to $m(\bot) = 0$. The inverse formula, giving m in terms of bel, has to be computed from (3), and depends only on the structure of L.

Remark that bel is an isotone function by nonnegativity of m, and hence a capacity.

Thanks to the definition of conjugation, if L is autodual and n is a $\vee$-negation, one can define *plausibility functions* as the $\vee$-conjugate of belief functions, which are again capacities.

3.1 k-monotone functions

Barthélemy defines belief function as totally monotone functions. To detail this point, we define k-monotone functions. For $k \geq 2$, a function $f : L \to \mathbb{R}$ is said to be *k-monotone* (called *weakly k-monotone* by Barthélemy) if it satisfies, for any family of elements $x_1, \dots, x_k \in L$:

$$f(\bigvee_{i \in K} x_i) \geq \sum_{I \subseteq K, I \neq \emptyset} (-1)^{|I|+1} f(\bigwedge_{i \in I} x_i) \tag{16}$$

where $K := \{1, \dots, k\}$. A function is said to be *totally monotone* if it is k-monotone for all $k \geq 2$. One can prove that in fact, if $|L| = n$, total monotonicity is equivalent to $(n-2)$-monotonicity [1].

For $k \geq 2$, a function is said to be a *k-valuation* if the inequality (16) degenerates into an equality (called also *Poincaré's inequality*). Similarly, a function is an *infinite valuation* or *total valuation* if it is a k-valuation for all $k \geq 2$. It is well known that monotone infinite valuations satisfying $f(\top) = 1$ and $f(\bot) = 0$ are probability measures.

The following lemma, cited in [1], summarizes well-known results from lattice theory (see Birkhoff [2]).

Lemma 3 *Let L be a lattice. Then*

(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 an infinite valuation.

Barthélemy showed in addition that any lattice admits a totally monotone function. In view of this result, Barthélemy defines belief functions as totally monotone function being monotone and satisfying $f(\top) = 1$ and $f(\bot) = 0$. In summary, a belief function can be defined on any lattice, while probability measures can live only on distributive lattices.

The following proposition shows the relation between both definitions. Before, we state a result from [1].

Lemma 4 *For any lattice L and any function $m : L \to [0,1]$ such that $m(\bot) = 0$ and $\sum_{x\in L} m(x) = 1$, the function $f^m : L \to [0,1]$ defined by $f^m(x) := \sum_{y\leq x} m(y)$ is totally monotone and satisfies $f^m(\top) = 1$ and $f^m(\bot) = 0$.*

Proposition 2 *Any belief function is totally monotone.*

Proof: Let bel be a belief function, and m its Möbius transform. We know that $m(\bot) = 0$ and $\sum_{x\in L} m(x) = 1$. Hence, by Lemma 4, bel is totally monotone. ■

A totally monotone function does not have necessarily a non negative Möbius function. Simple examples show that monotonicity is a necessary condition. The question to know whether monotonicity and total monotonicity imply non negativity of the Möbius function is still open.

3.2 Properties of belief functions

A first result shown by Barthélemy shows that capacities collapse to belief functions when L is linear [1].

Proposition 3 *Any capacity on L is a belief function if and only if L is a linear lattice.*

In the sequel, we address the combination of belief functions and their decomposition in terms of simple support functions. We will see that classical results generalize.

Definition 4 *Let* $\mathrm{bel}_1, \mathrm{bel}_2$ *be two belief functions on L, with Möbius transforms m_1, m_2. The* Dempster's rule of combination *of* $\mathrm{bel}_1, \mathrm{bel}_2$ *is defined through its Möbius transform m by:*

$$m(x) =: (m_1 \oplus m_2)(x) := \sum_{y_1\wedge y_2 = x} m_1(y_1)m_2(y_2), \quad \forall x \in L.$$

Since m defines unambiguously the belief function, we may write as well $\mathrm{bel} = \mathrm{bel}_1 \oplus \mathrm{bel}_2$ to denote the combination.

