Unifying practical uncertainty representations: I. Generalized p-boxes

S. Destercke [a,*] D. Dubois [b] E. Chojnacki [a]

[a] *Institut de radioprotection et de sûreté nucléaire (IRSN) 13115 St-Paul Lez Durance, France*

[b] *Institut de recherche en informatique de Toulouse (IRIT) Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse*

Abstract

There exist several simple representations of imprecise probability models of uncertainty that are easier to handle than the general approach. Among them are random sets, possibility distributions, probability intervals, and more recently Ferson's p-boxes and Neumaier's clouds. Both for theoretical and practical considerations, it is very useful to know whether one representation is equivalent to or can be approximated by other ones. In this paper, we define a generalized form of usual p-boxes. These generalized p-boxes have interesting connexions with other previously known representations. In particular, we show that they are equivalent to pairs of possibility distributions on finite or infinite sets, and that they are special kinds of random sets. They are also the missing link between p-boxes and clouds, which are the topic of the companion paper.

Key words: imprecise probability representations, p-boxes, possibility theory, random sets, clouds, probability intervals

1 Introduction

Different formal frameworks have been proposed to reason under uncertainty. The best known and oldest one is the probabilistic framework, where uncertainty is modeled by classical probability distributions. Although the probabilistic framework is of major importance in the treatment of uncertainty due

* Corresponding author

Email addresses: desterck@irit.fr (S. Destercke), dubois@irit.fr (D. Dubois), eric.chojnacki@irsn.fr (E. Chojnacki).

to variability, many arguments converge to the fact that a single probability distribution cannot adequately account for incomplete or imprecise information. Alternative theories and frameworks have been proposed to this end. The three main such frameworks, are, in decreasing order of generality, Imprecise probability theory [35], Random disjunctive sets [8,24,32] and Possibility theory [12,36].

Generally speaking, the more general a theory, the more expressive, but the more computationally demanding. In practice, simplified representations can greatly increase the efficiency of uncertainty handling. Among these simpler representations are possibility distributions [36], probability intervals [2], p-boxes [15] and, more recently, clouds [26, 27]. Such representations can be viewed as encoding special sets of probabilities, some of them special kinds of random sets. They are potentially important when trading expressiveness (possibly losing some information) against computational efficiency. Sometimes, the available information is simple enough to be faithfully captured by one of these representations. Moreover, simpler representations are instrumental in elicitation tasks. They also facilitate the presentation and interpretation of complex results.

Such a diversity of simplified representations motivates the study of their respective expressive power. Finding formal links between them also facilitates a unified handling and treatment of uncertainty, and suggest how tools used for one theory can eventually be useful in the setting of other theories. Finding such links is the purpose of the present study which, among other things, extends some results by Baudrit and Dubois [1] concerning the relations between p-boxes and possibility measures. This paper introduces a generalized form of p-box, that is also more general than a possibility distribution, thus bridging the gap between them. A generalized p-box basically consists of two comonotonic fuzzy sets, and is thus easy to represent. It is also the missing link between p-boxes, possibility distributions, and clouds, the latter being more general and studied in the companion paper.

The paper is divided in two main sections:

- Section 2 recalls the basic settings of imprecise probability theory, random disjunctive sets and possibility theory, as well as some of their previously known practical representation tools, and their already known relationships.
- Section 3 introduces generalized p-boxes. Section 3.2 bridges the gap between possibility theory and generalized p-boxes. The relationship between random sets and generalized p-boxes is explored in Section 3.3. It is shown that generalized p-boxes can be interpreted as a special case of random sets. Probability intervals and generalized p-boxes are less closely related, but Section 3.4 discusses some transformation methods to extract probability intervals from p-boxes, and vice-versa. We prove that the information

modeled by probability intervals correspond to a set of generalized p-boxes.

Generalized p-boxes are defined on arbitrary finite spaces. The infinite setting is studied in the companion paper. To make the paper easier to read, longer proofs have been moved to the appendix.

2 Non-additive uncertainty theories and some representation tools

Bayesian subjectivists advocate the use of single probability distributions in all circumstances. If enough statistical information is available, probability distributions are good models of uncertainty due to variability, but if the available information lacks precision or is incomplete, then choosing a unique probability distribution to represent uncertainty is debatable [1] . It generally forces to engage in too strong a commitment about the current situation than what is actually known.

Roughly speaking, alternative theories recalled here (imprecise probabilities, random sets, and possibility theory) have the potential to lay bare the existing imprecision or incompleteness in the information. They evaluate uncertainty on a particular event by means of a pair of (conjugate) lower an upper measures rather than by a single one. The difference between the upper and lower measures than reflecting the lack of precision in our knowledge.

In this section, we first recall basic mathematical notions used in the sequel, concerning capacities and the Moebius transform. Each theory mentioned above is then briefly introduced, with focus on practical representation tools available as of to-date, like possibility distributions p-boxes and probability intervals, their expressive power and complexity. Although no particular emphasis is given to interpretation issues in this paper, we nevertheless feel that is important to recall that such theories were often motivated by the modeling of specific phenomena such as variability, belief and so on, which still remains a matter of debate.

We recall that, in this paper, unless explicitly stated otherwise, we restrict ourselves to a finite arbitrary space X containing n elements denoted x or x_i.

[1] For instance, the following statement about a coin: "We are not sure that this coin is fair, so the probability for this coin to land on Heads (or Tails) lies between 1/4 and 3/4" cannot be faithfully modeled by a single probability.

Uncertainty is often represented by set-functions called *capacities*.

Definition 2.1 *Given a finite space X, a capacity on X is a function μ, defined on the set of subsets of X, such that:*

- $\mu(\emptyset) = 0, \mu(X) = 1$
- $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$.

A capacity such that

$$\forall A, B \subseteq X, A \cap B = \emptyset, \mu(A \cup B) \geq \mu(A) + \mu(B)$$

is said to be super-additive*. The dual notion, called* sub-additivity*, is obtained by reversing the inequality.*

A capacity is said to be *additive* if the inequality is turned into an equality. Then the capacity is called a probability measure and denoted P. As the set X is finite, a probability P can also be expressed by its probability distribution p defined on X such that $p(x) = P(\{x\})$. Then $\forall x \in X, p(x) \geq 0, \sum_{x \in X} p(x) = 1$ and $P(A) = \sum_{x \in A} p(x)$.

Capacities were first introduced in Choquet's work [4]. Given a capacity μ, one can also define a conjugate capacity μ^c, defined by $\mu^c(E) = \mu(X) - \mu(E^c) = 1 - \mu(E^c)$, for any subset E of X, E^c being its complement. Here we consider capacities such that either $\forall E \subset X, \mu(E) + \mu(E^c) \leq 1$ or $\forall E \subset X, \mu(E) + \mu(E^c) \geq 1$. The former can be called *cautious* capacities (since $\mu(E) \leq \mu^c(E), \forall E$) and the latter *bold* capacities. Super-additive capacities are cautious and sub-additive capacities are bold. A probability measure is self-conjugate ($P = P^c$), ie. is both cautious and bold.

In the following, the capacity of a subset evaluates the degree of confidence in the corresponding event. Cautious capacities are tailored for modelling the idea of certainty. Bold capacities may account for the weaker notion of plausibility. These kinds of capacities are also the basic tool for modelling lower and upper probabilities.

The core of a capacity μ is the (convex) set of probability measures that dominate μ, that is, $\mathcal{P}(\mu) = \{P \geq \mu\}$. This set may be empty even if the capacity is cautious. We need stronger properties to ensure a nonempty core.

Definition 2.2 *A super-additive capacity μ is $n-monotone$, where $n > 0$ and $n \in \mathbb{N}$, if and only if for any set $\mathcal{A} = \{A_i | i \in \mathbb{N}, 0 < i \leq n\}$ of events A_i,* *it holds that*

$$\mu(\bigcup_{A_i \in \mathcal{A}} A_i) \geq \sum_{I \subseteq \mathcal{A}} (-1)^{|I|+1} \mu(\bigcap_{A_i \in I} A_i)$$

An n-monotone capacity is also called a Choquet capacity of order n. If a capacity is n-monotone, then it is also $(n-1)$-monotone, but not necessarily $(n+1)$-monotone. An ∞-monotone capacity is a capacity that is n-monotone for every $n > 0$. On a finite space, a capacity is ∞-monotone if it is n-monotone with $n = |X|$. The two particular cases of 2-monotone (also called convex) capacities and ∞-monotone capacities have deserved special attention in the literature [3,23,35]. Indeed, 2-monotone capacities have a non-empty core. So, n-monotone capacities can be viewed as lower probabilities. Dual capacities, corresponding to upper probabilities, are called n-alternating capacities. ∞-monotone capacities have interesting mathematical properties that greatly increase computational efficiency. As we will see, many of the representations studied in this paper possess such properties. Extensions of the notion of capacity and of n-monotonicity have been studied by de Cooman *et al.* [7]

Given a capacity μ on X, one can obtain multiple equivalent representations by applying various (bijective) transformations [16] to it. One such transformation, useful in this paper, is the Möbius inverse:

Definition 2.3 *Given a capacity μ on X, its Möbius transform is a mapping $m : 2^{|X|} \to \mathbb{R}$ from the power set of X to the real line, which associates to any subset E of X the value*

$$m(E) = \sum_{\substack{B \subset E \\ B \in X}} (-1)^{|E-B|} \mu(B)$$

Since $\mu(X) = 1$, $\sum_{E \in X} m(E) = 1$ as well, and $m(\emptyset) = 0$. Moreover, it can be shown [32] that the values $m(E)$ are non-negative for all subsets E of X (hence $\forall E \in X, 1 \geq m(E) \geq 0$) if and only if the capacity μ is ∞-monotone. Then m is called a mass assignment. Otherwise, there are some (non-singleton) events E for which $m(E)$ is negative. Such a set-function m is actually the unique solution to the set of 2^n equations

$$\forall A \subseteq X, \mu(A) = \sum_{E \subseteq A} m(E),$$

given any capacity μ. The Möbius transform of a probability measure P coincides with its distribution p, assigning positive masses to singletons only.

