



Inclusion-exclusion principle for belief functions

Felipe Aguirre, Sebastien Destercke, Didier Dubois, Mohamed Sallak,
Christelle Jacob

► To cite this version:

Felipe Aguirre, Sebastien Destercke, Didier Dubois, Mohamed Sallak, Christelle Jacob. Inclusion-exclusion principle for belief functions. *International Journal of Approximate Reasoning*, Elsevier, 2014, 55 (8), pp.1708-1727. <10.1016/j.ijar.2014.04.018>. <hal-01056459>

HAL Id: hal-01056459

<https://hal.archives-ouvertes.fr/hal-01056459>

Submitted on 20 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Inclusion-exclusion principle for belief functions

F. Aguirre^a, S. Destercke^{b,*}, D. Dubois^c, M. Sallak^b, C. Jacob^{c,d}

^aPhimeca, 18/20 boulevard de Reuilly, F-75012 Paris

^bCNRS/UTC, UMR Heudiasyc, Centre de recherche de Royallieu, 60205 COMPIEGNE

^cIRIT CNRS, Université Paul Sabatier de Toulouse, France

^dISAE, Toulouse, France

Abstract

The inclusion-exclusion principle is a well-known property in probability theory, and is instrumental in some computational problems such as the evaluation of system reliability or the calculation of the probability of a Boolean formula in diagnosis. However, in the setting of uncertainty theories more general than probability theory, this principle no longer holds in general. It is therefore useful to know for which families of events it continues to hold. This paper investigates this question in the setting of belief functions. After exhibiting original sufficient and necessary conditions for the principle to hold, we illustrate its use on the uncertainty analysis of Boolean and non-Boolean systems in reliability.

Keywords. belief function, inclusion-exclusion principle, reliability analysis, boolean formula, independence

1. Introduction

Probability theory is the most well-known approach to model uncertainty. However, even when the existence of a single probability measure is assumed, it often happens that its distribution is only partially known. This is particularly the case in the presence of severe uncertainty (few samples, imprecise or unreliable data, etc.) or when subjective beliefs are elicited (e.g., from experts). Some authors use a selection principle that brings us back to a precise distribution (e.g., maximum entropy [23]). But other ones [28, 26, 16] have argued that in some situations involving imprecision or incompleteness, uncertainty cannot be modelled faithfully by a single probability measure. The same authors have advocated the need for frameworks accommodating imprecision, their efforts resulting in different frameworks such as possibility theory [16], belief functions [26], imprecise probabilities [28], info-gap theory [4], etc. that are formally connected [29, 17]. Regardless of interpretive issues, the formal setting of belief functions offers a good compromise between expressiveness and calculability, as it is

*Corresponding author

Email addresses: felipeam86@gmail.com (F. Aguirre), sebastien.destercke@hds.utc.fr (S. Destercke), dubois@irit.fr (D. Dubois), Mohamed.Sallak@hds.utc.fr (M. Sallak), jacob.christelle.11@gmail.com (C. Jacob)

more general than probability theory, yet in many cases remains more tractable than imprecise probability approaches.

Nevertheless using belief functions is often more computationally demanding than using probabilities. Indeed, its higher level of generality prevents the use of some properties, valid in probability theory, that help simplify calculations. This is the case, for instance, for the well known and useful inclusion-exclusion principle (also known as the Sylvester-Poincaré equality).

Given a space \mathcal{X} , a probability measure P over this space and any collection $\mathcal{A}_n = \{A_1, \dots, A_n | A_i \subseteq \mathcal{X}\}$ of measurable subsets of \mathcal{X} , the inclusion-exclusion principle states that

$$P(\cup_{i=1}^n A_i) = \sum_{\mathcal{J} \subseteq \mathcal{A}_n} (-1)^{|\mathcal{J}|+1} P(\cap_{A \in \mathcal{J}} A) \quad (1)$$

where $|\mathcal{J}|$ is the cardinality of \mathcal{J} . This equality allows us to easily compute the probability of $\cup_{i=1}^n A_i$, when the events A_i are stochastically independent, or when their intersections are disjoint. This principle has been applied to numerous problems, including the evaluation of the reliability of complex systems. It does not hold for belief functions, and only an inequality remains. However, it is useful to investigate whether or not an equality can be restored for specific families \mathcal{A}_n of events, in particular the ones encountered in applications to diagnosis and reliability. The main contribution of this paper is to give a positive answer to this question and to provide conditions characterizing the families of events for which the inclusion-exclusion principle still holds in the belief function setting.

This paper is organized as follows. First, Section 2 provides sufficient and necessary conditions under which the inclusion-exclusion principle holds for belief functions in general spaces; it is explained why the question may be more difficult for the conjugate plausibility functions. Section 3 then studies how the results apply to the practically interesting case where events A_i and focal elements are Cartesian products in a multidimensional space. Section 4 investigates the particular case of binary spaces, and considers the calculation of the degree of belief and plausibility of a Boolean formula expressed in Disjunctive Normal Form (DNF). Section 5 then shows that specific events described by means of monotone functions over a Cartesian product of totally ordered discrete spaces meet the conditions for the inclusion-exclusion principle to hold. Section 6 is devoted to illustrative applications of the preceding results to the field of reliability analysis (both for the binary and non-binary cases), in which the use of belief functions is natural and the need for efficient computation schemes is an important issue. Finally, Section 7 compares our results with those obtained when assuming stochastic independence between ill-known probabilities, displaying those cases for which these results coincide and those for which they disagree.

This work extends the results concerning the computation of uncertainty bounds within the belief function framework previously presented in [22, 1]. In particular, we provide full proofs as well as additional examples. We also discuss the application of the inclusion/exclusion principle to plausibilities, as well as a comparison of our approach with other types of independence notions proposed for imprecise probabilities (two issues not tackled in [22, 1]).

2. General Additivity Conditions for Belief Functions

After introducing some notation and the basics of belief functions (Section 2.1), we explore in Section 2.2 general conditions for families of subsets for which the inclusion-exclusion principle holds for belief functions. We then look more closely at the specific case where the focal elements of belief functions are Cartesian products of subsets. Readers not interested in technical details and familiar with belief functions may directly move to Section 3.

2.1. Setting

A mass distribution [26] defined on a (finite) space \mathcal{X} is a mapping $m : 2^{\mathcal{X}} \rightarrow [0, 1]$ from the power set of \mathcal{X} to the unit interval such that $m(\emptyset) = 0$ and $\sum_{E \subseteq \mathcal{X}} m(E) = 1$. A set E that receives a strictly positive mass is called a *focal element*, and the set of focal elements of m is denoted by \mathcal{F}_m . The mass function m can be seen as a probability distribution over sets, in this sense it captures both probabilities and sets: any probability p can be modelled by a mass m such that $m(\{x\}) = p(x)$ and any set E can be modelled by the mass $m(E) = 1$. In the setting of belief functions, a focal element is understood as a piece of incomplete information of the form $x \in E$ for some parameter x of interest. Then $m(E)$ can be understood as the probability that all that is known about x is that $x \in E$; in other words, $m(E)$ is a probability mass that should be divided over elements of E but is not, due to a lack of information.

From the mapping m are usually defined two set-functions, the belief and the plausibility functions, respectively defined for any $A \subseteq \mathcal{X}$ as

$$Bel(A) = \sum_{E \subseteq A} m(E), \quad (2)$$

$$Pl(A) = \sum_{E \cap A \neq \emptyset} m(E) = 1 - Bel(A^c), \quad (3)$$

with A^c the complement of A . They satisfy $Bel(A) \leq Pl(A)$. The belief function, which sums all masses of subsets that **imply** A , measures how much event A is certain, while the plausibility function, which sums all masses of subsets **consistent** with A , measures how much the event A is possible. Within the so-called theory of evidence [26], belief and plausibility functions are interpreted as confidence degrees about the event A , and are not necessarily related to probabilities. However, the mass distribution m can also be interpreted as the random set corresponding to an imprecisely observed random variable [12], and the measures Bel and Pl can be interpreted as describing a set of probabilities, that is, we can associate to them a set $\mathcal{P}(Bel)$ such that

$$\mathcal{P}(Bel) = \{P | \forall A, Bel(A) \leq P(A) \leq Pl(A)\}$$

is the set of all probabilities bounded by Bel and Pl . The belief function can then be computed as a lower probability $Bel(A) = \inf_{P \in \mathcal{P}(Bel)} P(A)$ and the plausibility function likewise as an upper probability. Note that, since Bel and Pl are conjugate ($Bel(A) = 1 - Pl(A^c)$), we can restrict our attention to one of them.

Consider now a collection of events $\mathcal{A}_n = \{A_1, \dots, A_n | A_i \subseteq \mathcal{X}\}$ of subsets of \mathcal{X} and a mass distribution m from which a belief function Bel can be computed. For any collection \mathcal{A}_n the inequality [26]

$$Bel(\cup_{i=1}^n A_i) \geq \sum_{\mathcal{S} \subseteq \mathcal{A}_n} (-1)^{|\mathcal{S}|+1} Bel(\cap_{A \in \mathcal{S}} A) \quad (4)$$

is valid. This property is called *order- n supermodularity*, and belief functions are super-modular for any $n > 0$. While the inclusion-exclusion property (1) of probabilities is a mere consequence of the additivity axiom (for $n = 2$), supermodularity of order n does not imply supermodularity of order $n + 1$; and the supermodularity property valid at any order is characteristic of belief functions.

If Equation (4) is an equality for some family \mathcal{A}_n , we say that the belief function is *additive* for this collection, or \mathcal{A}_n -additive, for short. Equation (4) is to be compared to Equation (1). Note that in the following we can assume without loss of generality that for any i, j , $A_i \not\subseteq A_j$, i.e. there is no pairwise inclusion relation between the sets of \mathcal{A}_n (otherwise A_i can be suppressed from Equation (4)). Then the family \mathcal{A}_n is said to be *proper*.

2.2. General necessary and sufficient conditions

In the case of two events A_1 and A_2 , neither of which is included in the other, the basic condition for the inclusion-exclusion law to hold is that focal elements in $A_1 \cup A_2$ should only lie (be included) in A_1 or A_2 . Indeed, otherwise, if there exists an event $E \subseteq A_1 \cup A_2$ with $E \not\subseteq A_1$, $E \not\subseteq A_2$ and $m(E) > 0$, then

$$\begin{aligned} Bel(A_1 \cup A_2) &\geq m(E) + Bel(A_1) + Bel(A_2) - Bel(A_1 \cap A_2) \\ &> Bel(A_1) + Bel(A_2) - Bel(A_1 \cap A_2). \end{aligned}$$

This means that, in order to ensure $\{A_1, A_2\}$ -additivity, one must check that

$$\mathcal{F}_m \cap 2^{A_1 \cup A_2} = \mathcal{F}_m \cap (2^{A_1} \cup 2^{A_2}) \quad (5)$$

where 2^C denotes the set of subsets of C . So, one must check that for all events $E \in \mathcal{F}_m$ such that $E \subseteq (A_1 \cup A_2)$, either $E \subseteq A_1$ or $E \subseteq A_2$, or equivalently

Lemma 1. *A belief function is additive for $\{A_1, A_2\}$ if and only if for all events $E \subseteq A_1 \cup A_2$ such that $(A_1 \setminus A_2) \cap E \neq \emptyset$ and $(A_2 \setminus A_1) \cap E \neq \emptyset$ then $m(E) = 0$.*

Proof. Immediate, as E overlaps A_1 and A_2 without being included in one of them if and only if $(A_1 \setminus A_2) \cap E \neq \emptyset$ and $(A_2 \setminus A_1) \cap E \neq \emptyset$. \square

Note that if $\{A_1, A_2\}$ is not proper, the belief function is trivially additive for it. Figure 1 provides an illustration of a focal element that makes a belief function non-additive for events A_1 and A_2 . This result can be extended to the case where $\mathcal{A}_n = \{A_1, \dots, A_n | A_i \subseteq \mathcal{X}\}$ in a quite straightforward way:

Proposition 1. $\mathcal{F}_m \cap 2^{A_1 \cup \dots \cup A_n} = \mathcal{F}_m \cap (2^{A_1} \cup \dots \cup 2^{A_n}) \Leftrightarrow \forall E \subseteq (A_1 \cup \dots \cup A_n)$, if $E \in \mathcal{F}_m$ then $\exists A_i, A_j$ with $(A_i \setminus A_j) \cap E \neq \emptyset$ and $(A_j \setminus A_i) \cap E \neq \emptyset$.

