

# Relating Imprecise Representations of imprecise Probabilities

Sébastien Destercke, Didier Dubois, Eric Chojnacki

► **To cite this version:**

Sébastien Destercke, Didier Dubois, Eric Chojnacki. Relating Imprecise Representations of imprecise Probabilities. fifth international symposium on imprecise probability: theories and applications, 2007, Prague, Czech Republic. pp.155-165, 2007. <irsn-00196674v2>

**HAL Id: irsn-00196674**

**<https://hal-irsn.archives-ouvertes.fr/irsn-00196674v2>**

Submitted on 2 Jan 2008

**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.

*Preprint of the paper presented at the ISIPTA 2007 conference*

# Relating practical representations of imprecise probabilities

**S. Destercke**

Institut de Radioprotection  
et de Sûreté Nucléaire,  
DPAM, SEMIC, LIMSI,  
Cadarache, France  
sebastien.destercke@irsn.fr

**D. Dubois**

Institut de Recherche en  
Informatique de Toulouse (IRIT)  
Toulouse, France  
dubois@irit.fr

**E. Chojnacki**

Institut de Radioprotection  
et de sûreté nucléaire,  
DPAM, SEMIC, LIMSI,  
Cadarache, France  
eric.chojnacki@irsn.fr

## Abstract

There exist many practical representations of probability families that make them easier to handle. Among them are random sets, possibility distributions, probability intervals, Ferson's p-boxes and Neumaier's clouds. Both for theoretical and practical considerations, it is important to know whether one representation has the same expressive power than other ones, or can be approximated by other ones. In this paper, we mainly study the relationships between the two latter representations and the three other ones.

**Keywords.** Random Sets, possibility distributions, probability intervals, p-boxes, clouds.

## 1 Introduction

There are many representations of uncertainty. When considering sets of probabilities as models of uncertainty, the theory of imprecise probabilities (including lower/upper previsions) [26] is the most general framework. It formally encompasses all the representations proposed by other uncertainty theories, regardless of their possible different interpretations.

The more general the theory, the more expressive it can be, and, usually, the more expensive it is from a computational standpoint. Simpler (but less flexible) representations can be useful if judged sufficiently expressive. They are mathematically and computationally easier to handle, and using them can greatly increase efficiency in applications.

Among these simpler representations are random sets [7], possibility distributions [27], probability intervals [2], p-boxes [15] and, more recently, clouds [20, 21]. With such a diversity of simplified representations, it is then natural to compare them from the standpoint of their expressive power. Building formal links between such representations also facilitates a unified handling of uncertainty, especially in propagation techniques exploiting uncertain data modeled

by means of such representations. This is the purpose of the present study. It extends some results by Baudrit and Dubois [1] concerning the relationships between p-boxes and possibility measures.

The paper is structured as follows: the first section briefly recalls the formalism of random sets, possibility distributions and probability intervals, as well as some existing results. Section 3 then focuses on p-boxes, first generalizing the notion of p-boxes to arbitrary finite spaces before studying the relationships of these generalized p-boxes with the three former representations. Finally, section 4 studies the relationships between clouds and the preceding representations. For the reader convenience, longer proofs are put in the appendix.

## 2 Preliminaries

In this paper, we consider that uncertainty is modeled by a family  $\mathcal{P}$  of probability distributions, defined over a finite referential  $X = \{x_1, \dots, x_n\}$ . We also restrict ourselves to families that can be represented by their lower and upper probability bounds, defined as follows:

$$\underline{P}(A) = \inf_{P \in \mathcal{P}} P(A) \text{ and } \overline{P}(A) = \sup_{P \in \mathcal{P}} P(A)$$

Let  $\mathcal{P}_{\underline{P}, \overline{P}} = \{P | \forall A \subseteq X, \underline{P}(A) \leq P(A) \leq \overline{P}(A)\}$ . In general, we have  $\mathcal{P} \subset \mathcal{P}_{\underline{P}, \overline{P}}$ , since  $\mathcal{P}_{\underline{P}, \overline{P}}$  can be seen as a projection of  $\mathcal{P}$  on events. Although they are already restrictions from more general cases, dealing with families  $\mathcal{P}_{\underline{P}, \overline{P}}$  often remains difficult.

### 2.1 Random Sets

Formally, a random set is a mapping  $\Gamma$  from a probability space to the power set  $\wp(X)$  of another space  $X$ , also called a multi-valued mapping. This mapping induces lower and upper probabilities on  $X$  [7]. In the continuous case, the probability space is of-

ten  $[0, 1]$  equipped with Lebesgue measure, and  $\Gamma$  is a point-to-interval mapping.

In the finite case, these lower and upper probabilities are respectively called belief and plausibility measures, and it can be shown that the belief measure is a  $\infty$ -monotone capacity [4]. An alternative (and useful) representation of the random set consists of a normalized distribution of positive masses  $m$  over the power set  $\wp(X)$  s.t.  $\sum_{E \subseteq X} m(E) = 1$  and  $m(\emptyset) = 0$  [24]. A set  $E$  that receives strict positive mass is said to be focal. Belief and plausibility functions are then defined as follows:

$$\begin{aligned} Bel(A) &= \sum_{E, E \subseteq A} m(E) \\ Pl(A) &= 1 - Bel(A^c) = \sum_{E, E \cap A \neq \emptyset} m(E). \end{aligned}$$

The set

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

is the probability family induced by the belief measure.

Although  $2^{|X|}$  values are still needed to fully specify a general random set, the fact that they can be seen as probability distributions over subsets of  $X$  allows for simulation by means of some sampling process.

## 2.2 Possibility distributions

A possibility distribution  $\pi$  [12] is a mapping from  $X$  to the unit interval such that  $\pi(x) = 1$  for some  $x \in X$ . Formally, a possibility distribution is the membership function of a fuzzy set. Several set-functions can be defined from a distribution  $\pi$  [11]:

- $\Pi(A) = \sup_{x \in A} \pi(x)$  (possibility measures);
- $N(A) = 1 - \Pi(A^c)$  (necessity measures);
- $\Delta(A) = \inf_{x \in A} \pi(x)$  (sufficiency measures).

Possibility degrees express the extent to which an event is plausible, i.e., consistent with a possible state of the world, necessity degrees express the certainty of events and sufficiency (also called guaranteed possibility) measures express the extent to which all states of the world where  $A$  occurs are plausible. They apply to so-called guaranteed possibility distributions [11] generally denoted by  $\delta$ .

