



HAL
open science

Handling bipolar knowledge with imprecise probabilities

Sébastien Destercke

► **To cite this version:**

Sébastien Destercke. Handling bipolar knowledge with imprecise probabilities. International Journal of Intelligent Systems, 2012, 26 (5), pp.426-443. 10.1002/int.20475 . hal-00652236

HAL Id: hal-00652236

<https://hal.science/hal-00652236>

Submitted on 15 Dec 2011

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

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

Handling bipolar knowledge with imprecise probabilities

Sébastien Destercke

December 15, 2011

Abstract

Information is said to be bipolar when it has a positive and a negative part. The problem of representing and processing such bipolar information has recently received a lot of attention in uncertainty theories. In this paper, we are concerned with the representation of asymmetric bipolarity, i.e., with situations where positive and negative information are unrelated and processed in parallel. In this latter case, positive information consists in observations of experiment results showing what values are possible, while negative information consists in constraints (e.g., provided by an expert) restricting the range of possible variable values. Up to now, there are no proposition as to how such bipolar information can be treated in the framework of imprecise probability theory, i.e., when information is represented by convex sets of probabilities. In this paper, we propose the basis of such a framework, and provide some illustrative examples.

1 Introduction

Information about a given variable usually comes in different forms and from various sources. Recently, there has been a growing interest in the handling of bipolar information. Information is bipolar when one can differentiate between a positive and a negative part in the information. Such information usually concerns either evidences about the true value assumed by an ill-known variable, or preferences expressed by one or more agent. In this paper, we are concerned with the first type of information. One can consider at least three different types of bipolarity (See Dubois and Prade [24] for more details). The first one, called symmetric univariate, models bipolarity by the use of an univariate scale and can be represented by the means of classical probability measures. The second one, called a symmetric bivariate scale, handle two separate unipolar scales (positive and negative) that refer to the same information and are usually linked by some duality relation. Lower and upper previsions [39, 33] are examples of this kind of bipolarity, as well as other uncertainty models encompassed by this representation (Belief and plausibility functions [35], possibility and necessity measures [20]).

The last type of bipolarity, coined as asymmetric or heterogeneous, is the one addressed in this paper. Such bipolarity arise when negative and positive information parts are two unrelated kinds of information that have to be processed in parallel: one asserting what is impossible (negative information), the other what can exist (positive information). Negative information can correspond to constraints over possible values expressed by physical laws, expert opinions, etc. Examples, observation and measurements are instances of positive information. Note that the two kinds of information are effectively unrelated, as an expert opinion (or a model prediction) declaring some values as impossible does not imply that all the others can or will be observed, hence the need for asymmetry. This need is further confirmed by some psychological studies [9] supporting the fact that the brain processes differently positive and negative information.

The notion of bipolarity have been declined in a number of application areas and theoretical frameworks: multicriteria decision making [29], conflict resolution in argumentative frameworks [2], uncertainty and preferences representation [24], spatial reasoning [8], database querying [23], The processing of bipolar information within uncertainty theories has already been discussed in the frameworks of possibility theory [24] and of the transferable belief model [36]. However, to the best of our knowledge, the processing of bipolar information when information is modelled by convex sets of probabilities, or so-called *credal sets* [31], has not been considered so far.

The purpose of this paper is to lay out a basic framework to represent and process bipolar information when information is modelled by such credal sets. The idea behind this framework is simple and can be summarised in two main steps:

1. the first step consists in collecting negative and positive information separately, and to represent each corpus of information by separate credal sets;
2. the second step consists in merging positive and negative information in a single representation, possibly coping with conflicting or new information in the process.

After recalling the basics of credal sets [31] and their relation with Walley's [39] lower previsions, Section 2 presents our proposition, i.e. how bipolar information can be modeled and processed with credal sets. Section 3 then provides some illustrative examples using some practical imprecise probabilistic representations (i.e., p-boxes, probability intervals and possibility distributions).

2 Handling bipolar information with credal sets

This section first recalls the basics of credal sets [31]. Our proposition as to how bipolar information should be processed with such representations is then detailed in other subsections.

Credal sets [31], or convex sets of probabilities, are very general models of uncertainty that encompass most known uncertainty models. They therefore

provide an attractive and unifying framework to model and reason under uncertainty. As uncertainty representations, they are equivalent to Walley's [39] coherent lower previsions, that extends de Finetti's previsions [26] by integrating imprecision to them.

2.1 Credal sets and lower previsions

In this paper, we consider a variable X assuming its values on a space \mathcal{X} made of mutually exclusive elements and whose exact value is ill-known or uncertain. We also assume that this uncertainty is modelled by the means of a credal set \mathcal{P} , or convex set of probability distributions over \mathcal{X} . We denote by $\mathcal{L}(\mathcal{X})$ the set of real-valued bounded functions on \mathcal{X} . Given a function $f \in \mathcal{L}(\mathcal{X})$, one can compute the lower and upper expectations $\underline{\mathbb{E}}_{\mathcal{P}}(f), \bar{\mathbb{E}}_{\mathcal{P}}(f)$ induced by \mathcal{P} such that

$$\underline{\mathbb{E}}_{\mathcal{P}}(f) = \inf_{p \in \mathcal{P}} \mathbb{E}_p(f) \quad \bar{\mathbb{E}}_{\mathcal{P}}(f) = \sup_{p \in \mathcal{P}} \mathbb{E}_p(f),$$

where p is a probability distribution over \mathcal{X} and $\mathbb{E}_p(f)$ the expected value of f w.r.t. p . These two values are dual, in the sense that $\underline{\mathbb{E}}_{\mathcal{P}}(f) = -\bar{\mathbb{E}}_{\mathcal{P}}(-f)$. Thanks to this duality, one can only work with one of the two mappings (usually $\underline{\mathbb{E}}$).

