A unified view of some representations of imprecise probabilities

Sébastien Destercke, Didier Dubois

Preprint of the paper presented at the SMPS 2006 conference

A unified view of some representations of imprecise probabilities

S. Destercke and D. Dubois

Institut de recherche en informatique de Toulouse (IRIT)
Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France
`desterck@irit.fr and dubois@irit.fr`

Summary. Several methods for the practical representation of imprecise probabilities exist such as Ferson's p-boxes, possibility distributions, Neumaier's clouds, and random sets . In this paper some relationships existing between the four kinds of representations are discussed. A cloud as well as a p-box can be modelled as a pair of possibility distributions. We show that a generalized form of p-box is a special kind of belief function and also a special kind of cloud.

1 Introduction

Many uncertainty calculi can be viewed as encoding families of probabilities. Representing such families in a practical way can be a real challenge, and several proposals have been made to do so, under various assumptions. Among these proposals are p-boxes[6], possibility distributions [3], clouds [8] and random sets [1].

Possibility theory, P-boxes, and clouds use nested confidence sets with upper and lower probability bounds. This way of representing imprecise subjective probabilistic knowledge is very natural, and correspond to numerous situations where an expert is asked for confidence intervals. In this paper, we investigate or recall various links existing between these representations, illustrating the fact that they are all closely related.

Section 2 reviews the different kinds of representations considered in this paper, and generalizes the notion of P-boxes. In section 3, we show that generalized P-boxes (which encompass usual P-boxes) can be encoded by a belief function, and we then give a practical method to build it. Finally, section 4 recalls briefly some results on clouds and possibility theory, before examining the relationship between clouds and generalized P-boxes more closely.

2 Imprecise probabilities representations

2.1 Upper and lower probabilities

A family $\mathcal{P}$ of probabilities on X induces lower and upper probabilities on sets A [12]. Namely $\underline{P}(A) = \inf_{P \in \mathcal{P}} P(A)$ and $\overline{P}(A) = \sup_{P \in \mathcal{P}} P(A)$. We have $\mathcal{P}_{\underline{P},\overline{P}}(A) = \{P | \forall A \subseteq X \text{measurable}, \underline{P}(A) \leq P(A) \leq \overline{P}(A)\}$. It should be noted that $\mathcal{P}_{\underline{P},\overline{P}}$ is convex and generally larger than the original family $\mathcal{P}$, since lower and upper probabilities are projections of $\mathcal{P}$ on sets A. Representing either $\mathcal{P}$ or $\mathcal{P}_{\underline{P},\overline{P}}$ on a computer can be tedious, even for one-dimension problems. Simpler representations can be very useful, even if it implies a loss in generality.

2.2 Random sets

Formally, a random set is a set-valued mapping from a (here finite) probability space to a set X. It induces lower and upper probabilities on X [1]. Here, we use mass functions [10] to represent random sets. A mass function m is defined by a mapping from the power set $\mathcal{P}(X)$ to the unit interval, s.t. $\sum_{E \subseteq X} m(E) = 1$. A set E with positive mass is called a focal set. Plausibility and belief measures can then be defined from this mass function :

$$Bel(A) = \sum_{E, E \subseteq A} m(E) \text{ and } Pl(A) = 1 - Bel(A^c) = \sum_{E, E \cap A} m(E).$$

The set $\mathcal{P}_{Bel} = \{P | \forall A \subseteq X \text{ measurable}, Bel(A) \leq P(A) \leq Pl(A)\}$ is the special probability family induced by the belief function.

2.3 Quantitative possibility theory

A possibility distribution π is a mapping from X to the unit interval (hence a fuzzy set) such that $\pi(x) = 1$ for some $x \in X$. Several set-functions can be defined from them [3]:

- Possibility measures: $\Pi(A) = \sup_{x \in A} \pi(x)$
- Necessity measures: $N(A) = 1 - \Pi(A^c)$
- Guaranteed possibility measures: $\Delta(A) = \inf_{x \in A} \pi(x)$

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 Δ-measures the extent to which all states of the world where A occurs are plausible. They apply to so-called guaranteed possibility distributions [3] generally denoted by δ.

A possibility degree can be viewed as an upper bound of a probability degree [4]. Let $\mathcal{P}_\pi = \{P, \forall A \subseteq X \text{ measurable}, P(A) \leq \Pi(A)\}$ be the set of probability measures encoded by π. A necessity measure is a special case of belief function when the focal sets are nested.