Proposition 4 *Let* $\mathrm{bel}_1, \mathrm{bel}_2$ *be two belief functions on L, with co-Möbius transforms q_1, q_2, and consider their Dempster combination. Then, if q denotes the co-Möbius transform of* $\mathrm{bel} := \mathrm{bel}_1 \oplus \mathrm{bel}_2$,

$$q(x) = q_1(x)q_2(x), \quad \forall x \in L.$$

Proof: We have:

$$q(x) = \sum_{y \geq x} \sum_{y_1 \wedge y_2 = y} m_1(y_1) m_2(y_2) = \sum_{y_1 \wedge y_2 \geq x} m_1(y_1) m_2(y_2).$$

One can decompose the above sum since if $y_1 \geq x$ and $y_2 \geq x$, then $y_1 \wedge y_2 \geq x$ and reciprocally. Thus,

$$q(x) = \sum_{y_1 \geq x} m(y_1) \sum_{y_2 \geq x} m(y_2) = q_1(x) q_2(x).$$

■

The above proposition generalizes (11), and gives a simple means to compute the Dempster combination.

> REMARK 1: In Def. 4, one may put as in the classical case $m(\bot) = 0$. This does not affect the validity of Prop. 4, except for $x = \bot$. Indeed, by Prop. 4, one obtains $q(\bot) = 1$, but $q(\bot) = \sum_{x \in L} m(x) < 1$ in general if one puts $m(\bot) = 0$ in Def. 4.

Definition 5 *Let* $y \in L$. *A* simple support function focused on y, *denoted by* y^w, *is a function on* L *such that its Möbius transform satisfies:*

$$m(x) = \begin{cases} 1 - w, & \text{if } x = y \\ w, & \text{if } x = \top \\ 0, & \text{otherwise.} \end{cases}$$

The decomposition of some belief function bel in terms of simple support functions is thus to write bel under the form:

$$\mathrm{bel}(x) = \bigoplus_{y \in L} y^{w_y}(x). \tag{17}$$

The following result generalizes the decomposition in the classical case (see Sec. 2.3).

Theorem 1 *Let* bel *be a belief function such that its Möbius transform* m *satisfies* $m(\top) \neq 0$. *The coefficients* w_y *of the decomposition (17) write*

$$w_y = \prod_{x \geq y} q(x)^{-\mu(x,y)}$$

where $\mu(x, y)$ *is the Möbius function of* L.

Proof: We try to find w_y such that

$$\mathrm{bel}(x) = \bigoplus_{y \in L} y^{w_y}.$$

This expression can be written in terms of the co-Möbius transform:

$$q(x) = \prod_{y \in L} q_y(x), \quad x \in L, \tag{18}$$

where q_y is the co-Möbius transform of y^{w_y}:

$$q_y(x) = \begin{cases} 1, & \text{if } x \leq y \\ w_y, & \text{otherwise.} \end{cases}$$

From (18), we obtain:

$$\begin{aligned} \log q(x) &= \sum_{y \in L} \log q_y(x) = \sum_{y \not\geq x} \log w_y \\ &= \sum_{y \in L} \log w_y - \sum_{y \geq x} \log w_y. \end{aligned}$$

On the other hand,

$$q(\top) = \prod_{y \in L} q_y(\top) = \prod_{y \in L} w_y.$$

We supposed that $q(\top) = m(\top) \neq 0$, hence:

$$\log q(x) = \log q(\top) - \sum_{y \geq x} \log w_y.$$

We set $Q(x) := \log q(x)$ and $W(y) := \log w_y$. The last equality becomes:

$$Q(x) = Q(\top) - \sum_{y \geq x} W(y).$$

If we define $Q'(x) = Q(\top) - Q(x)$, we finally obtain:

$$Q'(x) = \sum_{y \geq x} W(y).$$

We recognize here the equation defining the Möbius transform of Q', up to an inversion of the order (dual order)(see (1)). Hence, using (2):

$$W(y) = \sum_{x \geq y} \mu(x, y) Q'(x)$$

with μ defined by (3). Rewriting this with original notation, we obtain:

$$\log w_y = \sum_{x \geq y} \mu(x, y)[\log q(\top) - \log q(x)].$$

Remarking that $\sum_{x \geq y} \mu(x, y) \log q(\top)$ is zero, since it corresponds to the Möbius transform of a constant function, we finally get:

$$w_y = \prod_{x \geq y} q(x)^{-\mu(x,y)}.$$

■

Note that the above proof is much shorter and general than the original one by Shafer [16].

As in the classical case, these coefficients may be strictly greater than 1, hence corresponding simple support functions have negative Möbius transform and are no more belief functions.

4 Necessity functions

Definition 6 *A function* $\mathrm{N} : L \to [0,1]$ *is called a* necessity function *if it satisfies* $\mathrm{N}(x \wedge y) = \min(\mathrm{N}(x), \mathrm{N}(y))$*, for all* $x, y \in L$*, and* $\mathrm{N}(\perp) = 0$, $\mathrm{N}(\top) = 1$.