Alternatively, one can start from a lower mapping $\underline{P} : \mathcal{K} \rightarrow \mathbb{R}$ from a subset $\mathcal{K} \subseteq \mathcal{L}(\mathcal{X})$, and consider the induced credal set $\mathcal{P}(\underline{P})$ such that

$$\mathcal{P}_{\underline{P}} = \{p \in \mathbb{P}_{\mathcal{X}} \mid (\forall f \in \mathcal{K})(\mathbb{E}_p(f) \geq \underline{P}(f))\}.$$

with $\mathbb{P}_{\mathcal{X}}$ the set of all probability mass functions over $\mathbb{P}_{\mathcal{X}}$. In his theory of lower previsions [39], Walley starts from the mapping \underline{P} that he calls lower prevision. He interprets $\underline{P}(f)$ as the supremum buying price for the uncertain reward f . A lower prevision \underline{P} is then said to avoid sure loss iff $\mathcal{P}_{\underline{P}} \neq \emptyset$, and to be coherent if the lower expectation $\underline{\mathbb{E}}_{\mathcal{P}_{\underline{P}}}(f) = \underline{P}(f)$ coincides with \underline{P} for every $f \in \mathcal{K}$ (i.e., \underline{P} is the lower envelope of $\mathcal{P}_{\underline{P}}$). He also shows that coherent lower previsions and credal sets have the same expressive power (in the sense that any credal can be identified by a unique lower prevision, and vice versa). Given a credal set \mathcal{P} , its lower (resp. upper) probability of an event A , denoted by $\underline{P}_{\mathcal{P}}(A)$ (resp. $\bar{P}_{\mathcal{P}}(A)$), corresponds to the lower (resp. upper) expectation of the indicator function $\mathbf{1}_{(A)}$ of the event A , that takes value one on A and zero elsewhere. By duality, we have $\underline{P}_{\mathcal{P}}(A) = 1 - \bar{P}_{\mathcal{P}}(A^c)$.

Note that credal sets and coherent lower previsions are very general models, in the sense that they encompass most of the other uncertainty models proposed in the literature [40]. In particular both necessity measures of possibility theory [20] and belief measures of evidence theory [35] correspond to particular classes of lower probabilities inducing specific credal sets. We now detail the two main steps of bipolar information handling: information representation and merging.

2.2 Collecting and representing bipolar information

Similarly to what is done in possibility theory [24] and other frameworks [29, 36], we propose to model positive and negative information by using two separate models of our chosen framework. That is, positive information is modelled by a credal set \mathcal{P}^+ , while negative information is modelled by another credal set \mathcal{P}^- .

Negative information (\mathcal{P}^-): Negative information expresses constraints about the value X can assume. It rules out some values, considering them as impossible, or less likely than others. It can come from expert opinion about some particular events, or from the answer of some model for which X is the output value. The *negative credal set* \mathcal{P}^- corresponding to such information will typically be induced by a collection of expectation bounds over a set of chosen functions¹ $f_1, \dots, f_k \in \mathcal{L}(\mathcal{X})$, in the form

$$\underline{P}(f_i) \leq \sum_{x \in \mathcal{X}} f(x)p(x) \leq \overline{P}(f_i). \quad (1)$$

In many situations, functions f_1, \dots, f_k will be some indicator functions of events A_1, \dots, A_k , and the negative information will consists in lower and upper bounds over such events. Note that pieces of negative information are treated conjunctively, in the sense that we consider the credal set induced by all constraints (1) at once. This means that the more we accumulate negative information, the more precise is \mathcal{P}^- . We assume here that $\mathcal{P}^- \neq \emptyset$ (i.e., the lower prevision \underline{P} given by Eq. (1) avoids sure loss).

Positive information (\mathcal{P}^+): Positive information consists most of the time in a set of observations or of experiment results. Here, we consider that they consists in a set of m data or observations $\{x_1, \dots, x_m\}$. However, these data or observations alone are not sufficient by themselves to obtain a credal set \mathcal{P}^+ . A classical means to build a credal set \mathcal{P}^+ from these data is to associate the m observations with a statistical model and a learning process. For example, multinomial data can be associated to the well-known Imprecise Dirichlet model [7], or to some confidence intervals derived from empirical frequencies [17]. Again, such models usually become more and more precise as more data are accumulated and we can consider that positive information is accumulated conjunctively as well. Such a behaviour can be explained by the fact that the space \mathcal{X} is composed of mutually exclusive elements, meaning that observing one value more often makes the others less likely.

Remark 1 *There are cases where either positive or negative information should be combined disjunctively instead of conjunctively.*

Smets [36], when combining reasons to belief and reasons not to belief, propose a rule that combines disjunctively negative information and conjunctively

¹For example, functions corresponding to some chosen events, or to moments such as the mean or the variance.

positive information. However, he works at a different level from ours, since we work directly with knowledge (i.e., information) about variables, and not with knowledge about evidences (i.e., meta-information).

In their possibilistic approach, Dubois and Prade [24] also work directly with knowledge about variables, but propose to combine positive information disjunctively and negative information conjunctively. However, their proposition concerns variables taking their values on a conjunctive space \mathcal{X} , i.e., the true value of X can be several values of \mathcal{X} (in their example, the opening hours of a museum). In that case, it appears natural to combine disjunctively positive information, as observing a particular value does not make the others less likely.

2.3 Merging bipolar information

Once negative and positive information have been collected and represented, it is desirable to combine them into a unique representation, making the most of each information part. Such a unique representation then allows one to infer more precise conclusions than the conclusions one could have infer from each information part alone. This unique representation should also be coherent with both positive and negative information.

As both negative and positive information are modelled by the means of credal sets, it seems natural to merge them through a conjunctive combination operator. The final representation is then the credal set

$$\mathcal{P}^{+\cap-} := \mathcal{P}^+ \cap \mathcal{P}^-$$

resulting from the conjunction of \mathcal{P}^+ and \mathcal{P}^- , provided the set $\mathcal{P}^{+\cap-} \neq \emptyset$ is not empty. Note that this conjunction can again be associated to operations done on lower previsions (see [38, 16] for details). Note that a simple necessary condition for $\mathcal{P}^{+\cap-} \neq \emptyset$ is the following:

Proposition 1 $\mathcal{P}^{+\cap-} \neq \emptyset$ implies that, for any $f \in \mathcal{L}(\mathcal{X})$,

$$\max(\underline{\mathbb{E}}_{\mathcal{P}^-}(f), \underline{\mathbb{E}}_{\mathcal{P}^+}(f)) \leq \min(\overline{\mathbb{E}}_{\mathcal{P}^-}(f), \overline{\mathbb{E}}_{\mathcal{P}^+}(f))$$