2.4 Generalized Cumulative Distributions

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 $F^p(x) = \Pr((-\infty, x])$.

Interestingly the notion of cumulative distribution is based on the existence of the natural ordering of numbers. Consider a probability distribution (probability vector) $\alpha = (\alpha_1 \dots \alpha_n)$ defined over a finite domain X of cardinality n; α_i denotes the probability $\Pr(x_i)$ of the i-th element x_i, and $\sum_{j=1}^n \alpha_j = 1$. Then no obvious 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 $F_R^\alpha : X \to [0,1]$ defined by $F_R^\alpha(x) = \Pr((x]_R)$.*

Consider another probability distribution $\beta = (\beta_1 \dots \beta_n)$ on X. The corresponding R-dominance relation of α over β can be defined by the pointwise inequality $F_R^\alpha < F_R^\beta$. In other words, a generalized cumulative distribution can always be considered as a simple one, up to a reordering of elements.

In fact any generalized cumulative distribution F_R^α with respect to a weak order $>_R$ on X, of a probability measure Pr, with distribution α on X, can be viewed as a possibility distribution π_R whose associated measure dominates Pr, i.e. $\max_{x \in A} F_R^\alpha(x) \geq \Pr(A), \forall A \subseteq X$. This is because a (generalized) cumulative distribution is constructed by computing the probabilities of events $\Pr(A)$ in a nested sequence of downsets $(x_i]_R$. [2].

2.5 Generalized p-box

A P-box [6] is defined by a pair of cumulative distributions $\underline{F} \leq \overline{F}$ on the real line bounding the cumulative distribution of an imprecisely known probability function with density p. Using the results of section 2.4, we define a generalized p-box as follow

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

The probability family induced by a R-P-box is $\mathcal{P}_{p-box} = \{P | \forall x, F_R^\alpha(x) \leq F_R(x) \leq F_R^\beta(x)\}$ If we choose R and consider the sets $A_i = (x_i]_R, \forall x_i \in X$ with $x_i \leq_R x_j$ iff $i < j$, we define a family of nested confidence sets $\emptyset \subseteq A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subset X$. The family $\mathcal{P}_{p-box}$ can be encoded by the constraints

$$\alpha_i \leq P(A_i) \leq \beta_i \qquad i = 1, \dots, n \tag{1}$$

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

2.6 Clouds

This section recalls basic definitions and results due to Neumaier [8], 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 defined on a finite set X or it is an interval-valued fuzzy interval (IVFI) on the real line (then called a cloudy number). In the latter case each fuzzy set has cuts that are intervals. When the upper membership function coincides with the lower one, $(\delta = \pi)$ the cloud is called *thin*. When the lower membership function is identically 0, the cloud is said to be *fuzzy*.

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

$$P(\delta(x) \geq \alpha) \leq 1 - \alpha \leq P(\pi(x) > \alpha) \tag{2}$$

under all suitable measurability assumptions.

If X is a finite set of cardinality n, a *cloud* can be defined by the following constraints :

$$P(B_i) \leq 1 - \alpha_{i+1} \leq P(A_i) \text{ and } B_i \subseteq A_i \quad i = 1, \ldots, n \tag{3}$$

Where $1 = \alpha_1 > \alpha_2 > \ldots > \alpha_n = 0$ and $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n; B_1 \subseteq B_2 \subseteq \ldots \subseteq B_n$. The confidence sets A_i and B_i are respectively the α-cut of fuzzy sets π and δ ($A_i = \{x_i, \pi(x_i) > \alpha_{i+1}\}$ and $B_i = \{x_i, \delta(x_i) \geq \alpha_{i+1}\}$).