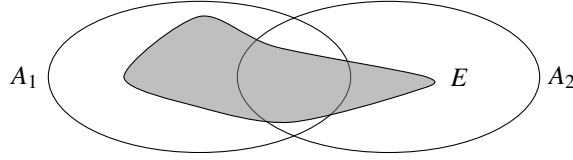


Figure 1: Focal element E of non-additive belief function for $\{A_1, A_2\}$

Proof. $\mathcal{F}_m \cap 2^{A_1 \cup \dots \cup A_n} = \mathcal{F}_m \cap (2^{A_1} \cup \dots \cup 2^{A_n})$
if and only if $\exists E \in \mathcal{F}_m \cap (2^{A_1 \cup \dots \cup A_n} \setminus (2^{A_1} \cup \dots \cup 2^{A_n}))$
if and only if $\exists E \subseteq (A_1 \cup \dots \cup A_n), E \in \mathcal{F}_m$ such that $\forall i = 1, \dots, n, E \not\subseteq A_i$
if and only if $\exists i \neq j, E \in \mathcal{F}_m, E \not\subseteq A_i, E \not\subseteq A_j, E \cap A_i \neq \emptyset, E \cap A_j \neq \emptyset$
if and only if $\exists i \neq j, E \in \mathcal{F}_m$, with $(A_i \setminus A_j) \cap E \neq \emptyset$ and $(A_j \setminus A_i) \cap E \neq \emptyset$ \square

So, based on Proposition 1, we have:

Theorem 2. *The equality*

$$Bel(\cup_{i=1}^n A_i) = \sum_{\mathcal{S} \subseteq \mathcal{A}_n} (-1)^{|\mathcal{S}|+1} Bel(\cap_{A \in \mathcal{S}} A) \quad (6)$$

holds if and only if for all $E \subseteq (A_1 \cup \dots \cup A_n)$, if $m(E) > 0$ then there is no pair A_i, A_j with $(A_i \setminus A_j) \cap E \neq \emptyset$ and $(A_j \setminus A_i) \cap E \neq \emptyset$.

Theorem 2 shows that going from \mathcal{A}_2 -additivity to \mathcal{A}_n -additivity is straightforward, as ensuring \mathcal{A}_n -additivity comes down to checking the conditions of \mathcal{A}_2 -additivity for every pair of subsets in \mathcal{A}_n . This feature makes the verification of the property rather inexpensive. Finally, note that if the family \mathcal{A}_n is not proper, it means $A_j \subseteq A_i$ for some $i \neq j$ and it is then impossible that $\exists E, (A_i \setminus A_j) \cap E \neq \emptyset$ and $(A_j \setminus A_i) \cap E \neq \emptyset$. So, we can dispense with checking the condition for those pairs of sets.

2.3. Inclusion-exclusion for plausibilities

Note that by duality one also can write a form of inclusion-exclusion property for plausibility functions:

$$Pl(\cap_{i=1}^n B_i) = \sum_{\mathcal{S} \subseteq \mathcal{B}_n} (-1)^{|\mathcal{S}|+1} Pl(\cup_{B \in \mathcal{S}} B) \quad (7)$$

for a family of sets $\mathcal{B}_n = \{A_i^c : A_i \in \mathcal{A}_n\}$ where \mathcal{A}_n satisfies the condition of Proposition 1. Although Equation (7) provides us with a kind of inclusion-exclusion property for plausibilities, it does not provide insight about the conditions under which the equality

$$Pl(\cup_{i=1}^n A_i) = \sum_{\mathcal{S} \subseteq \mathcal{A}_n} (-1)^{|\mathcal{S}|+1} Pl(\cap_{A \in \mathcal{S}} A) \quad (8)$$

holds. In this section, we will investigate this issue, concluding that the case of plausibility functions is harder to deal with, and less practically interesting than the case of belief functions.

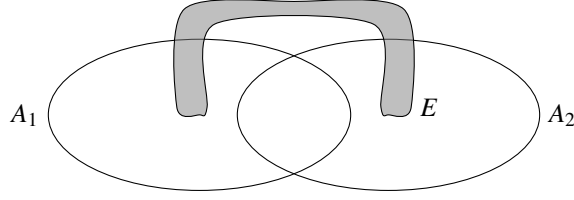


Figure 2: Focal element E of non-additive plausibility function for $\{A_1, A_2\}$

Let us first deal with two events A_1 and A_2 . In this case, any focal element E overlapping $A_1 \cup A_2$ should not overlap A_1 and A_2 without overlapping $A_1 \cap A_2$, otherwise let \mathcal{E} be the non-empty set of focal elements that overlap A_1 and A_2 without overlapping $A_1 \cap A_2$. It is then clear that Pl is strictly submodular, i.e.:

$$Pl(A_1 \cup A_2) + \sum_{E \in \mathcal{E}} m(E) = Pl(A_1) + Pl(A_2) - Pl(A_1 \cap A_2).$$

and additivity fails. This leads us to the following condition for a plausibility function to be \mathcal{A}_2 -additive.

Lemma 2. *A plausibility function is \mathcal{A}_2 -additive for $\mathcal{A}_2 = \{A_1, A_2\}$ if and only if $\forall E \cap (A_1 \cup A_2) \neq \emptyset$ such that $E \cap (A_1 \setminus A_2) \neq \emptyset$, $E \cap (A_2 \setminus A_1) \neq \emptyset$ and $E \cap (A_1 \cap A_2) = \emptyset$, then $m(E) = 0$.*

It should be noted that this condition is similar to, but quite different from the one in Lemma 1, as any set overlapping $A_1 \cap A_2$ but not included in $A_1 \cap A_2$ can receive a positive mass without leading to a violation of \mathcal{A}_2 -additivity for the associated plausibility function. This is not the case for belief functions: for instance, the focal element of Figure 1 is not in contradiction with Lemma 2 (plausibility could still be \mathcal{A}_2 -additive). Figure 2 pictures a focal element that would make the plausibility not \mathcal{A}_2 -additive.

Nevertheless, the condition for \mathcal{A}_2 -additivity in Lemma 2 can be equivalently expressed as follows

$$\mathcal{F}_m \cap 2^{(A_1^c \cup A_2^c)} = \mathcal{F}_m \cap \left(2^{A_1^c} \cup 2^{A_2^c} \right),$$

which can be deduced from Equation (5), using the fact that for two subsets A_1, A_2 , the \mathcal{A}_2 -additivity of a plausibility function is equivalent to the \mathcal{A}_2 -additivity of the dual belief function for A_1^c, A_2^c (clearly, $Pl(A_1 \cup A_2) = Pl(A_1) + Pl(A_2) - Pl(A_1 \cap A_2)$ is the same equation as $Bel(A_1^c \cup A_2^c) = Bel(A_1^c) + Bel(A_2^c) - Bel(A_1^c \cap A_2^c)$).

However, the situation for plausibility functions is more involved than the one for belief functions, and Lemma 2 cannot be straightforwardly extended to the case of n events, as the \mathcal{A}_3 -additivity of a plausibility function for A_1, A_2, A_3 :

$$Pl(A_1 \cup A_2 \cup A_3) = \underbrace{\sum_{A_i} Pl(A_i)}_{\text{term 1}} - \underbrace{\sum_{A_i, A_j} Pl(A_i \cap A_j)}_{\text{term 2}} + \underbrace{Pl(A_1 \cap A_2 \cap A_3)}_{\text{term 3}}. \quad (9)$$

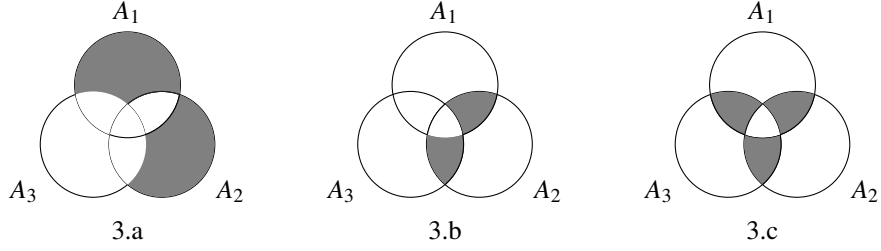


Figure 3: Plausibility \mathcal{A}_3 -additivity illustration

is no longer equivalent to \mathcal{A}_3 -additivity for the conjugate belief function for A_1^c, A_2^c, A_3^c . For one, the two equations are no longer the same. Moreover, the condition

$$\mathcal{F}_m \cap 2^{(A_1^c \cup A_2^c \cup A_3^c)} = \mathcal{F}_m \cap (2^{A_1^c} \cup 2^{A_2^c} \cup 2^{A_3^c})$$

no longer ensures the \mathcal{A}_3 -additivity of the plausibility function. For instance if $E = (A_1 \setminus (A_2 \cup A_3)) \cup (A_2 \setminus (A_1 \cup A_3))$, a subset of A_3^c pictured in Figure 3.a, is focal, the mass of this element is counted once in the left-hand side of Equation (9), twice in term 1 and zero times in terms 2 and 3 of the right-hand side, so \mathcal{A}_3 -additivity fails for this plausibility function on these sets.

Also, in the case of plausibility functions, we cannot expect \mathcal{A}_n -additivity to follow from \mathcal{A}_2 -additivity between all pairs of events. Consider indeed the following possible focal elements:

- E is included in $((A_1 \cap A_2) \setminus A_3) \cup ((A_2 \cap A_3) \setminus A_1)$ (pictured in Figure 3.b): such an element is counted once in the left-hand side of Equation (9), thrice in term 1, twice in term 2 and zero times in term 3 of the right-hand side, hence we can have $m(E) > 0$ without violating \mathcal{A}_3 -additivity. Yet, set E is such that $E \cap (A_1 \setminus A_3) \neq \emptyset$, $E \cap (A_3 \setminus A_1) \neq \emptyset$ and $E \cap (A_1 \cap A_3) = \emptyset$, showing that it does not satisfy Lemma 2 for the pair A_1, A_3 , and that this feature does not forbid \mathcal{A}_3 -additivity to hold for plausibilities;
- E is included in $\cup_{i \neq j \neq k} ((A_i \cap A_j) \setminus A_k)$ (pictured in Figure 3.c): such an element is counted once in the left-hand side of Eq. (9), thrice in term 1, thrice in term 2 and zero times in term 3 of the right-hand side, hence \mathcal{A}_3 -additivity does not hold if $m(E) > 0$. But E satisfies Lemma 2 for all three pairs, which proves not sufficient to ensure \mathcal{A}_3 -additivity for plausibilities.

This suggests that obtaining easy-to-check conditions for \mathcal{A}_n -additivity to hold for plausibility in a general setting will be difficult, if not impossible. Of course, one can check for a given collection \mathcal{A}_n that every focal element $E \in \mathcal{F}_m$ overlapping $\cup_{A \in \mathcal{A}_n} A$ is counted once in the right- and left-hand side of (8), yet such a tedious verification would defeat the purpose of using the inclusion-exclusion principle as a practical means to achieve efficient computations.

For this reason, we shall not deal with general conditions for the inclusion-exclusion principle to hold for plausibility functions. Yet we will mention those cases (which turn

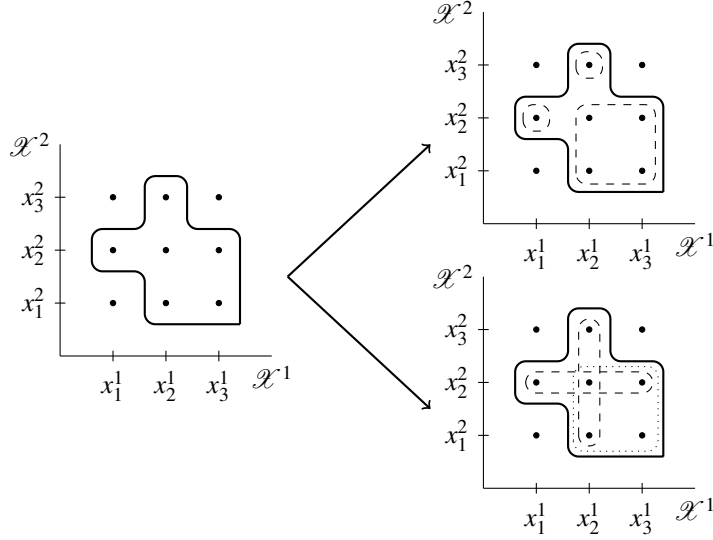


Figure 4: Two possible decompositions of an event A into rectangular subsets

out to be often met in practice) when the conjugacy Equation (3) with respect to belief functions can be exploited to compute $Pl(\cup_{i=1}^n A_i)$.