A possibility degree can be viewed as an upper bound of a probability degree [13]. Let

$$\mathcal{P}_\pi = \{P, \forall A \subseteq X, N(A) \leq P(A) \leq \Pi(A)\}$$

be the set of probability measures encoded by a possibility distribution  $\pi$ . A possibility distribution is

also equivalent to a random set whose realizations are nested.

From a practical standpoint, possibility distributions are the simplest representation of imprecise probabilities (as for precise probabilities, only  $|X|$  values are needed to specify them). Another important point is their interpretation in term of collection of confidence intervals [10], which facilitates their elicitation and makes them natural candidate for vague probability assessments (see [5]).

## 2.3 Probability intervals

Probability intervals are defined as lower and upper probability bounds restricted to singletons  $x_i$ . They can be seen as a collection of intervals  $L = \{[l_i, u_i], i = 1, \dots, n\}$  defining a probability family:

$$\mathcal{P}_L = \{P | l_i \leq p(x_i) \leq u_i \forall x_i \in X\}.$$

Such families have been extensively studied in [2] by De Campos et al.

In this paper, we consider non-empty families (i.e.  $\mathcal{P}_L \neq \emptyset$ ) that are reachable (i.e. each lower or upper bound on singletons can be reached by at least one distribution of the family  $\mathcal{P}_L$ ). Conditions of non-emptiness and reachability respectively correspond to avoiding sure loss and achieving coherence in Walley's behavioural theory.

Given intervals  $L$ , lower and upper probabilities  $\underline{P}(A), \overline{P}(A)$  are calculated by the following expressions

$$\begin{aligned} \underline{P}(A) &= \max(\sum_{x_i \in A} l_i, 1 - \sum_{x_i \notin A} u_i) \\ \overline{P}(A) &= \min(\sum_{x_i \in A} u_i, 1 - \sum_{x_i \notin A} l_i) \end{aligned} \quad (1)$$

De Campos et al. have shown that these bounds are Choquet capacities of order 2 ( $\underline{P}$  is a convex capacity).

The problem of approximating  $\mathcal{P}_L$  by a random set has been treated in [17] and [8]. While in [17], Lemmer and Kyburg find a random set  $m_1$  that is an inner approximation of  $\mathcal{P}_L$  s.t.  $Bel_1(x_i) = l_i$  and  $Pl_1(x_i) = u_i$ , Denoeux [8] extensively studies methods to build a random set that is an outer approximation of  $\mathcal{P}_L$ . The problem of finding a possibility distribution approximating  $\mathcal{P}_L$  is treated by Masson and Denoeux in [19].

Two common cases where probability intervals can be encountered as models of uncertainty are confidence intervals on parameters of multinomial distributions built from sample data, and expert opinions providing such intervals.

### 3 P-boxes

We first recall some usual notions on the real line that will be generalized in the sequel.

Let  $\Pr$  be a probability function on the real line with density  $p$ . The *cumulative* distribution of  $\Pr$  is denoted  $F^p$  and is defined by  $F^p(x) = \Pr((-\infty, x])$ .

Let  $F_1(x)$  and  $F_2(x)$  be two *cumulative* distributions. Then,  $F_1(x)$  is said to stochastically dominate  $F_2(x)$  iff  $F_1(x) \leq F_2(x) \forall x$ .

A P-box [15] is defined by a pair of cumulative distributions  $\underline{F} \leq \overline{F}$  ( $\underline{F}$  stochastically dominates  $\overline{F}$ ) on the real line. It brackets the cumulative distribution of an imprecisely known probability function with density  $p$  s.t.  $\underline{F}(x) \leq F^p(x) \leq \overline{F}(x) \quad \forall x \in \mathfrak{R}$ .

#### 3.1 Generalized Cumulative Distributions

Interestingly, the notion of cumulative distribution is based on the existence of the natural ordering of numbers. Consider a probability distribution (probability vector)  $\lambda = (\lambda_1 \dots \lambda_n)$  defined over the finite space  $X$ ;  $\lambda_i$  denotes the probability  $\Pr(x_i)$  of the  $i$ -th element  $x_i$ , and  $\sum_{j=1}^n \lambda_j = 1$ . In this case, no natural notion of cumulative distribution exists. In order to make sense of this notion over  $X$ , one must equip it with a complete preordering  $\leq_R$ , which is a reflexive, complete and transitive relation. An  $R$ -downset is of the form  $\{x_i : x_i \leq_R x\}$ , and denoted  $(x]_R$ .

**Definition 1.** *The generalized  $R$ -cumulative distribution of a probability distribution on a finite, completely preordered set  $(X, \leq_R)$  is the function  $F_R^\lambda : X \rightarrow [0, 1]$  defined by  $F_R^\lambda(x) = \Pr((x]_R)$ .*

The usual notion of stochastic dominance can also be defined for generalized cumulative distributions. Consider another probability distribution  $\kappa = (\kappa_1 \dots \kappa_n)$  on  $X$ . The corresponding  $R$ -dominance relation of  $\lambda$  over  $\kappa$  can be defined by the pointwise inequality  $F_R^\lambda < F_R^\kappa$ . Clearly, a generalized cumulative distribution can always be considered as a simple one, up to a reordering of elements.

Any generalized cumulative distribution  $F_R^\lambda$  with respect to a complete preorder  $\leq_R$  on  $X$ , of a probability measure  $\Pr$ , with distribution  $\lambda$  on  $X$ , can also be used as a possibility distribution  $\pi_R$  whose associated measure dominates  $\Pr$ , i.e.  $\max_{x \in A} F_R^\lambda(x) \geq \Pr(A), \forall A \subseteq X$ . This is because a (generalized) cumulative distribution is constructed by computing the probabilities of events  $\Pr(A)$  in a nested sequence of downsets  $(x_i]_R$ . [10].

#### 3.2 Generalized p-box

Using the generalizations of the notions of cumulative distributions and of stochastic dominance described in section 3.1, we define a generalized p-box as follows

**Definition 2.** *A  $R$ -P-box on a finite, completely preordered set  $(X, \leq_R)$  is a pair of  $R$ -cumulative distributions  $F_R^\lambda(x)$  and  $F_R^\kappa(x)$ , s.t.  $F_R^\lambda(x) \leq F_R^\kappa(x)$  (i.e.  $\kappa$  is a probability distribution  $R$ -dominated by  $\lambda$ )*

The probability family induced by a  $R$ -P-box is

$$\mathcal{P}_{p\text{-box}} = \{P | \forall x, F_R^\lambda(x) \leq F_R(x) \leq F_R^\kappa(x)\}$$

If we choose a relation  $R$  with  $x_i \leq_R x_j$  iff  $i < j$ , and,  $\forall x_i \in X$ , consider the sets  $A_i = (x_i]_R$ , it comes down to a family of nested confidence sets  $\emptyset \subseteq A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subset X$ . The family  $\mathcal{P}_{p\text{-box}}$  can then be represented by the following restrictions on probability measures

$$\alpha_i \leq P(A_i) \leq \beta_i \quad i = 1, \dots, n \quad (2)$$

with  $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_n \leq 1$  and  $\beta_1 \leq \beta_2 \leq \dots \leq \beta_n \leq 1$ . Choosing  $X = \mathfrak{R}$  and  $A_i = (-\infty, x_i]$ , it is easy to see that we find back the usual definition of P-boxes.