3 Generalized p-boxes are belief functions

In this section, we show that $\mathcal{P}_{p-box}$, the probability family described in section 2.5 can be encoded by a belief function. In order to achieve this, we reformulate the constraints given by equations (1).
Consider the following partition : $E_1 = A_1, E_2 = A_2 \setminus A_1, \ldots, E_n = A_n \setminus A_{n-1}, E_{n+1} = X \setminus A_n$

The constraints on the confidence sets A_i can be rewritten

$$\alpha_i \leq \sum_{k=1}^{i} P(E_i) \leq \beta_i \qquad i = 1, \ldots, n \tag{4}$$

The proof that a belief function encoding $\mathcal{P}_{p-box}$ exists follows in four points

1. The family $\mathcal{P}_{p-box}$ is always non-empty
2. Constraints induce $\underline{P}(\bigcup_{k=i}^{j} E_k) = max(0, \alpha_j - \beta_{i-1})$
3. Construction of a belief function s.t. $Bel(\bigcup_{k=i}^{j} E_k) = \underline{P}(\bigcup_{k=i}^{j} E_k)$
4. For any subset A of X, $Bel(A) = \underline{P}(A)$, then $\mathcal{P}_{p-box} = \mathcal{P}_{Bel}$ follows.

3.1 $\mathcal{P}$ is non-empty

Consider the case where $\alpha_i = \beta_i, \quad i = 1, \ldots, n$ in equation (4). Any probability distribution s.t. $P(E_1) = \alpha_1; P(E_2) = \alpha_2 - \alpha_1; \ldots; P(E_n) = \alpha_n - \alpha_{n-1}; P(E_{n+1}) = 1 - \alpha_n$ always exists and is in $\mathcal{P}_{p-box}$. Hence, $\mathcal{P}_{p-box} \neq \emptyset$. Every other cases being a relaxation of this one, $\mathcal{P}_{p-box}$ always contains at least one probability.

3.2 Lower probabilities on sets $(\bigcup_{k=i}^{j} E_k)$

Using partition given in section 3, we have $P(\bigcup_{k=i}^{j} E_k) = \sum_{k=i}^{j} P(E_k)$. Equations (4) induce the following lower and upper bounds on $P(\bigcup_{k=i}^{j} E_k)$

Proposition 1 $\underline{P}(\bigcup_{k=i}^{j} E_k) = \max(0, \alpha_j - \beta_{i-1}); \overline{P}(\bigcup_{k=i}^{j} E_k) = \beta_j - \alpha_{i-1}$

Proof To obtain $\underline{P}(\bigcup_{k=i}^{j} E_k)$, we must find min $(\sum_{k=i}^{j} P(E_k))$. From equation (4), we have

$$\alpha_j \leq \sum_{k=1}^{i-1} P(E_k) + \sum_{k=i}^{j} P(E_k) \leq \beta_j \text{ and } \alpha_{i-1} \leq \sum_{k=1}^{i-1} P(E_k) \leq \beta_{i-1}$$

Hence $\sum_{k=i}^{j} P(E_k)) \geq \max(0, \alpha_j - \beta_{i-1})$ and this lower bound $\max(0, \alpha_j - \beta_{i-1})$ is always reachable : if $\alpha_j > \beta_{i-1}$, take P s.t. $P(A_{i-1}) = \beta_{i-1}, P(\bigcup_{k=i}^{j} E_k) = \alpha_j - \beta_{i-1}, P(\bigcup_{k=j+1}^{n+1} E_k) = 1 - \alpha_j$. If $\alpha_j \leq \beta_{i-1}$, take P s.t. $P(A_{i-1}) = \beta_{i-1}, P(\bigcup_{k=i}^{j} E_k) = 0, P(\bigcup_{k=j+1}^{n+1} E_k) = 1 - \beta_{i-1}$. Proof for $\overline{P}(\bigcup_{k=i}^{j} E_k) = \beta_j - \alpha_{i-1}$ follows the same line.

3.3 Building the belief function

We now build a belief function s.t. $Bel(\bigcup_{k=i}^{j} E_k) = \underline{P}(\bigcup_{k=i}^{j} E_k)$, and in section 3.4, we show that this belief function is equivalent to the lower envelope of $\mathcal{P}_{p-box}$. We rank the α_i and β_i increasingly and rename them as

$$\alpha_0 = \beta_0 = \gamma_0 = 0 \leq \gamma_1 \leq \ldots \leq \gamma_{2n} \leq 1 = \gamma_{2n+1} = \beta_{n+1} = \alpha_{n+1}$$

and the successive focal elements F_l with $m(F_l) = \gamma_l - \gamma_{l-1}$. The construction of the belief function can be summarized as follow :

$$\text{If } \gamma_{l-1} = \alpha_i, \text{ then } F_l = F_{l-1} \cup E_{i+1} \tag{5}$$

$$\text{If } \gamma_{l-1} = \beta_i, \text{ then } F_l = F_{l-1} \setminus E_i \tag{6}$$

equation (5) means that element E_{i+1} is added to the previous focal set after reaching α_i, and equation (6) means that element E_i is deleted from the previous focal set after reaching β_i.

