



HAL
open science

Extreme Points of the Credal Sets Generated by Elementary Comparative Probabilities

Enrique Miranda, Sébastien Destercke

► **To cite this version:**

Enrique Miranda, Sébastien Destercke. Extreme Points of the Credal Sets Generated by Elementary Comparative Probabilities. Proceedings of ECSQARU 2013, Jul 2013, Utrecht, Netherlands. pp.424-435. hal-00861991

HAL Id: hal-00861991

<https://hal.science/hal-00861991v1>

Submitted on 14 Sep 2013

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.

Extreme Points of the Credal Sets Generated by Elementary Comparative Probabilities

Enrique Miranda¹ and Sébastien Destercke²

¹ University of Oviedo, Dept. of Statistics and O.R., Oviedo, Spain
mirandaenrique@uniovi.es

² CNRS, HEUDIASYC Joint Research Unit, Compiègne, France
sebastien.destercke@hds.utc.fr

Abstract. When using convex probability sets (or, equivalently, lower previsions) as models of uncertainty, identifying extreme points can be useful to perform various computations or to use some algorithms. In general, sets induced by specific models such as possibility distributions, linear vacuous mixtures or 2-monotone measures may have extreme points easier to compute than generic convex sets. In this paper, we study extreme points of another specific model: comparative probability orderings between the elements of a finite space. We use these extreme points to study the properties of the lower probability induced by this set, and connect comparative probabilities with other uncertainty models.

Keywords: Comparative probabilities, credal sets, 2-monotone capacities, belief functions, regular extension, imprecise mass functions.

1 Introduction

In the last decades, there has been a growing interest on imprecise probability models as alternative models to probability in situations where the available information is vague or scarce. This type of models include for instance belief functions [1], possibility measures [2], 2- and n-monotone capacities [3] or probability boxes [4]. All the above examples can be seen as instances of coherent lower and upper previsions [5].

The adequacy of each of these models for a particular problem depends, among other things, on the interpretation we are giving to our uncertainty. In this paper, we consider a *robust Bayesian* interpretation [6]: we assume the existence of a precise, but unknown, probability model, and work with the set of probability measures that are compatible with the available information. This gives rise to a *credal set*, as considered by Levi in [7].

Here, we consider the case where the information is expressed by means of a *comparative probability model* [8]: we consider a finite probability space Ω and assume that we are given judgements of the type “the probability of A is at least as great as that of B ”. Comparative probabilities have been deemed of particular interest within the context of subjective probability theory [9,10,11]; see also [5, Section 4.5] for a study from the point of view of coherent lower previsions. One of their advantages is that they seem well suited for modelling qualitative judgements.

In spite of this, there are only few works dealing with the numerical and practical aspects of comparative probabilities [12]. One reason for this is that it is not easy to summarize the set of probabilities associated to the comparative assessments, for instance by means of a lower and an upper probability, and this renders it difficult to summarize the information about the probability of an event of interest. In this paper, we solve this problem by characterizing the comparative probability models by means of the extreme points of their associated credal sets. This is a problem that has been studied for other types of imprecise probability models, such as 2-monotone capacities [13], possibility measures [14], probability intervals [15] and belief functions [16]. In this paper, we focus on probability sets generated by comparisons between singletons. Focusing on this particular case allows us to derive nice graphical characterizations, and we provide some practical examples where this special case may be useful. Such There is only one partial result for this type of assessments [17], and we generalize it in this paper.

After giving some preliminary results in Section 2, we shall see in Section 3 that, when the comparison judgements are made on the probabilities of the singletons, a graphical representation of these judgements makes it easy to derive the extreme points of the associated credal sets. In Section 4, we use this result to discuss some practical aspects of these models: we establish tight lower and upper bounds on the number of extreme points; investigate their relationship with other imprecise probability models; provide algorithms for the computation of these extreme points; and discuss the computation of conditional lower probabilities and the merging of multiple comparison judgements. Some additional remarks related to the practical use of these models and their extensions are provided in Section 5.