2.2 Imprecise probability theory

The theory of imprecise probabilities has been systematized and popularized by Walley's book [35]. In this theory, uncertainty is modeled by closed convex sets $\mathcal{P}$ of (finitely additive) probability measures P on X. In the rest of the paper, such convex sets will be named *credal sets* (as is often done). It is important to stress that, even if they share similarities, Walley's behavioral interpretation of imprecise probabilities is different from the one of classical robust statistics [2] [18].

Imprecise probability theory is very general, and, from a purely mathematical point of view, it encompasses all theories studied in this paper. Thus, in all approaches presented here, a corresponding credal set can be generated, which facilitates their comparison.

2.2.1 Lower/upper probabilities

Credal sets in Walley's theory can be represented by lower bounds (called lower previsions) on the mean values of so-called gambles (bounded real-valued functions with domain X). In this paper, credal sets induced by lower probabilities (lower previsions assigned to events) are sufficient to our purpose. A *lower probability* $\underline{P}$ is a super-additive capacity. Its conjugate $\overline{P}(A) = 1 - \underline{P}(A^c)$ is called an upper probability. The credal set $\mathcal{P}_{\underline{P},\overline{P}}$ induced by a given lower probability is its core:

$$\mathcal{P}_{\underline{P},\overline{P}} = \{P | \forall A \subset X,\ P(A) \geq \underline{P}(A)\}.$$

Conversely a credal set $\mathcal{P}$ induces a lower envelope P_* on events, defined by $\forall A, P_*(A) = \inf_{P \in \mathcal{P}} P(A)$. As a lower envelope is a super-additive capacity, it is a lower probability. The upper envelope $P*(A) = \sup_{P \in \mathcal{P}} P(A)$ is the conjugate of P_*. In general, a credal set $\mathcal{P}$ is included in the core of its lower envelope: $\mathcal{P} \subset \mathcal{P}_{\underline{P},\overline{P}}$, since $\mathcal{P}_{\underline{P},\overline{P}}$ can be seen as a projection of $\mathcal{P}$ on events.

In this paper, we consider so-called *coherent* lower probabilities $\underline{P}$, that is, lower probabilities that coincide with the lower envelopes of their core, i.e. for all events A of X,

$$\underline{P}(A) = \inf_{P \in \mathcal{P}_{\underline{P},\overline{P}}} P(A).$$

In other words, $\mathcal{P}_{\underline{P},\overline{P}}$ is such that for every event A, there is a probability

[2] Roughly speaking, in Walley's approach, the primitive notions are lower and upper previsions or sets of so-called desirable gambles, and the fact that there always exists a "true" precise probability distribution is not assumed.

distribution P in $\mathcal{P}_{\underline{P},\overline{P}}$ such that $P(A) = \underline{P}(A)$. The same property holds for the associated upper probability ($\sup_{P\in\mathcal{P}} P(A) = \overline{P}(A)$).

A credal set $\mathcal{P}_{\underline{P},\overline{P}}$ can also be described by a set of constraints on probability assignments to elements of X:

$$\underline{P}(A) \leq \sum_{x\in A} p(x) \leq \overline{P}(A).$$

Finally, let us note that, since the measures $\underline{P}, \overline{P}$ are conjugate, specifying one of them on all events is enough to completely characterize the credal set $\mathcal{P}_{\underline{P},\overline{P}}$. This means that, when X is finite, $2^{|X|} - 2$ values ($|X|$ being the cardinality of X), are needed in addition to constraints $\underline{P}(X) = 1, \underline{P}(\emptyset) = 0$ to completely specify $\mathcal{P}_{\underline{P},\overline{P}}$.

2.2.2 Simplified representations

Although credal sets induced by lower probabilities are already a restriction compared to more general representations, they can still be difficult to handle or to elicit. This is why, in practice, simpler representations are used. P-boxes and interval probabilities are two such simplified tools.

P-boxes

Let us first recall some background on cumulative distributions. Let P be a probability measure on the real line $\mathbb{R}$ and p its probability density. Its *cumulative distribution* $P((-\infty, r]), r \in \mathbb{R}$ is denoted F^p.

Let F_1 and F_2 be two cumulative distributions. Then, F_1 is said to stochastically dominate F_2 if only if F_1 is point-wise lower than F_2: $F_1 \leq F_2$.

A p-box [15] is then defined as a pair of (discrete) cumulative distributions $[\underline{F}, \overline{F}]$ such that $\underline{F}$ stochastically dominates $\overline{F}$ ($\underline{F} \leq \overline{F}$). A p-box induces a credal set $\mathcal{P}_{[\underline{F},\overline{F}]}$ such that:

$$\mathcal{P}_{[\underline{F},\overline{F}]} = \{P|\underline{F}(r) \leq P((-\infty, r]) \leq \overline{F}(r)\ \forall r \in \mathbb{R}\} \quad (1)$$

We can already notice that since sets $(-\infty, x]$ are nested, $\mathcal{P}_{[\underline{F},\overline{F}]}$ is described by constraints that are lower and upper bounds on a collection of nested sets. This interesting characteristic will be crucial in the generalized form of p-box we introduce in section 3. Conversely we can extract a p-box from a credal set $\mathcal{P}$ by considering its lower and upper envelopes restricted to events of the form $(-\infty, x]$, namely, letting $\underline{F}(x) = P_*((-\infty, x]), \overline{F}(x) = P^*((-\infty, x])$. The core of this p-box is a gross approximation of $\mathcal{P}$.

Cumulative distributions are often used to elicit probabilistic knowledge from experts [5]. P-boxes can thus directly benefit from such methods and tools, with the advantages of allowing some imprecision in the representation (e.g., allowing experts to give imprecise percentiles). P-boxes are also sufficient to represent final results produced by imprecise probability models when only a threshold violation has to be checked.

Probability intervals

Another example of a simple representation of imprecise probability is provided by *probability intervals* on a finite set X. They are defined as lower and upper bounds of probability distributions. They are defined by a set of intervals $L = \{[l(x), u(x)] | x \in X\}$ such that $l(x) \leq p(x) \leq u(x), \forall x \in X$, where $p(x) = P(\{x\})$. Probability intervals are extensively studied by De Campos *et al.* [2]. A probability interval induces the credal set

$$\mathcal{P}_L = \{P | l(x) \leq p(x) \leq u(x),\ x \in X\}$$

$\mathcal{P}_L$ is thus totally determined by $2|X|$ values only. De campos *et al.* [2] have shown that probability intervals have numerous computational advantages.

A probability interval L is called reachable if its credal set is not empty and, for each x, we can find at least one probability measure $P \in \mathcal{P}_L$ s.t. $p(x) = l(x)$ and one for which $p(x) = u(x)$. In other words, each bound can be reached by a probability measure in $\mathcal{P}_L$). Non-emptiness and reachability respectively correspond to the conditions [2]:

$$\sum_{x \in X} l(x) \leq 1 \leq \sum_{x \in X} u(x) \qquad \text{non-emptiness}$$

$$u(x) + \sum_{y \in X \setminus \{x\}} l(y) \leq 1 \text{ and } l(x) + \sum_{y \in X \setminus \{x\}} u(y) \geq 1 \qquad \text{reachability}$$

If a probability interval L is non-reachable, it can be transformed into a probability interval L', by letting $l'(x) = \inf_{P \in \mathcal{P}_L}(p(x))$ and $u'(x) = \sup_{P \in \mathcal{P}_L}(p(x))$. More generally, coherent lower and upper probabilities $\underline{P}(A), \overline{P}(A)$ induced by $\mathcal{P}_L$ on all events $A \subset X$ are easily calculated by the following expressions

$$\underline{P}(A) = \max(\sum_{x \in A} l(x), 1 - \sum_{x \in A^c} u(x)), \overline{P}(A) = \min(\sum_{x \in A} u(x), 1 - \sum_{x \in A^c} l(x)). \quad (2)$$

De Campos *et al.* [2] have shown that these lower and upper probabilities are Choquet capacities of order 2.

Probability intervals are very convenient tools to model uncertainty on multinomial data, where they can express lower and upper confidence probability bounds. They can thus be easily used when one has a sample of small size [22]. On the real line, discrete probability intervals correspond to imprecisely known histograms. Probability intervals can be extracted, as useful information, from any credal set $\mathcal{P}$ on a finite set X, by constructing $L_{\mathcal{P}} = \{[\underline{P}(\{x\}, \overline{P}(\{x\}], x \in X\}$.

2.3 *Random disjunctive sets*

A more specialized setting for representing partial knowledge is that of a random set, where each set represents an incomplete observation, and the probability bearing on this set could potentially be shared among its elements, but is not by lack of sufficient information.

2.3.1 *Belief and Plausibility functions*

Formally, a random set is a mapping $\Gamma : \Omega \to \wp(X)$ from a probability space $(\Omega, \mathcal{A}, P)$ to the power set $\wp(X)$ of another space X (here finite). It is also called a multi-valued mapping Γ. Insofar as sets $\Gamma(\omega)$ represent incomplete knowledge about a random variable, such sets contain mutually exclusive elements and are called *disjunctive* sets [3]. Then this mapping induces the following coherent lower and upper probabilities on X for all events A [8] (representing all probability functions on X that could be found if the available information were complete):

$$\underline{P}(A) = P(\{\omega \in \Omega | \Gamma(\omega) \subseteq A\}) \tag{3}$$
$$\overline{P}(A) = P(\{\omega \in \Omega | \Gamma(\omega) \cap A \neq \emptyset\}) \tag{4}$$

When X is finite, a random set can also be represented as a mass assignment m over the power set $\wp(X)$ of X, letting $m(E) = P(\{\omega, \Gamma(\omega) = E\}), \forall E \in X$. Then, $\sum_{E \subseteq X} m(E) = 1$ and $m(\emptyset) = 0$. A set E that receives strict positive mass is called a focal set, and the mass $m(E)$ can be interpreted as the probability that the most precise description of the actual solution to the problem is of the form "$x \in E$". From this mass assignment , Shafer [32] define two set functions, called *belief and plausibility functions*, respectively:

$$Bel(A) = \sum_{E, E \subseteq A} m(E); \qquad Pl(A) = 1 - Bel(A^c) = \sum_{E, E \cap A \neq \emptyset} m(E).$$

On finite spaces, since the mass assignment is positive, the belief function derived from a random set is an ∞-monotone capacity. The mass assignment

[3] as opposed to sets as collections of objects, i.e. sets whose elements are jointly present, such as a region in a digital image.

m is indeed the Möbius transform of the capacity Bel. Conversely, any ∞-monotone capacity is induced by one and only one random set. We can thus speak of the random set underlying Bel. In the sequel, we will use this notation for lower probabilities stemming from random sets (Dempster and Shafer definitions being equivalent on finite spaces). Smets [34] has studied the case of continuous random intervals defined on the real line $\mathbb{R}$, where the mass function is replaced by a mass density bearing on pairs of interval endpoints.

