

Handling Bipolar Knowledge with Credal Sets

Sébastien Destercke

Abstract. How to represent and handle bipolar information has recently received a lot of attention. Being bipolar means that the information has a positive and negative part. In this paper, we consider asymmetric bipolar information (i.e. situations where positive and negative information are unrelated and should be processed separately). We propose a framework to represent and handle it with so-called credal sets, i.e., with convex sets of probability distributions. We also provide some illustrative examples.

Keywords: Bipolarity, Imprecise probabilities, Information fusion.

1 Introduction

Bipolarity consists in differentiating between positive and negative information. This information usually concerns either evidences about the true value assumed by an ill-known variable or preferences expressed by some agents. In this paper, we are concerned with the first type of information. One can consider at least three different types of bipolarity (See [9] 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 symmetric bivariate, handles two separate unipolar scales (positive and negative) that refer to the same information and are usually linked by some duality relation. Lower and dual upper previsions [16], whose expressiveness is equivalent to credal sets, are examples of such kind of bipolarity, as well as other models encompassed by this representation (lower/upper probabilities, belief functions, possibility distributions).

The last type of bipolarity, coined as asymmetric or heterogeneous, is the one addressed in this paper. Such bipolarity is used when considering two

Sébastien Destercke

UMR IATE, Campus Supagro, 34060 Montpellier, France

e-mail: sebastien.destercke@irsn.fr

unrelated kinds of information that have to be processed in parallel: one constraining the possible values of a variable (negative information), the other exhibiting what is likely to be observed (positive information). The first kind of information corresponds (for example) to constraints, physical laws, expert opinions, while examples, observations and measurements are instances of the second type. Note that the two kinds of information are effectively unrelated (for instance, an expert may judge as possible a value that will never be observed), hence the need for asymmetry. Also, some psychological studies [3] support the fact that the brain processes differently positive and negative information.

Notions of bipolarity have been declined in a number of frameworks: multi-criteria decision making [11], conflict resolution in argumentation [1], uncertainty and preferences representation in possibility theory [9]. In this paper, we propose a framework to model, represent and treat bipolar information when uncertainty is modelled by convex sets of probabilities, here called credal sets [12], which constitute very generic uncertainty models. The idea behind this framework is quite simple: we propose to represent each corpus of positive and negative information as two separate credal sets, and then to conjunctively merge them in a single credal set. We also propose some solutions to deal with conflicting negative and positive information. After recalling the basics of credal sets, Section 2 presents our proposal. Section 3 then provides some illustrative examples, using the popular imprecise probabilistic representations that are p-boxes and probability intervals.

2 Handling Bipolar Information with Credal Sets

In this paper, we consider that information regarding a variable X assuming its values on a space \mathcal{X} made of mutually exclusive elements is modelled by the means of a credal set \mathcal{P} . Let us 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 $\underline{E}_{\mathcal{P}}(f), \overline{E}_{\mathcal{P}}(f)$ induced by \mathcal{P} such that

$$\underline{E}_{\mathcal{P}}(f) = \inf_{p \in \mathcal{P}} E_p(f) \quad \overline{E}_{\mathcal{P}}(f) = \sup_{p \in \mathcal{P}} E_p(f),$$

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

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

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

with $\mathbb{P}_{\mathcal{X}}$ the set of all probability mass functions over $\mathbb{P}_{\mathcal{X}}$. In his theory of lower previsions [16], 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 $\underline{E}_{\mathcal{P}(\underline{P})}(f) = \underline{P}(f)$ coincides with \underline{P} for every $f \in \mathcal{X}$ (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. $\overline{P}_{\mathcal{P}}(A)$), 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 $\underline{P}_{\mathcal{P}}(A) = 1 - \overline{P}_{\mathcal{P}}(A^c)$.

Credal sets are very general uncertainty models, in the sense that they encompass most of the other known uncertainty models, in particular both necessity measures of possibility theory [7] and belief measures of evidence theory [13] correspond to particular classes of lower probabilities inducing specific credal sets.

2.1 Collecting and Representing Bipolar Information

As what is done in possibility theory [9] and evidence theory [14], 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 possible values of X , considering them impossible or less likely than others (expert opinions are an example of such information). 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

$$\underline{P}(f_i) \leq \sum_{n=1}^N f_i(x_n) p(x_n) \leq \overline{P}(f_i) \quad (1)$$

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 in a set of M observations (experiments), coming in the form of data in our case. To obtain a *positive credal set* \mathcal{P}^+ from these data, one can use a model or a learning process. For instance, multinomial data can be associated to the well-known

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

Imprecise Dirichlet model [2]. Again, positive information is accumulated conjunctively, since the more data we have, the more precise is \mathcal{P}^+ . This is due to the fact that \mathcal{X} is made of mutually exclusive elements, meaning that observing a value more often makes the observation of others less likely.