2 Preliminaries

Consider a finite space $\mathcal{X} = \{x_1, \dots, x_n\}$, modelling the set of outcomes of some experiment. In this paper, we assume that our information about these outcomes can be modelled by means of *comparative probability orderings of the states*, i.e., statements of the type “the probability of x_i is at least as great as that of x_j ”. Hence, we shall represent the available information by means of a subset \mathcal{L} of $\mathcal{X} \times \mathcal{X}$.

The set of probability measures compatible with this information is given by

$$\mathcal{P}(\mathcal{L}) = \{p \in \mathbb{P}_{\mathcal{X}} : \forall (i, j) \in \mathcal{L}, p(x_i) \geq p(x_j)\}, \quad (1)$$

where $\mathbb{P}_{\mathcal{X}}$ denotes the set of all probabilities on the power set of \mathcal{X} .

For the purposes of this paper, it shall be useful to represent these assessments by means of a graph $\mathcal{G} = (\mathcal{X}, \mathcal{L})$ where the nodes are the elements of \mathcal{X} and we draw an edge between x_i and x_j when $(i, j) \in \mathcal{L}$.

Example 1. Consider the space $\mathcal{X} = \{x_1, \dots, x_5\}$ and the set of assessments $\mathcal{L} = \{(1, 3), (1, 4), (2, 5), (4, 5)\}$. Its associated graph \mathcal{G} is given by Figure 1. ♦

Note that the set $\mathcal{P}(\mathcal{L})$ determined by Eq. (1) is always non-empty, because it includes for instance the uniform probability distribution. It is interesting to compare it

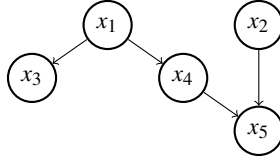


Fig. 1. Graph \mathcal{G} of Example 1

with the set

$$\mathcal{P}(\mathcal{K}) = \{p \in \mathbb{P}_{\mathcal{X}} : \forall (i, j) \in \mathcal{L}, p(x_i) > p(x_j)\},$$

i.e., with the credal set associated by *strict* elementary probability comparisons, which appear also sometimes in the literature. Since $\mathcal{P}(\mathcal{L})$ is a closed convex polytope in \mathbb{R}^n , it follows from basic convex analysis that $\mathcal{P}(\mathcal{L})$ corresponds to the closure of $\mathcal{P}(\mathcal{K})$ when the latter set is non-empty, and that $\mathcal{P}(\mathcal{K})$ is the topological interior of $\mathcal{P}(\mathcal{L})$. The non-emptiness of $\mathcal{P}(\mathcal{K})$ is easy to characterise.

Proposition 1. $\mathcal{P}(\mathcal{K}) \neq \emptyset$ if and only if \mathcal{G} is acyclic.

Hence, our results in this paper will also allow us to characterize the set $\mathcal{P}(\mathcal{K})$. As we shall see in Remark 1, we can also deal with assessments of equality between the probabilities, which correspond to a cycle in \mathcal{G} .

3 Extreme points of $\mathcal{P}(\mathcal{L})$.

In this section, we characterise the extreme points of the credal set $\mathcal{P}(\mathcal{L})$ associated with a number of elementary probability comparisons. Consider a space \mathcal{X} and a subset \mathcal{L} of $\mathcal{X} \times \mathcal{X}$, and let $\mathcal{P}(\mathcal{L})$ be the set it determines by means of Eq. (1). Any of the probability measures in $\mathcal{P}(\mathcal{L})$ is completely determined by its mass function, and as a consequence it can be seen as an element of the n -th dimensional Euclidean space. Then, $\mathcal{P}(\mathcal{L})$ is a closed convex subset of \mathbb{R}^n in the Euclidean topology, that corresponds thus to the closed convex hull of its set of extreme points. We shall determine these extreme points by means of the graphical representation we have established in Section 2.