Belief functions can be considered as special cases of coherent lower probabilities, since they are ∞-monotone capacities. This lower probability is given by equation (3). A random set thus induces the credal set $\mathcal{P}_{Bel} = \{P | \forall A \subseteq X, Bel(A) \leq P(A) \leq Pl(A)\}$.

Note that Shafer [32] does not refer to an underlying probability space, nor does he uses the fact that a belief function is a lower probability: in his view, extensively taken over by Smets [33], $Bel(A)$ is supposed to quantify an agent's belief per se with no reference to a probability. However, the primary mathematical tool common to Dempster's upper and lower probabilities and to the Shafer-Smets view is the notion of (generally finite) random disjunctive set.

2.3.2 Practical aspects

In general, $2^{|X|} - 2$ values are still needed to completely specify a random set, thus not forcefully reducing the complexity of the model representation with respect to capacities. However, simple belief functions defined by only a few positive focal elements do not exhibit such complexity. For instance, a simple support belief function is a natural model of an unreliable testimony, namely an expert stating that the value of a parameter x belong to set $A \subseteq X$. Le α be the reliability of the expert testimony, i.e.the probability that the information is irrelevant. The corresponding mass assignment is $m(A) = \alpha, m(X) = 1-\alpha$. More generally, imprecise results from statistical experiments are easily expressed by means of random sets, $m(A)$ being the probability of an observation of the form $x \in A$.

As practical models of uncertainty, random sets have many advantages. First, as they can be seen as probability distributions over subsets of X, they can be easily simulated by classical methods such as Monte-Carlo sampling, which is not the case for other Choquet capacities. On the real line, a random set is often restricted to a finite collection of closed intervals with associated weights, and one can then easily extend results from interval analysis [25] to random intervals [14, 17].

2.4 Possibility theory

The primary mathematical tool of possibility theory is the possibility distribution, which is a set-valued piece of information where some elements are more plausible than others. To a possibility distribution are associated specific measures of certainty and plausibility.

2.5 Possibility and necessity measures

A possibility distribution is a mapping $\pi : X \to [0,1]$ from a (here finite) space X to the unit interval such that $\pi(x) = 1$ for at least one element x in X. Formally, a possibility distribution is equivalent to the membership function of a fuzzy set [36] Twenty years earlier, Shackle [31] had introduced an equivalent notion called distribution of potential surprise (corresponding to $1 - \pi(x)$) with a view to represent non-probabilistic uncertainty.

Several set-functions can be defined from a possibility distribution π [11]:

$$\text{Possibility measures: } \Pi(A) = \sup_{x \in A} \pi(x). \tag{5}$$

$$\text{Necessity measures: } N(A) = 1 - \Pi(A^c). \tag{6}$$

$$\text{Sufficiency measures: } \Delta(A) = \inf_{x \in A} \pi(x). \tag{7}$$

The possibility degree of an event A evaluates the extent to which this event is plausible, i.e., consistent with the available. Necessity degrees express the certainty of events, by duality. In this context, distribution π is potential (in the spirit of Shackle's), i.e. $\pi(x) = 1$ does not guarantee the existence of x. Their characteristic property are: $N(A \cap B) = \min(N(A), N(B))$ and $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$ for any pair of events A, B of X.

On the contrary $\Delta(A)$ measures the extent to which all states of the world where A occurs are plausible. Sufficency (or guaranteed possibility) distributions [11] generally denoted by δ, are understood as degree of empirical support and obey an opposite convention: $\delta(x) = 1$ guarantees (= is sufficient for) the existence of x.

Given a possibility distribution π and a degree $\alpha \in [0,1]$, strong and regular α-cuts are subsets respectively defined as $A_{\overline{\alpha}} = \{x \in X | \pi(x) > \alpha\}$ and $A_\alpha = \{x \in X | \pi(x) \geq \alpha\}$. These α-cuts are nested, since if $\alpha > \beta$, then $A_\alpha \subset A_\beta$. On finite spaces, the set $\{\pi(x), x \in X\}$ is of the form $\alpha_0 = 0 < \alpha_1 < \ldots < \alpha_M = 1$. Then, there will only be M distinct α-cuts.

A necessity measure (resp a possibility measure)can be viewed as a belief function (resp. a plausibility function), with nested focal elements (already in [32]). A possibility distribution π, defines a random set having the following focal sets E_i order on Xes $m(E_i)$, $i = 1, \ldots, m$:

$$\begin{cases} E_i = \{x \in X | \pi(x) \geq \alpha_i\} = A_{\alpha_i} \\ \qquad m(E_i) = \alpha_i - \alpha_{i-1} \end{cases} \tag{8}$$

In this nested situation, the same amount of information is contained in the mass function m and the possibility distribution $\pi(x) = Pl(\{x\})$. For instance a simple support belief function such that $m(A) = \alpha, m(X) = 1 - \alpha$ forms a nested structure, and yields the possibility distribution $\pi(x) = 1$ if $x \in A$ and $1 - \alpha$ otherwise. In the general case, m cannot be reconstructed from π. Outer and inner approximations of general random sets in terms of possibility distributions have been studied by Dubois and Prade in [13].

Since the necessity measure is a particular belief function it is also an ∞-monotone capacity, hence a particular coherent lower probability. If the necessity measure is viewed as a coherent lower probability, its possibility distribution induces the credal set $\mathcal{P}_\pi = \{P | \forall A \subseteq X,\ N(A) \leq P(A) \leq \Pi(A)\}$.

We recall here a result, proved by Dubois *et al.* [10], and which links probabilities P that are in $\mathcal{P}_\pi$ with constraints on α-cuts, that will be useful in the sequel:

Proposition 2.4 *Given a possibility distribution π and the induced convex set $\mathcal{P}_\pi$, we have for all α in $(0,1]$, $P \in \mathcal{P}_\pi$ if and only if*

$$1 - \alpha \leq P(\{x \in X | \pi(x) > \alpha\})$$

This result means that the probabilities P in the credal set $\mathcal{P}_\pi$ can also be described in terms of constraints on strong α-cuts of π (i.e. $1 - \alpha \leq P(A_{\overline{\alpha}})$).

2.6.1 Practical aspects

At most $|X|$ values are needed to fully assess a possibility distribution, which makes it the simplest uncertainty representation explicitly coping with imprecise or incomplete knowledge. This simplicity makes this representation very easy to handle. This also implies less expressive power, in the sense that, for any event A , either $\Pi(A) = 1$ or $N(A) = 0$ (i.e. intervals $[N(A), \Pi(A)]$ are of the form $[0, \alpha]$ or $[\beta, 1]$). This means that, in several situations, possibility distributions will be insufficient to reflect the available information.

Nevertheless, the expressive power of possibility distributions fits various practical situations. Indeed, they can be interpreted as a set of nested sets with different confidence degrees (the bigger the set, the highest the confidence degree). Moreover, a recent psychological study [29] shows that possibility distributions are convenient in elicitation procedures. On the real line [10], possibility distributions can model, for example, an expert opinion concerning the value of a badly known parameter by means of a finite collection of nested confidence intervals. Similarly, it is natural to view nested confidence intervals coming from statistics as a possibility distribution. Another practical case where uncertainty can be modeled by possibility distributions is the case of vague linguistic assessments concerning probabilities [6].

2.7 P-boxes and probability intervals in the imprecise probability landscape

P-boxes, reachable probability intervals, random sets and possibility distributions can be modeled by credal sets and define coherent upper and lower probabilities. Kriegler and Held [20] show that the credal set of any discrete p-box can be described by an equivalent random-set. However, there are different random sets inducing a given p-box. So p-boxes are a special case of random sets.

There is no inclusion relationship between the sets of possibility distributions, p-boxes and probability intervals. None can be seen as a special case of the other. Baudrit and Dubois [1] showed that a possibility distribution π induces a p-box whose credal set is larger than $\mathcal{P}_\pi$. The rest of our paper is devoted to a generalized version of p-boxes that covers possibility distributions as a special case, and is also representable by a random set.

There is no direct relationship between probability intervals and random sets. Indeed upper and lower probabilities induced by reachable probability intervals are order 2 capacities only, while belief functions are ∞-monotone. In general, one can only approximate one representation by the other.

Transforming a belief function Bel into the tightest set L of probability intervals such that $\mathcal{P}_{Bel} \subset \mathcal{P}_L$ (i.e. L is an outer approximation of the random set) is simple, and consists of taking for all $x \in X$:

$$l(x) = Bel(x) \text{ and } u(x) = Pl(x)$$

and since belief and plausibility functions are the lower envelope of the induced set $\mathcal{P}_{Bel}$, we are sure that the so-built probability interval L is reachable. This

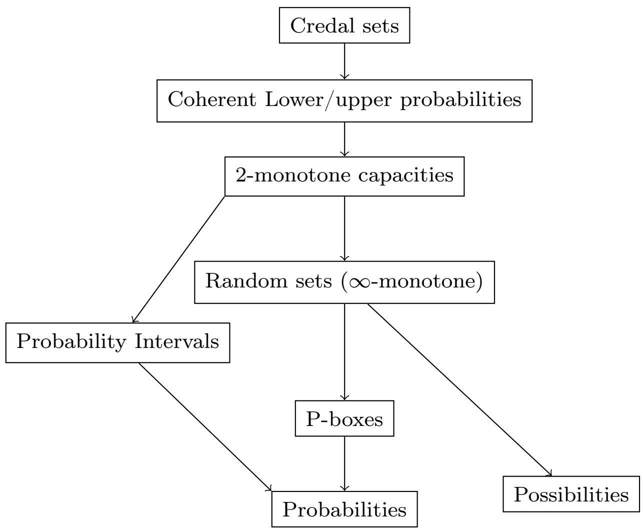

Figure 1. Representation relationships: summary $A \longrightarrow B$: B is a special case of A

method can be used to extract probability intervals from any credal set as shown earlier.