A generalized cumulative distribution being fully specified by  $|X|$  values, it follows that  $2|X|$  values must be given to completely determine a generalized p-box. Moreover, we can interpret p-boxes as a collection of nested confidence intervals with upper and lower probability bounds (which could come, for example, from expert elicitation). In order to make notation simpler, the upper and lower cumulative distributions will respectively be noted  $F^*$ ,  $F_*$  in the sequel and, unless stated otherwise, we will consider (without loss of generality) the order  $R$  s.t.  $x_i \leq_R x_j$  iff  $i < j$  with the associated nested sets  $A_i$ . The notion of generalized p-box is orthogonal to the notion of probability intervals in the sense that, in the former, probability bounds are assigned to a nested family of events, while for the latter, probability bounds are assigned to disjoint elementary events.

#### 3.3 Generalized P-boxes in the setting of possibility theory

Given that sets  $A_i$  can be interpreted as nested confidence intervals with upper and lower bounds, it is natural to search a connection with possibility theory, since possibility distributions can be interpreted as a collection of nested confidence intervals (a natural way of expressing expert knowledge). We thus have the following proposition

**Proposition 1.** A family  $\mathcal{P}_{p\text{-box}}$  described by a generalized P-box can be encoded by a pair of possibility distributions  $\pi_1, \pi_2$  s.t.  $\mathcal{P}_{p\text{-box}} = \mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2}$  with  $\pi_1(x) = F^*(x)$  and  $\pi_2(x) = 1 - F_*(x)$

**Proof of proposition 1.** Consider the definition of a generalized p-box and the fact that a generalized cumulative distribution can be used as a possibility distribution  $\pi_R$  dominating the probability distribution  $\text{Pr}$  (see section 3.1). Then, the set of constraints  $(P(A_i) \geq \alpha_i)_{i=1,n}$  from equation (2) generates a possibility distribution  $\pi_1$  and the set of constraints  $(P(A_i) \geq 1 - \beta_i)_{i=1,n}$  generates a possibility distribution  $\pi_2$ . Clearly  $\mathcal{P}_{p\text{-box}} = \mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2}$ .  $\square$

### 3.4 Generalized P-boxes are special case of random sets

The following proposition was proved in [9]

**Proposition 2.** A family  $\mathcal{P}_{p\text{-box}}$  described by a generalized P-box can be encoded by a random set s.t.  $\mathcal{P}_{p\text{-box}} = \mathcal{P}_{\text{Bel}}$ .

---

#### Algorithm 1: R-P-box $\rightarrow$ random set

---

**Input:** Nested sets  $\emptyset, A_1, \dots, A_n, X$  and bounds  $\alpha_i, \beta_i$

**Output:** Equivalent random set

**for**  $k = 1, \dots, n + 1$  **do**

    Build partition  $F_i = A_i \setminus A_{i-1}$

Rank  $\alpha_i, \beta_i$  increasingly

**for**  $k = 0, \dots, 2n + 1$  **do**

    Rename  $\alpha_i, \beta_i$  by  $\gamma_l$  s.t.  
     $\alpha_0 = \gamma_0 = 0 \leq \gamma_1 \leq \dots \leq \gamma_l \leq \dots \leq \gamma_{2n} \leq 1 =$   
     $\gamma_{2n+1} = \beta_{n+1}$

Define focal set  $E_0 = \emptyset$

**for**  $k = 1, \dots, 2n + 1$  **do**

**if**  $\gamma_{k-1} = \alpha_i$  **then**  
            Build  $E_k = E_{k-1} \cup F_{i+1}$

**if**  $\gamma_{k-1} = \beta_i$  **then**  
            Build  $E_k = E_{k-1} \setminus F_i$

    Set  $m(E_k) = \gamma_k - \gamma_{k-1}$

---

Algorithm 1 provides an easy way to build the random set encoding a generalized p-box. It is similar to algorithms given in [16, 23], and extends them to more general spaces. The main idea of the algorithm is to use the fact that a generalized p-box can be seen as a random set whose focal elements are unions of adjacent sets in a partition. Thanks to the nested nature of sets  $A_i$ , we can build a partition of  $X$  made of  $F_i = A_i \setminus A_{i-1}$ , and then add or subtract consecutive elements of this partition to build the focal sets (of the form  $\bigcup_{j \leq i \leq k} F_i$ ) of the random set equivalent to the generalized p-box.

### 3.5 Generalized P-boxes and probability intervals

Provided an order  $R$  has been defined on elements  $x_i$ , a method to build a p-box from probability intervals  $L$  can be easily derived from equations (1). Lower and upper generalized cumulative distributions can be computed as follows

$$\begin{aligned} F_*(x_i) &= \underline{P}(A_i) = \max\left(\sum_{x_j \in A_i} l_j, 1 - \sum_{x_j \notin A_i} u_j\right) \\ F^*(x_i) &= \overline{P}(A_i) = \min\left(\sum_{x_j \in A_i} u_j, 1 - \sum_{x_j \notin A_i} l_j\right) \end{aligned} \quad (3)$$

Transforming a p-box into probability intervals is also an easy task. First, let us assume that each element  $F_i$  of the partition used in algorithm 1 is reduced to a singleton  $x_i$ . Corresponding probability intervals are then given by the two following formulas:

$$\begin{aligned} \underline{P}(F_i) &= \underline{P}(x_i) = l_i = \max(0, \alpha_i - \beta_{i-1}) \\ \overline{P}(F_i) &= \overline{P}(x_i) = u_i = \beta_i - \alpha_{i-1} \end{aligned}$$

if a set  $F_i$  is made of  $n$  elements  $x_{i1}, \dots, x_{in}$ , it is easy to see that  $l(x_{ij}) = 0$  and that  $u(x_{ij}) = \overline{P}(F_i)$ , since  $x_{ij} \subset F_i$ .