The following result is due to Barthélemy [1].

Proposition 5 N *is a necessity function if and only if it is belief function whose Möbius transform* m *is such that its focal elements form a chain in* L.

We define possibility functions as $\vee$-conjugates of necessity functions.

Definition 7 *Let* L *be an autodual lattice, and* n *a* $\vee$*-negation on* L*. For any necessity function* N *on* L*, its* $\vee$*-conjugate is called a* possibility function.

Let Π be a possibility function. Then ${}^{\wedge}\overline{\Pi}$ is its corresponding necessity function by Lemma 2 (ii).

Proposition 6 *Let* L *be an autodual lattice, and* n *a* $\vee$*-negation on* L*. The mapping* $\Pi : L \to [0,1]$ *is a possibility function if and only if*

$$\Pi(x \vee y) = \max(\Pi(x), \Pi(y)), \quad \forall x, y \in L. \tag{19}$$

Proof: Let Π be a possibility function being the $\vee$-conjugate of some necessity function N. Then:

$$\begin{aligned}\Pi(x \vee y) &= 1 - \mathrm{N}(n(x \vee y)) = 1 - \mathrm{N}(n(x) \wedge n(y))\\ &= 1 - \min(\mathrm{N}(n(x)), \mathrm{N}(n(y))) = \max(1 - \mathrm{N}(n(x)), 1 - \mathrm{N}(n(y)))\\ &= \max(\Pi(x), \Pi(y)).\end{aligned}$$

Conversely, let Π satisfy (19) and consider its $\wedge$-conjugate ${}^{\wedge}\overline{\Pi}$. We have:

$$\begin{aligned}{}^{\wedge}\overline{\Pi}(x \wedge y) &= 1 - \Pi(n^{-1}(x \wedge y)) = 1 - \Pi(n^{-1}(x) \vee n^{-1}(y))\\ &= 1 - \max(\Pi(n^{-1}(x)), \Pi(n^{-1}(y)))\\ &= \min({}^{\wedge}\overline{\Pi}(x), {}^{\wedge}\overline{\Pi}(y)).\end{aligned}$$

Hence ${}^{\wedge}\overline{\Pi}$ is a necessity function, which implies that Π is a possibility function since ${}^{\vee}\overline{{}^{\wedge}\overline{\Pi}} = \Pi$ by Lemma 2 (ii). ■

The next topic we address concerns distributions. Since we need the property of decomposition of elements into supremum of join-irreducible elements, we impose that L is lower locally distributive. Since L has to be autodual, then it is also upper locally distributive, and so it is distributive. We propose the following definition.

Definition 8 *Let* L *be an autodual distributive lattice, some* $\vee$*-negation* n *on* L*, and* N *a necessity function. The* possibility distribution $\pi : \mathcal{J}(L) \to [0,1]$ *associated to* N *is defined by* $\pi(j) := \Pi(\{j\})$, $j \in \mathcal{J}(L)$*, with* Π *the possibility function which is* $\vee$*-conjugate of* N.

The necessity distribution $\nu : \mathcal{M}(L) \to [0,1]$ *associated to* N *is defined by* $\nu(m) := \mathrm{N}(\{m\})$, $m \in \mathcal{M}(L)$.

Then, the value of Π and N at any $x \in L$ can be computed as follows:

$$\Pi(x) = \max(\pi(j) \mid j \in \eta^*(x)), \quad \mathrm{N}(x) = \min(\nu(m) \mid m \in \mu^*(x)).$$

Remark that due to isotonicity of Π and N, and hence of π and ν, one can replace as well η^*, μ^* by η, μ. The above formulas are well-defined since the decomposition is unique for distributive lattices. Lastly, remark that necessarily there exists $j_0 \in \mathcal{J}(L)$ such that $\pi(j_0) = 1$, and $m_0 \in \mathcal{M}(L)$ such that $\nu(m_0) = 0$, since $\Pi(\top) = 1$ and $\mathrm{N}(\bot) = 0$.

π and ν are related through conjugation since n maps join-irreducible elements to meet-irreducible elements and vice-versa for n^{-1} (see Lemma 1 (iii)). Hence, for $j \in \mathcal{J}(L)$ and $m \in \mathcal{M}(L)$:

$$\begin{aligned} \pi(j) &= 1 - \mathrm{N}(n(j)) = 1 - \nu(m_j) \\ \nu(m) &= 1 - \Pi(n^{-1}(m)) = 1 - \pi(j_m), \end{aligned}$$

where $m_j := n(j)$, and $j_m := n^{-1}(m)$.