Proof Immediate, since for any probability distribution $p \in \mathcal{P}^{+\cap-}$, we have

$$\max(\underline{\mathbb{E}}_{\mathcal{P}^-}(f), \underline{\mathbb{E}}_{\mathcal{P}^+}(f)) \leq \mathbb{E}_p(f) \leq \min(\overline{\mathbb{E}}_{\mathcal{P}^-}(f), \overline{\mathbb{E}}_{\mathcal{P}^+}(f))$$

Now, it may be possible that positive and negative information conflict with each other, i.e. that $\mathcal{P}^{+\cap-} = \emptyset$. In such a case, it is desirable to restore consistency through some revision process. As in [24], we propose to weaken one type of information to restore consistency. Given a parameter $\epsilon \in [0, 1]$ and a credal set \mathcal{P} , let us first define the ϵ -discounted credal set \mathcal{P}_ϵ as

$$\mathcal{P}^\epsilon = \{\epsilon p_{\mathcal{P}} + (1 - \epsilon)p \mid p_{\mathcal{P}} \in \mathcal{P}, p \in \mathbb{P}_{\mathcal{X}}\}. \quad (2)$$

When positive and negative information conflict, then it seems reasonable to weaken one of the two state of information by a value ϵ such that they are no

longer conflicting. Note that the value ϵ is often interpreted as the reliability of the given information. When dealing with knowledge about the value of a variable, direct observations are usually judged more reliable than negative information (as this latter one often comes from models or experts), and it seems more reasonable to revise \mathcal{P}^- rather than \mathcal{P}^+ . A solution to restore consistency is to consider the minimal value ϵ^* such that $\mathcal{P}^{-,\epsilon^*}$ is consistent with \mathcal{P}^+ , i.e.,

$$\epsilon^* = \min \{ \epsilon \in [0, 1] \mid \mathcal{P}^{-,\epsilon^*} \cap \mathcal{P}^+ \neq \emptyset \} \quad (3)$$

and then take $\mathcal{P}^{-,\epsilon^*} \cap \mathcal{P}^-$ as our final state of knowledge. However, note that the above (minimal) revision can lead to an overly precise final information state (as will be shown in the examples of Section 3), and one may well consider some value $\epsilon \geq \epsilon^*$ (for instance, coming from some previous reliability assessment about the information sources).

In principle, the same revision process can be applied to \mathcal{P}^+ , instead of \mathcal{P}^- . As pointed out in [24], this strategy is often more suited to cases where information represent preferences rather than knowledge about a variable. However, it could be used in the case of knowledge representation when the reliability of observations is questionable.

2.4 Revising knowledge with new piece of information

Another case where keeping track of positive and negative information separately can be of importance is when we learn a new piece of information from some source. This new piece of information can be positive information, i.e., new observations, or negative information, given as a credal set \mathcal{P}_{new}^- described by new constraints.

In both cases, we can argue that it is desirable to add this new piece of information to its proper corpus of knowledge before merging the information, rather than adding the new piece of information to the merged information. Namely, it is desirable to revise the model \mathcal{P}^+ by considering old and new observations when the new information is positive, and to revise \mathcal{P}^- into \mathcal{P}'^- such that $\mathcal{P}'^- = \mathcal{P}_{new}^- \cap \mathcal{P}^-$ when the new information is negative.

Indeed, the following situation may happen: the new piece of information \mathcal{P}_{new}^- is not conflicting with \mathcal{P}^- , i.e. $\mathcal{P}_{new}^- \cap \mathcal{P}^- \neq \emptyset$, while it is conflicting with the merged knowledge, in the sense that $\mathcal{P}_{new}^- \cap \mathcal{P}^{+\cap-} = \emptyset$. Thus, there would be no problem in adding this new piece of information to \mathcal{P}^- , since it does not conflict with it, while it would be conflicting with the merged information $\mathcal{P}^{+\cap-}$.

In this latter case, a safe behaviour would be to merge disjunctively the new piece of information with the merged one (i.e., compute $\mathcal{P}_{new}^- \cup \mathcal{P}^{+\cap-}$), as one would no longer be able to tell the difference between negative and positive information in $\mathcal{P}^{+\cap-}$. This could lead to a possible bigger loss of information than if we had added \mathcal{P}_{new}^- directly to its proper corpus of information (again, this will be illustrated in the examples of Section 3). Moreover, the two final representations would not be the same.

3 Illustrative examples

The above framework to deal with bipolar information is quite general, in the sense that it can be applied to any credal set. However, computations with such generic models can be fastidious, and it is often desirable to work with simple models for which computations are more tractable. In this section, we provide some illustrative examples using some popular imprecise probabilistic models, for which the above scheme can be easily applied. The chosen examples represent situations that are likely to be encountered in practical situations (e.g., risk analysis, classification problems, ...).

We provide for each example a means to build a representation from positive and negative information. We then detail how the information merging can be done with the representations, including how revision can be performed in case of conflicting information, and finally we provide a numerical example.

3.1 p-boxes

A p-box [25], denoted by $[\underline{F}, \overline{F}]$ and defined on the real line \mathbb{R} , is a pair of lower and upper cumulative distributions describing our uncertainty about the value of a given random variable. They consist in lower and upper probabilities given over events of the type $(-\infty, x]$. P-boxes are very often used in risk analysis where variable values are ill-known. A p-box induce the credal set $\mathcal{P}_{[\underline{F}, \overline{F}]}$ such that

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

Positive information: Following [25], it is possible to derive a p-box from a set of (i.i.d.) observations (x_1, \dots, x_m) by using Kolmogorov-Smirnov confidence limits to define bounds around the empirical distribution F_m , making no assumption about the distribution form. The distribution F_m is defined as

$$F_m(x) = \begin{cases} 0 & \text{for } x \leq x_{(1)} \\ \vdots & \\ i/n & \text{for } x_{(i)} \leq x \leq x_{(i+1)} \\ \vdots & \\ 1 & \text{for } x_{(m)} \leq x \end{cases}$$

where $x_{(i)}$ are the ordered sample values (i.e., $x_{(i)} \leq x_{(j)}$ if $(i) \leq (j)$). Given the unknown distribution F , we denote by D_{KS} the maximal deviation such that

$$D_{KS} = \max \{|F(x_{(i)}) - i/m|, |F(x_{(i)}) - i^{-1}/m| | i = 1, \dots, m\}.$$

D_{KS} is a random variable whose exact distribution is unknown, but Kolmogorov has shown that $\sqrt{m}D_{KS}$ has a limiting distribution that allows to define, for each confidence level $\alpha \in [0, 1]$, a value $D_m(\alpha)$ such that $P(D_{KS} \leq D_m(\alpha)) \leq 1 - \alpha$. Given a level α , we then define a p-box $[\underline{F}_m, \overline{F}_m]$ such that

$$\underline{F}_m = \max(0, F_m - D_m(\alpha)) \quad \text{and} \quad \overline{F}_m = \min(1, F_m + D_m(\alpha))$$

We denote by $[\underline{F}, \overline{F}]^+$ the p-box generated by the accumulation of positive information, and $\mathcal{P}_{[\underline{F}, \overline{F}]}^+$ the credal sets induced by this p-box.