Let us note that transforming probability intervals into p-boxes (and inversely) generally loses information, except in the degenerated cases of precise probability distribution and of total ignorance. If no obvious order relation  $R$  between elements  $x_i$  is to be privileged, and if one wants to transform probability intervals into generalized p-boxes, we think that a good choice for the order  $R$  is the one s.t.

$$\sum_{i=1}^n F^*(x_i) - F_*(x_i)$$

is minimized, so that a minimal amount of information is lost in the process.

Another interesting fact to pinpoint is that both cumulative distributions given by equations (3) can be interpreted as possibility distributions dominating the family  $\mathcal{P}_L$  (for  $F_*$ , the associated possibility distribution is  $1 - F_*$ ). Thus, computing either  $F^*$  or  $F_*$  is a method to find a possibility distribution approximating  $\mathcal{P}_L$ , which is different from the one proposed by Masson and Denoeux [19].

## 4 Clouds

We begin this section by recalling basic definitions and results due to Neumaier [20], cast in the terminology of fuzzy sets and possibility theory. A

cloud is an Interval-Valued Fuzzy Set  $F$  such that  $(0, 1) \subseteq \cup_{x \in X} F(x) \subseteq [0, 1]$ , where  $F(x)$  is an interval  $[\delta(x), \pi(x)]$ . In the following, it is either defined on a finite space  $X$ , or it is a continuous interval-valued fuzzy interval (IVFI) on the real line (a “cloudy” interval). In the latter case each fuzzy set has cuts that are closed intervals. When the upper membership function coincides with the lower one, ( $\delta = \pi$ ) the cloud is called *thin*, and when the lower membership function is identically 0, the cloud is called *fuzzy* by Neumaier. Let us note that these names are somewhat counter-intuitive, since a *thin* cloud correspond to a fuzzy set with precise membership function, while a fuzzy cloud is equivalent to a probability family modeled by a possibility distribution.

A random variable  $x$  with values in  $X$  is said to belong to a cloud  $F$  if and only if  $\forall \alpha \in [0, 1]$ :

$$P(\delta(x) \geq \alpha) \leq 1 - \alpha \leq P(\pi(x) > \alpha) \quad (4)$$

under all suitable measurability assumptions.

If  $X$  is a finite space of cardinality  $n$ , a *cloud* can be defined by the following restrictions :

$$P(B_i) \leq 1 - \alpha_i \leq P(A_i) \text{ and } B_i \subseteq A_i, \quad (5)$$

where  $1 = \alpha_0 > \alpha_1 > \alpha_2 > \dots > \alpha_n > \alpha_{n+1} = 0$  and  $\emptyset = A_0 \subset A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subseteq A_{n+1} = X$ ;  $\emptyset = B_0 \subseteq B_1 \subseteq B_2 \subseteq \dots \subseteq B_n \subseteq B_{n+1} = X$ . The confidence sets  $A_i$  and  $B_i$  are respectively the strong and regular  $\alpha$ -cut of fuzzy sets  $\pi$  and  $\delta$  ( $A_i = \{x_i, \pi(x_i) > \alpha_{i+1}\}$  and  $B_i = \{x_i, \delta(x_i) \geq \alpha_{i+1}\}$ ).

As for probability intervals and p-boxes, eliciting a cloud requires  $2|X|$  values.

#### 4.1 Clouds in the setting of possibility theory

Let us first recall the following result regarding possibility measures (see [10]):

**Proposition 3.**  $P \in \mathcal{P}_\pi$  if and only if  $1 - \alpha \leq P(\pi(x) > \alpha), \forall \alpha \in (0, 1]$

The following proposition directly follows

**Proposition 4.** A probability family  $\mathcal{P}_{\delta, \pi}$  described by the cloud  $(\delta, \pi)$  is equivalent to the family  $\mathcal{P}_\pi \cap \mathcal{P}_{1-\delta}$  described by the two possibility distributions  $\pi$  and  $1 - \delta$ .

**Proof of proposition 4.** Consider a cloud  $(\delta, \pi)$ , and define  $\bar{\pi} = 1 - \delta$ . Note that  $P(\delta(x) \geq \alpha) \leq 1 - \alpha$  is equivalent to  $P(\bar{\pi} > \beta) \geq 1 - \beta$ , letting  $\beta = 1 - \alpha$ . So it is clear from equation (4) that probability measure  $P$  is in the cloud  $(\delta, \pi)$  if and only if it is in  $\mathcal{P}_\pi \cap \mathcal{P}_{\bar{\pi}}$ . So a cloud is a family of probabilities dominated by two possibility distributions (see [14]).  $\square$

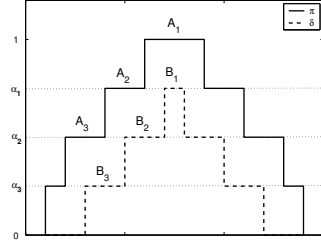


Figure 1: Comonotonic cloud

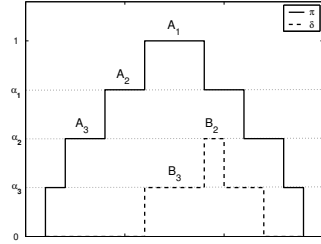


Figure 2: Non-Comonotonic cloud

This property is common to generalized p-boxes and clouds: they define probability families upper bounded by two possibility measures. It is then natural to investigate their relationships.

#### 4.2 Finding clouds that are generalized p-boxes

**Proposition 5.** A cloud is a generalized p-box iff  $\{A_i, B_i, i = 1, \dots, n\}$  form a nested sequence of sets (i.e. there is a linear preordering with respect to inclusion)

**Proof of proposition 5.** Assume the sets  $A_i$  and  $B_j$  form a globally nested sequence whose current element is  $C_k$ . Then the set of constraints defining a cloud can be rewritten in the form  $\gamma_k \leq P(C_k) \leq \beta_k$ , where  $\gamma_k = 1 - \alpha_i$  and  $\beta_k = \min\{1 - \alpha_j : A_i \subseteq B_j\}$  if  $C_k = A_i$ ;  $\beta_k = 1 - \alpha_i$  and  $\gamma_k = \max\{1 - \alpha_j : A_j \subseteq B_i\}$  if  $C_k = B_i$ .