3. When focal elements are Cartesian products

The previous section has formulated general conditions for \mathcal{A}_n -additivity to hold for a given family of sets. In this section, we investigate a practically important particular case where focal elements and events $A_i, i = 1, \dots, n$ are Cartesian products. That is, we assume that $\mathcal{X} = \mathcal{X}^1 \times \dots \times \mathcal{X}^D := \mathcal{X}^{1:D}$ is the product of *finite* spaces $\mathcal{X}^i, i = 1, \dots, D$. We will call the spaces \mathcal{X}^i *dimensions*. We will denote by x^i the value of a variable (e.g., the state of a component, the value of a propositional variable) on \mathcal{X}^i .

Given $A \subseteq \mathcal{X}$, we will denote by A^i the projection of A on \mathcal{X}^i . Let us call *rectangular* a subset $A \subseteq \mathcal{X}$ that can be expressed as the Cartesian product $A = A^1 \times \dots \times A^D$ of its projections (in general, only $A \subseteq A^1 \times \dots \times A^D$ holds for all subsets A). A rectangular subset A is completely characterized by its projections.

In the following, we derive conditions for the n -additivity property over families \mathcal{A}_n containing rectangular sets only, when the focal elements of mass functions defined on \mathcal{X} are also rectangular (to simplify the proofs, we will also assume that all rectangular sets are focal elements).

Assuming focal elements to be rectangular is a restrictive assumption, as they cannot be freely manipulated and decomposed in different ways, but we discuss in Section 3.3 those (many) practical situations where such focal elements will appear. However, assuming that the collection of sets $\mathcal{A}_n = \{A_1, \dots, A_n\}$ over which the belief value $Bel(\cup_{i=1}^n A_n)$ must be evaluated is not very restrictive, at least in the finite case. Indeed, any such set $A_i \subseteq \mathcal{X}$ can then be decomposed into a (non-unique) finite union of (not

necessarily disjoint) rectangular subsets. To see this, note that there exists an elementary way to always achieve such a decomposition: one can always decompose A as the union of its singletons, each of them being a degenerate rectangular subset. Figure 4 illustrates two possible decompositions of the same subset A . However, we shall see that such a decomposition is not always very interesting for applying the inclusion-exclusion principle. The results of this section also provide conditions under which such a decomposition will allow one to apply the inclusion-exclusion principle.

3.1. Two sets, two dimensions

Let us first explore the case $n = 2$ and $D = 2$, that is $\mathcal{A}_2 = \{A_1, A_2\}$ with $A_i = A_i^1 \times A_i^2$ for $i = 1, 2$. The main idea in this case is that if $A_1 \setminus A_2$ and $A_2 \setminus A_1$ are rectangular with disjoint projections, then 2-additivity holds for belief functions, and this property is characteristic.

Lemma 3. *If A_1 and A_2 are rectangular and have disjoint projections on dimensions $\mathcal{X}^1, \mathcal{X}^2$, then there is no rectangular subset of $A_1 \cup A_2$ overlapping both A_1 and A_2 .*

Proof. Consider $C = C^1 \times C^2$ overlapping both A_1 and A_2 . So there is $(a^1, a^2) \in A_1 \cap C$ and $b^1 \times b^2 \in A_2 \cap C$. Since C is rectangular, (a^1, b^2) and $(b^1, a^2) \in C$. However if $C \subseteq A_1 \cup A_2$ then $(a^1, b^2) \in A_1 \cup A_2$ and either $b^2 \in A_1^2$ or $a^1 \in A_2^1$. Since $a^1 \in A_1^1$ and $b^2 \in A_2^2$ by assumption, it would mean that projections of A_1 and A_2 are not disjoint, which leads to a contradiction. \square

We can now characterise under which conditions 2-additivity holds for belief functions

Theorem 3. *2-additivity applied to a proper family $\mathcal{A}_2 = \{A_1, A_2\}$ of rectangular sets holds for belief functions having rectangular focal elements if and only if one of the following conditions holds*

1. $A_1^1 \cap A_2^1 = A_1^2 \cap A_2^2 = \emptyset$
2. $A_1^1 \subseteq A_2^1$ and $A_2^2 \subseteq A_1^2$ (or $A_1^1 \supseteq A_2^1$ and $A_2^2 \supseteq A_1^2$, changing both inclusion directions).

Proof. First note that inclusions of Condition 2 can be considered as strict, as we have assumed A_1, A_2 to not be included in each other, in which case the result immediately follows from $A_1 \cup A_2 = A_1$ if $A_1 \subseteq A_2$ or from $A_1 \cup A_2 = A_2$ if $A_2 \subseteq A_1$.

\Leftarrow :

1. If $A_1^1 \cap A_2^1 = A_1^2 \cap A_2^2 = \emptyset$, A_1 and A_2 are disjoint, as well as their projections. Then by Lemma 3, all rectangular subsets included in $A_1 \cup A_2$ are either included in A_1 or in A_2 , hence Lemma 1 applies and 2-additivity holds for belief functions.
2. $A_1^1 \subseteq A_2^1$ and $A_2^2 \subseteq A_1^2$ imply that $A_1 \setminus A_2 = A_1^1 \times (A_1^2 \setminus A_2^2)$ and $A_2 \setminus A_1 = (A_2^1 \setminus A_1^1) \times A_2^2$. As $A_1 \setminus A_2$ and $A_2 \setminus A_1$ are rectangular and have disjoint projections, Lemma 3 and Lemma 1 apply (as above) and 2-additivity holds for belief functions.

\Rightarrow :

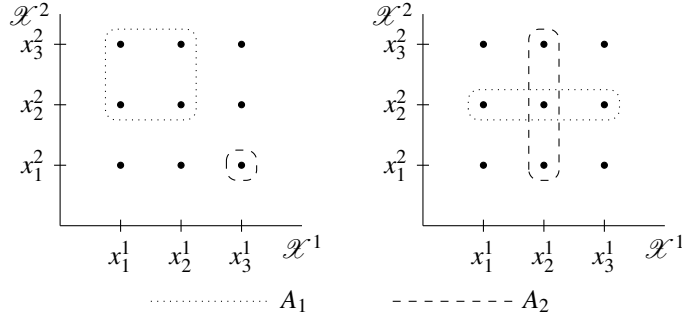


Figure 5: Situations satisfying Theorem 3

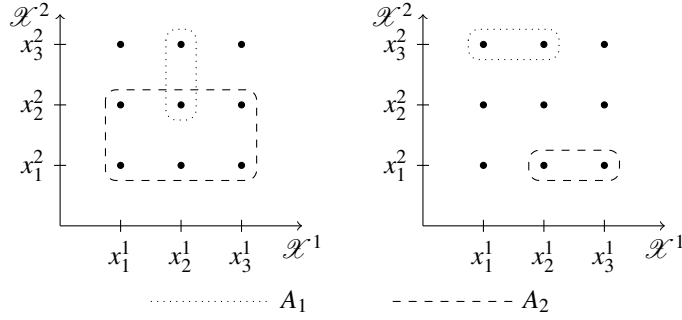


Figure 6: Situations not satisfying Theorem 3

1. Suppose $A_1 \cap A_2 = \emptyset$ with $A_1^1 \cap A_2^1 \neq \emptyset$. Then $(A_1^1 \cap A_2^1) \times (A_1^2 \cup A_2^2)$ is rectangular, not contained in A_1 nor A_2 but contained in $A_1 \cup A_2$, showing that the focal element $(A_1^1 \cap A_2^1) \times (A_1^2 \cup A_2^2)$ does not satisfy Lemma 1 and therefore that 2-additivity does not hold.
2. Suppose $A_1^1 \subset A_2^1$ but $A_2^2 \not\subset A_1^2$. Again, $(A_1^1 \cap A_2^1) \times (A_1^2 \cup A_2^2) = A_1^1 \times (A_1^2 \cup A_2^2)$ is rectangular, neither contained in A_1 nor A_2 but contained in $A_1 \cup A_2$, hence 2-additivity does not hold.

□

Figure 5 and 6 show various situations where conditions of Theorem 3 are satisfied and not satisfied, respectively.

3.2. The multidimensional case

We can now proceed to extend Theorem 3 to the case of any number D of dimensions. However, this extension will not be as straightforward as going from Lemma 1 to Proposition 1, and we need first to characterize when the union of two singletons is

rectangular. We will call such rectangular unions *minimal rectangles*. A singleton is a degenerate example of a minimal rectangle.

Lemma 4. *Let $a = (a^1, \dots, a^D)$ and $b = (b^1, \dots, b^D)$ be two distinct elements in \mathcal{X} . Then, $\{a, b\}$ forms a minimal rectangle if and only if there is only one $i \in [1 : D]$ such that $a^i \neq b^i$*

Proof. \Rightarrow : If $a^i \neq b^i$ for only one i , then $\{a, b\} = \{a^1\} \times \dots \times \{a^i, b^i\} \times \dots \times \{a^D\}$ is rectangular.

\Leftarrow : Let us now consider the case where $a^i \neq b^i$ and $a^j \neq b^j$ for $i \neq j$. In this case,

$$\{a, b\} = \{(a^1, \dots, a^i, \dots, a^j, \dots, a^D), (a^1, \dots, b^i, \dots, b^j, \dots, a^D)\}.$$

The projections of $\{a, b\}$ on the dimensions of \mathcal{X} are $\{a^k, b^k\}$, and we know that $\{a^i, b^i\}$ as well as $\{a^j, b^j\}$, do not reduce to singletons. Hence, the Cartesian product of the projections of $\{a, b\}$ contains the set $\{a^1\} \times \dots \times \{a^i, b^i\} \times \{a^j, b^j\} \times \dots \times \{a^D\}$, that contains elements not in $\{a, b\}$ (e.g. $(a^1, \dots, b^i, \dots, a^j, \dots, a^D)$). Since $\{a, b\}$ is not characterised by its projections on dimensions \mathcal{X}_i , it is not rectangular, and this finishes the proof. \square

As mentioned before, any set can be decomposed into rectangular sets, and in particular any rectangular set can be decomposed into minimal rectangles. Also, any rectangular set that is not a singleton will at least contain one minimal rectangle, implying that there always exist at least two singletons of a rectangular set forming a minimal rectangle (we will use this in subsequent proofs). Let us now show how Theorem 3 can be extended to D dimensions.

Theorem 4. *2-additivity holds for a proper family $\mathcal{A}_2 = \{A_1, A_2\}$ of rectangular sets for belief functions having rectangular focal elements if and only if one of the following conditions holds*

1. \exists distinct $p, q \in \{1, \dots, D\}$ such that $A_1^p \cap A_2^p = A_1^q \cap A_2^q = \emptyset$
2. $\forall i \in \{1, \dots, D\}$ either $A_1^i \subseteq A_2^i$ or $A_2^i \subseteq A_1^i$

Proof. Again, we can consider that there is at least two distinct $p, q \in \{1, \dots, D\}$ such that inclusions $A_1^p \subset A_2^p$ and $A_2^q \subset A_1^q$ of Condition 2 are strict, as we have assumed A_1, A_2 to not be included in each other (as in Theorem 3 and for the same reasons).

\Leftarrow :

1. Any two $a_1 \in A_1$ and $a_2 \in A_2$ will be such that $a_1^i \in A_1^i$ and $a_2^i \in A_2^i$ must be distinct for $i = p, q$ since $A_1^p \cap A_2^p = A_1^q \cap A_2^q = \emptyset$. By Lemma 4, this means that there is no pair $a_1 \in A_1$ and $a_2 \in A_2$ forming a minimal rectangle. This implies that there is no minimal rectangle included in $A_1 \cup A_2$, and therefore no rectangular subset. It follows that each rectangular subset included in $A_1 \cup A_2$ is either included in A_1 or in A_2 , hence Lemma 1 applies and 2-additivity holds for belief functions.
2. Let us denote by P the set of indices p such that $A_1^p \subset A_2^p$ and by Q the set of indices q such that $A_2^q \subset A_1^q$. Now, let us consider two singletons $a_1 \in A_1 \setminus A_2$ and $a_2 \in A_2 \setminus A_1$. Then

- $\exists p \in P$ such that $a_1^p \in A_1^p \setminus A_2^p$, otherwise a_1 is included in $A_1 \cap A_2$
- $\exists q \in Q$ such that $a_2^q \in A_2^q \setminus A_1^q$, otherwise a_2 is included in $A_1 \cap A_2$

but since $a_1^q \in A_1^q$ and $a_2^p \in A_2^p$ by definition, a_1 and a_2 must differ at least on two dimensions, hence by Lemma 4 one cannot form a minimal rectangle outside $A_1 \cap A_2$, that is, by picking pairs of singletons in $A_1 \setminus A_2$ and $A_2 \setminus A_1$. As above, this implies that Lemma 1 is satisfied and that 2-additivity holds.