We shall make two assumptions on the graph \mathcal{G} associated to \mathcal{L} :

- (G1) The first one is that \mathcal{G} is acyclic, meaning that there are no assumptions of equality between the probability of two different states.
- (G2) The second is that \mathcal{G} is connected, so for every $i \neq j$ there is an undirected path in \mathcal{G} that connects the nodes x_i and x_j .

Remark 1. The results we obtain can be used to characterise the general case. On the one hand, when \mathcal{G} has cycles, we have some assumptions of equality $P(x_i) = P(x_j)$ between the probabilities of two elements x_i, x_j in our possibility space. It is not difficult to determine the structure of the set $\mathcal{P}(\mathcal{L})$ in that case: for each of the assumptions of equality, we consider one of the elements x_i in the corresponding set and store the

number of elements n_i in \mathcal{X} that are assumed to have the same probability as x_i ; from this we derive the simplified space $\mathcal{X}' \subset \mathcal{X}$ for which the graph \mathcal{G} satisfies (G1).

By this, we can establish a one-to-one correspondence between the sets $\mathcal{P}(\mathcal{L}) \subseteq \mathbb{P}_{\mathcal{X}}$ and $\mathcal{P}(\mathcal{L}') \subseteq \mathbb{P}_{\mathcal{X}'}$: any probability $P := (p_1, \dots, p_n)$ in $\mathcal{P}(\mathcal{L})$ induces the probability P' on $\mathcal{P}(\mathcal{L}')$, with $P'(x_i) = P(x_i) \cdot n_i$. Then, once we determine the distributions of the extreme points associated to the graph \mathcal{G}' , we just have to ‘expand’ this graph by reversing the above correspondence between the probabilities.

On the other hand, if \mathcal{G} does not satisfy (G2), we can decompose it as a union of its weakly connected components $\mathcal{G}_1, \dots, \mathcal{G}_k$. For each of these components we can characterise their associated extreme points in the form we shall give below, and then the extreme points associated to \mathcal{G} will be the union of the sets of extreme points in each of these subgraphs. \blacklozenge

In order to characterise the extreme points of $\mathcal{P}(\mathcal{L})$, we are going to consider a number of lemmas:

Lemma 1. *Any extreme point p of $\mathcal{P}(\mathcal{L})$ corresponds to a uniform probability measure over some subset $A \subseteq \mathcal{X}$.*

For every subset A of \mathcal{X} , we shall denote by P_A the uniform probability measure on A , that is associated to the mass function

$$P_A(x_i) = \begin{cases} \frac{1}{|A|} & \text{if } x_i \in A \\ 0 & \text{otherwise} \end{cases}$$

for any $i \in \{1, \dots, n\}$. Using the acyclic graph \mathcal{G} , we can now characterize those subsets $A \subseteq \mathcal{X}$ for which P_A is an extreme point of $\mathcal{P}(\mathcal{L})$. For every $x_j \in \mathcal{X}$, we shall denote by $H(x_j)$ the set of ancestors of x_j , i.e., those nodes x_i such that there is a directed path from x_i to x_j in \mathcal{G} . By an abuse of notation, we shall also consider that x_j is an ancestor of itself, i.e., we shall assume that $x_j \in H(x_j)$ for all j . Finally, for every $A \subseteq \mathcal{X}$, we shall denote $H(A) := \cup_{x \in A} H(x)$.

The following lemma gives further insight onto which uniform probabilities may be extreme points of the credal set $\mathcal{P}(\mathcal{L})$.