The converse problem, i.e. to transform a set L of probability intervals into a random set was studied by Lemmer and Kyburg [21]. They concentrate on transforming the set L into an inner approximation (i.e., $\mathcal{P}_{Bel} \subset \mathcal{P}_L$). On the contrary, Denoeux [9], extensively studies the problem of transforming a set L of probability intervals into a random set that is an outer approximation (i.e., $\mathcal{P}_L \subset \mathcal{P}_{Bel}$). The transformation of a given set L of probability intervals into a possibility distribution is studied by Masson and Denoeux [22], who propose efficient methods to achieve such a transformation.

The main relations existing between imprecise probabilities, lower/upper probabilities, random sets, probability intervals, p-boxes and possibility distributions, are pictured on Figure 1. From top to bottom, it goes from the more general, expressive and complex theories to the less general, less expressive but simpler representations. An arrow is directed from a general representation to a less general one.

3 Generalized p-boxes

As recalled in Section 2.2, p-boxes are useful representations of uncertainty in many practical applications. So far, they only make sense on the (discretized) real line and their definition requires the natural ordering of numbers. This is a bit restrictive, and since the model is already quite useful in this restrictive setting, it would be interesting to extend the model to more general settings. Moreover, such extensions can give a better understanding of the nature of this representation.

This extension, called here *generalized p-boxes*, and its study is the subject of the present section. We first define the extension of p-boxes to arbitrary finite spaces in Section 3.1. Then we show that this model has very close link with possibility distributions and random sets. Actually, we show in Section 3.2 that a generalized p-box corresponds to a pair of possibility distributions. In Section 3.3 that any generalized p-box is a particular case of random set. We then explore in Section 3.4 transformations between generalized p-boxes and probability intervals.

3.1 Definition of generalized p-boxes

Two mappings f and f' from a finite set $X = \{x_1, \dots, x_n\}$ to the real line are said to be comonotonic if there is a common permutation σ of $\{1, 2, \dots, n\}$ such that $f(\sigma(1)) \geq f(\sigma(2)) \geq \cdots \geq f(\sigma(n))$ and $f'(\sigma(1)) \geq f'(\sigma(2)) \geq \cdots \geq f'(\sigma(n))$. We define a generalized p-box as follows:

Definition 3.1 *A generalized p-box $[\underline{F}, \overline{F}]$ over a finite space X is a pair of comonotonic mappings $\underline{F}, \overline{F}$, $\underline{F} : X \to [0,1]$ and $\overline{F} : X \to [0,1]$ from X to $[0,1]$ such that $\underline{F}$ is pointwise less than $\overline{F}$ (i.e. $\underline{F} \leq \overline{F}$) and there is at least one element x in X for which $\overline{F}(x) = \underline{F}(x) = 1$.*

Since each distribution $\underline{F}, \overline{F}$ is fully specified by $|X| - 1$ values, it follows that $2|X| - 2$ values completely determine a generalized p-box. Contrary to usual p-boxes, no notion of ordering is used in this definition. In order to relate this definition with usual p-boxes, we must notice that, given a generalized p-box $[\underline{F}, \overline{F}]$, we can always define a *complete* pre-ordering $\leq_{[\underline{F},\overline{F}]}$ on X such that $x \leq_{[\underline{F},\overline{F}]} y$ if $\underline{F}(x) \leq \underline{F}(y)$ and $\overline{F}(x) \leq \overline{F}(y)$, due to the comonotonicity condition. If X is the real line and if $\leq_{[\underline{F},\overline{F}]}$ is the natural ordering of numbers, then we retrieve usual p-boxes.

To simplify notations in the sequel, we will consider that, given a generalized p-box $[\underline{F}, \overline{F}]$, elements x of X are indexed such that $x_i \leq_{[\underline{F},\overline{F}]} x_j$ if and only if $i < j$. A $[\underline{F}, \overline{F}]$-downset, denoted $(x]_{[\underline{F},\overline{F}]}$, will be of the form $\{x_i | x_i \leq_{[\underline{F},\overline{F}]} x\}$.

The credal set induced by a generalized p-box $[\underline{F}, \overline{F}]$ can now be defined as

$$\mathcal{P}_{[\underline{F},\overline{F}]} = \{P | \underline{F}(x_i) \leq P((x_i]_{[\underline{F},\overline{F}]}) \leq \overline{F}(x_i)\}.$$

It induces coherent upper and lower probabilities such that $\underline{F}(x_i) = \underline{P}((x_i]_{[\underline{F},\overline{F}]})$ and $\overline{F}(x_i) = \overline{P}((x_i]_{[\underline{F},\overline{F}]})$. Again, if we consider real numbers $\mathbb{R}$ and the natural ordering on them, then $\forall r \in \mathbb{R}$, $(r]_{[\underline{F},\overline{F}]} = (-\infty, r]$, and the above equation coincides with Equation (1).

In the following, the sets $(x_i]_{[\underline{F},\overline{F}]}$ are denoted A_i, for all $i = 1, \ldots, n$. These sets are nested, since $\emptyset \subset A_1 \subseteq \ldots \subseteq A_n = X$[4]. For all $i = 1, \ldots, n$, let $\underline{F}(x_i) = \alpha_i$ and $\overline{F}(x_i) = \beta_i$. With these conventions, the credal set $\mathcal{P}_{[\underline{F},\overline{F}]}$ can now be described by the following constraints bearing on probabilities of nested sets A_i:

$$i = 1, \ldots, n \qquad \alpha_i \leq P(A_i) \leq \beta_i \tag{9}$$

with $0 = \alpha_0 \leq \alpha_1 \leq \ldots \leq \alpha_n = 1$, $0 = \beta_0 < \beta_1 \leq \beta_2 \leq \ldots \leq \beta_n = 1$ and $\alpha_i \leq \beta_i$.

As a consequence, a generalized p-box can be generated in two different ways:

- Either we start from two comonotone functions $\underline{F}, \overline{F}$ on the space X, and the order on X is then induced by the values taken by these functions,
- or a generalized p-box is built by assigning upper and lower bounds on probabilities of nested sets, (i.e. sets A_i built or not from a complete ordering on X).

The second approach is likely to be more useful in practical assessments of generalized p-boxes.

Example 3.2 *All along this section, we will use this example to illustrate results on generalized p-boxes. Consider a space X made of six elements $\{x_1, \ldots, x_6\}$. These elements could be, for instance, successive components on a production line. For various reasons (cost, production constraints, ...), when a component breaks down, the safety system only informs us whether the broken component is in the set $A_1 = \{x_1, x_2\}$, $A_2 = \{x_1, x_2, x_3\}$, $A_3 = \{x_1, x_2, x_3, x_4, x_5\}$, or the whole $X(= A_4)$. Asking an expert to evaluate breakdowns, he can only give us lower and upper probability bounds for each of these sets:*

$$P(A_1) \in [0, 0.3] \qquad P(A_2) \in [0.2, 0.7] \qquad P(A_3) \in [0.5, 0.9]$$

Since these sets are nested, the uncertainty can be modeled by the generalized p-box pictured on Figure 2:

$$\overline{F}(x_1) = 0.3 \; \overline{F}(x_2) = 0.3 \; \overline{F}(x_3) = 0.7 \; \overline{F}(x_4) = 0.9 \; \overline{F}(x_5) = 0.9 \; \overline{F}(x_6) = 1$$

$$\underline{F}(x_1) = 0 \quad \underline{F}(x_2) = 0 \quad \underline{F}(x_3) = 0.2 \; \underline{F}(x_4) = 0.5 \; \underline{F}(x_5) = 0.5 \; \underline{F}(x_6) = 1.$$

[4] Since there is a complete pre-order on X, we can have $x_i =_{[\underline{F},\overline{F}]} x_{i+1}$ and $A_i = A_{i+1}$, which explains the non-strict inclusions. They would be strict if $<_{[\underline{F},\overline{F}]}$ were a linear order.

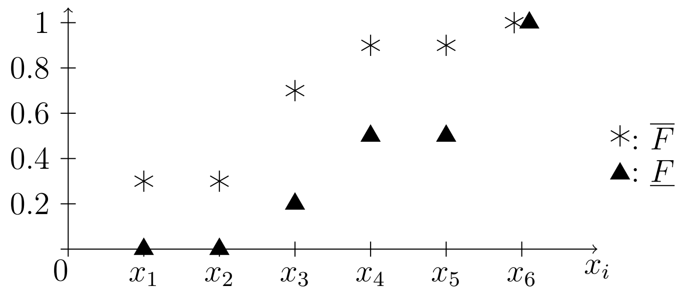

Figure 2. Generalized p-box $[\underline{F}, \overline{F}]$ of Example 3.2

3.2 Connecting generalized p-boxes with possibility distributions

Kozine and Utkin [19] discussed the problem of building p-boxes from partial information. They already noticed that, for usual p-boxes, sets A_i can be interpreted as nested confidence intervals with upper and lower bounds. It is therefore natural to search a connection with possibility theory, since possibility distributions can be interpreted as a collection of nested sets with lower probability bounds. Given a generalized p-box $[\underline{F}, \overline{F}]$, the following proposition holds:

Proposition 3.3 *A generalized p-box $[\underline{F}, \overline{F}]$ on a finite set X can be encoded by a pair of possibility distributions $\pi_{\overline{F}}, \pi_{\underline{F}}$, such that $\mathcal{P}_{[\underline{F},\overline{F}]} = \mathcal{P}_{\pi_{\overline{F}}} \cap \mathcal{P}_{\pi_{\underline{F}}}$, where:*

$$\pi_{\overline{F}}(x_i) = \beta_i \text{ and } \pi_{\underline{F}}(x_i) = 1 - \max\{\alpha_j | j = 0, \dots, i \ \alpha_j < \alpha_i\}$$

for $i = 1, \dots, n$, with $\alpha_0 = 0$.