\Rightarrow :

1. Suppose $A_1 \cap A_2 = \emptyset$ with $A_1^q \cap A_2^q \neq \emptyset$ only for q . Then the following rectangular set contained in $A_1 \cup A_2$

$$(A_1^1 \cap A_2^1) \times \cdots \times (A_1^{q-1} \cap A_2^{q-1}) \times (A_1^q \cup A_2^q) \times (A_1^{q+1} \cap A_2^{q+1}) \cdots \times (A_1^D \cap A_2^D)$$

is neither contained in A_1 nor A_2 , so 2-additivity will not hold (by Lemma 1)

2. suppose $A_1 \cap A_2 \neq \emptyset$ and $A_1^q \subsetneq A_2^q$, $A_1^q \not\supseteq A_2^q$ for some q . Again,

$$(A_1^1 \cap A_2^1) \times \cdots \times (A_1^{q-1} \cap A_2^{q-1}) \times (A_1^q \cup A_2^q) \times (A_1^{q+1} \cap A_2^{q+1}) \cdots \times (A_1^D \cap A_2^D)$$

is rectangular, neither contained in A_1 nor A_2 but contained in $A_1 \cup A_2$, so 2-additivity will not hold (by Lemma 1).

□

Using Proposition 1, the extension to n -additivity in D dimensions is straightforward:

Theorem 5. *n -additivity holds for a proper family $\mathcal{A}_n = \{A_1, \dots, A_n\}$ of rectangular sets for belief functions having rectangular focal elements if and only if, for each pair A_i, A_j , one of the following conditions holds*

1. \exists distinct $p, q \in \{1, \dots, D\}$ such that $A_i^p \cap A_j^p = A_i^q \cap A_j^q = \emptyset$
2. $\forall \ell \in \{1, \dots, D\}$ either $A_i^\ell \subseteq A_j^\ell$ or $A_j^\ell \subseteq A_i^\ell$

Note that the second condition is insensitive to set-complements, hence the following result:

Corollary 6. *n -additivity holds on both $\mathcal{A}_n = \{A_1, \dots, A_n\}$ and $\mathcal{A}_n^c = \{A_1^c, \dots, A_n^c\}$ for belief functions whenever for each pair A_i, A_j , $\forall \ell \in \{1, \dots, D\}$ either $A_i^\ell \subseteq A_j^\ell$ or $A_j^\ell \subseteq A_i^\ell$.*

3.3. On the practical importance of rectangular focal elements

While limiting ourselves to rectangular subsets in \mathcal{A} is not especially restrictive, the assumption that focal elements have to be restricted to rectangular sets may seem restrictive (as we are not free to cut any focal element into rectangular subsets). However, such mass assignments actually appear in many practical situations. They can

result for example from the combination of marginal masses m^i defined on each dimension \mathcal{X}^i , $i = 1, \dots, D$ under an assumption of (random set) independence [10]. In this case, the joint mass assigned to each rectangular set E is

$$m(E) = \prod_{i=1}^D m^i(E^i). \quad (10)$$

Additionally, the random set independent assumption makes the computation of the belief and plausibility functions of any rectangular set A easier, as they factorize in the following way:

$$Bel(A) = \prod_{i=1}^D Bel^i(A^i), \quad (11)$$

$$Pl(A) = \prod_{i=1}^D Pl^i(A^i), \quad (12)$$

where Bel^i, Pl^i are the belief/plausibility measures induced by m^i .

An interesting fact is that since the proofs of Section 3 only require focal elements and events to be Cartesian products, they also apply to the cases of unknown or partially known dependence, as long as these latter cases can be expressed by linear constraints imposed on the joint mass [2]. Considering more generic models than belief functions is also possible, e.g. lower probabilities. Then the positivity of the mass functions m^i no longer holds, but the approach can be carried out without modifying our results since the product of (possibly negative) masses in such approaches preserves the approximation properties of random set independence [13].

The following sections explore and discuss specific cases of interest where the inclusion-exclusion property applies.

4. The case of Boolean formulas

In this section, we explore the case where spaces \mathcal{X}^i are binary. In particular, conditions are laid bare for applying the inclusion-exclusion property to Boolean formulas expressed in Disjunctive Normal Form (DNF). We also discuss the problem of estimating plausibilities of Boolean formulas using the inclusion-exclusion property.

In propositional logic, each dimension \mathcal{X}^i is of the form $\{x^i, \neg x^i\}$. It can be associated to a Boolean variable also denoted by x^i , and $\mathcal{X}^{1:D}$ is also called the set of interpretations of the propositional language generated by the set of variables x^i . In this case, x^i is understood as an atomic proposition, while $\neg x^i$ denotes its negation. An element of \mathcal{X}^i is called a literal (x^i is a positive one and $\neg x^i$ a negative one). Any rectangular set $A \subseteq \mathcal{X}^{1:D}$ can then be interpreted as a conjunction of literals (it is often called a partial model), and given a collection of n such partial models $\mathcal{A}_n = \{A_1, \dots, A_n\}$, the event $A_1 \cup \dots \cup A_n$ is a Boolean formula expressed in Disjunctive Normal Form (DNF - a disjunction of conjunctions). All Boolean formulas can be written in such a form.

A convenient representation of a partial model A is in the form of an orthopair [8] (P, N) of disjoint subsets of indices of variables $P, N \subseteq [1 : D]$ such that $A_{(P, N)} =$

$\bigwedge_{k \in P} x^k \wedge \bigwedge_{k \in N} \neg x^k$. Then an element in $\mathcal{X}^{1:D}$ is of the form $\bigwedge_{k \in P} x^k \wedge \bigwedge_{k \in P^c} \neg x^k$, i.e. corresponds to an orthopair (P, P^c) .

We consider belief functions generated by focal elements having the form of partial models. To this end, we consider that the uncertainty over each Boolean variable x^i is described by a belief function Bel^i . As \mathcal{X}^i is binary, its mass function m^i only needs two numbers to be defined. Indeed, it is enough to know $l^i = Bel^i(\{x^i\})$ and $u^i = Pl^i(\{x^i\}) \geq l^i$ (for instance a probability interval [11]) to characterise the marginal mass function m_i since:

- $Bel^i(\{x^i\}) = l^i = m^i(\{x^i\})$;
- $Pl^i(\{x^i\}) = 1 - Bel^i(\{\neg x^i\}) = u^i \implies m^i(\{\neg x^i\}) = Bel^i(\{\neg x^i\}) = 1 - u^i$;
- The sum of masses is $m^i(\{x^i\}) + m^i(\{\neg x^i\}) + m^i(\mathcal{X}^i) = 1$, so $m^i(\mathcal{X}^i) = u^i - l^i$.

Given D independent marginal masses m^i on \mathcal{X}^i , $i = 1, \dots, D$, the joint mass m on $\mathcal{X}^{1:D}$ can be computed as follows for any partial model $A_{(P,N)}$, applying Equation (10):

$$m(A_{(P,N)}) = \left(\prod_{i \in P} l_i \right) \left(\prod_{i \in N} (1 - u_i) \right) \left(\prod_{i \notin P \cup N} (u_i - l_i) \right). \quad (13)$$

We can then give explicit expressions for the belief and plausibility of conjunctions or disjunctions of literals in terms of marginal mass functions:

Proposition 7. *The belief of a conjunction $C_{(P,N)} = \bigwedge_{k \in P} x^k \wedge \bigwedge_{k \in N} \neg x^k$, and that of a disjunction $D_{(P,N)} = \bigvee_{k \in P} x^k \vee \bigvee_{k \in N} \neg x^k$ of literals forming an orthopair (P, N) are respectively given by:*

$$Bel(C_{(P,N)}) = \prod_{i \in P} l^i \prod_{i \in N} (1 - u^i), \quad (14)$$

$$Bel(D_{(P,N)}) = 1 - \prod_{i \in P} (1 - l^i) \prod_{i \in N} u^i. \quad (15)$$

Proof. $Bel(C_{(P,N)})$ can be obtained by applying Equation (11) to $C_{(P,N)}$.

For $Bel(D_{(P,N)})$, we have

$$\begin{aligned} Pl(C_{(N,P)}) &= Pl(\bigwedge_{i \in N} x^i \wedge \bigwedge_{i \in P} \neg x^i) \\ &= \prod_{i \in N} (1 - l^i) \prod_{i \in P} u^i \\ &= 1 - (1 - \prod_{i \in N} (1 - l_i) \prod_{i \in P} u_i) \\ &= 1 - Bel(\bigvee_{i \in N} \neg x^i \vee \bigvee_{i \in P} x^i) \\ &= 1 - Bel(D_{(P,N)}) \end{aligned}$$

with the second equality following from Equation (12). □

Using the fact that $Bel(C_{(N,P)}) = 1 - Pl(D_{(P,N)})$, we can deduce

$$Pl(D_{(P,N)}) = 1 - \prod_{i \in P} l^i \prod_{i \in N} (1 - u^i).$$

We can particularize Theorem 5 to the case of Boolean formulas, and identify conditions under which the belief or the plausibility of a DNF can be easily estimated using the inclusion-exclusion Equality (1). Let us see how the conditions exhibited in this theorem can be expressed in the Boolean case.

Consider the first condition of Theorem 5

$$\exists p \neq q \in \{1, \dots, D\} \text{ such that } A_i^p \cap A_j^q = A_i^q \cap A_j^p = \emptyset.$$

Note that when spaces are binary, $A_i^p = \{x^p\}$ (if $p \in P_i$), or $A_i^p = \{\neg x^p\}$ (if $p \in N_i$), or yet $A_i^p = \mathcal{X}^i$ (if $p \notin P_i \cup N_i$). $A_i \cap A_j = \emptyset$ therefore means that for some index p , $p \in (P_i \cap N_j) \cup (P_j \cap N_i)$ (there are two opposite literals in the conjunction).

The condition can thus be rewritten as follows, using orthopairs (P_i, N_i) and (P_j, N_j) :

$$\exists p \neq q \in \{1, \dots, D\} \text{ such that } p, q \in (P_i \cap N_j) \cup (P_j \cap N_i).$$

Example 1. Consider the equivalence connective $x^1 \Leftrightarrow x^2 = (x^1 \wedge x^2) \vee (\neg x^1 \wedge \neg x^2)$ so that $A_1 = x^1 \wedge x^2$ and $A_2 = \neg x^1 \wedge \neg x^2$. We have $P_1 = \{1, 2\}, N_1 = \emptyset, P_2 = \emptyset, N_2 = \{1, 2\}$. So, $p = 1 \in P_1 \cap N_2, q = 2 \in P_1 \cap N_2$, hence the condition is satisfied and $Bel(x^1 \Leftrightarrow x^2) = Bel(x^1 \wedge x^2) + Bel(\neg x^1 \wedge \neg x^2)$ (the remaining term is $Bel(\emptyset)$).

Likewise, the exclusive or : $x^1 \oplus x^2 = (x^1 \wedge \neg x^2) \vee (\neg x^1 \wedge x^2)$ so that $A_1 = x^1 \wedge \neg x^2$ and $A_2 = \neg x^1 \wedge x^2$. We have $P_1 = \{1\}, N_1 = \{2\}, P_2 = \{2\}, N_2 = \{1\}$. So, $p = 1 \in P_1 \cap N_2, q = 2 \in N_1 \cap P_2$ and $Bel(x^1 \oplus x^2) = Bel(x^1 \wedge \neg x^2) + Bel(\neg x^1 \wedge x^2)$ (again, the remaining term is $Bel(\emptyset)$).

The second condition of Theorem 5 reads

$$\forall \ell \in \{1, \dots, D\} \text{ either } A_i^\ell \subseteq A_j^\ell \text{ or } A_j^\ell \subseteq A_i^\ell$$

and the condition $A_i^\ell \subseteq A_j^\ell$ can be expressed in the Boolean case as:

$$\ell \in (P_i \cap N_j^c) \cup (N_i \cap P_j^c) \cup (P_i^c \cap N_i^c \cap P_j^c \cap N_j^c).$$

The condition can thus be rewritten as follows, using orthopairs (P_i, N_i) and (P_j, N_j) :

$$P_i \cap N_j = \emptyset \text{ and } P_j \cap N_i = \emptyset$$

Example 2. Consider the disjunction $x^1 \vee x^2$, where $A_1 = x^1$ and $A_2 = x^2$, so that $P_1 = \{1\}, P_2 = \{2\}, N_1 = N_2 = \emptyset$. So $Bel(x^1 \vee x^2) = Bel(x^1) + Bel(x^2) - Bel(x^1 \wedge x^2)$. Likewise for implication, $x^1 \rightarrow x^2 = \neg x^1 \vee x^2$, where $A_1 = \neg x^1$ and $A_2 = x^2$, so that $N_1 = \{1\}, P_2 = \{2\}, P_1 = N_2 = \emptyset$. So $Bel(x^1 \rightarrow x^2) = Bel(\neg x^1) + Bel(x^2) - Bel(\neg x^1 \wedge x^2)$.