Note that there are cases where either positive or negative information should be combined disjunctively instead of conjunctively. Smets [14], when combining reasons to believe and reasons not to believe, proposes a rule that combines disjunctively negative information and conjunctively positive information. He works at a different level from ours, since we work directly with knowledge about variables, and not with evidences from which this knowledge is inferred. In their possibilistic approach, Dubois and Prade [9] also work directly with knowledge about variables, but propose to combine positive information disjunctively and negative information conjunctively. However, their proposal 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.

2.2 Merging Bipolar Information

Once negative and positive information have been collected, it is desirable to combine them into a unique credal set. This unique credal set should be non-empty (i.e., consistent) and more precise than the positive and negative credal sets considered separately. Given these requirements, it seems natural to merge them through a conjunctive combination operator, namely to consider as our final information the merged credal set $\mathcal{P}^{+\cap-} := \mathcal{P}^+ \cap \mathcal{P}^-$, when this intersection is not empty.

When positive and negative information conflict with each other (i.e., $\mathcal{P}^{+\cap-} = \emptyset$), it is desirable to restore consistency through some revision process. As in [9], we propose to weaken one type of information to restore consistency. Given a parameter $\varepsilon \in [0, 1]$ and a credal set \mathcal{P} , let us first define the ε -discounted credal set \mathcal{P}_ε as

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

When dealing with bipolar knowledge, observations are usually judged more reliable than negative information, thus it seems more reasonable to weaken \mathcal{P}^- rather than \mathcal{P}^+ . A solution to restore consistency is to consider the minimal value ε^* such that $\mathcal{P}_{\varepsilon^*}^-$ is consistent with \mathcal{P}^+ , i.e.,

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

and then take $\mathcal{P}_{\varepsilon^*}^- \cap \mathcal{P}^+$ as our final state of knowledge. However, as the above revision can lead to a very precise final information state, one may consider some value $\varepsilon \leq \varepsilon^*$. The same revision process can be applied to \mathcal{P}^+ .

Even if this strategy makes more sense when bipolar information represent preferences [9], it could also be used in knowledge representation when data reliability is questionable.

2.3 Revising Knowledge with new Pieces of Information

Another case where differentiating positive and negative information rather than directly considering the merged representation $\mathcal{P}^{+\cap-}$ is useful is the case when one receives new pieces of information to be incorporated into its knowledge. For example, consider new negative information, possibly provided by an additional (reliable) expert, and modelled as a credal set \mathcal{P}_{new}^- . The information conveyed by \mathcal{P}_{new}^- should be first added to \mathcal{P}^- , e.g., by computing $\mathcal{P}'^- = \mathcal{P}_{new}^- \cap \mathcal{P}^-$, before merging negative and positive information in a single representation. Note that making this distinction can be important, as \mathcal{P}_{new}^- may be non-conflicting with \mathcal{P}^- (i.e., $\mathcal{P}_{new}^- \cap \mathcal{P}^- \neq \emptyset$), while it may be conflicting with the current positive and negative information taken together (i.e., $\mathcal{P}_{new}^- \cap \mathcal{P}^+ \cap \mathcal{P}^- = \emptyset$).

3 Illustrative Examples

Let us now provide some illustrative examples of the proposed way to deal with bipolar knowledge. The examples concern two popular imprecise probabilistic models: p-boxes [10] and probability intervals [4].

3.1 p-Boxes

A p-box $[F, \bar{F}]$ defined on the (here discretized) real line \mathbb{R} is a pair of lower and upper cumulative distributions describing our uncertainty about the value of a variable. They consists in lower and upper probabilities given over events of the type $(-\infty, x]$, inducing a credal set $\mathcal{P}_{[F, \bar{F}]}$ such that

$$\mathcal{P}_{[F, \bar{F}]} = \{p \in \mathbb{P}_{\mathbb{R}} \mid \forall x \in \mathbb{R}, F(x) \leq F_p(x) = P([-\infty, x]) \leq \bar{F}(x)\},$$

where F_p is the cumulative distribution of p .