3.4 $\mathcal{P}_{Bel}$ is equivalent to $\mathcal{P}_{p-box}$

To show that $\mathcal{P}_{Bel} = \mathcal{P}_{p-box}$, we show that $Bel(A) = \underline{P}(A)\ \forall A \subseteq X$

Lower probability on sets A_i

Looking at equations (5,6) and taking $\gamma_l = \alpha_i$, we see that focal elements $F_1, \ldots, F_l$ only contain E_k s.t. $k \leq i$, hence we have $(F_1, \ldots, F_l) \subset A_i$. After γ_l, the focal elements $F_{l+1}, \ldots, F_{2n}$ contain at least one element E_k s.t. $k > i$. Summing the weights $m(F_1), \ldots, m(F_l)$, we have $Bel(A_i) = \gamma_l = \alpha_i$.

Sets of the type $P(\bigcup_{k=i}^{j} E_k)$

From section 3.2, we have $\underline{P}(\bigcup_{k=i}^{j} E_k) = max(0, \alpha_j - \beta_{i-1})$. Considering equations (5,6) and taking $\gamma_l = \alpha_j$, we have that focal elements $F_{l+1}, \ldots, F_{2n}$ contain at least one element E_k s.t. $k > j$, hence the focal elements $(F_{l+1}, \ldots, F_{2n}) \not\subset (\bigcup_{k=i}^{j} E_k)$. Taking then $\gamma_m = \beta_{i-1}$, we have that the focal elements $F_1, \ldots, F_m$ contain at least one element E_k s.t. $k < i$, hence the focal elements $(F_1, \ldots, F_m) \not\subset (\bigcup_{k=i}^{j} E_k)$.

If $m < l$ (i.e. $\gamma_l = \alpha_j \geq \beta_{i-1} = \gamma_m$), then the focal elements $(F_{m+1}, \ldots, F_l) \subset (\bigcup_{k=i}^{j} E_k)$ and we have $Bel(\bigcup_{k=i}^{j} E_k) = \gamma_l - \gamma_m = \alpha_j - \beta_{i-1}$. Otherwise, there is no focal element $F_l,\ l = 1, \ldots, 2n$ s.t. $F_l \subset (\bigcup_{k=i}^{j} E_k)$ and we have $Bel(\bigcup_{k=i}^{j} E_k) = \underline{P}(\bigcup_{k=i}^{j} E_k) = 0$.

Sets made of non-successive E_k

Consider a set of the type $A = (\bigcup_{k=i}^{i+l} E_k\ \cup\ \bigcup_{k=i+l+m}^{j} E_k)$ with $m > 1$ (i.e. there's a "hole" in the sequence, since at least $E_{i+l+1} \notin A$).

Proposition 2 *We have* $\underline{P}(\bigcup_{k=i}^{i+l} E_k\ \cup \bigcup_{k=i+l+m}^{j} E_k) = Bel(\bigcup_{k=i}^{i+l} E_k)) + Bel(\bigcup_{k=i+l+m}^{j} E_k)$

Sketch of proof The following inequalities gives us a lower bound on $\underline{P}$

$$min\left(P(\bigcup_{k=i}^{i+l} E_k\ \cup \bigcup_{k=i+l+m}^{j} E_k)\right) \geq \min\ P(\bigcup_{k=i}^{i+l} E_k) + \min\ P(\bigcup_{k=i+l+m}^{j} E_k)$$

we then use a reasoning similar to the one of section 3.2 to show that this lower bound is always reachable. The result can then be easily extended to a number n of "holes" in the sequence of E_k. This completes the proof and shows that $Bel(A) = \underline{P}(A)\ \forall A \in X$, so $\mathcal{P}_{Bel} = \mathcal{P}_{p-box}$.