Given a mass allocation defining some necessity function, it is easy to derive the corresponding possibility distribution. The converse problem, i.e., given a possibility distribution, find (if possible) the corresponding chain of focal elements and mass allocation giving rise to this possibility distribution, is less simple. Interestingly enough, this problem has always a unique solution, which is very close to the classical case.

Theorem 2 *Let L be autodual, distributive, and n be a $\vee$-negation on L. Let π be a possibility distribution, and assume that the join-irreducible elements of L are numbered such that $\pi(j_1) < \cdots < \pi(j_n) = 1$. Then there is a unique maximal chain of focal elements generating π, given by the following procedure:*

> *Going from j_n to j_1, at each step $k = n, n-1, \ldots, 1$, select the unique join-irreducible element ι_k such that:*
>
> $$\iota_k \notin \eta(n(j_k)), \quad \iota_k \in \bigcap_{l=1}^{k-1} \eta(n(j_l)). \tag{20}$$
>
> *Then the maximal chain is defined by $C_\pi := \{\iota_n, \iota_n \vee \iota_{n-1}, \ldots, \iota_n \vee \cdots \vee \iota_2, \top\}$, and*
>
> $$m(\iota_n \vee \iota_{n-1} \vee \cdots \vee \iota_k) = \pi(j_k) - \pi(j_{k-1}), \quad k = 1, \ldots, n, \tag{21}$$
>
> *with $\pi(j_0) := 0$. Moreover, at each step k, it is equivalent to choose ι_k as the smallest in $\eta(n(j_{k-1})) \setminus \eta(n(j_k))$.*

Proof: For ease of notation, denote $n(j_k)$ by m_k (meet-irreducible).

We first show that such a procedure can always work and leads to a unique solution for C_π. Assume that the poset $\mathcal{J}(L)$ has q connected components $J_1, \ldots, J_q$. By definition, j_n is one of the maximal elements of one of the connected components, say J_{q_0}. Clearly, $\bigwedge_{k=1}^n m_k = n(\bigvee_{k=1}^n j_k) = n(\top) = \bot$. But $\bigvee_{k=1}^{n-1} j_k \neq \top$, otherwise $\big(\bigcup_{l=1,\ldots,q, l \neq q_0} J_l\big) \cup J'_{q_0}$, where J'_{q_0} is a maximal downset of $J_{q_0} \setminus \{j_n\}$, would be another downset corresponding to $\top$, which is impossible since L is distributive (Birkhoff's theorem). This implies that there exists $\iota_n \in \bigcap_{k=1}^{n-1} \eta(m_k)$, and $\iota_n \notin \eta(m_n)$. Let us show that ι_n is unique. Since L is distributive, it is ranked and any maximal chain has length $|\mathcal{J}(L)| = n$. Hence, $\bigvee_{k=1}^{n-1} j_k$ has height $n-1$ (it is a co-atom), and $\bigwedge_{k=1}^{n-1} m_k$ is an atom. Therefore, $\bigcap_{k=1}^{n-1} \eta(m_k)$ is a singleton.

For ι_{n-1} and subsequent ones, we apply the same reasoning on the lattice $\mathcal{O}(\mathcal{J}(L) \setminus \{\iota_n\})$, then on $\mathcal{O}(\mathcal{J}(L) \setminus \{\iota_n, \iota_{n-1}\})$, etc., instead of $L = \mathcal{O}(\mathcal{J}(L))$. Hence, there will be n steps, and at each step one join-irreducible element is chosen in a unique way.

We prove now that the sequence $\{\iota_n, \iota_n \vee \iota_{n-1}, \dots, \iota_n \vee \cdots \vee \iota_2, \top\}$ is a maximal chain, denoted C_π. It suffices to prove that $\iota_n \vee \iota_{n-1} \vee \cdots \vee \iota_k \succ \iota_n \vee \iota_{n-1} \vee \cdots \vee \iota_{k+1}$, $k = 1, \dots, n-1$. The fact that the former is greater or equal to the latter is obvious, hence C_π is a chain. To prove that it is maximal, we have to show that equality cannot occur among any two subsequent elements. To see this, observe that at each step k:

$$\iota_k \not\leq m_k, \quad \iota_k \leq m_{k-1}, \iota_k \leq m_{k-2}, \dots, \iota_k \leq m_1. \tag{22}$$

Hence $\iota_{k-1} \not\leq \iota_k$, otherwise $\iota_{k-1} \leq m_{k-1}$ would hold, a contradiction. Hence, the sequence $\iota_n, \iota_{n-1}, \dots, \iota_1$ is non decreasing, and equality cannot occur.