Since  $1 = \alpha_0 > \alpha_1 > \dots > \alpha_n < \alpha_{n+1} = 0$ , these constraints are equivalent to those of a generalized p-box. But if  $\exists B_j, A_i$  with  $j > i$  s.t.  $B_j \not\subseteq A_i$  and  $A_i \not\subseteq B_j$ , then the cloud is not equivalent to a p-box, since confidence sets would no more form a complete preordering with respect to inclusion.  $\square$

In term of pairs of possibility distributions, it is now easy to see that a cloud  $(\delta, \pi)$  is a generalized p-box if and only if  $\pi$  and  $\delta$  are comonotonic. We will thus call such clouds comonotonic clouds. If a cloud is comonotonic, we can thus directly adapt the various results obtained for generalized p-boxes. In partic-

ular, because comonotonic clouds are generalized p-boxes, algorithm 1 can be used to get the corresponding random set. Notions of comonotonic and non-comonotonic clouds are respectively illustrated by figures 1 and 2

### 4.3 Characterizing and approximating non-comonotonic clouds

The following proposition characterizes probability families represented by most non-comonotonic clouds, showing that the distinction between comonotonic and non-comonotonic clouds makes sense (since the latter cannot be represented by random sets).

**Proposition 6.** *If  $(\delta, \pi)$  is a non-comonotonic cloud for which there are two overlapping sets  $A_i, B_j$  that are not nested (i.e.  $A_i \cap B_j \neq \{A_i, B_j, \emptyset\}$ ), then the lower probability of the induced family  $\mathcal{P}_{\delta, \pi}$  is not even 2-monotone.*

and the proof can be found in the appendix.

**Remark 1.** *The case for which we have  $B_j \cap A_i \in \{A_i, B_j\}$  for all pairs  $A_i, B_j$  is the case of comonotonic clouds. Now, if a cloud is such that for all pairs  $A_i, B_j : B_j \cap A_i \in \{A_i, B_j, \emptyset\}$  with at least one empty intersection, then it is still a random set, but no longer a generalized p-box. Let us note that this special case can only occur for discrete clouds.*

Since it can be computationally difficult to work with capacities that are not 2-monotone, one could wish to work either with outer or inner approximations. We propose two such approximations, which are easy to compute and respectively correspond to necessity (possibility) measures and belief (plausibility) measures.

**Proposition 7.** *If  $\mathcal{P}_{\delta, \pi}$  is the probability family described by the cloud  $(\delta, \pi)$  on a referential  $X$ , then, the following bounds provide an outer approximation of the range of  $P(A)$  :*

$$\begin{aligned} \max(N_\pi(A), N_{1-\delta}(A)) \leq P(A) \leq \\ \min(\Pi_\pi(A), \Pi_{1-\delta}(A)) \quad \forall A \subset X \end{aligned} \quad (6)$$

**Proof of proposition 7.** Since we have that  $\mathcal{P}_{\delta, \pi} = \mathcal{P}_{1-\delta} \cap \mathcal{P}_\pi$ , and given the bounds defined by each possibility distributions, it is clear that equation 6 give bounds of  $P(A)$ .  $\square$

Nevertheless, these bounds are not, in general, the infimum and the supremum of  $P(A)$  over  $\mathcal{P}_{\delta, \pi}$ . To see this, consider a discrete cloud made of four non-

empty elements  $A_1, A_2, B_1, B_2$ . It can be checked that

$$\begin{aligned} \pi(x) &= 1 \text{ if } x \in A_1; \\ &= \alpha_1 \text{ if } x \in A_2 \setminus A_1; \\ &= \alpha_2 \text{ if } x \notin A_2. \\ \delta(x) &= \alpha_1 \text{ if } x \in B_1; \\ &= \alpha_2 \text{ if } x \in B_2 \setminus B_1; \\ &= 0 \text{ if } x \notin B_2. \end{aligned}$$

Since  $P(A_2) \geq 1 - \alpha_2$  and  $P(B_1) \leq 1 - \alpha_1$ , from (5), we can easily check that  $\underline{P}(A_2 \setminus B_1) = \underline{P}(A_2 \cap B_1^c) = \alpha_1 - \alpha_2$ . Now,  $N_\pi(A_2 \cap B_1^c) = \min(N_\pi(A_2), N_\pi(B_1^c)) = 0$  since  $\Pi_\pi(B_1) = 1$  because  $B_1 \subseteq A_1$ . Considering distribution  $\delta$ , we can have  $N_{1-\delta}(A_2 \cap B_1^c) = \min(N_{1-\delta}(A_2), N_{1-\delta}(B_1^c)) = 0$  since  $N_{1-\delta}(A_2) = \Delta_\delta(A_2^c) = 0$  since  $B_2 \subseteq A_2$ . Equation (6) can thus result in a trivial lower bound, different from  $\underline{P}(A_2 \setminus B_1)$ .

We can check that the bounds given by equation (6) are the one considered by Neumaier in [20]. Since these bounds are, in general, not the infimum and supremum of  $P(A)$  on  $\mathcal{P}_{\delta, \pi}$ , Neumaier's claim that clouds are only vaguely related to Walley's previsions or random sets is not surprising. Nevertheless, if we consider the relationship between clouds and possibility distributions, taking this outer approximation, that is very easy to compute, seems very natural. The next proposition provides an inner approximation of  $\mathcal{P}_{\delta, \pi}$

**Proposition 8.** *Given the sets  $\{B_i, A_i, i = 1, \dots, n\}$  constituting the distributions  $(\delta, \pi)$  of a cloud and the corresponding  $\alpha_i$ , the belief and plausibility measures of the random set s.t.  $m(A_i \setminus B_{i-1}) = \alpha_{i-1} - \alpha_i$  are inner approximations of  $\mathcal{P}_{\delta, \pi}$ .*

It is easy to see that this random set can always be defined. We can see that it is always an inner approximation by using the contingency matrix advocated in the proof of proposition 6 (see appendix). In this matrix, the random set defined above comes down to concentrating weights on diagonal elements.

This inner approximation is exact in case of comonotonicity or when we have  $A_i \cap B_j \in \{A_i, B_j, \emptyset\}$  for any pair of sets  $A_i, B_j$  defining the clouds.

### 4.4 A note on thin and continuous clouds

Thin clouds ( $\delta = \pi$ ) constitute an interesting special case of clouds. In this latter case, conditions defining clouds are reduced to

$$P(\pi(x) \geq \alpha) = P(\pi(x) > \alpha) = 1 - \alpha, \forall \alpha \in (0, 1).$$

On finite sets these constraints are generally contradictory, because  $P(\pi(x) \geq \alpha) > P(\pi(x) > \alpha)$  for some  $\alpha$ , hence the following theorem:

**Proposition 9.** *If  $X$  is finite, then  $\mathcal{P}(\pi) \cap \mathcal{P}(1 - \pi)$  is empty.*

which is proved in [14], where it is also shown that this emptiness is due to finiteness. A simple shift of indices solves the difficulty. Let  $\pi(u_i) = \alpha_i$  such that  $\alpha_1 = 1 > \dots > \alpha_n > \alpha_{n+1} = 0$ . Consider  $\delta(u_i) = \alpha_{i+1} < \pi_1(u_i)$ . Then  $\mathcal{P}(\pi) \cap \mathcal{P}(1 - \delta)$  contains the unique probability measure  $P$  such that the probability weight attached to  $u_i$  is  $p_i = \alpha_i - \alpha_{i+1}, \forall i = 1 \dots n$ . To see it, refer to equation (5), and note that in this case  $A_i = B_i$ .