4 Clouds and generalized p-boxes

Let us recall the following result regarding possibility measures (see [2]):

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

Consider a cloud (δ, π), and define $\overline{\pi} = 1-\delta$. Note that $P(\delta(x) \geq \alpha) \leq 1-\alpha$ is equivalent to $P(\overline{\pi} \geq \beta) \geq 1-\beta$, letting $\beta = 1-\alpha$. So it is clear from equation (2) that probability measure P is in the cloud (δ, π) if and only if it is in $\mathcal{P}_\pi \cap \mathcal{P}_{\overline{\pi}}$. So a cloud is a family of probabilities dominated by two possibility distributions (see [5]) . It follows that

Proposition 4 *A generalized p-box is a cloud*

Consider the definition of a generalized p-box and the fact that a generalized cumulative distribution can be viewed as a possibility distribution π_R dominating the probability distribution Pr (see section 2.4). Then, the set of constraints $(P(A_i) \geq \alpha_i)_{i=1,n}$ from equation (1) generates a possibility distribution π_1 and the set of constraints $(P(A_i^c) \geq 1-\beta_i)_{i=1,n}$ generates a possibility distribution π_2. Clearly $\mathcal{P}_{p-box} = \mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2}$, and corresponds to the cloud $(1-\pi_2, \pi_1)$. The converse is not true.

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

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 $\gamma_k = 1-\alpha_{i+1}$ and $\beta_k = \min\{1-\alpha_{j+1} : A_i \subseteq B_j\}$ if $C_k = A_i$; $\beta_k = 1-\alpha_{i+1}$ and $\gamma_k = \max\{1-\alpha_{j+1} : A_j \subseteq B_i\}$ if $C_k = B_i$.

Since $1 = \alpha_1 > \alpha_2 > \ldots > \alpha_n = 0$, these constraints are equivalent to those of a generalized p-box. But if $\exists\ B_j, A_i$ with $j > i$ s.t. $B_j \not\subset A_i$ and $A_i \not\subset B_j$, then the cloud is not equivalent to a p-box.

In term of pairs of possibility distributions, a cloud is a p-box iff π_1 and π_2 are comonotonic.

When the cloud is thin ($\delta = \pi$), cloud constraints reduce to $P(\pi(x) \geq \alpha) = P(\pi(x) > \alpha) = 1-\alpha$. On finite sets these constraints are contradictory. The closest approximation corresponds to the generalized p-box such that $\alpha_i = P(A_i), \forall i$. It allocates fixed probability weights to elements E_i of the induced partition. In the continuous case, a thin cloud is non trivial. A cumulative distribution function defines a thin cloud containing the only random variable having this cumulative distribution. A continuous unimodal possibility distribution π on the real line induces a thin cloud ($\delta = \pi$) which 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 $\alpha \longrightarrow \{x(\alpha), y(\alpha)\}$. It is indeed clear that $Bel(\pi(x) \geq \alpha) = 1-\alpha$.

5 Conclusions and open problems

There are several concise representations of imprecise probabilities. This paper highlights some links existing between clouds, possibility distributions, p-boxes and belief functions. We generalize p-boxes and show that they can be encoded by a belief function (extending results from [7, 9]). Another interesting result is that generalized p-boxes are a particular case of clouds, which are themselves equivalent to a pair of possibility distributions.

This paper shows that at least some clouds can be represented by a belief function. Two related open questions are : can a cloud be encoded by a belief function as well? can a set of probabilities dominated by two possibility measures be encoded by a belief function ? and if not, can we find inner or outer approximations following a principle of minimal commitment? Another issue is to extend these results to the continuous framework of Smets [11].

References

1. A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
2. D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable computing*, 10:273–297, 2004.
3. D. Dubois, P. Hajek, and H. Prade. Knowledge-driven versus data-driven logics. *Journal of logic, Language and information*, 9:65–89, 2000.
4. D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy sets and systems*, 49:65–74, 1992.
5. 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.
6. S. Ferson, L. Ginzburg, V. Kreinovich, D. Myers, and K. Sentz. Construction probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories, 2003.
7. E. Kriegler and H. Held. Utilizing belief functions for the estimation of future climate change. *International journal of approximate reasoning*, 39:185–209, 2005.
8. A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10:249–272, 2004.
9. H. Regan, S. Ferson, and D. Berleant. Equivalence of methods for uncertainty propagation of real-valued random variables. *International journal of approximate reasoning*, 36:1–30, 2004.
10. G. Shafer. *A mathematical Theory of Evidence*. Princeton University Press, 1976.
11. P. Smets. Belief functions on real numbers. *International journal of approximate reasoning*, 40:181–223, 2005.
12. P. Walley. *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, 1991.