We can summarize the above results as

Proposition 8. *The set of partial models $\mathcal{A}_n = \{A_1, \dots, A_n\}$ satisfies the inclusion-exclusion principle if and only if, for any pair A_i, A_j one of the two following conditions is satisfied:*

- $\exists p \neq q \in \{1, \dots, D\}$ such that $p, q \in (P_i \cap N_j) \cup (P_j \cap N_i)$.
- $P_i \cap N_j = \emptyset$ and $P_j \cap N_i = \emptyset$

This condition tells us that for any pair of partial models, :

- either conjunctions A_i, A_j contain at least two opposite literals,
- or events A_i, A_j have a non-empty intersection and have a common model.

As a consequence we can compute the belief of any logical formula that obeys the conditions of Proposition 8 in terms of the belief and plausibilities of atoms x^i .

Example 3. Consider the formula $(x^1 \wedge \neg x^2) \vee (\neg x^1 \wedge x^2) \vee x^3$, with $A_1 = x^1 \wedge \neg x^2$, $A_2 = \neg x^1 \wedge x^2$, $A_3 = x^3$. We have $P_1 = \{1\}, N_1 = \{2\}, P_2 = \{2\}, N_2 = \{1\}, P_3 = \{3\}, N_3 = \emptyset$. Thus it satisfies Proposition 8, and

$$\begin{aligned} & Bel((x^1 \wedge \neg x^2) \vee (\neg x^1 \wedge x^2) \vee x^3) \\ &= Bel(x^1 \wedge \neg x^2) + Bel(\neg x^1 \wedge x^2) + Bel(x^3) - Bel(x^1 \wedge \neg x^2 \wedge x^3) - Bel(\neg x^1 \wedge x^2 \wedge x^3) \\ & \text{(other belief values are equal to 0 since referring to contradictory Boolean expressions)} \\ &= l_1(1 - u_2) + (1 - u_1)l_2 + l_3(1 - l_1(1 - u_2) - (1 - u_1)l_2) \end{aligned}$$

The conditions of Proposition 8 allow us to check, once a formula has been put in DNF, whether or not the inclusion-exclusion principle applies. Important particular cases where it applies are disjunctions of partial models C_i having only positive (resp. negative) literals, of the form $C_1 \vee \dots \vee C_n$, where $N_1 = \dots = N_n = \emptyset$ (resp. $P_1 = \dots = P_n = \emptyset$). This is the typical Boolean formula obtained in fault tree analysis, where elementary failures are modelled by positive literals, and the general failure event is due to the simultaneous occurrence of some subsets of elementary failures (see Section 6.1). Namely, we have

$$\begin{aligned} Bel(C_1 \vee \dots \vee C_n) &= \sum_{i=1}^n Bel(C_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n Bel(C_i \wedge C_j) \\ &+ \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n Bel(C_i \wedge C_j \wedge C_k) - \dots + (-1)^{m+1} Bel(C_1 \wedge \dots \wedge C_n), \end{aligned} \quad (16)$$

where the terms on the right-hand side can be computed from belief values of atoms as $Bel(C_{(P,\emptyset)}) = \prod_{i \in P} l^i$ as per Proposition 7.

More generally, the inclusion-exclusion principle applies to disjunctions of partial models which can, via a renaming, be rewritten as a disjunction of conjunctions of positive literals: namely, whenever a single variable never appears in a positive and negative form in two of the conjunctions. This is equivalent to the second condition of Proposition 8. Then, of course, values $1 - u^i$ must be used in place of l^i for negative literals.

For such Boolean formulas, the inclusion-exclusion principle can also be used to also estimate the plausibility of $C_1 \vee \dots \vee C_n$. Indeed, consider the formula $\bigvee_{i \in [1:n]} (\bigwedge_{k \in P_i} x^k)$ possibly obtained after a renaming, then

$$\begin{aligned} \neg \left(\bigvee_{i \in [1:n]} \left(\bigwedge_{k \in P_i} x^k \right) \right) &= \bigwedge_{i \in [1:n]} \neg \left(\bigwedge_{j \in P_i} x^j \right) \\ &= \bigwedge_{i \in [1:n]} \bigvee_{j \in P_i} \neg x^j \\ &= \bigvee_{\vec{k} \in P_1 \times \dots \times P_n} \bigwedge_{j \in [1:n]} \neg x^{k_j} \end{aligned}$$

using distributivity, where \vec{k} ranges on n -tuples of indices (one component per conjunction C_i). Namely, starting with a DNF involving conjunctions of positive literals, $\bigwedge_{i \in [1:n]} \bigvee_{j \in P_i} \neg x^j$ is turned into a DNF with only negative literals, to which the second condition of Proposition 8 applies, and

$$Pl\left(\bigvee_{i \in [1:n]} \left(\bigwedge_{k \in P_i} x^k\right)\right) = 1 - Bel\left(\bigvee_{\vec{k} \in P_1 \times \dots \times P_n} \bigwedge_{j \in [1:n]} \neg x^{k_j}\right).$$

On the other hand, it is not always possible to put the complement of every formula satisfying the second condition of Proposition 8 in a DNF form that also satisfies Proposition 8.

Example 4. Consider the equivalence formula between three elements, that is the formula $F = (\neg x^1 \wedge \neg x^2 \wedge \neg x^3) \vee (x^1 \wedge x^2 \wedge x^3)$. It satisfies Proposition 8, but its negation

$$\neg F = (x^1 \wedge \neg x^2) \vee (x^2 \wedge \neg x^3) \vee (x^3 \wedge \neg x^1), \quad (17)$$

once put in DNF form, does not satisfy Proposition 8 (each pair of conjunctions possesses only one variable with opposite literal). However, the negation of other formulas such as logical equivalence between two elements possesses a DNF form $\neg(x^1 \Leftrightarrow x^2) = (x^1 \wedge \neg x^2) \vee (\neg x^1 \wedge x^2)$ that satisfies Proposition 8.

Example 5. As another example where the inclusion-exclusion principle cannot be applied, consider the formula $x^1 \vee (\neg x^1 \wedge x^2)$ (which is just the disjunction $x^1 \vee x^2$ we already considered above). It does not hold that $Bel(x^1 \vee (\neg x^1 \wedge x^2)) = Bel(x^1) + Bel(\neg x^1 \wedge x^2) = l^1 + (1 - u_1)l^2$. Indeed the latter sum neglects $m(x^2) = (u^1 - l^1)l^2$, since x^2 is a focal element that implies $x^1 \vee x^2$ but neither x^1 nor $\neg x^1 \wedge x^2$. However, computing $Bel(x^1 \vee x^2)$ is obvious as $1 - (1 - l^1)(1 - l^2)$ from Proposition 7.

The last remark suggests that normal forms that are very useful to compute the probability of a Boolean formula efficiently, such as BDD [6] may be useless to speed up the computation of its belief and plausibility degrees. For instance, $x^1 \vee (\neg x^1 \wedge x^2)$ is a binary decision diagram (BDD) for the disjunction, and this form prevents $Bel(x^1 \vee x^2)$ from being properly computed by standard methods as the inclusion-exclusion principle fails in this case. The question whether any Boolean formula can be re-expressed in a form satisfying Proposition 8 is answered to the negative by Formula (17), which provides a counterexample to this claim.

5. The case of events defined by monotone functions

In this section, we show that the inclusion-exclusion principle can be applied to evaluate some events of interest defined by means of monotone functions on Cartesian products of discrete linearly ordered spaces. Such functions are commonly used in problems such as multi-criteria decision making [20], reliability assessments [14] or optimization problems [19].

We assume that we have some function $\phi : \mathcal{X}^{1:D} \rightarrow \mathcal{Y}$ where variables x^j , $j = 1, \dots, D$ take their values on a finite linearly ordered space $\mathcal{X}^j = \{x_1^j, \dots, x_{k_j}^j\}$ of k_j elements. We denote by \leq_j the order relation on \mathcal{X}^j and assume (without loss of generality) that elements are indexed such that $x_i^j <_j x_k^j$ iff $i < k$. We also assume that

the output space \mathcal{Y} is ordered and we denote by $\leq_{\mathcal{Y}}$ the order on \mathcal{Y} , assuming an indexing such that $y_i <_{\mathcal{Y}} y_k$ iff $i < k$. Given two elements $x, z \in \mathcal{X}^{1:D}$, we simply write $x \leq z$ if $x^j \leq_j z^j$ for $j = 1, \dots, n$, and $x < z$ if moreover $x^j <_j z^j$ for at least one j .

We assume that the function is non-decreasing in each of its variables x^j , that is

$$\phi(x_{i_1}^1, \dots, x_{i_\ell}^\ell, \dots, x_{i_D}^D) \leq_{\mathcal{Y}} \phi(x_{i'_1}^1, \dots, x_{i'_\ell}^\ell, \dots, x_{i'_D}^D) \quad (18)$$

iff $i_\ell \leq i'_\ell$. Note that a function monotone in each variable x^j can always be transformed into a non-decreasing one, since if ϕ is decreasing in X_i , it becomes non-decreasing in x^i when considering the reverse ordering of \leq_j (i.e., $x_i^j < x_k^j$ iff $k < i$).

We now consider the problem where we want to estimate the uncertainty of some event $\{\phi \geq d\}$ (or $\{\phi < d\}$, that can be obtained by duality). Evaluating the uncertainty over such events is instrumental in a number of applications, from checking whether a threshold can be trespassed in risk analysis [3] to computing level sets when solving the Choquet integral, e.g., in multi-criteria decision making [20]. Given a value $d \in \mathcal{Y}$, let us define the concept of minimal path and minimal cut vectors.

Definition 1. A minimal path (MP) vector p for value d , induced by a function ϕ , is an element $p \in \mathcal{X}^{1:D}$ such that $\phi(p) \geq d$ and $\phi(y) < d$ for any $y < p$.

Definition 2. A minimal cut (MC) vector c for value d , induced by a function ϕ , is an element $c \in \mathcal{X}^{1:D}$ such that $\phi(c) < d$ and $\phi(y) \geq d$ for any $y > c$.

Let $\{p_1, \dots, p_n\}$ be the set of all minimal path vectors of some function ϕ for a given threshold demand d . We denote by $A_{p_i} = \{x \in \mathcal{X}^{1:D} | x \geq p_i\}$ the event corresponding to the set of configurations dominating the minimal path vector p_i and by $\mathcal{A}_n = \{A_{p_1}, \dots, A_{p_n}\}$ the set of events induced by minimal path vectors. Note that each set

$$A_{p_i} = \times_{j=1}^D \{x^j | x^j \geq_j p_i^j\} \quad (19)$$

is rectangular, hence we can use results from Section 3.

Lemma 5. *The collection of rectangular sets \mathcal{A}_n induced by minimal path vectors satisfies Theorem 5*

Proof. Consider two events A_{p_i}, A_{p_j} induced by minimal path vectors and a dimension ℓ , then clearly either $\{x^\ell \geq_\ell p_i^\ell\} \subseteq \{x^\ell \geq_\ell p_j^\ell\}$ ($A_{p_i}^\ell \subseteq A_{p_j}^\ell$) or $\{x^\ell \geq_\ell p_i^\ell\} \supseteq \{x^\ell \geq_\ell p_j^\ell\}$ ($A_{p_i}^\ell \supseteq A_{p_j}^\ell$). \square

It can be checked that $\{x \in \mathcal{X}^{1:D} | \phi(x) \geq d\} = \cup_{i=1}^n A_{p_i}$. We can therefore write the inclusion-exclusion formula for belief functions:

$$\begin{aligned} Bel(\phi(x) \geq d) &= Bel(A_{p_1} \cup \dots \cup A_{p_n}) = \sum_{\mathcal{J} \subseteq \mathcal{A}_n} (-1)^{|\mathcal{J}|+1} Bel(\cap_{A \in \mathcal{J}} A), \\ &= 1 - Pl(\phi(x) < d) \end{aligned}$$

Under the assumption of random set independence, computing each term simplifies into

$$\begin{aligned} Bel(A_{p_j}) &= \prod_{i=1}^D Bel(\{x^i \geq p_j^i\}) \\ Bel(A_{p_j} \cap \dots \cap A_{p_k}) &= \prod_{i=1}^D Bel(\{x^i \geq \max(p_j^i, \dots, p_k^i)\}) \end{aligned}$$

The computation of $Bel(\phi(x) < d)$ can be carried out similarly by using minimal cut vectors. Let c_1, \dots, c_m be the set of all minimal cut vectors of ϕ for threshold d . Then $A_{c_i} = \{x \in \mathcal{X}^{1:D} | x \leq c_i\} = \times_{j=1}^D \{x^j | x^j \leq_j c_i^j\}$ is rectangular and we have the following result, whose proof is similar to the one of Lemma 5.