In the continuous case, a thin cloud is non-trivial. The inclusions  $[\delta(x) \geq \alpha] \subseteq [\pi(x) > \alpha]$  (corresponding to  $B_i \subseteq A_i$ ) again do not work but we may have  $P(\pi(x) \geq \alpha) = P(\pi(x) > \alpha) = 1 - \alpha, \forall \alpha \in (0, 1)$ . For instance, a cumulative distribution function, viewed as a tight p-box, defines a thin cloud containing the only random variable having this cumulative distribution (the "right" side of the cloud is rejected to  $\infty$ ). In fact, it was suggested in [14] that a thin cloud contains in general an infinity of probability distributions.

Insofar as Proposition 5 can be extended to the reals (this could be shown, for instance, by proving the convergence of some finite outer and inner approximations of the continuous model, or by using the notion of directed set [5] to prove the complete monotonicity of the model), then a thin cloud can be viewed as a generalized p-box and is thus a (continuous) belief function with uniform mass density, whose focal sets are doubletons of the form  $\{x(\alpha), y(\alpha)\}$  where  $\{x : \pi(x) \geq \alpha\} = [x(\alpha), y(\alpha)]$ . It is defined by the Lebesgue measure on the unit interval and the multimapping  $\alpha \rightarrow \{x(\alpha), y(\alpha)\}$ . This result gives us a nice way to characterize the infinite quantity of random variables contained in a thin cloud. In particular, concentrating the mass density on elements  $x(\alpha)$  or on elements  $y(\alpha)$  would respectively give the upper and lower cumulative distributions that would have been associated to the possibility distribution  $\pi$  alone (let us note that every convex mixture of those two cumulative distributions would also be in the thin cloud). It is also clear that  $Bel(\pi(x) \geq \alpha) = 1 - \alpha$ . More generally, if Proposition 5 holds in the continuous case, a comonotonic cloud can be characterized by a continuous belief function [25] with uniform mass density, whose focal sets would be disjoint sets of the form  $[x(\alpha), u(\alpha)] \cup [v(\alpha), y(\alpha)]$  where  $\{x : \pi(x) \geq \alpha\} = [x(\alpha), y(\alpha)]$  and  $\{x : \delta(x) \geq \alpha\} = [u(\alpha), v(\alpha)]$ .

## 4.5 Clouds and probability intervals

Since probability intervals are 2-monotone capacities, while clouds are either  $\infty$ -monotone capacities or not even 2-monotone capacities, there is no direct correspondence between probability intervals and clouds. Nevertheless, given previous results, we can easily build a cloud approximating a family  $\mathcal{P}_L$  defined by a set  $L$  of probability intervals (but perhaps not the most "specific" one): indeed, any generalized p-box built from the probability intervals is a comonotonic cloud encompassing the family  $\mathcal{P}_L$ .

Although finding the "best" (i.e. keeping as much information as possible, given some information measure) method to transform probability intervals into cloud is an open problem. Any such transformation should follow some basic requirements such as:

1. Since clouds can model precise probability distributions, the method should insure that a precise probability distribution will be transformed into the corresponding *thin* cloud.
2. Given a set  $L$  of probability intervals, the transformed cloud  $[\delta, \pi]$  should contain  $\mathcal{P}_L$  (i.e.  $\mathcal{P}_{\delta, \pi},$  while being as close to it as possible.  $\subset \mathcal{P}_L$ ).

Let us note that using the transformation proposed in section 3.5 for generalized p-boxes satisfies these two requirements. Another solution is to extend Masson and Denoeux's [19] method that builds a possibility distribution covering a set of probability intervals, completing it by a lower distribution  $\delta$  (due to lack of space, we do not explore this alternative here).

## 5 Conclusions

Figure 3 summarizes our results cast in a more general framework of imprecise probability representations (our main contributions in boldface).

In this paper, we have considered many practical representations of imprecise probabilities, which are easier to handle than general probability families. They often require less data to be fully specified and they allow many mathematical simplifications, which may prove to increase computational efficiency (except, perhaps, for non-comonotonic clouds).

Some clarifications have been brought concerning the properties of the cloud formalism. The fact that non-comonotonic clouds are not even 2-monotone capacities tends to indicate that, from a computational standpoint, they sound less interesting than the other formalisms. Nevertheless, as far as we know, they are the only simple model generating capacities that are not 2-monotone.



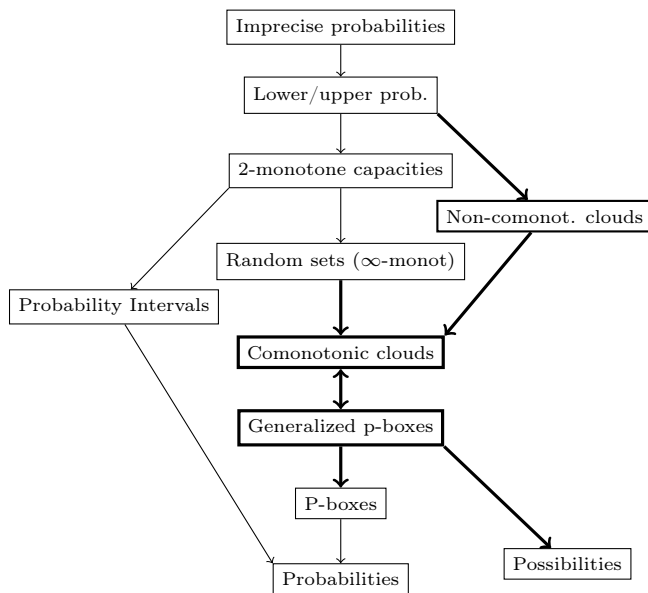


Figure 3: Representations relationships.  $A \longrightarrow B$  : B is a special case of A

A work that remains to be done to a large extent is to evaluate the validity and the usefulness of these representations, particularly from a psychological standpoint (even if some of it has already been done [22, 18]). Another issue is to extend presented results to continuous spaces or to general lower/upper previsions (by using results from, for example [25, 6]). Finally, a natural continuation to this work is to explore various aspects of each formalisms in a manner similar to the one of De campos et al. [2]. What becomes of random sets, possibility distributions, generalized p-boxes and clouds after fusion, marginalization, conditioning or propagation? Do they preserve the representation? and under which assumptions? To what extent are these representations informative? Can they easily be elicited or integrated? If many results already exist for random sets and possibility distributions, there are fewer results for generalized p-boxes or clouds, due to their novelty.