Positive information. Following [10], it is possible to derive a p-box from a limited set of observations (x_1, \dots, x_m) by using Kolmogorov-Smirnov confidence limits to define bounds around the empirical distribution F_m , thus making no assumption about the distribution form. The distribution F_m is defined as

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

where $x_{(i)}$ are the ordered sampled values. Given the samples and a confidence level $\alpha \in [0, 1]$, one can use KS confidence limits to obtain 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 $\mathcal{P}_{[\underline{E}, \overline{F}]}^+$ the credal set obtained from this positive information.

Negative information. Negative information forming p-boxes usually comes from experts evaluating some percentiles for a set of fixed values. We denote by $\mathcal{P}_{[\underline{E}, \overline{F}]}^-$ the credal set induced by negative information.

Merging. In the particular case of p-boxes, the credal set $\mathcal{P}_{[\underline{E}, \overline{F}]}^{-\cap+} = \mathcal{P}_{[\underline{E}, \overline{F}]}^+ \cap \mathcal{P}_{[\underline{E}, \overline{F}]}^-$ is also induced by a p-box $[\underline{E}, \overline{F}]^{-\cap+}$ such that

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

In case of conflict, applying Eq. (2) does not usually result in a credal set induced by a p-box. However, given a value ε , the p-box $[\underline{E}, \overline{F}]^\varepsilon$ such that $\underline{F}^\varepsilon = \varepsilon \underline{E}$ and $\overline{F}^\varepsilon = \varepsilon \overline{F} + 1 - \varepsilon$ induces an outer approximation of \mathcal{P}_ε .

Example 1. Assume $X \in [0, 16]$. 10 samples (1; 1.5; 3; 3.5; 4; 6; 10; 11; 14; 15) provide an empirical cumulative distribution. For a confidence level of 0.95, the value $D_{10}(0.95) = 0.40925$. An expert also provides its opinion about the probabilities that the variable value is lower than values 4, 8, 12, 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 $[\underline{E}, \overline{F}]^+$ and $[\underline{E}, \overline{F}]^-$ resulting from these two types of information as well as the merging result.

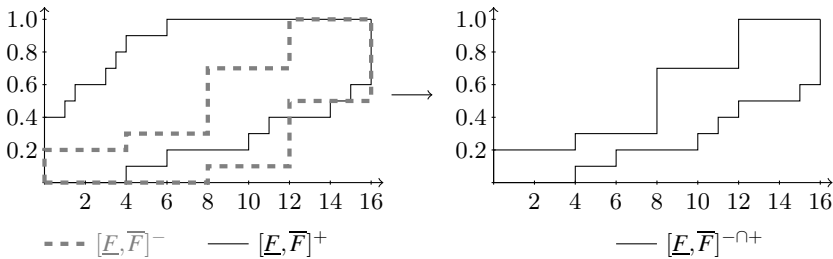


Fig. 1 Illustrative example: p-boxes

3.2 Probability Intervals

Probability intervals [4] are a set of lower and upper probabilistic bounds given over singletons $x \in \mathcal{X}$. They can be described by a set $L = \{[l(x), u(x)] \mid x \in \mathcal{X}\}$ of intervals. They induce a credal set \mathcal{P}_L such that

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

Necessary and sufficient conditions for probability intervals to induce a non-empty credal set are provided by [4]. They can be summarized by the conditions that, $\forall x \in \mathcal{X}$,

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

Positive information. There are multiple models to compute confidence bounds on multinomial data with a limited number of samples. This can be done, for instance, by considering statistical confidence intervals over multinomial data [6] or by using the so-called Imprecise Dirichlet Model (IDM) [2]. Here, we consider the IDM. Let $\{x_1, \dots, x_N\}$ be an arbitrary indexing of elements of \mathcal{X} , M the total number of observations, m_k the number of times x_k has been observed, and s a positive real value determining the quickness of convergence of the IDM. Then, the probability intervals derived from the IDM are such that, for $x_k, k = 1, \dots, N$

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

We denote by L^+ the obtained probability intervals, and \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 (e.g., a credal network [5]). We denote by L^- the obtained probability intervals, and \mathcal{P}_L^- the induced credal set.