Lemma 6. *The collections of rectangular sets \mathcal{A}_m induced by minimal cut vectors satisfy Theorem 5*

Denoting $\mathcal{A}_m = \{A_{c_1}, \dots, A_{c_m}\}$ the set of events induced by minimal cut vectors, we have that $\{x \in \mathcal{X}^{1:D} | \phi(x) < d\} = \cup_{i=1}^m A_{c_i}$, hence applying the inclusion-exclusion formula for belief functions gives

$$\begin{aligned} Bel(\phi(x) < d) &= Bel(A_{c_1} \cup \dots \cup A_{c_m}) = \sum_{\mathcal{J} \subseteq \mathcal{A}_m} (-1)^{|\mathcal{J}|+1} Bel(\cap_{A \in \mathcal{J}} A), \\ &= 1 - Pl(\phi(x) \geq d). \end{aligned}$$

This also shows that the inclusion-exclusion formula can be used to estimate both belief and plausibilities of events of the type $\{\phi \geq d\}$ and $\{\phi < d\}$ when ϕ is monotone.

6. Application in reliability analysis

In this section, we illustrate how our results can be used in the particular field of reliability analysis, as this field is typically concerned with monotone and potentially large systems for which marginal uncertainty models are specified on components. We first deal with binary systems before addressing the case of Multi-State Systems (MSS).

6.1. Reliability of binary systems

In classical system reliability, \mathcal{X}^i are the states of some component, and the system state depends on the joint state of elements. We will consider the most common case in which $\mathcal{X}^i = \{x^i, \neg x^i\}$ is binary, x^i being a Boolean variable meaning that component i works, $\neg x^i$ that it failed. The structure function $\phi : \mathcal{X} \rightarrow \{0, 1\}$ specifies when the system works ($\phi = 1$) and when it does not. The problem is then to evaluate, from the joint mass m , the values $Bel(\phi^{-1}(1)) = 1 - Pl(\phi^{-1}(0))$ and $Bel(\phi^{-1}(0)) = 1 - Pl(\phi^{-1}(1))$. This is a special case of the one addressed in Section 4.

In the binary case, a minimal path p can be expressed as a subset $S_p \subseteq [1 : D]$ (the counterpart of a minimal vector path in Section 5). It indicates a minimal set of elements that must be in working state, in the sense that if only those components are working, then the system is guaranteed to work but will fail if one of them fails. For

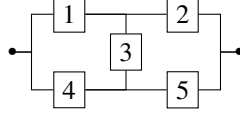


Figure 7: Bridge system

example, $S_p = \{1, 2\}$ states that $\phi(x^1, x^2, \dots) = 1$ whatever the values of the other components. Note that two minimal paths p_1 and p_2 are such that $S_{p_1} \not\subseteq S_{p_2}$ and $S_{p_2} \not\subseteq S_{p_1}$, otherwise one of the two is not minimal.

A minimal path p specifies a partial model (a conjunction of literals) A_p with no negative literal (an orthopair $(P, N) = (S_p, \emptyset)$ in the notations of Section 4). If p_1, \dots, p_n are the minimal paths of a system, then $\phi^{-1}(1)$ is the disjunction $A_{p_1} \vee \dots \vee A_{p_n}$. This means that computing our belief in the fact that a system will work is given by

$$Bel(\phi^{-1}(1)) = Bel(A_{p_1} \vee \dots \vee A_{p_n}). \quad (20)$$

That $A_{p_1} \vee \dots \vee A_{p_n}$ satisfies Proposition 8 is immediate, as only positive literals appear in the formulas.

The same reasoning can be carried out for minimal cuts c to obtain $Bel(\phi^{-1}(0))$. In this case we can specify a set of m minimal cuts, a minimal cut c being encoded as a subset $S_c \subseteq [1 : D]$ indicating a minimal set of components such that, if all of them fail, then the system is guaranteed to fail. For example, $S_c = \{1, 2\}$ states that $\phi(-x^1, -x^2, \dots) = 0$ whatever the values of the other components. As for minimal paths, a minimal cut c specifies partial models (P, N) such that $P = \emptyset$ and $N = S_c$.

Given the minimal paths p_1, \dots, p_n and cuts c_1, \dots, c_m of a system, the whole reliability of the system can be computed as

$$Bel(\phi^{-1}(1)) = \sum_{\mathcal{J} \subseteq [1:n]} (-1)^{|\mathcal{J}|+1} Bel(\cap_{i \in \mathcal{J}} A_{p_i}), \quad (21)$$

$$Bel(\phi^{-1}(0)) = \sum_{\mathcal{J} \subseteq [1:m]} (-1)^{|\mathcal{J}|+1} Bel(\cap_{i \in \mathcal{J}} A_{c_i}) \quad (22)$$

These equations are particularly easy to evaluate using Equation (14).

Example 6. Let us take the example of a bridge system, pictured in Figure 7. Note that this system is complex, in the sense that it cannot be reduced to a set of parallel and series connections.

The minimal paths and cuts of this system are the following (we only write p_i for S_{p_i} , etc.):

$$\begin{array}{ll} p_1 = \{1, 2\} & c_1 = \{1, 4\} \\ p_2 = \{4, 5\} & c_2 = \{2, 5\} \\ p_3 = \{1, 3, 5\} & c_3 = \{1, 3, 5\} \\ p_4 = \{2, 3, 4\} & c_4 = \{2, 3, 4\} \end{array}$$

	x^1	x^2	x^3	x^4	x^5
$Pl(x^i)$	0.72	0.79	0.88	0.92	0.91
$Bel(x^i)$	0.70	0.75	0.80	0.85	0.90

Table 1: Binary reliability uncertainty.

The uncertainty information on component states is given in Table 1. Equations (21) and (22) can then be computed efficiently. For example, the degree of belief associated to the event $A_{p_1} \cap A_{p_2}$ induced by the minimal paths p_1, p_2 is

$$Bel(A_{p_1} \cap A_{p_2}) = Bel(\{x_1\})Bel(\{x_2\})Bel(\{x_4\})Bel(\{x_5\}).$$

The final reliability of the bridge system is $[Bel(\phi^{-1}(1)), Pl(\phi^{-1}(1))] = [0.92, 0.96]$ using the results of the two previous sections.

6.2. Multi-State Systems (MSS) reliability

In the previous subsection, we made the usual assumption in system reliability analysis that components can assume 2 states: failed or working. Multi-State Systems (MSS) reliability goes beyond this assumption. It allows each component to be in one of multiple (exclusive) states. For example, a power station may have four different states corresponding to generating electricity at 0, 25, 50, 75 and 100 percent of its full capacity.

In recent years, multi-state system reliability analysis has received considerable attention, yet less than binary systems. The complexity in MSS analysis is due to the non-binary nature of the system and its components. There are many solutions to reduce this complexity, such as Markov methods [27, 7], discrete event simulation, among others. We refer to Lisnianski and Levitin [25] for a detailed review of the problem. MSS analysed in this section are such that

- the components are s-independent, meaning that the occurrence of one component state change event has no influence on the occurrence of the other state change event;
- the states of each component of the MSS are mutually exclusive, i.e. at any time, any component is in one of its states;
- the MSS is coherent (if one state component efficiency increases, the overall efficiency increases).

Let us now show that for such systems, we can define minimal path sets and minimal cut sets that satisfy the inclusion-exclusion principle.

6.2.1. Minimal path sets and minimal cut sets of MSS

In reliability analysis, variables x^j , $j = 1, \dots, D$ correspond to the D components of the system and the value x_i^j is the i th state of component j . Typically, states are ordered according to their performance rates, hence we can assume the spaces \mathcal{X}^j

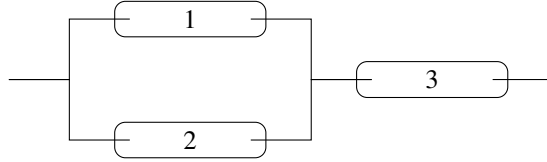


Figure 8: Flow transmission system

to be ordered. $\mathcal{X}^{1:D}$ corresponds to the system states and $\mathcal{Y} = \{y_1, \dots, y_{|\mathcal{Y}|}\}$ is the ordered set of global performance rates of the system.

The structure function $\phi : \mathcal{X}^{1:D} \rightarrow \mathcal{Y}$ maps the system states to the global system performance. Our assumption that the system is coherent means that the function ϕ is non-decreasing, in the sense of Equation (18). Note that, in this work and for simplicity, the term “multi-state system” is used to designate systems where both components and system performance take several possible states (such systems being usually called multi-state systems with multi-state components).

As the structure function of a MSS is monotone, we can directly apply the results from Section 5 to estimate uncertainty bounds about the event $\{\phi \geq d\}$, where $d \in \mathcal{Y}$. Estimating such an uncertainty is a typical task in multi-state reliability analysis, as it amounts to estimating the degree of certainty that a system will guarantee a level d of performance. This is illustrated in the next section.

6.2.2. Example

Let us now illustrate our approach on a complete example, inspired from Ding and Lisnianski [15]. The results of the belief function approach will be compared to the ones obtained using the UGF¹ probability interval approach proposed by Li *et al.* [24]².

In this example, we aim to evaluate the availability of a flow transmission system design presented in Figure 8 and made of three pipes. The flow is transmitted from left to right and the performance of a pipe is measured by its transmission capacity (tons per minute). It is supposed that elements 1 and 2 have three states: a state of total failure corresponding to a capacity of 0, a state of full capacity and a state of partial failure. Element 3 only has two states: a state of total failure and a state of full capacity. All performance levels are precise.

The state performance levels and the state probabilities of the flow transmitter system are given in Table 3. In Li *et al.* [24], these probabilities are obtained with the imprecise Dirichlet model [5]. We aim to estimate the availability of the system when $d = 1.5$.

The minimal paths are

$$p_1 = \{x_1^1, x_2^2, x_2^3\} = \{0, 1.5, 4\}, \quad p_2 = \{x_3^1, x_1^2, x_2^3\} = \{1.5, 0, 4\}.$$

¹Universal Generating Function

²A comparison of Li *et al.* [24] and Ding and Lisnianski [15] can be found in Li *et al.* [24]

x^1	0	0	0	0	0	0	1	1	1	1	1	1	1.5	1.5	1.5	1.5	1.5	1.5
x^2	0	0	1.5	1.5	2	2	0	0	1.5	1.5	2	2	0	0	1.5	1.5	2	2
x^3	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4
$y = \phi(x^1, x^2, x^3)$	0	0	0	1.5	0	2	0	1	0	2.5	0	3	0	1.5	0	3	0	3.5

Table 2: Performance rates of the oil transmission system

The set A_{p_1} and A_{p_2} of vectors a such that $a \geq p_1$, $b \geq p_2$ are

$$A_{p_1} = \{0, 1, 1.5\} \times \{1.5, 2\} \times \{4\} \text{ and } A_{p_2} = \{1.5\} \times \{0, 1.5, 2\} \times \{4\},$$

and their intersection $A_{p_1} \cap A_{p_2}$ of vectors c such that $c \geq p_1 \wedge p_2$ (with $\wedge = \min$) is

$$A_{p_1} \cap A_{p_2} = \{1.5\} \times \{1.5, 2\} \times \{4\}.$$

Applying the inclusion-exclusion formula for a demand level $d = 1.5$, we obtain

$$Bel(\phi \geq 1.5) = Bel(A_{p_1}) + Bel(A_{p_2}) - Bel(A_{p_1} \cap A_{p_2})$$

For example, we have

$$\begin{aligned} Bel(A_{p_1}) &= Bel(\{0, 1, 1.5\} \times \{1.5, 2\} \times \{4\}) \\ &= Bel(\{0, 1, 1.5\}) \cdot Bel(\{1.5, 2\}) \cdot Bel(\{4\}) \\ &= 1 \cdot 0.895 \cdot 0.958 \\ &= 0.8574 \end{aligned}$$

and $Bel(A_{p_2})$, $Bel(A_{p_1} \cap A_{p_2})$ can be computed similarly. Finally we get

$$Bel(\phi \geq 1.5) = 0.8574 + 0.7654 - 0.6851 = 0.9377$$

and by duality with $Bel(\phi < 1.5)$, we get

$$Pl(\phi \geq 1.5) = 1 - Bel(\phi < 1.5) = 0.9523.$$

The availability A_s of the flow transmission system for a demand level $d = 1.5$ is given by $A_s = [0.9377, 0.9523]$. The use of the interval UGF method proposed by Li *et al.* [24] leads to $A_{UGF}^s = [0.9215, 0.9855]$. Note that we always have $[Bel(A), Pl(A)] \subseteq A_{UGF}^s$, as the Li *et al.* approach uses an interval arithmetic approach, which is known to provide quite conservative approximations in the presence of repeated variables (as is often the case when using the inclusion-exclusion principle).