Proof of Proposition 3.3 Consider the set of constraints given by Equation (9) and describing the convex set $\mathcal{P}_{[\underline{F},\overline{F}]}$. These constraints can be separated into two distinct sets: $(P(A_i) \leq \beta_i)_{i=1,n}$ and $(P(A_i^c) \leq 1 - \alpha_i)_{i=1,n}$. Now, rewrite constraints of Proposition 2.4, in the form $\forall \alpha \in (0, 1]$: $P \in \mathcal{P}_\pi$ if and only if $P(\{x \in X | \pi(x) \leq \alpha\}) \leq \alpha$.

The first set of constraints $(P(A_i) \leq \beta_i)_{i=1,n}$ defines a credal set $\mathcal{P}_{\pi_{\overline{F}}}$ that is induced by the possibility distribution $\pi_{\overline{F}}$, while the second set of constraints $(P(A_i^c) \leq 1 - \alpha_i)_{i=1,n}$ defines a credal set $\mathcal{P}_{\pi_{\underline{F}}}$ that is induced by the possibility distribution $\pi_{\underline{F}}$, since $A_i^c = \{x_k, \dots x_n\}$, where $k = \max\{j | \alpha_j < \alpha_i\}$. The credal set of the generalized p-box $[\underline{F}, \overline{F}]$, resulting from the two sets of constraints, namely $i = 1, \dots, n, \quad \beta_i \leq P(A_i) \leq \alpha_i$, is thus $\mathcal{P}_{\pi_{\overline{F}}} \cap \mathcal{P}_{\pi_{\underline{F}}}$. □

Note that if $\underline{F}$ is injective, it induces a complete order $<_{[\underline{F},\overline{F}]}$, and then $\pi_{\underline{F}}(x_i) = 1 - \alpha_{i-1}$

Example 3.4 *The possibility distributions $\pi_{\overline{F}}, \pi_{\underline{F}}$ for the generalized p-box* *defined in Example 3.2 are:*

$$\pi_{\overline{F}}(x_1) = 0.3 \ \pi_{\overline{F}}(x_2) = 0.3 \ \pi_{\overline{F}}(x_3) = 0.7 \ \pi_{\overline{F}}(x_4) = 0.9 \ \pi_{\overline{F}}(x_5) = 0.9 \ \ \pi_{\overline{F}}(x_6) = 1$$

$$\pi_{\underline{F}}(x_1) = 1 \quad \pi_{\underline{F}}(x_2) = 1 \quad \pi_{\underline{F}}(x_3) = 1 \ \ \pi_{\underline{F}}(x_4) = 0.8 \ \pi_{\underline{F}}(x_5) = 0.8 \ \pi_{\underline{F}}(x_6) = 0.5$$

So, generalized p-boxes allow to model uncertainty in terms of pairs of comonotone possibility distributions. In this case, contrary to the case of only one possibility distribution, the two bounds describing uncertainty on a particular event A can be tighter, i.e. no longer restricted to the form $[0, \alpha]$ or $[\beta, 1]$, since the corresponding probability interval containing $P(A)$ will be contained in the intersection of intervals of this form.

An interesting case is the one where, for all $i = 1, \ldots, n-1$, $\underline{F}(x_i) = 0$ and $\underline{F}(x_n) = 1$. Then, $\pi_{\underline{F}} = 1$ and $\mathcal{P}_{\pi_{\overline{F}}} \cap \mathcal{P}_{\pi_{\underline{F}}} = \mathcal{P}_{\pi_{\overline{F}}}$ and we retrieve the single distribution $\pi_{\overline{F}}$. We recover $\pi_{\underline{F}}$ if we take for all $i = 1, \ldots, n$, $\overline{F}(x_i) = 1$. This means that any possibility distribution can be viewed a generalized cumulative distribution function F (it an be understood either as an upper or a lower function of a generalized p-box) associated to the specific ordering the possibility degrees induce on X.

3.3 Connecting Generalized p-boxes and random sets

The existing result relating p-boxes to random sets can be extended to generalized p-boxes.

Proposition 3.5 *For any generalized p-box $[\underline{F}, \overline{F}]$, there always exist a belief function Bel that encodes the same information as the generalized p-box. In particular, the credal set $\mathcal{P}_{[\underline{F},\overline{F}]}$ induced by the generalized p-box and the credal set $\mathcal{P}_{Bel}$ induce the same lower probabilites.*

In order to prove Proposition 3.5, we show that the lower probabilities on events induced by a generalized p-box are the same as the belief function given by Algorithm 1. To do that, we first build the partition of the space X, induced by sets A_i, and we formulate lower probabilities on events by means of elements of this partition. We then calculate lower bounds of these lower probabilities on all events, and show that these bounds are reached. We then check that the lower probabilities on all events coincide with the belief function induced by the algorithm. The detailed proof can be found in the appendix.

Algorithm 1 below provides an easy way to build the random set encoding a given generalized p-box. It is similar to algorithms given in [20, 30], 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 sets are obtained by thresholding the cumulative distributions (as in Figure 2). Since the sets A_i are nested, they induce a partition of X whose elements are of the form $G_i = A_i \setminus A_{i-1}$. The focal sets of the random set equivalent to the generalized p-box are made of unions of consecutive elements of this partition.

Basically, the procedure comes down to considering a threshold $\theta \in [0, 1]$. When $\alpha_{i+1} > \theta \geq \alpha_i$ and $\beta_{j+1} > \theta \geq \beta_j$, then, the corresponding focal set is $A_{i+1} \setminus A_j$, with mass

$$m(A_{i+1} \setminus A_j) = \min(\alpha_{i+1}, \beta_{j+1}) - \max(\alpha_i, \beta_j). \quad (10)$$

Example 3.6 illustrates the application of Algorithm 1.

Algorithm 1: R-P-box → random set transformation

Input: Generalized p-box $[\underline{F}, \overline{F}]$ and corresponding nested sets
$\emptyset = A_0, A_1, \ldots, A_n = X$, lower bounds α_i and upper bounds β_i

Output: Equivalent random set

```
for i = 1,...,n+1 do
  Build partition G_i = A_i \ A_{i-1}
Build set of values
{γ_l | l = 1,...,2n−1} = {α_i | i = 1,...,n} ∪ {β_i | i = 1,...,n}
With γ_l indexed such that γ_1 ≤ ... ≤ γ_l ≤ ... ≤ γ_{2n−1} = β_n = α_n = 1
Set α_0 = β_0 = γ_0 = 0
Set focal set E_0 = ∅
for k = 1,...,2n−1 do
  if γ_{k−1} = α_i then
    E_k = E_{k−1} ∪ G_{i+1}
  if γ_{k−1} = β_i then
    E_k = E_{k−1} \ G_i
  Set m(E_k) = γ_k − γ_{k−1}
```

Example 3.6 *Consider again the generalized p-box given in Example 3.2 and let us build the associated random set by applying Algorithm 1. We have:*

$$G_1 = \{x_1, x_2\} \qquad G_2 = \{x_3\} \qquad G_3 = \{x_4, x_5\} \qquad G_4 = \{x_6\}$$

and

$$\begin{array}{l} 0 \leq 0 \leq 0.2 \leq 0.3 \leq 0.5 \leq 0.7 \leq 0.9 \leq 1 \\ \alpha_0 \leq \alpha_1 \leq \alpha_2 \leq \beta_1 \leq \alpha_3 \leq \beta_2 \leq \beta_3 \leq \alpha_4 \\ \gamma_0 \leq \gamma_1 \leq \gamma_2 \leq \gamma_3 \leq \gamma_4 \leq \gamma_5 \leq \gamma_6 \leq \gamma_7 \end{array}$$

which finally yields the following random set

$$\begin{aligned} m(E_1) = m(G_1) = 0 \qquad & m(E_2) = m(G_1 \cup G_2) = 0.2 \\ m(E_3) = m(G_1 \cup G_2 \cup G_3) = 0.1 \qquad & m(E_4) = m(G_2 \cup G_3) = 0.2 \\ m(E_5) = m(G_2 \cup G_3 \cup G_4) = 0.2 \qquad & m(E_6) = m(G_3 \cup G_4) = 0.2 \\ & m(E_7) = m(G_4) = 0.1 \end{aligned}$$

This random set can then be used as an alternative representation of the provided information. This representation lays bare the high imprecision of the information This imprecision can only be alleviated by seeking more information.

Proposition 3.5 shows that generalized p-boxes can be seen as a particular case of general random sets. Generalized p-boxes are thus more expressive than single possibility distributions and less expressive than random sets, but, as emphasized in the introduction, less expressive (and, in this sense, simpler) models are often easier to handle in practice. As shown by the following remark, we can expect it to be the case for generalized p-boxes.

Remark 3.7 *Let $[\underline{F}, \overline{F}]$ be a generalized p-box over X, and, for $i = 1, \ldots, n$, G_i be the elements of the partition induced by nested subsets A_i. For an event A, let $A_* = \bigcup_{G_i \subseteq A} G_i$. We know that $\underline{P}(A) = \underline{P}(A_*)$. Then, there is an explicit expression of this lower probability in terms of bounds α_k, β_k. If $A_* = \bigcup_{k=i}^{j} G_k$ then, we have*

$$\underline{P}(A_*) = \max(0, \alpha_j - \beta_{i-1}). \tag{11}$$

And in the case were A_* is a union of "disjoint" subsets, each of this subset being a union of elements G_k whose indices k are consecutive, then $\underline{P}(A)$ remains simple to compute and just becomes a sum of lower probabilities of those subsets formed of unions of consecutive G_k included in A.

This simple remark shows the potential advantages of using generalized p-boxes rather than general random sets, since the computation of lower probabilities is faster than checking which focal elements E_i are included in a given event A.