Negative information: Asking percentile estimations to an expert is a classical elicitation method [10]. However, the reliability of precisely estimated percentiles may be questionable, and an expert may be more comfortable in providing interval estimations rather than point estimations. Here, we consider that an expert can give imprecise evaluations of percentiles, that is a set of increasing values (x_1, \dots, x_m) with $x_i \in \mathbb{R}$ is fixed and the expert provides estimations of the probabilities of events² $[-\infty, x_i]$ for $i = 1, \dots, m$. Other ways by which a p-box could be built using negative information include using bounds resulting from some uncertainty propagation through a physical model [30, 5]. We denote by $[\underline{F}, \overline{F}]^-$ the p-box shaped after negative information, and $\mathcal{P}_{[\underline{F}, \overline{F}]}^-$ the induced credal set.

Merging: When both $\mathcal{P}_{[\underline{F}, \overline{F}]}^+$ and $\mathcal{P}_{[\underline{F}, \overline{F}]}^-$ are induced by p-boxes, it is known [18] that the credal set $\mathcal{P}_{[\underline{F}, \overline{F}]}^+ \cap \mathcal{P}_{[\underline{F}, \overline{F}]}^-$ is still induced by a p-box $[\underline{F}, \overline{F}]^{-\cap+}$ such that

$$[\underline{F}, \overline{F}]^{-\cap+} = [\max\{\underline{F}^-, \underline{F}^+\}, \min\{\overline{F}^-, \overline{F}^+\}].$$

This means that in the particular case of p-boxes, calculations of the final representation is straightforward, at least in the case where positive and negative information are not conflicting. Treating exactly conflicting information is more difficult, as applying Eq. (2) to a credal set induced by a p-box does not usually result in another credal set induced by a p-box. However, given a value ϵ , the p-box $[\underline{F}, \overline{F}]^\epsilon = [\underline{F}^\epsilon, \overline{F}^\epsilon]$ such that $\underline{F}^\epsilon = \epsilon \underline{F}$ and $\overline{F}^\epsilon = \epsilon \overline{F} + 1 - \epsilon$ gives an outer approximation of \mathcal{P}^ϵ . Such a procedure simply comes down to only retain from \mathcal{P}^ϵ the information concerning the lower and upper probabilities on events $[-\infty, x]$.

Numerical example: Assume that variable X evolves between $[0, 16]$ and that 10 samples (1; 1.5; 3; 3.5; 4; 6; 10; 11; 14; 15) have been collected concerning this variable. Given a classical confidence level α of 0.95, the value $D_{10}(0.95) = 0.40925$.

An expert also provides his opinion about the probabilities that the variable is lower than the values 4, 8, 12, 16, in the form of the following lower and upper bounds: $[0, 0.2]$, $[0.1, 0.3]$, $[0.5, 0.7]$.

Figure 1 displays the p-boxes resulting from negative and positive information, as well as the resulting merged p-box. In this case, they are non-conflicting.

²Note that the reverse can also be done, i.e., the expert is given a set of probability values from 0 to 1 and is asked to provide intervals in which these percentiles fall.

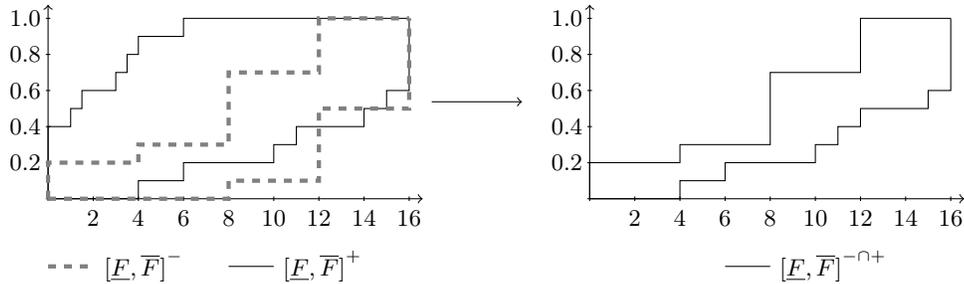


Figure 1: Illustrative example: p-boxes

3.2 probability intervals

Probability intervals [13] are lower/upper probability bounds given on singletons $x \in \mathcal{X}$. Probability intervals can thus be described by a set $L = \{[l(x), u(x)] | x \in \mathcal{X}\}$ of intervals. They induce a credal set \mathcal{P}_L such that

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

Necessary and sufficient conditions for probability intervals to induce a non-empty credal set and to be exact lower/upper probabilistic bounds are provided by [13]. They can be summarized by the conditions that, $\forall x \in \mathcal{X}$, $u(x) + \sum_{y \in \mathcal{X} \setminus x} l(y) \leq 1$ and $l(x) + \sum_{y \in \mathcal{X} \setminus x} u(y) \geq 1$

Positive information: There exist multiple models to compute confidence bounds on multinomial data with only a limited number of samples. This can be done, for instance, by considering statistical confidence intervals over multinomial data [28] or by using the so-called Imprecise Dirichlet Model (IDM) [7]. In this paper, we have retained the latter option, which is the commonest when working with imprecise probabilities. Let m be the total number of observations, $m(x)$ the number of times an element $x \in \mathcal{X}$ has been observed, and $s \in [0, \infty]$ a positive hyperparameter. Probability intervals derived from the IDM are such that, for any $x \in \mathcal{X}$,

$$l(x) = \frac{m(x)}{m+s} \quad \text{and} \quad u(x) = \frac{m(x)+s}{m+s}. \quad (4)$$

We will denote by L^+ the set of probability intervals obtained in this way, and by \mathcal{P}_L^+ the induced credal set.