Let us prove that it suffices to choose ι_k as the smallest in $\eta(n(j_{k-1}) \setminus \eta(n(j_k))$. If at step k, a smallest ι_k is not chosen in $\eta(n(j_{k-1}) \setminus \eta(n(j_k))$, it will be taken after, and the sequence $\iota_n, \iota_{n-1}, \dots, \iota_1$ will be no more non decreasing, a contradiction.

It remains to prove that π is strictly increasing and to verify the expression of m. Let us prove by induction that

$$\pi(j_k) = 1 - m(\iota_n) - m(\iota_n \vee \iota_{n-1}) - \cdots - m(\iota_n \vee \cdots \vee \iota_{k+1}), \quad k = n, \dots, 1. \tag{23}$$

We show it for $k = n$. We have

$$\pi(j_n) = 1 - \nu(m_n) = 1 - \sum_{\substack{x \leq m_n \\ x \in C_\pi}} m(x).$$

Since $\iota_n \not\in \eta(m_n)$, no x in C_π can be smaller than m_n. Hence $\pi(j_n) = 1$. Let us assume (23) is true from n up to some k, and prove it is still true for $k-1$. Using (22), we have:

$$\begin{aligned}\pi(j_{k-1}) &= 1 - \nu(m_{k-1}) = 1 - \sum_{\substack{x \leq m_{k-1} \\ x \in C_\pi}} m(x) \\ &= 1 - \sum_{\substack{x \leq m_k \\ x \in C_\pi}} m(x) - m(\iota_n \vee \cdots \vee \iota_k) = \pi(j_k) - m(\iota_n \vee \cdots \vee \iota_k),\end{aligned}$$

which proves (23). Lastly, remark that the linear system of n equations (23) is triangular, with no zero on the diagonal. Hence it has a unique solution, which is easily seen to be (21). ■

As illustration of the theorem, we give an example.

Example 1: Let us consider the distributive autodual lattice given on Fig. 4. Join-irreducible elements are a, b, c, d, e, f, while meet-irreducible ones are $\alpha, b, \gamma, \delta, \epsilon, f$. We propose as $\vee$-negation the following:

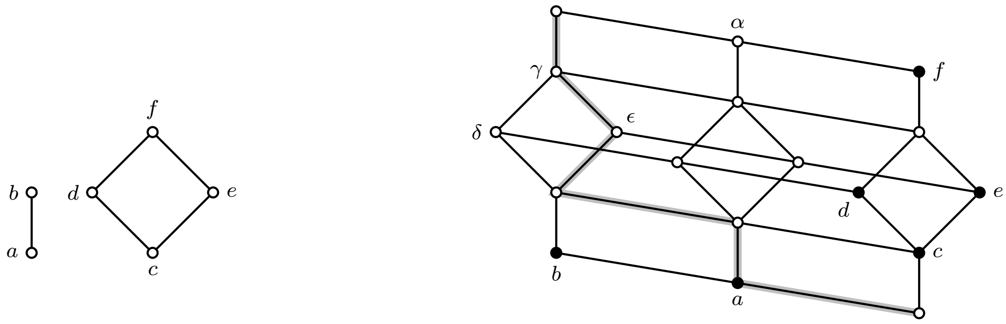

Figure 4: Example of autodual distributive lattice L (right), with $\mathcal{J}(L)$ (left)

x	$n(x)$	x	$n(x)$
a	α	d	ϵ
b	f	e	δ
c	γ	f	b

Let us consider a possibility distribution satisfying

$$\pi(c) < \pi(d) < \pi(e) < \pi(a) < \pi(f) < \pi(b) = 1.$$

(observe that the sequence c, d, e, a, f, b is non decreasing, as requested). We apply the procedure of Th. 2. For b, we have $n(b) = f = c \vee d \vee e \vee f$, and for f, we have $n(f) = b = a \vee b$. Hence the first join-irreducible element of the sequence, ι_6, is a (not in $\eta(f)$, and minimal in $\eta(b)$). Table 1 summarizes all the steps. The maximal chain is in gray on Fig. 4. We deduce that:

step k	x	$n(x)$	$\eta(n(x))$	ι_k	chain
6	b	f	c, d, e, f	a	a
5	f	b	a, b	c	$a \vee c$
4	a	α	a, c, d, e, f	b	$a \vee c \vee b$
3	e	δ	a, b, c, d	e	$a \vee c \vee b \vee e$
2	d	ϵ	a, b, c, e	d	$a \vee c \vee b \vee e \vee d$
1	c	γ	a, b, c, d, e	f	$\top$