7. Comparison with strong independence

It should be noticed, as shown by Jacob *et al.* [22], that using random set independence should not be confused with an assumption of stochastic independence between ill-known probabilities. In this section, we will compare the previously used notion of random set independence to the one of strong independence, which can be interpreted as a robust version (i.e., applied to sets of probabilities) of the notion of stochastic independence. We will then discuss the effects of using one independence notion in place of the other in the previously treated problems.

Component G_j	1	2	3
p_1^j	[0.096,0.106]	[0.095,0.105]	-
p_2^j	[0.095,0.105]	[0.195,0.205]	[0.032,0.042]
p_3^j	[0.799,0.809]	[0.7,0.71]	[0.958,0.968]
g_1^j	0	0	-
g_2^j	1	1.5	0
g_3^j	1.5	2	4

Table 3: Parameters of the flow transmission system

7.1. Two distinct independence notions

So far, we have mainly considered that the joint mass over $\mathcal{X}^{1:D}$ was obtained by combining marginal masses m^i using Equation (10). It corresponds to the notion of random set independence. Yet, within the imprecise probabilistic literature, there are many other notions of independence available [10, 9], and it is out of the scope of this paper to discuss all of them. In this section, we will compare our results with those that would be obtained using a robust version of stochastic independence, usually called *strong independence*.

If $\mathcal{P}(Bel^i)$ denotes the set of probabilities compatible with m^i on dimension i , then the joint model \mathcal{P}_{SI} obtained by applying stochastic independence to elements of $\mathcal{P}(Bel^i), i = 1, \dots, n$ is

$$\mathcal{P}_{SI} = \left\{ \prod_{i=1}^D P^i \mid P^i \in \mathcal{P}(Bel^i) \right\}. \quad (23)$$

Assuming strong independence consists in representing our knowledge by means of the convex hull of \mathcal{P}_{SI} , that we will denote by $\mathcal{P}_{\overline{SI}}$. Particularly interesting elements of $\mathcal{P}_{\overline{SI}}$ are its *extreme points*, that are obtained by computing the product of extreme points of $\mathcal{P}(Bel^i), i = 1, \dots, n$ (hence extreme points of $\mathcal{P}_{\overline{SI}}$ are also in \mathcal{P}_{SI}).

In the particular case where a probability set $\mathcal{P}(Bel)$ is induced by a mass function m on $2^{\mathcal{X}}$, its extreme points can be obtained in the following way: specify (select), for each focal element $A \in \mathcal{F}_m$, an element $s_A \in A$ and take the probability measure P such that $P(\{x\}) = \sum_{A \in \mathcal{F}_m} m(A) \mathbf{1}_{(s_A=x)}$. This comes down to taking a convex mixture of Dirac measures located at $s_A \in A$, weighted by masses $m(A)$.

Let us denote by $\mathcal{P}(Bel^{1:D})$ the set of probabilities induced by considering the joint mass $m^{1:D}$ obtained by Equation (10). This corresponds to the random set independence assumption. To build extreme points $\mathcal{P}(Bel^{1:D})$, a D -tuple s_E has to be specified (selected) in each set $E = \times_{i=1}^D E^i$ with $E^i \in \mathcal{F}_{m^i}$, while to build extreme points of $\mathcal{P}_{\overline{SI}}$, one has to specify elements s_{E^i} within each marginal model m^i , then take the product of corresponding probabilities. One can check [18] that the latter

construction is more constrained than the former, hence $\mathcal{P}_{\overline{SI}} \subseteq \mathcal{P}(Bel^{1:D})$ (already in Couso *et al.* [9]). Among other things, this implies that the lower probabilities

$$\underline{P}_{\overline{SI}}(A) = \inf_{P \in \mathcal{P}_{\overline{SI}}} P(A) \text{ and } \underline{P}_{m^{1:D}}(A) = Bel^{1:D}(A) = \inf_{P \in \mathcal{P}(Bel^{1:D})} P(A)$$

are such that $\underline{P}_{\overline{SI}}(A) \geq \underline{P}_{m^{1:D}}(A)$ for any $A \subseteq \mathcal{X}^{1:D}$, meaning that random set independence can be used to outer-approximate strong independence. Also recall that such lower probabilities are obtained for extreme points of the set, hence for such inferences working with \mathcal{P}_{SI} or its convex closure $\mathcal{P}_{\overline{SI}}$ makes no difference.

The two notions also have different interpretations: random set independence can be associated to an independence of sources providing the uncertainty, making no claim about the possible interaction between variables of different marginal spaces, while strong independence can be interpreted as an extension of stochastic independence between random variables when probabilities are partially known. The next example illustrates the inequality $\underline{P}_{\overline{SI}}(A) \geq \underline{P}_{m^{1:D}}(A)$ as well as how the selection process to obtain a probability reaching the lower bound is different in the two cases.

Example 7. Let $\mathcal{X}^1 = \{x^1, \neg x^1\}$ and $\mathcal{X}^2 = \{x^2, \neg x^2\}$ be two Boolean frames with $[l^1, u^1] = [0.6, 0.8]$ and $[l^2, u^2] = [0.2, 0.4]$ corresponding to the masses

$$\begin{aligned} m^1(x^1) &= 0.6, & m^1(\neg x^1) &= 0.2, & m^1(\mathcal{X}^1) &= 0.2, \\ m^2(x^2) &= 0.2, & m^2(\neg x^2) &= 0.6, & m^2(\mathcal{X}^2) &= 0.2, \end{aligned}$$

which themselves induce probability sets $\mathcal{P}(Bel^1)$ and $\mathcal{P}(Bel^2)$. Consider now the problem of finding the lower probability of the event $E = \{(x^1, \neg x^2), (\neg x^1, x^2)\}$ (corresponding to the Boolean formula $(x^1 \wedge \neg x^2) \vee (\neg x^1 \wedge x^2)$). In the case of strong independence, this comes down to find the extreme points within $\mathcal{P}(Bel^1)$ and $\mathcal{P}(Bel^2)$ whose stochastic product induces the lowest value on $P(E)$. This is obtained by considering $p^1(x^1) = l^1 = 1 - p^1(\neg x^1) = 0.6$ and $p^2(x^2) = u^2 = 1 - p^2(\neg x^2) = 0.4$ that induce

$$\underline{P}_{\overline{SI}}(E) = p^1(x^1) \cdot p^2(\neg x^2) + p^1(\neg x^1) \cdot p^2(x^2) = 0.6 \cdot 0.6 + 0.4 \cdot 0.4 = 0.52.$$

Within the focal elements of m^1 and m^2 , these two extreme points corresponds to selections

- $s_{\mathcal{X}^1} = \neg x^1$ that transfers $m^1(\mathcal{X}^1)$ to $\neg x^1$ and
- $s_{\mathcal{X}^2} = x^2$ that transfers $m^2(\mathcal{X}^2)$ to x^2 .

When considering random set independence, one first have to build the joint random set $m^{1:2}$ such that

$$\begin{aligned} m^{1:2}(x^1 \times x^2) &= 0.12, & m^1 : 2(\neg x^1 \times x^2) &= 0.04, & m^1 : 2(\mathcal{X}^1 \times x^2) &= 0.04, \\ m^{1:2}(x^1 \times \neg x^2) &= 0.36, & m^1 : 2(\neg x^1 \times \neg x^2) &= 0.12, & m^1 : 2(\mathcal{X}^1 \times \neg x^2) &= 0.12, \\ m^{1:2}(\neg x^1 \times x^2) &= 0.12, & m^1 : 2(\neg x^1 \times \mathcal{X}^2) &= 0.04, & m^1 : 2(\mathcal{X}^1 \times \mathcal{X}^2) &= 0.04, \end{aligned}$$

and that induces the credal set $\mathcal{P}(Bel^{1:2})$. We have $\underline{P}_{m^{1:2}}(E) = Bel^{1:2}(E) = m^{1:2}(\{x^1\} \times \{\neg x^2\}) + m^1(\neg x^1 \times \{x^2\}) = 0.4$ and we indeed have $\underline{P}_{\overline{SI}}(E) > \underline{P}_{m^{1:D}}(E)$. The extreme point of $\mathcal{P}(Bel^{1:2})$ whose probability $P(E) = \underline{P}_{m^{1:2}}(E)$ is obtained for the selection

- $s_{x^1 \times \mathcal{X}^2} = (x^1, x^2)$ that transfers $m^{1:2}(x^1 \times \mathcal{X}^2)$ to (x^1, x^2) ,
- $s_{\neg x^1 \times \mathcal{X}^2} = (\neg x^1, \neg x^2)$ that transfers $m^{1:2}(\neg x^1 \times \mathcal{X}^2)$ to $(\neg x^1, \neg x^2)$,
- $s_{\mathcal{X}^1 \times x^2} = (x^1, x^2)$ that transfers $m^{1:2}(\mathcal{X}^1 \times x^2)$ to (x^1, x^2) ,
- $s_{\mathcal{X}^1 \times \neg x^2} = (\neg x^1, \neg x^2)$ that transfers $m^{1:2}(\mathcal{X}^1 \times \neg x^2)$ to $(\neg x^1, \neg x^2)$ and
- $s_{\mathcal{X}^1 \times \mathcal{X}^2} = (x^1, x^2)$ that transfers $m^{1:2}(\mathcal{X}^1 \times \mathcal{X}^2)$ to (x^1, x^2) .

The joint probability $p^{1:2}$ resulting from this selection is

$$p^{1:2}(x^1 \times x^2) = 0.32, \quad p^{1:2}(\neg x^1 \times x^2) = 0.04,$$

$$p^{1:2}(x^1 \times \neg x^2) = 0.36, \quad p^{1:2}(\neg x^1 \times \neg x^2) = 0.28,$$

which cannot be expressed as a product of extreme points of $\mathcal{P}(Bel^1)$ and $\mathcal{P}(Bel^2)$, and is therefore not included in \mathcal{P}_{SI} .

Example 7 clearly shows the difference of meaning between the two notions: in the strong independence case, fixing the element of \mathcal{X}^1 does not influence the selection on \mathcal{X}^2 , while in the random set independence case, obtaining the lower bound implies considering a very strong relation between the selections (e.g., fixing x^1 implies selecting x^2 whenever possible).

7.2. Consequences for Boolean formulas

Assume we have a formula F in DNF $A_1 \vee \dots \vee A_n$ with $A_i \wedge A_j = \perp$, i.e. the sets of models of A_i, A_j are disjoint. Such formulas, in the form of a disjunction of exclusive conjunctions of literals (it can be at worst, just the disjunction of models of F) can be obtained by using the Shannon decomposition of F (the basic notion from which BDDs are derived). In this case, assuming stochastic independence between variables/atoms, the probability of F reads

$$P(A_1 \vee \dots \vee A_n) = P(F) = \sum_{i=1}^n \left[\prod_{j \in P_i} P(x^j) \prod_{j \notin N_i} (1 - P(x^j)) \right]. \quad (24)$$

When the probabilities $P(x^i) \in [l^i, u^i]$ of atoms are incompletely known, bounds $[P_*(F), P^*(F)]$ for $P(F)$ can be obtained by interval analysis of (24) [21]. As Equation (24) is a multilinear function³, it is locally monotone in each of its variables (it is either increasing or decreasing in $P(x^i)$ once the probabilities of other atoms are fixed). This means that each bound $P_*(F), P^*(F)$ is attained for some vertex of the hypercube $\times_{i=1}^n [l^i, u^i]$. We have that $[P_*(F), P^*(F)] = [\underline{P}_{SI}(F), \overline{P}_{SI}(F)]$, since selecting the right vertices comes down to making the right selection for each marginal, selection l^i corresponding to $s_{\mathcal{X}^i} = \neg x^i$ and u^i to $s_{\mathcal{X}^i} = x^i$.

This also means that, in practice, we will have $[P_*(F), P^*(F)] = [\underline{P}_{SI}(F), \overline{P}_{SI}(F)] \subseteq [Bel^{1:D}(F), Pl^{1:D}(F)]$. However, a noticeable exception is when each variable will always appear either in a positive or negative way in the expression of a Boolean formula.