Merging. The credal set $\mathcal{P}_L^{-\cap+} = \mathcal{P}_L^+ \cap \mathcal{P}_L^-$ is again induced by a probability interval $L^{+\cap-}$ which is 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, $\forall x \in \mathcal{X}$,

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

Example 2. We consider a 3-elements space $\mathcal{X} = \{x_1, x_2, x_3\}$ on which are defined our probability intervals. The observed samples are such that $m = 8$ with $m_1 = 1, m_2 = 7, m_3 = 0$. To model positive information, we use the IDM with a parameter $s = 2$ and apply Eq. (4) to obtain the probability intervals L^+ such that

$$u^+(x_1) = 0.3, u^+(x_2) = 0.9, u^+(x_3) = 0.2; l^+(x_1) = 0.1, l^+(x_2) = 0.7, l^+(x_3) = 0.$$

Negative information is assumed to be an expert opinion given as a set L^- such that

$$u^-(x_1) = 0.4, u^-(x_2) = 0.5, u^-(x_3) = 0.3; l^-(x_1) = 0.2, l^-(x_2) = 0.4, l^-(x_3) = 0.$$

In this case, negative and positive information are conflicting ($u(x_2)^- \leq l(x_2)^+$), as $\mathcal{P}_L^+ \cap \mathcal{P}_L^- = \emptyset$. Using Eq. (3), we obtain $\varepsilon^* = 0.6$ and $L_{\varepsilon^*}^-$ such that

$$u^-(x_1) = 0.64, u^-(x_2) = 0.7, u^-(x_3) = 0.58; l^-(x_1) = 0.12, l^-(x_2) = 0.24, l^-(x_3) = 0.$$

Finally giving the merged structure $L_{\varepsilon^*}^{+\cap-}$

$$u^-(x_1) = 0.3, u(x_2)^- = 0.7, u(x_3)^- = 0.18; l^-(x_1) = 0.12, l(x_2)^- = 0.7, l(x_3)^- = 0$$

which indeed gives a very precise evaluation of the uncertainty of having $X = x_2$.

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 of a given variable. The proposal is illustrated with some credal sets induced by specific probability bounds often used in practice. This work is a first step towards the modelling and handling of bipolar information within the recent theory of imprecise probabilities. It still has to be compared in a deeper way with other approaches made in possibility theory and evidence theory, possibly by making sense of the concept of guaranteed possibility [8] or of commonality function in the context of imprecise probabilities.

Another interesting problem is how to handle bipolarity when credal sets or lower previsions are used not to express uncertainty but imprecise preferences or utilities. An idea would be to consider the alternative model of desirable gambles, recently considered as a solution to multicriteria decision problems [15].

References

1. Amgoud, L., Cayrol, C., Lagasque-Schiex, M., Livet, P.: On bipolarity in argumentation frameworks. *Int. J. Intell. Syst.* 23, 1062–1093 (2008)
2. Bernard, J.M.: An introduction to the imprecise Dirichlet model. *Internat. J. Approx. Reason.* 39, 123–150 (2008)
3. Cacioppo, J., Bernston, G.: The affect system: architecture and operating characteristics. *Curr. Dir. Psychol. Sci.* 8, 133–137 (1999)
4. de Campos, L., Huete, J., Moral, S.: Probability intervals: a tool for uncertain reasoning. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems* 2, 167–196 (1994)
5. Cozman, F.: Credal networks. *Artificial Intelligence* 120, 199–233 (2000)

6. Denoeux, T.: Constructing belief functions from sample data using multinomial confidence regions. *Internat. J. Approx. Reason.* 42, 228–252 (2006)
7. Dubois, D., Prade, H.: *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York (1988)
8. Dubois, D., Prade, H.: Interval-valued fuzzy sets, possibility theory and imprecise probability. In: *Proceedings of International Conference in Fuzzy Logic and Technology, EUSFLAT 2005, Barcelona, Spain (2005)*
9. Dubois, D., Prade, H.: An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. *Fuzzy Sets Syst.* 160, 1355–1366 (2009)
10. Ferson, S., Ginzburg, L., Kreinovich, V., Myers, D., Sentz, K.: *Constructing probability boxes and Dempster-Shafer structures*. Tech. rep., Sandia National Laboratories (2003)
11. Grabisch, M., Labreuche, C.: Bi-capacities I: definition, Möbius transform and interaction. II: The choquet integral. *Fuzzy Sets Syst.* 151, 211–259 (2005)
12. Levi, I.: *The Enterprise of Knowledge*. MIT Press, London (1980)
13. Shafer, G.: *A mathematical Theory of Evidence*. Princeton University Press, New Jersey (1976)
14. Smets, P.: The canonical decomposition of a weighted belief. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 1995, Montréal, Qu'ebec, Canada*, pp. 1896–1901 (1995)
15. Utkin, L.: Multi-criteria decision making with a special type of information about importance of groups of criteria. In: *Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications, ISIPTA 2009, Durham, UK*, pp. 411–420 (2009)
16. Walley, P.: *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, New York (1991)