Negative information: As for p-boxes, negative information can be provided by some experts or by a propagation through a model such as a decision tree [1] or a credal network [12]. We denote by L^- the obtained probability intervals and by \mathcal{P}_L^- the induced credal set.

	x_1	x_2	x_3
u	0.3	0.9	0.2
l	0.1	0.7	0

	x_1	x_2	x_3
u	0.4	0.8	0.3
l	0.2	0.4	0

	x_1	x_2	x_3
u	0.3	0.8	0.1
l	0.2	0.7	0

Table 1: Probability intervals L^+ (above left), L^- (above right) and $L^{+\cap-}$ (bottom middle)

Merging: As in the case of p-boxes, the credal set $\mathcal{P}_L^+ \cap \mathcal{P}_L^-$ resulting from the merging of two probability intervals is again induced by a probability interval. Probability intervals $L^{+\cap-}$ inducing $\mathcal{P}_L^+ \cap \mathcal{P}_L^-$ are such that, $\forall x \in \mathcal{X}$,

$$l^{+\cap-}(x) = \max\{l^+(x), l^-(x), 1 - \sum_{y \in \mathcal{X} \setminus x} u^+(y), 1 - \sum_{y \in \mathcal{X} \setminus x} u^-(y)\}$$

$$u^{+\cap-}(x) = \min\{u^+(x), u^-(x), 1 - \sum_{y \in \mathcal{X} \setminus x} l^+(y), 1 - \sum_{y \in \mathcal{X} \setminus x} l^-(y)\}.$$

Also note that the result of Eq (2), when applied to probability intervals L , result in a credal set still induced by probability intervals L^ϵ such that, $\forall x \in \mathcal{X}$,

$$l^\epsilon(x) = \epsilon l(x) \text{ and } u^\epsilon(x) = \epsilon u(x) + 1 - \epsilon,$$

therefore, in the specific case of probability intervals, the proposed framework can be exactly applied without much computational costs.

Numerical example: We consider a 3-element space $\mathcal{X} = \{x_1, x_2, x_3\}$ on which are defined our probability intervals, as such a space gives us the opportunity to represent credal sets in barycentric coordinates (the space of all probability mass functions is represented by a simplex where each vertex correspond to an element of the space. Each point in the simplex then represents a probability mass p , where the mass $p(x)$ allocated to an element x is proportional to the distance from the point p to the edge opposed to the vertex corresponding to x).

Assume that the observed samples are such that $m = 8$ with $m(x_1) = 1, m(x_2) = 7, m(x_3) = 0$. To model positive information, we use the IDM with a parameter $s = 2$ and apply Eq. (4) to obtain the probability intervals L^+ . Negative information is assumed to come from an expert opinion given as a set of lower and upper bounds L^- . The two probability intervals and the resulting merged intervals $L^{+\cap-}$ are summarised in Table 1 and pictured in Figure 2.

To illustrate the problem of coping with new information, consider now that a new expert (i.e. negative information) provides some complementary opinion in the shape of probability intervals L_{new}^- that are summarized in Table 2. This

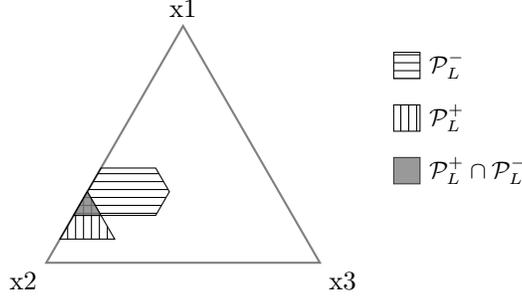


Figure 2: Initial probability intervals and merged information

	x_1	x_2	x_3
u	0.7	0.5	0.2
l	0.2	0.1	0

	x_1	x_2	x_3
u	0.4	0.5	0.3
l	0.2	0.4	0

	x_1	x_2	x_3
u	0.64	0.7	0.58
l	0.12	0.24	0

	x_1	x_2	x_3
u	0.3	0.7	0.18
l	0.12	0.7	0

Table 2: New probability intervals L_{new}^- (above left), new information state $L_{new}^- \cap L^-$ (above right), discounted intervals $(L_{new}^- \cap L^-)^{\epsilon^*}$ (below left) and merged intervals $\mathcal{P}_L^+ \cap (\mathcal{P}_L^- \cap \mathcal{P}_{L_{new}^-}^-)^{\epsilon^*}$ (below right)

new probability intervals are in conflict with $L^{+\cap-}$, but not with L^- alone, and we can thus revise our negative information with this new piece of information by computing $L_{new}^- \cap L^-$. Positive and negative information now conflicts. Using Eq (3), we obtain the value $\epsilon^* = 0.6$. The discounted probability intervals $(L_{new}^- \cap L^-)^{\epsilon^*}$ and the merged one are summarized in Table 2. Figure 3 illustrates the complete process that makes them consistent again. As pointed out in Section 2, the result is very precise (the probability of x_2 is precisely known) and it would perhaps be safer to adopt a strategy providing more cautious inferences (such as taking $\epsilon \geq \epsilon^*$ or simply ignoring the new piece of information if its reliability is questionable).

3.3 comonotonic possibility distributions and clouds

A possibility distribution is a mapping $\pi : \mathcal{X} \rightarrow [0, 1]$ such that $\exists x, \pi(x) = 1$. From a possibility distribution [15, 21] can be defined a possibility measure such that $\Pi(A) = \sup_{x \in A} \pi(x)$. In this paper, possibility measures are interpreted as upper probabilities, and a possibility distribution induces the credal set

$$\mathcal{P}_\pi = \{P \in \mathbb{P}_{\mathcal{X}} | P(A) \leq \Pi(A)\}.$$

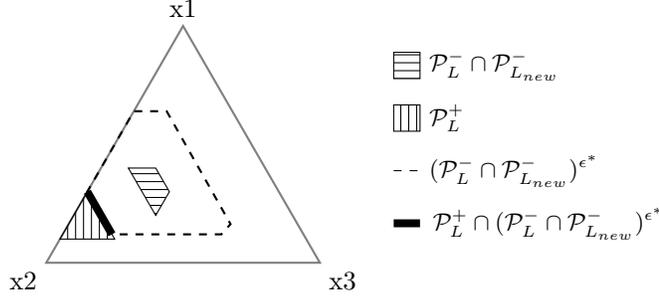


Figure 3: New information and merging of conflicting information

The lower probability induced by a possibility measure is called a necessity measure. Possibility measures are popular imprecise probabilistic models, due to their simplicity and the fact that they can be interpreted (and elicited) as sets of lower confidence bounds given over collection of nested sets, as recalls the next proposition [11]:

Proposition 2 *Given a distribution π , a probability distribution P is in \mathcal{P}_π if and only if*

$$\forall \alpha \in [0, 1], 1 - \alpha \leq P(\{x \in \mathcal{X} | \pi(x) > \alpha\}). \quad (5)$$

However, possibility distributions alone can hardly be used in our bipolar context. Indeed the intersection of two credal sets induced by possibility distributions does not usually result in a credal set induced by another possibility distribution, nor in a credal set that can be approximated by a possibility distribution without much loss of information. Also possibility distributions, although attractive, remain of limited expressiveness as they do not allow for instance to treat the case where an expert provides both lower and upper confidence bounds over nested intervals. This is why we will need a richer representation, namely clouds [34], to deal with this example and situation. Clouds are recent imprecise probabilistic representation, defined as follows:

Definition 1 *A cloud on a space \mathcal{X} is a pair of mappings $[\pi, \delta]$ such that $\delta \leq \pi$, and there is at least one element $x \in \mathcal{X}$ s.t. $\pi(x) = 1$ and one element $y \in \mathcal{X}$ s.t. $\delta(y) = 0$.*

After Neumaier [34], a cloud $[\pi, \delta]$ induces a probability family $\mathcal{P}_{[\pi, \delta]}$ s.t.

$$\mathcal{P}_{[\pi, \delta]} = \{P \in \mathbb{P}_{\mathcal{X}} | P(\delta(x) \geq \alpha) \leq 1 - \alpha \leq P(\pi(x) > \alpha)\}.$$

Note that we retrieve Eq. (5) when $\delta = 0$, in which case the cloud $[\pi, \delta]$ is equivalent to the possibility distribution π alone. It is also known [19] that $\mathcal{P}_{[\pi, \delta]} = \mathcal{P}_\pi \cap \mathcal{P}_{1-\delta}$, where $1-\delta$ is a possibility distribution and $\mathcal{P}_{1-\delta}$ its induced credal set. Clouds therefore allow us to represent the merging of two distinct possibility distributions, provided they satisfy Definition 1. However, this does

not solve the problem of information merging when either negative or positive information is represented originally by clouds. In this case, the subfamily of comonotonic clouds presents important practical advantages.

Definition 2 *A cloud is said to be comonotonic if $\forall x, y \in \mathcal{X}, \pi(x) \leq \pi(y) \Rightarrow \delta(x) \leq \delta(y)$.*

From a practical standpoint, comonotonic clouds are particularly attractive models. When they take a finite number of values, the credal set $\mathcal{P}_{[\pi, \delta]}$ is induced by a set of lower and upper probabilistic bounds given over a collection of nested subsets $\emptyset = A_0 \subset A_1 \subset \dots \subset A_N = \mathcal{X}$ that have the following forms

$$\alpha_i \leq P(A_i) \leq \beta_i$$

and the corresponding cloud is such that, for all $x \in \mathcal{X}$, $\delta(x) = 1 - \beta_i$ and $\pi(x) = 1 - \alpha_{i-1}$ with $x \in A_i \setminus A_{i-1}$.

Positive information: As possibility distributions can represent nested confidence intervals, they are natural candidate to represent information derived from statistical inequalities such as Chebyshev inequality [4]. They are also adequate representations when only few measurements are available [32]. In this paper, we consider that enough data are available so as to provide good estimations of the mean value μ and the variance σ of variable X , and that Chebyshev inequality can be used. The possibility distribution π^+ induced by this inequality is such that

$$\pi^+(\mu - k \cdot \sigma) = \pi^+(\mu + k \cdot \sigma) = \begin{cases} 1 & \text{if } k < 1 \\ \frac{1}{k^2} & \text{if } k \geq 1. \end{cases} \quad (6)$$

We denote by \mathcal{P}_π^+ the credal set induced by the distribution π^+ .

Negative information: Clouds and their interpretation in terms of bounds over nested sets gives a convenient way to elicit information from experts [27]. We consider here that a collection of intervals centred around x^* are provided to the expert, and that he is asked for confidence bounds around these intervals. Such information generates a cloud $[\pi^-, \delta^-]$, and we denote by $\mathcal{P}_{[\pi, \delta]}^-$ the induced credal set.

Merging: When both π^+ , δ^- and π^- are all comonotonic, the credal set $\mathcal{P}_\pi^+ \cap \mathcal{P}_{[\pi, \delta]}^-$ is again a cloud $[\pi^{-\cap+}, \delta^{-\cap+}]$ [18] such that $\delta^{-\cap+} = \delta^-$ and $\pi^{-\cap+} = \min(\pi^+, \pi^-)$. Conflict happens when there is an element $x \in \mathcal{X}$ such that $\delta^{-\cap+}(x) > \pi^{-\cap+}(x)$.

In case of conflict, applying Eq.(2) to a cloud $[\pi, \delta]$ does not result in a credal set induced by a cloud, while it does in the case of possibility distribution (hence, if the cloud is reduced to a single distribution, this problem does not happen). However, as for p-boxes, given a value ϵ , the cloud $[\pi, \delta]^\epsilon$ such that $\pi^\epsilon = \epsilon\pi + 1 - \epsilon$

and $\delta^\epsilon = \epsilon\delta$ gives an outer approximation of $\mathcal{P}_{[\pi, \delta]}^\epsilon$. Other revision processes have been proposed in literature [6], but they seem difficult to interpret in terms of imprecise probabilities and credal sets.

Note that assuming comonotonicity of all functions π^+ , δ^- and π^- may appear unreasonable in some situations: for instance, an expert most plausible value for X (i.e., the value x^* around which nested intervals are built) may differ from a measurement obtained by a sensor or an estimated mean value. In such a case, the merging result is no longer representable by a cloud but still remains a lower probability (computing the merged representation then requires heavier computational efforts). However, it is always possible to consider a weakening of the information so that clouds are made comonotonic, but the incurred information loss can be important.