So far, results in this section mainly exploit the fact that a collection of nested subsets on a space X induces a partition on this space, useful when computing lower probabilites of events. In the following we explain the links between this partition and the complete pre-ordering $\leq_{[\underline{F}, \overline{F}]}$ as well as the two possibility distributions $\pi_{\overline{F}}, \pi_{\underline{F}}$. First notice that Equation (11) can be restated in terms of the two possibility distributions $\pi_{\overline{F}}, \pi_{\underline{F}}$, rewriting $\underline{P}(A_*)$ as

$$\underline{P}(A_*) = \max(0, N_{\pi_{\underline{F}}}(\bigcup_{k=1}^{j} F_k) - \Pi_{\pi_{\overline{F}}}(\bigcup_{k=1}^{i-1} F_k)),$$

where $N_{\pi_i}(A), \Pi_{\pi_i}(A)$ are respectively the necessity and possibility degree of event A (given by Equations (5)) with respect to π_i. It makes $\underline{P}(A_*)$ even easier to compute.

We can also directly derive the random set equivalent to a given generalized p-box $[\underline{F}, \overline{F}]$: let us note $0 = \gamma_0 < \gamma_1 < \ldots < \gamma_M = 1$ the distinct values taken by $\underline{F}, \overline{F}$ over elements x_i of X (note that M is finite and $M < 2n$). Then, for $j = 1, \ldots, m$, the random set defined as:

$$\begin{cases} E_j = \{x_i \in X | (\overline{F}(x_i) \geq \gamma_j) \wedge (\max\{\underline{F}(x_k) | j=1,\ldots,i \ \underline{F}(x_k) < \underline{F}(x_i)\} < \gamma_j)\} \\ m(E_j) = \gamma_j - \gamma_{j-1} \end{cases} \tag{12}$$

is the same as the one built by using Algorithm 1, but this formulation lays bare the link between Equation (8) and the possibility distributions $\pi_{\overline{F}}, \pi_{\underline{F}}$.

3.4 Probability intervals and generalized p-boxes

As in the case of random sets, there is no direct relationship between probability intervals and generalized p-boxes. The two representations have comparable complexities, but do not involve the same kind of events. Nevertheless, given previous results, we can state how a set L of probability intervals can be transformed into a generalized p-box $[\underline{F}, \overline{F}]$, and vice-versa.

Let us first consider a set L of probability intervals on a space X and some indexing of elements in X. For all $i = 1, \ldots, n$, let $l(x_i) = l_i$ and $u(x_i) = u_i$. A generalized p-box $[\underline{F}', \overline{F}']$ covering the set L of probability intervals can be computed by means of Equations (2) of Section 2.2.2 in the following way:

$$\begin{aligned} \underline{F}'(x_i) &= \underline{P}(A_i) = \alpha'_i = \max(\sum_{x_i \in A_i} l_i, 1 - \sum_{x_i \notin A_i} u_i) \\ \overline{F}'(x_i) &= \overline{P}(A_i) = \beta'_i = \min(\sum_{x_i \in A_i} u_i, 1 - \sum_{x_i \notin A_i} l_i) \end{aligned} \tag{13}$$

where $\underline{P}, \overline{P}$ are respectively the lower and upper probabilities of $\mathcal{P}_L$ for events A_i, given by Equations (2). Each permutation of elements of X would provide a different generalized p-box. There is no tightest covering among them.

Now we consider a generalized p-box $[\underline{F}, \overline{F}]$ with nested sets $A_1 \subseteq \ldots \subseteq A_n$.

The set L' of probability intervals on elements x_i corresponding to $[\underline{F}, \overline{F}]$ is given by:

$$\begin{aligned} \underline{P}(F_i) = \underline{P}(x_i) = l'_i = \max(0, \alpha_i - \beta_{i-1}) \\ \overline{P}(F_i) = \overline{P}(x_i) = u'_i = \beta_i - \alpha_{i-1} \end{aligned} \tag{14}$$

where $\underline{P}, \overline{P}$ are the lower and upper probabilities of $\mathcal{P}_{[\underline{F},\overline{F}]}$, given by Equation (11), and with $\beta_0 = \alpha_0 = 0$. This is the tightest set of probability intervals induced by the generalized p-box.

Of course, transforming a set L of probability intervals into a p-box $[\underline{F}, \overline{F}]$ and vice-versa generally induces a loss of information, as shown by the two following propositions:

Proposition 3.8 *Given an initial set L of probability intervals over a space X, and given the two consecutive transformations*

$$\textit{Prob. Intervals } L \rightarrow_{(13)} \textit{ p-box } [\underline{F}', \overline{F}'] \rightarrow_{(14)} \textit{ Prob. Intervals } L''$$

we have $\mathcal{P}_L \subseteq \mathcal{P}_{L''}$*, and the differences between bounds of intervals in the sets* L'' *and* L *are given, for* $i = 1, \ldots, n$*, by*

$$l_i - l''_i = \min(l_i, 0 + \sum_{x_i \in A_{i-1}} (u_i - l_i), 0 + \sum_{x_i \in A_i^c} (u_i - l_i), (l_i + \sum_{\substack{x_j \neq x_i \\ x_j \in X}} u_j) - 1, 1 - \sum_{x_i \in X} l_i) \tag{15}$$

$$u''_i - u_i = \min(0 + \sum_{x_i \in A_{i-1}} (u_i - l_i), 0 + \sum_{x_i \in A_i^c} (u_i - l_i), 1 - (u_i + \sum_{\substack{x_j \neq x_i \\ x_j \in X}} l_j), \sum_{x_i \in X} u_i - 1)$$

with $A_0 = \emptyset$*. Under the assumptions that set L is non-empty and reachable, these differences are positive.*

Proposition 3.9 *Given an initial generalized p-box* $[\underline{F}, \overline{F}]$ *over a space X, and given the two consecutive transformations*

$$\textit{p-box } [\underline{F}, \overline{F}] \rightarrow_{(14)} \textit{ Prob. Intervals } L' \rightarrow_{(13)} \textit{ p-box } [\underline{F}'', \overline{F}'']$$

we have that $\mathcal{P}_{[\underline{F},\overline{F}]} \subseteq \mathcal{P}_{[\underline{F}'',\overline{F}'']}$*, and the differences between values of* $[\underline{F}, \overline{F}]$ *and* $[\underline{F}'', \overline{F}'']$ *are, for* $i = 1, \ldots, n$

$$\begin{aligned} \underline{F}(x_i) - \underline{F}''(x_i) = \min(\sum_{j=1}^{i-1}(\alpha_j - \beta_j), \sum_{j=i+1}^{n-1}(\alpha_j - \beta_j)) \\ \overline{F}''(x_i) - \overline{F}(x_i) = \min(\sum_{j=1}^{i-1}(\alpha_j - \beta_j), \sum_{j=i+1}^{n-1}(\alpha_j - \beta_j)) \end{aligned} \tag{16}$$

The proofs can be found in the appendix. Example 3.10 illustrates both the transformation procedure and the fact that this procedure implies an information loss.

Example 3.10 *Let us take the same four probability intervals as in the example given by Masson and Denoeux [22], on the space* $X = \{w, x, y, z\}$*, and summarized in the following table*

	w	x	y	z
l	0.10	0.34	0.25	0
u	0.28	0.56	0.46	0.08

if we consider the order R *such that* $w <_R x <_R y <_R z$*. After application of Equations (13), we have the following generalized p-box*

	$\underline{F}'$	$\overline{F}'$
$A_1 = \{w\}$	0.10	0.28
$A_2 = \{w, x\}$	0.46	0.75
$A_3 = \{w, x, y\}$	0.92	1
$A_4 = X$	1	1

and if generate probability intervals from this generalized p-box by applying Equations (14), we obtain the set L''

	w	x	y	z
l''	0.10	0.18	0.17	0
u''	0.28	0.65	0.54	0.08

a result which is less informative than the first probability intervals.

As no natural order exists on X, as many as $n!$ generalized p-boxes can be generated from a set of probability intervals. Let Σ_σ the set of all possible permutations σ of elements of X, each defining a linear order. A generalized p-box according to permutation σ is denoted $[\underline{F}', \overline{F}']_\sigma$ and called a σ-p-box. We then have the following proposition:

Proposition 3.11 *Let* L *be a set of probability intervals, and let* $[\underline{F}', \overline{F}']_\sigma$ *be the* σ*-p-box obtained from* L *by applying Equations (13). Moreover, let* L''_σ *denote the set of probability intervals obtained from the* σ*-p-box* $[\underline{F}', \overline{F}']_\sigma$ *by applying Equations (14). Then, the various credal sets thus defined satisfy the*

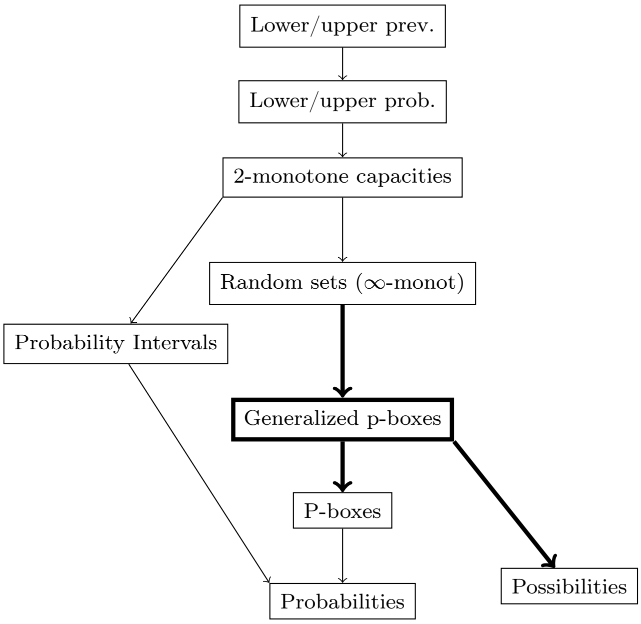

Figure 3. Representation relationships: summary with generalized p-boxes $A \longrightarrow B$: B is a special case of A

following property:

$$\mathcal{P}_L = \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{[\underline{F}', \overline{F}']_\sigma} = \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{L''_\sigma} \tag{17}$$

Concretely, this means that, given initial information modeled by a set L of probability intervals, this information can be entirely recovered by considering the set of all σ-p-boxes, varying all permutations. Since there are $|X|!$ such permutations, representing a set of probability intervals L by a set of generalized p-boxes does not look very interesting at first glance. In practice, L can be exactly recovered if a reduced set $\mathcal{S}$ of $|X|/2$ permutations is used to generate the generalized p-boxes, provided that $\{x_{\sigma(1)}, \sigma \in \mathcal{S}\} \cup \{x_{\sigma(n)}, \sigma \in \mathcal{S}\} = X$. Since $\mathcal{P}_{[\underline{F}, \overline{F}]} = \mathcal{P}_{\pi_{\underline{F}}} \cap \mathcal{P}_{\pi_{\overline{F}}}$, then it is immediate from Proposition 3.11,that , in terms of credal sets, $\mathcal{P}_L = \bigcap_{\sigma \in \Sigma_\sigma} \left(\mathcal{P}_{\pi_{\underline{F}_\sigma}} \cap \mathcal{P}_{\pi_{\overline{F}_\sigma}} \right)$, where $\pi_{\underline{F}_\sigma}, \pi_{\overline{F}_\sigma}$ are respectively the possibility distributions corresponding to $\underline{F}_\sigma$ and $\overline{F}_\sigma$.