³A multivariate function is multilinear if it is linear in each of its variables.

Proposition 9. *If the logical expression F is a disjunction of conjunctive terms $\{A_1, \dots, A_n\}$ such that $\forall i \neq j, P_i \cap N_j = \emptyset$ and $P_j \cap N_i = \emptyset$, then $[Bel^{1:D}(F), Pl^{1:D}(F)] = [P_{\overline{SI}}(F), \overline{P}_{\overline{SI}}(F)]$*

Proof. Without loss of generality, assume that all variables appear in a positive way (we just have to rename those appearing in a negative way). Consider the case of two dimensions, then the lower bound $P_{\overline{SI}}(F)$ is reached by selecting $s_{\mathcal{X}^i} = \neg x^i$ for $i = 1, 2$. Similarly, the extreme point reaching $Bel^{1:2}(F)$ for any joint set $E = E^1 \times E^2$ is obtained by selecting negative literals whenever possible.

A similar reasoning can be used for higher dimensions and for the upper bounds. \square

Such a situation occurs for connectives like the conjunction, disjunction or implication, and more generally for all expressions that obey condition 2 of Proposition 8.

In such cases the choice of the dependency assumption (between variables or sources) has no influence on the output interval. The fact that in such cases the same results are obtained by both approaches does not make the belief function analysis redundant: it shows that the results induced by the stochastic independence assumption are valid even when this assumption is relaxed (the independence assumption of mass functions is indeed weaker), for some kinds of Boolean formulas.

On the contrary, expressions satisfying condition 1 of Proposition 8 correspond to non-monotonic functions, like for the *Equivalence* and the *Exclusive Or* [21]. In this case, an exhaustive computation for all combinations of interval boundaries must be carried out in the case of interval analysis and strong independence, while the computation of the belief and plausibility are still very simple, but provably less precise than the result of interval analysis.

Example 8. Namely, consider the Exclusive Or $x^1 \oplus x^2 = (x^1 \wedge \neg x^2) \vee (\neg x^1 \wedge x^2)$ and $P(x^1) \in [0.3, 0.8]$, $P(x^2) \in [0.4, 0.6]$. Then

$$\begin{aligned} Bel(x^1 \oplus x^2) &= l_1(1 - u_2) + l_2(1 - u_1) = 0.2 \\ Pl(x^1 \oplus x^2) &= 1 - Bel(x^1 \iff x^2) = u_1 + u_2 - l_1 l_2 - u_1 u_2 = 0.8 \\ P_*(x^1 \oplus x^2) &= \min(l_1(1 - u_2) + u_2(1 - l_1), l_1(1 - l_2) + l_2(1 - l_1), \\ &\quad u_1(1 - u_2) + u_2(1 - u_1), l_2(1 - u_1) + l_1(1 - u_2)) = 0.44 \\ P_*(x^1 \oplus x^2) &= \max(l_1(1 - u_2) + u_2(1 - l_1), l_1(1 - l_2) + l_2(1 - l_1), \\ &\quad u_1(1 - u_2) + u_2(1 - u_1), l_2(1 - u_1) + l_1(1 - u_2)) = 0.56 \end{aligned}$$

Example 7 is also of this kind.

7.3. Extension to the multivariate case

Let us now deal with the non-binary case and with conditions of Theorem 5. Using the fact that $\mathcal{P}_{\overline{SI}} \subseteq \mathcal{P}(Bel^{1:D})$ and Example 7, it is clear that bounds computed using $\mathcal{P}_{\overline{SI}}$ and $\mathcal{P}(Bel^{1:D})$ will not coincide if events in \mathcal{A}_n satisfy condition 2 of Theorem 5, as they already fail to coincide in the binary case. In the case of the first condition, however, we can give a result similar to Proposition 9.

Proposition 10. *Let \mathcal{A}_n be a collection of events satisfying condition 2 of Theorem 5. Then, the following equality*

$$[Bel^{1:D}(\bigcup_{i=1}^n A_i), Pl^{1:D}(\bigcup_{i=1}^n A_i)] = [P_{\overline{SI}}(\bigcup_{i=1}^n A_i), \overline{P}_{\overline{SI}}(\bigcup_{i=1}^n A_i)]$$

holds, with $Bel^{1:D}$, $Pl^{1:D}$ and $P_{\overline{SI}}, \overline{P}_{\overline{SI}}$ the lower/upper bounds obtained using, respectively, the joint models $\mathcal{P}(Bel^{1:D})$ and $\mathcal{P}_{\overline{SI}}$.

Proof. If \mathcal{A}_n satisfies condition 2 of Theorem 5, we have that

$$Bel^{1:D}(\bigcup_{i=1}^n A_i) = \sum_{\mathcal{S} \subseteq \mathcal{A}_n} (-1)^{|\mathcal{S}|+1} Bel(\bigcap_{A \in \mathcal{S}} A) = \sum_{\mathcal{S} \subseteq \mathcal{A}_n} (-1)^{|\mathcal{S}|+1} \prod_{j=1}^D Bel^j(\bigcap_{A \in \mathcal{S}} A^j). \quad (25)$$

Given the relation recalled in Section 7.1 between random set independence and strong independence, this expression coincides with $P_{\overline{SI}}(\bigcup_{i=1}^n A_i)$ if and only if the lower bounds $Bel^j(\bigcap_{A \in \mathcal{S}} A^j)$ on a given dimension \mathcal{X}^j are all be obtained using the same extreme point of the marginal model $\mathcal{P}(Bel^j)$, irrespectively of the subset \mathcal{S} . This comes down to showing that these lower bounds can be attained, for any subset \mathcal{S} , by a unique selection s_{E_k} in the focal elements $E_k \in \mathcal{F}_{m^j}$ of m^j .

Consider a given dimension j . According to condition 2 of Theorem 5, the sets A_i^j are nested (since for a given j and two $i, k \in [1 : D]$ either $A_i^j \subseteq A_k^j$ or $A_i^j \supseteq A_k^j$). Let us re-order them increasingly according to the permutation denoted by $()$ so that

$$\emptyset = A_{(0)}^j \subseteq A_{(1)}^j \subseteq \dots \subseteq A_{(n)}^j \subseteq A_{(n+1)}^j = \mathcal{X}^j$$

Consider now a focal element E_k of \mathcal{F}_{m^j} , then there is some i such that $A_{(i)}^j \subset E_k \subseteq A_{(i+1)}^j$, and we then choose a selection $s_{E_k} \in A_{(i+1)}^j \setminus A_{(i)}^j$. This implies that $s_{E_k} \in A_{(i)}$ if and only if $E_k \subseteq A_{(i)}$. If we do such a selection for any E_k and take the corresponding probability measure P^j , we have that

$$P^j(A_i^j) = \sum_{s_{E_k} \in A_i^j} m^j(E_k) = \sum_{E_k \subseteq A_i^j} m^j(E_k) = Bel^j(A_i^j).$$

Since P^j corresponds to a single extreme point of $\mathcal{P}(Bel^j)$ that does not depend on the subset \mathcal{S} , this finishes the proof. \square

Although such a situation will not always appear, this result directly applies to the case of monotone functions ϕ treated in Section 5, showing that in this particular case, choosing between assumptions of random set independence or of strong independence will not change our inferences about events of the kind $\{\phi \geq d\}$ and $\{\phi < d\}$. In practice, this means that in those cases we can either use tools originating from evidence theory, imprecise probability theory or interval analysis to carry out computations (whichever is the most suited to the situation).

8. Conclusion

In this work, we have studied families of events for which the principle of exclusion/inclusion applies to belief functions. Although the framework we have retained may look restrictive at first glance, it can be applied to a number of practical situations, and we have shown that one particular application is the evaluation of system reliability (both in the binary and multi-state cases).

Such results facilitate computations and are particularly useful when probabilistic data are imprecise. An interesting perspective to this study is to look for conditions under which other uncertainty theories (e.g., general lower probabilities) satisfy the exclusion/inclusion principle. Further potential applications of our results include the study of other problems of reliability analysis, such as importance measures used to detect critical components.

Acknowledgments

This work was partially carried out in the framework of the Labex MS2T (Reference ANR-11-IDEX-0004-02). C. Jacob acknowledges support by @MOST Prototype, a joint project of Airbus, IRIT, LAAS, ONERA and ISAE.

- [1] F. Aguirre, C. Jacob, S. Destercke, D. Dubois, and M. Sallak. Inclusion/exclusion principle for belief functions. In *Proc. of the Eighth International Symposium on Imprecise Probability: Theories and Applications*, pages 3–13, Compiègne, France, 2013.
- [2] C. Baudrit and D. Dubois. Comparing methods for joint objective and subjective uncertainty propagation with an example in a risk assessment. In *Proc. Fourth International Symposium on Imprecise Probabilities and Their Application (ISIPTA'05)*, pages 31–40, Pittsburg (USA, Pennsylvania), 2005.
- [3] C. Baudrit, D. Guyonnet, and D. Dubois. Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. *IEEE Trans. Fuzzy Systems*, 14:593–608, 2006.
- [4] Y. Ben-Haim. *Info-gap Decision Theory: Decisions Under Severe Uncertainty*. Academic Press, 2006.
- [5] J.-M. Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39(2):123–150, 2005.
- [6] R. Bryant. Symbolic Boolean manipulation with ordered binary-decision diagrams. *ACM Computing Surveys (CSUR)*, 24(3):293–318, 1992.
- [7] Y.-R. Chang, S. V. Amari, and S.-Y. Kuo. Reliability evaluation of multi-state systems subject to imperfect coverage using OBDD. In *Proc. of the 2002 Pacific Rim International Symposium on Dependable Computing (PRDC'02)*, pp. 193–200, 2002.

- [8] D. Ciucci. Orthopairs: A simple and widely used way to model uncertainty. *Fundam. Inform.*, 108(3-4):287–304, 2011.
- [9] I. Couso and S. Moral. Independence concepts in evidence theory. *Int. J. Approx. Reasoning*, 51(7):748–758, 2010.
- [10] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.
- [11] L. De Campos, J. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2(02):167–196, 1994.
- [12] A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
- [13] S. Destercke. Independence and 2-monotonicity: Nice to have, hard to keep. *International Journal of Approximate Reasoning*, 54:478–490, 2013.
- [14] S. Destercke and M. Sallak. An extension of universal generating function in multi-state systems considering epistemic uncertainties. *IEEE Transactions on Reliability*, 62(2):504–514, 2013.
- [15] Y. Ding, M. J. Zuo, A. Lisnianski, and Z. G. Tian. Fuzzy multi-state system: General definition and performance assessment. *IEEE Transactions on Reliability*, 57:589 – 594, 2008.
- [16] D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.
- [17] D. Dubois and H. Prade. Formal representations of uncertainty. In D. Bouyssou, D. Dubois, M. Pirlot, and H. Prade, editors, *Decision-making - Concepts and Methods*, chapter 3, pages 85–156. Wiley, <http://www.wiley.com/>, 2009.
- [18] T. Fetz. Sets of joint probability measures generated by weighted marginal focal sets. In F. Cozman, R. Nau, and T. Seidenfeld, editors, *Proc. 2nd International Symposium on Imprecise Probabilities and Their Applications*, 2001.
- [19] J. Fortin, P. Zielinski, D. Dubois, and H. Fargier. Criticality analysis of activity networks under interval uncertainty. *J. Scheduling*, 13(6):609–627, 2010.
- [20] M. Grabisch and C. Labreuche. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *4OR: A Quarterly Journal of Operations Research*, 6(1):1–44, 2008.
- [21] C. Jacob, D. Dubois, and J. Cardoso. Uncertainty handling in quantitative BDD-based fault-tree analysis by interval computation. In *SUM*, pages 205–218, 2011.
- [22] C. Jacob, D. Dubois, and J. Cardoso. Evaluating the uncertainty of a Boolean formula with belief functions. *Advances in Computational Intelligence*, pages 521–531, 2012.

- [23] E. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.
- [24] C.-Y. Li, X. Chen, X.-S. Yi, and J. y. Tao. Interval-valued reliability analysis of multi-state systems. *IEEE Transactions on Reliability*, 60:323 – 330, 2011.
- [25] A. Lisnianski and G. Levitin. *Multi-State System Reliability: Assessment, Optimization and Applications*. World Scientific Publishing Co Pte Ltd, 2003.
- [26] G. Shafer. *A mathematical Theory of Evidence*. Princeton University Press, New Jersey, 1976.
- [27] I. Ushakov. *Handbook of Reliability Engineering*. Wiley-Interscience, 1994.
- [28] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.
- [29] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.