Numerical example: We assume that estimations from data of the mean and variance are respectively $\mu = 30$ and $\sigma = 5$. Distribution π^+ can then be obtained by using Equation (6).

Let us now assume that an expert has provided the following constraints about the value of X (knowing the estimated mean, one can propose this mean as a starting point to the expert):

$$\begin{aligned} 0.3 &\leq P(X \in [28, 32]) \leq 0.6, \\ 0.7 &\leq P(X \in [20, 40]) \leq 0.9, \\ 1 &\leq P(X \in [15, 45]) \leq 1. \end{aligned}$$

These opinions and constraints can be represented as a cloud $[\pi^-, \delta^-]$, and both information can then be merged. The whole process is illustrated in Fig. 4. In this case, there is no conflict between positive and negative information.

Linking clouds and bipolar possibility theory: let us explore a bit more the links between clouds and bipolar possibility theory [24]. In bipolar possibility theory, two distributions are defined, π and δ , and to them are associated four measures: the classical possibility and necessity measures, respectively defined on any event $A \subseteq \mathcal{X}$ as

$$\Pi(A) = \sup_{x \in A} \pi(x) \quad \text{and} \quad N(A) = 1 - \Pi(A^c) = \inf_{x \in A^c} (1 - \pi(x)),$$

and two other measures [22], called *guaranteed possibility* (Δ) and *potential certainty* (∇), and respectively defined on any event $A \subseteq \mathcal{X}$ as

$$\Delta(A) = \inf_{x \in A} \delta(x) \quad \text{and} \quad \nabla(A) = 1 - \Delta(A^c) = 1 - \inf_{x \in A^c} \delta(x).$$

Now, let us consider that a cloud $[\pi, \delta]$ models the intersection of two credal sets modelled by possibility distributions π (negative information) and $1 - \delta$ (positive information). We can then compute lower and upper probabilities of \mathcal{P}_π for any event $A \subseteq \mathcal{X}$

$$\bar{P}_{\mathcal{P}_\pi}(A) = \Pi_\eta(A) = \sup_{x \in A} \eta(x) = \Pi(A),$$

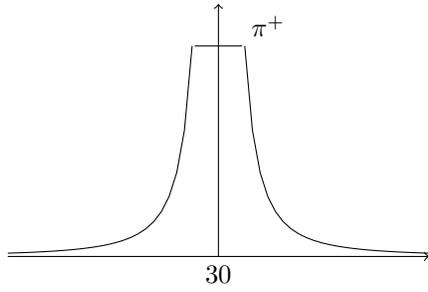


Fig. 4.A Distribution π^+ using Chebyshev inequality ($\mu = 30$ and $\sigma = 5$)

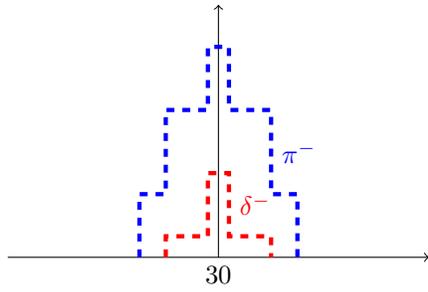


Fig. 4.B Distributions obtained from expert opinion

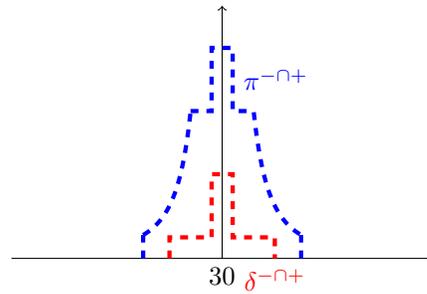


Fig. 4.C Distribution after merging

Figure 4: Cloud and possibility distributions and their merging result.

$$\underline{P}_{\mathcal{P}_\eta}(A) = N_\eta(A) = \inf_{x \in A^c} 1 - \eta(x) = N(A),$$

where Π and N are the measures induced by the initial distribution π . On the other hand we have that the lower and upper probabilities of $\mathcal{P}_{1-\delta}$ are such that

$$\overline{P}_{\mathcal{P}_{1-\delta}}(A) = \Pi_{1-\delta}(A) = \sup_{x \in A} (1 - \delta(x)) = 1 - \inf_{x \in A} \delta(x) = 1 - \Delta(A) = \nabla(A^c).$$

$$\underline{P}_{\mathcal{P}_{1-\delta}}(A) = N_{1-\delta}(A) = \inf_{x \in A^c} (\delta(x)) = 1 - \nabla(A) = \Delta(A^c).$$

Clearly, ∇ and Δ respectively play the role of upper and lower probability measures of positive information, and provided $\mathcal{P}_{[\pi, \delta]} \neq \emptyset$, we have the inequality (using Prop 1)

$$\max(N(A), \Delta(A^c)) \leq \min(\nabla(A^c), \Pi(A)).$$

Note that this inequality is different from the following inequality

$$\max(N(A), \Delta(A)) \leq \min(\Pi(A), \nabla(A)),$$

proved in [22] in the case where $\delta = \pi$ in the setting of possibility theory. However, both inequalities confirms the role of Δ as a lower uncertainty measure and of ∇ as an upper uncertainty measure.

Although clouds are formally equivalent to interval-valued fuzzy sets intuitionistic fuzzy sets [3], their link with the pair $\pi, 1 - \delta$ is only apparent (as noted in [24]). This comes from the fact that the two approaches models different things (e.g., the so-called mirror cloud [34] $[1 - \delta, 1 - \pi]$ corresponds to negation for intuitionistic fuzzy sets, while in the current approach it models exactly the same information as the cloud $[\pi, \delta]$).

4 Conclusion

We have proposed a framework to handle bipolar asymmetric information in the framework of imprecise probabilities, when this information concerns knowledge about the value assumed by a variable. This framework mainly comes down to consider two separate credal sets, one for positive information and one for negative information, and to merge them conjunctively to obtain a final representation. This is in the spirit of other propositions made in other uncertainty theories to handle bipolar information, where two separate (positive and negative) representations are also merged to give a final representation. We then have illustrated our proposition with some specific uncertainty representations that are easier to handle than generic ones, and for which the proposed framework applies easily.