Figure 3 summarizes the results of this paper, by placing generalized p-boxes inside the graph of Figure 1. New relationships and representations obtained in this paper are in bold lines.

4 Conclusions

This paper has proposed a generalized notion of p-box, that remains a special kind of random set but subsumes possibility distributions. In particular, we have shown that a generalized p-box can be represented by means of a pair of comonotone possibility distributions, and the equivalent random set has been laid bare. Generalized p-boxes are thus more expressive than single possibility distributions and likely to be more tractable than general random sets. Moreover, the fact that they can be interpreted as lower and upper confidence bounds over nested sets makes them quite attractive tools for subjective elicitation. Finally, we showed the gap existing between generalized p-boxes and sets of probability intervals.

Computational aspects of calculations with generalized p-boxes need to be explored in greater detail (as is done by De Campos *et al.* [2] for probability intervals) as well as their psychological relevance (as done by Raufaste *et al.* [29] for possibility distributions). Another issue is to extend presented results to more general spaces, to general lower/upper previsions or to cases not considered here (e.g. continuous p-boxes with discontinuity points), possibly using existing results [7, 34].

Interestingly, the key condition when representing generalized p-boxes by two possibility distributions is their comonotonicity. In a companion paper, we pursue the present study by dropping this assumption. We get very close to the so-called clouds, recently proposed by Neumaier [26].

References

[1] C. Baudrit, D. Dubois, Practical representations of incomplete probabilistic knowledge, Computational Statistics and Data Analysis 51 (1) (2006) 86–108.

[2] L. Campos, J. Huete, S. Moral, Probability intervals: a tool for uncertain reasoning, I. J. of Uncertainty, Fuzziness and Knowledge-Based Systems 2 (1994) 167–196.

[3] A. Chateauneuf, J.-Y. Jaffray, Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, Mathematical Social Sciences 17 (3) (1989) 263–283.

[4] G. Choquet, Theory of capacities, Annales de l'institut Fourier 5 (1954) 131–295.

[5] R. Cooke, Experts in uncertainty, Oxford University Press, Oxford, UK, 1991.

[6] G. de Cooman, A behavioural model for vague probability assessments, Fuzzy Sets and Systems 154 (2005) 305–358.

[7] G. de Cooman, M. Troffaes, E. Miranda, n-monotone lower previsions and lower integrals, in: F. Cozman, R. Nau, T. Seidenfeld (eds.), Proc. 4th International Symposium on Imprecise Probabilities and Their Applications, 2005.

[8] A. Dempster, Upper and lower probabilities induced by a multivalued mapping, Annals of Mathematical Statistics 38 (1967) 325–339.

[9] T. Denoeux, Constructing belief functions from sample data using multinomial confidence regions, I. J. of Approximate Reasoning 42 (2006) 228–252.

[10] D. Dubois, L. Foulloy, G. Mauris, H. Prade, Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities, Reliable Computing 10 (2004) 273–297.

[11] D. Dubois, P. Hajek, H. Prade, Knowledge-driven versus data-driven logics, J. of Logic, Language and Information 9 (2000) 65–89.

[12] D. Dubois, H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, New York, 1988.

[13] D. Dubois, H. Prade, Consonant approximations of belief functions, I.J. of Approximate reasoning 4 (1990) 419–449.

[14] D. Dubois, H. Prade, Random sets and fuzzy interval analysis, Fuzzy Sets and Systems (42) (1992) 87–101.

[15] S. Ferson, L. Ginzburg, V. Kreinovich, D. Myers, K. Sentz, Constructing probability boxes and dempster-shafer structures, Tech. rep., Sandia National Laboratories (2003).

[16] M. Grabisch, J. Marichal, M. Roubens, Equivalent representations of set functions, Mathematics on operations research 25 (2) (2000) 157–178.

[17] J. Helton, W. Oberkampf (eds.), Alternative Representations of Uncertainty, Special issue of Reliability Engineering and Systems Safety, vol. 85, Elsevier, 2004.

[18] P. Huber, Robust statistics, Wiley, New York, 1981.

[19] I. Kozine, L. Utkin, Constructing imprecise probability distributions, I. J. of General Systems 34 (2005) 401–408.

[20] E. Kriegler, H. Held, Utilizing belief functions for the estimation of future climate change, I. J. of Approximate Reasoning 39 (2005) 185–209.

[21] J. Lemmer, H. Kyburg, Conditions for the existence of belief functions corresponding to intervals of belief, in: Proc. 9th National Conference on A.I., Anaheim, 1991.

[22] M. Masson, T. Denoeux, Inferring a possibility distribution from empirical data, Fuzzy Sets and Systems 157 (3) (2006) 319–340.

[23] E. Miranda, I. Couso, P. Gil, Extreme points of credal sets generated by 2-alternating capacities, I. J. of Approximate Reasoning 33 (2003) 95–115.

[24] I. Molchanov, Theory of Random Sets, Springer, London, 2005.

[25] R. Moore, Methods and applications of Interval Analysis, SIAM Studies in Applied Mathematics, SIAM, Philadelphia, 1979.

[26] A. Neumaier, Clouds, fuzzy sets and probability intervals, Reliable Computing 10 (2004) 249–272.

[27] A. Neumaier, On the structure of clouds, Available on http://www.mat.univie.ac.at/∼neum (2004).

[28] Z. Pawlak, Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic, Dordrecht, 1991.

[29] E. Raufaste, R. Neves, C. Mariné, Testing the descriptive validity of possibility theory in human judgments of uncertainty, Artificial Intelligence 148 (2003) 197–218.

[30] H. Regan, S. Ferson, D. Berleant, Equivalence of methods for uncertainty propagation of real-valued random variables, I. J. of Approximate Reasoning 36 (2004) 1–30.

[31] G. Shackle, Decision, Order and Time in Human Affairs, Cambridge University Press, UK, 1961.

[32] G. Shafer, A mathematical Theory of Evidence, Princeton University Press, New Jersey, 1976.

[33] P. Smets, The normative representation of quantified beliefs by belief functions, Artificial Intelligence 92 (1997) 229–242.

[34] P. Smets, Belief functions on real numbers, I. J. of Approximate Reasoning 40 (2005) 181–223.

[35] P. Walley, Statistical reasoning with imprecise Probabilities, Chapman and Hall, New York, 1991.

[36] L. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28.

Appendix

Proof of Proposition 3.5 From the nested sets $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n = X$ we can build a partition s.t. $G_1 = A_1, G_2 = A_2 \setminus A_1, \ldots, G_n = A_n \setminus A_{n-1}$. Once we have a finite partition, every possible set $B \subseteq X$ can be approximated from above and from below by pairs of sets $B_* \subseteq B^*$ [28]:

$$B^* = \bigcup\{G_i, G_i \cap B \neq \emptyset\}$$
$$B_* = \bigcup\{G_i, G_i \subseteq B\}$$

made of a finite union of the partition elements intersecting or contained in this set B. Then $\underline{P}(B) = \underline{P}(B_*)$,$\overline{P}(B) = \overline{P}(B^*)$, so we only have to care about unions of elements G_i in the sequel. Especially, for each event $B \subset G_i$ for some i, it is clear that $\underline{P}(B) = 0 = Bel(B)$ and $\overline{P}(B) = \overline{P}(G_i) = Pl(B)$.

Let us first consider union of consecutive elements $\bigcup_{k=i}^{j} G_k$ (when $k = 1$, we retrieve the sets A_j). Finding $\underline{P}(\bigcup_{k=i}^{j} G_k)$ is equivalent to computing the minimum of $\sum_{k=i}^{j} P(G_k)$ under the constraints

$$i = 1, \dots, n \qquad \alpha_i \leq P(A_i) = \sum_{k=1}^{i} P(G_k) \leq \beta_i$$

which reads

$$\alpha_j \leq P(A_{i-1}) + \sum_{k=i}^{j} P(G_k) \leq \beta_j$$

so $\sum_{k=i}^{j} P(G_k) \geq \max(0, \alpha_j - \beta_{i-1})$. This lower bound is optimal, since it is always reachable: if $\alpha_j > \beta_{i-1}$, take P s.t. $P(A_{i-1}) = \beta_{i-1}$, $P(\bigcup_{k=i}^{j} G_k) = \alpha_j - \beta_{i-1}$, $P(\bigcup_{k=j+1}^{n} G_k) = 1 - \alpha_j$. If $\alpha_j \leq \beta_{i-1}$, take P s.t. $P(A_{i-1}) = \beta_{i-1}$, $P(\bigcup_{k=i}^{j} G_k) = 0$, $P(\bigcup_{k=j+1}^{n} E_k) = 1 - \beta_{i-1}$.

And we can see, by looking at Algorithm 1, that $Bel(\bigcup_{k=i}^{j} G_k) = \max(0, \alpha_j - \beta_{i-1})$: focal elements of Bel are subsets of $\bigcup_{k=i}^{j} G_k$ if $\beta_{i-1} < \alpha_j$ only.