## Acknowledgements

This paper has been supported by a grant from the Institut de Radioprotection et de Sûreté Nucléaire (IRSN). Scientific responsibility rests with the authors.

## References

[1] C. Baudrit and D. Dubois. Practical representations of incomplete probabilistic knowl-

edge. *Computational Statistics and Data Analysis*, 51(1):86–108, 2006.

- [2] L. de Campos, J. Huete, and S. Moral. Probability intervals : a tool for uncertain reasoning. *I. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.
- [3] A. Chateaufneuf. Combination of compatible belief functions and relation of specificity. In *Advances in the Dempster-Shafer theory of evidence*, pages 97–114. John Wiley & Sons, Inc, New York, NY, USA, 1994.
- [4] G. Choquet. Theory of capacities. *Annales de l’institut Fourier*, 5:131–295, 1954.
- [5] G. de Cooman. A behavioural model for vague probability assessments. *Fuzzy sets and systems*, 154:305–358, 2005.
- [6] G. de Cooman, M. Troffaes, and E. Miranda.  $n$ -monotone lower previsions and lower integrals. In F. Cozman, R. Nau, and T. Seidenfeld, editors, *Proc. 4th International Symposium on Imprecise Probabilities and Their Applications*, 2005.
- [7] A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
- [8] T. Denoeux. Constructing belief functions from sample data using multinomial confidence regions. *I. J. of Approximate Reasoning*, 42, 2006.
- [9] S. Destercke and D. Dubois. A unified view of some representations of imprecise probabilities. In J. Lawry, E. Miranda, A. Bugarin, and S. Li, editors, *Int. Conf. on Soft Methods in Probability and Statistics (SMPS)*, Advances in Soft Computing, pages 249–257, Bristol, 2006. Springer.
- [10] D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10:273–297, 2004.
- [11] D. Dubois, P. Hajek, and H. Prade. Knowledge-driven versus data-driven logics. *Journal of logic, Language and information*, 9:65–89, 2000.
- [12] D. Dubois and H. Prade. *Possibility Theory : An Approach to Computerized Processing of Uncertainty*. Plenum Press, 1988.
- [13] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49:65–74, 1992.

- [14] D. Dubois and H. Prade. Interval-valued fuzzy sets, possibility theory and imprecise probability. In *Proceedings of International Conference in Fuzzy Logic and Technology (EUSFLAT'05)*, Barcelona, September 2005.
- [15] S. Ferson, L. Ginzburg, V. Kreinovich, D. Myers, and K. Sentz. Construction probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories, 2003.
- [16] E. Kriegler and H. Held. Utilizing belief functions for the estimation of future climate change. *I. J. of Approximate Reasoning*, 39:185–209, 2005.
- [17] J. Lemmer and H. Kyburg. Conditions for the existence of belief functions corresponding to intervals of belief. In *Proc. 9th National Conference on A.I.*, pages 488–493, 1991.
- [18] G. N. Linz and F. C. de Souza. A protocol for the elicitation of imprecise probabilities. In *Proceedings 4th International Symposium on Imprecise Probabilities and their Applications*, Pittsburgh, 2005.
- [19] M. Masson and T. Denoeux. Inferring a possibility distribution from empirical data. *Fuzzy Sets and Systems*, 157(3):319–340, february 2006.
- [20] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10:249–272, 2004.
- [21] A. Neumaier. On the structure of clouds. available on [www.mat.univie.ac.at/~neum](http://www.mat.univie.ac.at/~neum), 2004.
- [22] E. Raufaste, R. Neves, and C. Mariné. Testing the descriptive validity of possibility theory in human judgments of uncertainty. *Artificial Intelligence*, 148:197–218, 2003.
- [23] H. Regan, S. Ferson, and D. Berleant. Equivalence of methods for uncertainty propagation of real-valued random variables. *I. J. of Approximate Reasoning*, 36:1–30, 2004.
- [24] G. Shafer. *A mathematical Theory of Evidence*. Princeton University Press, 1976.
- [25] P. Smets. Belief functions on real numbers. *I. J. of Approximate Reasoning*, 40:181–223, 2005.
- [26] P. Walley. *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, 1991.
- [27] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1:3–28, 1978.

## Appendix

**Proof of proposition 6 (sketch).** Our proof uses the following result by Chateaufneuf [3]: Let  $m_1, m_2$  be two random sets with focal sets  $\mathcal{F}_1, \mathcal{F}_2$ , each of them respectively defining a probability family  $\mathcal{P}_{Bel_1}, \mathcal{P}_{Bel_2}$ . Here, we assume that those families are "compatible" (i.e.  $\mathcal{P}_{Bel_1} \cap \mathcal{P}_{Bel_2} \neq \emptyset$ ).

Then, the result from Chateaufneuf states the following : the lower probability  $\underline{P}(E)$  of the event  $E$  on  $\mathcal{P}_{Bel_1} \cap \mathcal{P}_{Bel_2}$  is equal to the least belief measure  $Bel(E)$  that can be computed on the set of joint normalized random sets with marginals  $m_1, m_2$ . More formally, let us consider a set  $\mathcal{Q}$  s.t.  $Q \in \mathcal{Q}$  iff

- $Q(A, B) > 0 \Rightarrow A \times B \in \mathcal{F}_1 \times \mathcal{F}_2$  (masses over the cartesian product of focal sets)
- $A \cap B = \emptyset \Rightarrow Q(A, B) = 0$  (normalization constraints)
- $m_1(A) = \sum_{B \in \mathcal{F}_2} Q(A, B)$  and  $m_2(B) = \sum_{A \in \mathcal{F}_1} Q(A, B)$  (marginal constraints)

and the lower probability  $\underline{P}(E)$  is given by the following equation

$$\underline{P}(E) = \min_{Q \in \mathcal{Q}} \sum_{(A \cap B) \subseteq E} Q(A, B) \quad (7)$$

where  $\mathcal{Q}$  is the set of joint normalized random sets. This result can be applied to clouds, since the family described by a cloud is the intersection of two families modeled by possibility distributions.