Table 1: Computation of C_π

$$
\begin{aligned}
&\pi(b) = 1\\
&\pi(f) = 1 - m(a)\\
&\pi(a) = 1 - m(a) - m(a \vee c)\\
&\pi(e) = 1 - m(a) - m(a \vee c) - m(a \vee c \vee b)\\
&\pi(d) = 1 - m(a) - m(a \vee c) - m(a \vee c \vee b) - m(a \vee c \vee b \vee e)\\
&\pi(c) = 1 - m(a) - m(a \vee c) - m(a \vee c \vee b) - m(a \vee c \vee b \vee e) - m(a \vee c \vee b \vee e \vee d)
\end{aligned}
$$

from which we deduce

$$
\begin{aligned}
m(a) &= \pi(b) - \pi(f) \\
m(a \vee c) &= \pi(f) - \pi(a) \\
m(a \vee c \vee b) &= \pi(a) - \pi(e) \\
m(a \vee c \vee b \vee e) &= \pi(e) - \pi(d) \\
m(a \vee c \vee b \vee e \vee d) &= \pi(d) - \pi(c)
\end{aligned}
$$

and $m(\top) = 1 - m(a) - m(a \vee c) - m(a \vee c \vee b) - m(a \vee c \vee b \vee e) - m(a \vee c \vee b \vee e \vee d) = \pi(c)$.

5 Acknowledgment

The author addresses all his thanks to Bruno Leclerc and Bernard Monjardet for fruitful discussions on negations.

References

[1] J.-P. Barthélemy. Monotone functions on finite lattices: an ordinal approach to capacities, belief and necessity functions. In J. Fodor, B. De Baets, and P. Perny, editors, *Preferences and Decisions under Incomplete Knowledge*, pages 195–208. Physica Verlag, 2000.

[2] G. Birkhoff. *Lattice Theory*. American Mathematical Society, 3d edition, 1967.

[3] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Orders*. Cambridge University Press, 1990.

[4] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38:325–339, 1967.

[5] T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics*, 25(05):804–813, 1995.

[6] T. Denœux. A neural network classifier based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics A*, 30(2):131–150, 2000.

[7] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, 1988.

[8] B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer Verlag, 1999.

[9] M. Grabisch. The interaction and Möbius representations of fuzzy measures on finite spaces, k-additive measures: a survey. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals — Theory and Applications*, pages 70–93. Physica Verlag, 2000.

[10] M. Grabisch and Ch. Labreuche. The symmetric and asymmetric Choquet integrals on finite spaces for decision making. *Statistical Papers*, 43:37–52, 2002.

[11] M. Grabisch and Ch. Labreuche. Bi-belief functions and bi-possibility measures. In *Proc. of the Int. Fuzzy Systems Association World Congress (IFSA 2003)*, pages 155–158, Istanbul, Turkey, June 2003.

[12] I. Kramosil. Degrees of belief in partially ordered sets. *Neural Network World*, 4:363–389, 2001.

[13] I. Kramosil. *Probabilistic analysis of belief functions.* Kluwer Academic/Plenum Publishers, New York, 2001.

[14] B. Monjardet. The presence of lattice theory in discrete problems of mathematical social sciences. why. *Math. Soc. Sci.*, 46:103–144, 2003.

[15] G. C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:340–368, 1964.

[16] G. Shafer. *A Mathematical Theory of Evidence.* Princeton Univ. Press, 1976.

[17] Ph. Smets. The combination of evidence in the transferable belief model. *IEEE Tr. On Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990.

[18] Ph. Smets. What is Dempster-Shafer's model? In R. R. Yager, M. Fedrizzi, and J. Kacprzyk, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 5–34. J.Wiley & Sons, 1994.

[19] Ph. Smets. The canonical decomposition of a weighted belief. In *Proc. of the 14th Int. Joint Conf. on Artificial Intelligence (IJCAI'95)*, pages 1896–1901, Montreal, August 1995.

[20] R. M. Thrall and W. F. Lucas. n-person games in partition function form. *Naval Research Logistic Quarterly*, 10:281–298, 1963.