Now, let us consider a union A of non-consecutive elements s.t. $A = (\bigcup_{k=i}^{i+l} G_k \ \cup \ \bigcup_{k=i+l+m}^{j} G_k)$ with $m > 1$. As in the previous case, we must compute $\min\Big(\sum_{k=i}^{i+l} P(G_k) + \sum_{k=i+l+m}^{j} P(G_k)\Big)$ to find the lower probability on $\underline{P}(A)$. An obvious lower bound is given by

$$\min\Big(\sum_{k=i}^{i+l} P(G_k) + \sum_{k=i+l+m}^{j} P(G_k)\Big) \geq \min\Big(\sum_{k=i}^{i+l} P(G_k)\Big) + \min\Big(\sum_{k=i+l+m}^{j} P(G_k)\Big)$$

and this lower bound is equal to

$$\max(0, \alpha_{i+l} - \beta_{i-1}) + \max(0, \alpha_j - \beta_{i+l+m-1}) = Bel(A)$$

Consider the two following cases and the probability assignments showing that bounds are attained:

- $\alpha_{i+l} < \beta_{i-1}$, $\alpha_j < \beta_{i+l+m-1}$ and probability masses $P(A_{i-1} = \beta_{i-1})$, $P(\bigcup_{k=i}^{i+l} G_k) = \alpha_{i+l} - \beta_{i-1}$, $P(\bigcup_{k=i+l+1}^{i+l+m-1} G_k) = \beta_{i+l+m-1} - \alpha_{i+l}$, $P(\bigcup_{k=i+l+m}^{j} G_k) = \alpha_j - \beta_{i+l+m-1}$ and $P(\bigcup_{k=j+1}^{n} G_k) = 1 - \alpha_j$
- $\alpha_{i+l} > \beta_{i-1}$, $\alpha_j > \beta_{i+l+m-1}$ and probability masses $P(A_{i-1} = \beta_{i-1}$, $P(\bigcup_{k=i}^{i+l} G_k) = 0$, $P(\bigcup_{k=i+l+1}^{i+l+m-1} G_k) = \alpha_j - \beta i - 1$, $P(\bigcup_{k=i+l+m}^{j} E_k) = 0$ and $P(\bigcup_{k=j+1}^{n} G_k) = 1 - \alpha_j$

A same line of thought can be followed for the two remaining cases. As in the consecutive case, the lower bound is reachable without violating any of the restrictions associated to the generalized p-box. We have $\underline{P}(A) = Bel(A)$ and this result can be easily extended to any number n of "discontinuities" in the sequence of G_k.

The proof is complete, since for every possible union A of elements G_k, we have $\underline{P}(A) = Bel(A)$ □

Proof of Proposition 3.8 Let X be a finite set and define a ranking of their elements $x_i < x_j$ if and only if $i < j$. Given this ranking, and to prove Proposition 3.8, we start from a set L with, for $i = 1, \dots, n$, initial bounds u_i, l_i. We then apply successively Equations (13) and (14), with the aim of expressing bounds u''_i, l''_i of the set L'' in terms of initial bounds u_i, l_i. Computation of $(l_i - l''_i)$ and $(u''_i - u_i)$ then follows. The positiveness of these two differences is sufficient to prove inclusion between credal sets $\mathcal{P}_L$ and $\mathcal{P}_{L''}$ To shorten the proof, we focus on lower bounds (proof for upper bounds is similar).

Let us then consider the p-box $[\underline{F}', \overline{F}']$ built from a given reachable non-empty set L of probability intervals, given, for $i = 1, \dots, n$, by equations

$$\underline{P}(A_i) = \alpha'_i = \max(\sum_{x_i \in A_i} l_i, 1 - \sum_{x_i \notin A_i} u_i)$$
$$\overline{P}(A_i) = \beta'_i = \min(\sum_{x_i \in A_i} u_i, 1 - \sum_{x_i \notin A_i} l_i)$$

with $\underline{P}, \overline{P}$ the lower and upper probabilities of $\mathcal{P}_L$. Now, given these bounds, we can compute the set L'' of probability intervals s.t.

$$l''_i = \underline{P}'(x_i) = \max(0, \alpha'_i - \beta'_{i-1})$$

with $\underline{P}'$ the lower probability of $\mathcal{P}_{[\underline{F}', \overline{F}']}$. When expressed in term of values l_i, u_i of the original set L, l''_i is given by

$$l''_i = \max(0, \sum_{x_i \in A_i} l_i - \sum_{x_i \in A_{i-1}} u_i, \sum_{x_i \in A_i} l_i + \sum_{x_i \in A^c_{i-1}} l_i - 1,$$
$$1 - \sum_{x_i \in A^c_i} u_i - \sum_{x_i \in A_{i-1}} u_i, \sum_{x_i \in A^c_{i-1}} l_i - \sum_{x_i \in A^c_i} u_i)$$

and, given that the set L is reachable and non-empty, we have that $l''_i \leq l_i$. Equation (15) giving $(l_i - l''_i)$ then follows.

The same procedure can be followed for the bounds u''_i, and we have $\mathcal{P}_L \subseteq \mathcal{P}_{L''}$. The set L'' is non-empty (since $\mathcal{P}_L \subseteq \mathcal{P}_{L''}$) and reachable (by construction, the new bounds $[l''_i, u''_i]$ are reached by one distribution in the p-box $[\underline{F}', \overline{F}']$, and this distribution is also in $\mathcal{P}_{L''}$, thus set L'' is reachable) □

Proof of Proposition 3.9 Proof of proposition 3.9 follows the same line of thought as the proof of Proposition 3.8.

Let us consider an original generalized p-box $[\underline{F}, \overline{F}]$ with bounds α_i, β_i on sets A_i. The set L' of probability intervals corresponding to this generalized p-box is given by equations

$$\underline{P}(x_i) = l'_i = \max(0, \alpha_i - \beta_{i-1})$$
$$\overline{P}(x_i) = u'_i = \beta_i - \alpha_{i-1},$$

where $\underline{P}, \overline{P}$ are the lower and upper probabilities of $\mathcal{P}_{[\underline{F},\overline{F}]}$. From the set L', we can get the lower bound $\underline{F}''$ of $[\underline{F}'', \overline{F}'']$ by using equations

$$\underline{P}'(A_i) = \alpha''_i = \max(\sum_{x_i \in A_i} l'_i, 1 - \sum_{x_i \notin A_i} u'_i)$$

with $\underline{P}'$ the lower probability of $\mathcal{P}_{L'}$. In terms of the original p-box bounds α_i, β_i, this gives us

$$\alpha''_i = \max(\sum_{j=1}^{i} \alpha_j - \sum_{j=1}^{i-1} \beta_j, 1 + \sum_{j=i}^{n-1} \alpha_j - \sum_{j=i+1}^{n} \beta_j)$$
$$\alpha''_i = \max(\sum_{j=1}^{i} \alpha_j - \sum_{j=1}^{i-1} \beta_j, \alpha_i + \sum_{j=i+1}^{n-1} \alpha_j - \sum_{j=i+1}^{n-1} \beta_j)$$

Given that $\forall j; \alpha_j \leq \beta_j$ by definition of a generalized p-box, we have $\alpha''_i \leq \alpha_i$ and Equation (16) follows. The same procedure can again be done for the upper bound to check that $\beta''_i \geq \beta_i$, and we get $\mathcal{P}_{[\underline{F},\overline{F}]} \subseteq \mathcal{P}_{[\underline{F}'',\overline{F}'']}$. □

Proof of Proposition 3.11 To prove this proposition, we must first recall a result given by De Campos *et al.* [2]: given two sets of probability intervals L and L' defined on a space X and the induced credal sets $\mathcal{P}_L$ and $\mathcal{P}_{L'}$, the conjunction $\mathcal{P}_{L \cap L'} = \mathcal{P}_L \cap \mathcal{P}_{L'}$ of these two sets can be modeled by the set $(L \cap L')$ of probability intervals that is such that for every element x of X,

$$l_{(L \cap L')}(x) = \max(l_L(x), l_{L'}(x)) \text{ and } u_{(L \cap L')}(x) = \min(u_L(x), u_{L'}(x))$$

and these formulas extend directly to the conjunction of any number of set of probability intervals on X.

To prove Proposition 3.11, we will show, by using the above conjunction, that $\mathcal{P}_L = \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{L''_\sigma}$. Since, by Proposition 3.8 and for any $\sigma \in \Sigma_\sigma$, $\mathcal{P}_L \subset \mathcal{P}_{[\underline{F}',\overline{F}']_\sigma} \subset \mathcal{P}_{L''_\sigma}$, showing this equality is sufficient to prove the whole proposition.

Let us note that the above inclusion relationships alone ensure us that $\mathcal{P}_L \subseteq \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{[\underline{F}',\overline{F}']_\sigma} \subseteq \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{L''_\sigma}$. So, all we have to show is that the inclusion relationship is in fact an equality.

Since we know that both $\mathcal{P}_L$ and $\bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{L''_\sigma}$ can be modeled by set of probability intervals, we will show that the lower bounds l on every element x in these two sets coincide (and the proof for upper bounds is similar).

For all x in X, $l_{L''_\Sigma}(x) = \max_{\sigma \in \Sigma_\sigma}\{l_{L''_\sigma}(x)\}$, with L''_Σ the set of probability intervals corresponding to $\bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{L''_\sigma}$ and L''_σ the set of probability intervals corresponding to a particular permutation σ. We must now show that, for all x in X, $l_{L''_\Sigma}(x) = l_L(x)$.

From Proposition 3.11, we already know that, for any permutation σ and for all x in X, we have $l_{L''_\sigma}(x) \leq l_L(x)$. So we must now show that, for a given x in X, there is one permutation σ such that $l_{L''_\sigma}(x) = l_L(x)$. Let us consider the permutation placing the given element at the front. If x is the first element $x_{\sigma(1)}$, then Equation (15) has value 0 for this element, and we thus have $l_{L''_\sigma}(x) = l_L(x)$. Since if we consider every possible ranking, every element x of X will be first in at least one of these rankings, this completes the proof. $\square$