To illustrate the general proof, we will restrict ourselves to a 4-set cloud (the most simple non-trivial cloud that can be found). We thus consider four sets  $A_1, A_2, B_1, B_2$  s.t.  $A_1 \subset A_2, B_1 \subset B_2, B_i \subset A_i$  together with two values  $\alpha_1, \alpha_2$  s.t.  $1(= \alpha_0) > \alpha_1 > \alpha_2 > 0(= \alpha_3)$  and the cloud is defined by enforcing the inequalities  $P(B_i) \leq 1 - \alpha_i \leq P(A_i)$   $i = 1, 2$ . The random sets equivalent to the possibility distributions  $\pi, 1 - \delta$  are summarized in the following table:

$\pi$	$1 - \delta$
$m(A_1) = 1 - \alpha_1$	$m(B_0^c = X) = 1 - \alpha_1$
$m(A_2) = \alpha_1 - \alpha_2$	$m(B_1^c) = \alpha_1 - \alpha_2$
$m(A_3 = X) = \alpha_2$	$m(B_2^c) = \alpha_2$

Furthermore, we add the constraint  $A_1 \cap B_2 \neq \{A_1, B_2, \emptyset\}$ , related to the non-monotonicity of the cloud. We then have the following contingency matrix, where the mass  $m_{ij}$  is assigned to the intersection of the corresponding sets at the beginning of line

$i$  and the top of column  $j$ :

	$B_0^c = X$	$B_1^c$	$B_2^c$	$\sum$
$A_1$	$m_{11}$	$m_{12}$	$m_{13}$	$1 - \alpha_1$
$A_2$	$m_{21}$	$m_{22}$	$m_{23}$	$\alpha_1 - \alpha_2$
$A_3 = X$	$m_{31}$	$m_{32}$	$m_{33}$	$\alpha_2$
$\sum$	$1 - \alpha_1$	$\alpha_1 - \alpha_2$	$\alpha_2$	$1$

We now consider the four events  $A_1, B_2^c, A_1 \cap B_2^c, A_1 \cup B_2^c$ . Given the above contingency matrix, we immediately have  $\underline{P}(A_1) = 1 - \alpha_1$  and  $\underline{P}(B_2^c) = \alpha_2$ , since  $A_1$  only includes the (joint) focal sets in the first line and  $B_2^c$  in the third column.

It is also easy to see that  $\underline{P}(A_1 \cap B_2^c) = 0$ , by considering the mass assignment  $m_{ii} = \alpha_{i-1} - \alpha_i$  (we then have  $m_{13} = 0$ , which is the mass of the only joint focal set included in  $A_1 \cap B_2^c$ ).

Now, concerning  $\underline{P}(A_1 \cup B_2^c)$ , let us consider the following mass assignment:

$$\begin{aligned} m_{22} &= \alpha_1 - \alpha_2 \\ m_{31} &= \min(1 - \alpha_1, \alpha_2) \\ m_{11} &= 1 - \alpha_1 - m_{31} \\ m_{33} &= \alpha_2 - m_{31} \\ m_{13} &= m_{31} \end{aligned}$$

it can be checked that this mass assignment satisfies the constraints of the contingency matrix, and that the only joint focal sets included in  $A_1 \cup B_2^c$  are those with masses  $m_{11}, m_{33}, m_{13}$ . Summing these masses, we have  $\underline{P}(A_1 \cup B_2^c) = \max(\alpha_2, 1 - \alpha_1)$ . Hence:

$$\begin{aligned} \underline{P}(A_1 \cup B_2^c) + \underline{P}(A_1 \cap B_2^c) &< \underline{P}(B_2^c) + \underline{P}(A_1) \\ \max(\alpha_2, 1 - \alpha_1) &< 1 - \alpha_1 + \alpha_2 \end{aligned}$$

an inequality that clearly violates the 2-monotonicity property. We have thus shown that in the 4-set case, 2-monotonicity never holds for families modeled by non-comonotonic clouds.

Now, in the general case, we have the following contingency matrix

	$B_0^c$	$B_j^c$	$B_n^c$	$\sum$
$A_1$	$m_{11}$			$1 - \alpha_1$
$\cdot$		$\cdot$		
$A_i$		$m_{i(j+1)}$		$\alpha_{i-1} - \alpha_i$
$\cdot$			$\cdot$	
$A_{n+1}$			$m_{(n+1)(n+1)}$	$\alpha_n$
$\sum$	$1 - \alpha_1$	$\alpha_j - \alpha_{j+1}$	$\alpha_n$	

Under the hypothesis of proposition 6, there are two sets  $A_i, B_j$  s.t.  $A_i \cap B_j \neq \{A_i, B_j, \emptyset\}$ . Due to the

inclusion relationships between the sets, and similarly to what was done in the 4-set case, we have

$$\begin{aligned} \underline{P}(A_i) &= 1 - \alpha_i \\ \underline{P}(B_j^c) &= \alpha_j \\ \underline{P}(A_i \cap B_j^c) &= 0 \end{aligned}$$

Next, let us concentrate on event  $A_i \cup B_j^c$  (which is different from  $X$  by hypothesis). Let us suppose that  $m_{kk} = \alpha_{k-1} - \alpha_k$ , except for masses  $m_{(j+1)i}, m_{ii}, m_{i(j+1)}, m_{(j+1)(j+1)}$ . This is similar to the 4-set case with masses  $m_{(j+1)i}, m_{ii}, m_{i(j+1)}, m_{(j+1)(j+1)}$  and we get the following assignment

$$\begin{aligned} q_{i(j+1)} &= \min(\alpha_{i-1} - \alpha_i, \alpha_j - \alpha_{j+1}) \\ q_{ii} &= \alpha_{i-1} - \alpha_i - q_{i(j+1)} \\ q_{(j+1)(j+1)} &= \alpha_j - \alpha_{j+1} - m_{12} \\ q_{(j+1)i} &= \min(\alpha_{i-1} - \alpha_i, \alpha_j - \alpha_{j+1}) \end{aligned}$$

Given this specific mass assignment (which is always inside the set  $\mathcal{Q}$ ), and by considering every subsets of  $A_i \cup B_j^c$ , the following inequality results:

$$\underline{P}(A_i \cup B_j^c) \leq \alpha_{j+1} + 1 - \alpha_{i-1} + \max(\alpha_{i-1} - \alpha_i, \alpha_j - \alpha_{j+1})$$

so,

$$\underline{P}(A_i \cup B_j^c) + \underline{P}(A_i \cap B_j^c) < \underline{P}(A_i) + \underline{P}(B_j^c),$$

which clearly violates the 2-monotonicity property.  $\square$