This work is a first step towards the modelling and handling of bipolar information within the theory of imprecise probabilities. It still has to be compared with other approaches proposed within the frameworks of possibility theory and evidence theory (although some elements have been given here, a deeper investigation would be needed). One of the main difference with these two latter

approaches is that both positive and negative information are here combined conjunctively, and give more precise models as more information becomes available. However, since we're working with variable taking their values on spaces made of mutually exclusive elements, such a behaviour is not counter-intuitive.

There are other situations where the use of bipolar information in imprecise probabilities could be of interest, among which:

- the case where information concerns probabilities themselves, for example in the case of linguistic assessments [14]. This would require to connect the present proposition with axioms governing lower previsions, possibly progressing towards an operational definition of bipolarity in terms of betting strategy;
- the case where information concerns not knowledge but preferences, that is when credal sets or lower previsions are used not to express uncertainty but rather to express some preferences or utilities between different criteria of an agent. With this objective in mind, the model of desirable gambles, which have been recently considered as a solution to multicriteria decision problems [37] and extends both credal sets and lower previsions, seems particularly interesting to study.

References

- [1] J. Abellan and S. Moral. Using the total uncertainty criterion for building classification trees. *Int. J. of Intelligent Systems*, 18:1215–1225, 2003.
- [2] L. Amgoud, C. Cayrol, M. Lagasquie-Schiex, and P. Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23:1062–1093, 2008.
- [3] K. Atanassov. Intuitionistic fuzzy sets. *Fuzzy Sets and System*, 20:87–96, 1986.
- [4] C. Baudrit and D. Dubois. Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis*, 51(1):86–108, 2006.
- [5] C. Baudrit, D. Guyonnet, and D. Dubois. Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. *IEEE Trans. Fuzzy Systems*, 14:593–608, 2006.
- [6] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Bipolar possibility theory in preference modelling: representation, fusion and optimal solutions. *Information Fusion*, 7:135–150, 2006.
- [7] J.-M. Bernard. An introduction to the imprecise dirichlet model. *Int. J. of Approximate Reasoning*, 39:123–150, 2008.

- [8] I. Bloch. Fuzzy and bipolar mathematical morphology, applications in spatial reasoning. In *Proc. of Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, 2009.
- [9] J. Cacioppo and G. Bernston. The affect system: architecture and operating characteristics. *Current Directions in Psychological Sciences*, 8:133–137, 1999.
- [10] R. Cooke. *Experts in uncertainty*. Oxford University Press, Oxford, UK, 1991.
- [11] I. Couso, S. Montes, and P. Gil. The necessity of the strong alpha-cuts of a fuzzy set. *Int. J. on Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:249–262, 2001.
- [12] F. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.
- [13] L. de Campos, J. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *I. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.
- [14] G. de Cooman. A behavioural model for vague probability assessments. *Fuzzy Sets and Systems*, 154:305–358, 2005.
- [15] G. de Cooman and D. Aeyels. Supremum-preserving upper probabilities. *Information Sciences*, 118:173–212, 1999.
- [16] G. de Cooman and M. Troffaes. Coherent lower previsions in system modelling : products and aggregation rules. *Reliability Engineering and System Safety*, 85:113–134, 2004.
- [17] T. Denoeux. Constructing belief functions from sample data using multinomial confidence regions. *I. J. of Approximate Reasoning*, 42:228–252, 2006.
- [18] S. Destercke and D. Dubois. The role of generalised p-boxes in imprecise probability models. In *ISIPTA'09 - Proceedings of the sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 179–188, 2009.
- [19] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: II clouds. *Int. J. of Approximate Reasoning*, 49(3):664–677, 2008.
- [20] D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.
- [21] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49:65–74, 1992.

- [22] D. Dubois and H. Prade. Possibility theory: qualitative and quantitative aspects. In *Quantified Representation of Uncertainty and Imprecision*, volume 1 of *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 196–226. Kluwer Academic Publisher, 1998.
- [23] D. Dubois and H. Prade. Bipolarity in flexible querying. In *Proc. 5th Int. Conf. Flexible Query Answering Systems (FQAS)*, pages 174–182, 2002.
- [24] D. Dubois and H. Prade. An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. *Fuzzy Sets and Systems*, 160:1355–1366, 2009.
- [25] S. Ferson, L. Ginzburg, V. Kreinovich, D. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical report, Sandia National Laboratories, 2003.
- [26] B. Finetti. *Theory of probability*, volume 1-2. Wiley, NY, 1974. Translation of 1970 book.
- [27] M. Fuchs and A. Neumaier. Potential based clouds in robust design optimization. *Journal of statistical theory and practice*, To appear, 2008.
- [28] L. Goodman. On simultaneous confidence intervals for multinomial proportions. *Technometrics*, 7:247–254, 1965.
- [29] M. Grabisch and C. Labreuche. Bi-capacities I: definition, Möbius transform and interaction. II: The Choquet integral. *Fuzzy Sets and Systems*, 151:211–259, 2005.
- [30] Karanki D.R., K. H.S., and A. S. Uncertainty analysis based on probability bounds (p-box) approach in probabilistic safety assessment. *Risk Analysis*, 29(5):662–675, 2009.
- [31] I. Levi. *The Enterprise of Knowledge*. MIT Press, London, 1980.
- [32] G. Mauris. Inferring a possibility distribution from very few measurements. In *Soft Methods in Probability, Statistics and Data Analysis*, pages 92–99, 2008.
- [33] E. Miranda. A survey of the theory of coherent lower previsions. *Int. J. of Approximate Reasoning*, In press, 2008.
- [34] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10:249–272, 2004.
- [35] G. Shafer. *A mathematical Theory of Evidence*. Princeton University Press, New Jersey, 1976.
- [36] P. Smets. The canonical decomposition of a weighted belief. In *Proc. Int. Joint. Conf. on Artificial Intelligence*, pages 1896–1901, Montreal, 1995.

- [37] L. Utkin. Multi-criteria decision making with a special type of information about importance of groups of criteria. In *Proc. 6th Int. Symp. on Imprecise Probabilities: Theories and Applications*, pages 411–420, 2009.
- [38] P. Walley. The elicitation and aggregation of beliefs. Technical report, University of Warwick, 1982.
- [39] P. Walley. *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, New York, 1991.
- [40] P. Walley. Towards a unified theory of imprecise probability. In *Proc. of the first Int. Symp. on Imprecise Probabilities and Their Applications*, 1999.