Lemma 2. *1. If $A \neq H(A)$, then P_A is not an extreme point of $\mathcal{P}(\mathcal{L})$.
2. If there are $C_1, C_2 \subseteq A$ such that $H(C_1) \cap H(C_2) = \emptyset$ and $H(C_1) \cup H(C_2) = H(A)$, then $P_{H(A)}$ is not an extreme point on $\mathcal{P}(\mathcal{L})$.*

Next, if B is a subset of A and $H(B) = H(A)$, both A and B give rise to the same probability measure $P_{H(B)} = P_{H(A)}$. This is related to the notion of strongly connected nodes:

Definition 1. *Two nodes x_i, x_j in the graph \mathcal{G} are said to be strongly connected when there is a directed path from x_i to x_j , or viceversa, and are called strongly disconnected otherwise.*

Equivalently, x_i, x_j are strongly connected when either $x_i \in H(x_j)$ or $x_j \in H(x_i)$. This allows us to establish the following result:

Theorem 1. *If $P_{H(A)}$ is an extreme point of $\mathcal{P}(\mathcal{L})$, then there is some $B \subseteq A$ with $H(B) = H(A)$ and such that any two nodes in B are strongly disconnected. Thus, the set of extreme points coincide with the set of probabilities $P_{H(A)}$ generated by sets A*

(EXT1) *composed of strongly disconnected nodes of \mathcal{G} and*

(EXT2) *that cannot be decomposed as in Lemma 2.*

Remark 2. An interesting related result has been established in [17], in the context of credal classification. The author considers the credal set determined by the comparisons of the probabilities of the states, and computes the lower probability of the set A of elements with no predecessor in \mathcal{G} . In order to do this, she provides results analogous to our Lemmas 1 and 2, and then in [17, Theorem B.2.2] she establishes which of the elements in $\mathcal{P}(\mathcal{L})$ attain the lower probability of A .

Our previous result subsumes these results, in the sense that we give the explicit form of the extreme points (from which we may determine also the lower probability of any other set, as well as the lower prevision induced by a comparative probability model). Note moreover that we have showed that not all the uniform probability distributions $P_{H(A)}$ determine an extreme point of $\mathcal{P}(\mathcal{L})$. ♦

Example 2. The extreme points generated by Example 1 are summarised in Table 1.

Table 1. Extreme points of Example 1

A	$H(A)$	P				
		x_1	x_2	x_3	x_4	x_5
$\{x_1\}$	$\{x_1\}$	1	0	0	0	0
$\{x_2\}$	$\{x_2\}$	0	1	0	0	0
$\{x_3\}$	$\{x_1, x_3\}$	1/2	0	1/2	0	0
$\{x_4\}$	$\{x_1, x_4\}$	1/2	0	0	1/2	0
$\{x_5\}$	$\{x_1, x_2, x_4, x_5\}$	1/4	1/4	0	1/4	1/4
$\{x_1, x_2\}$	$\{x_1, x_2\}$	1/2	1/2	0	0	0
$\{x_2, x_3\}$	$\{x_1, x_2, x_3\}$	1/3	1/3	1/3	0	0
$\{x_2, x_4\}$	$\{x_1, x_2, x_4\}$	1/3	1/3	0	1/3	0
$\{x_3, x_4\}$	$\{x_1, x_3, x_4\}$	1/3	0	1/3	1/3	0
$\{x_3, x_5\}$	$\{x_1, x_2, x_3, x_4, x_5\}$	1/5	1/5	1/5	1/5	1/5
$\{x_2, x_3, x_4\}$	$\{x_1, x_2, x_3, x_4\}$	1/4	1/4	1/4	1/4	0

4 Practical aspects

4.1 Number of extreme points

Since extreme points correspond to uniform distributions over certain subsets $A \subseteq \mathcal{X}$, we immediately see that an upper bound of the number of extreme points is $2^{|\mathcal{X}|}$. Note that this is significantly lower than the maximal number of extreme points generated by lower coherent probabilities, known to be $|\mathcal{X}|!$ [18]. We next show that this number of extreme points can be reduced even further (recall that we are assuming throughout that the graph \mathcal{G} associated with \mathcal{L} satisfies (G1) and (G2)):

Theorem 2. *The maximal number of extreme points of $\mathcal{P}(\mathcal{L})$ is $2^{(|\mathcal{X}|-1)}$, and the minimum number is $|\mathcal{X}|$. Each of these bounds can be attained.*

To see that the upper bound given by the above theorem can indeed be reached, consider the case where a single modal value is provided. Figure 2 illustrates the situation.

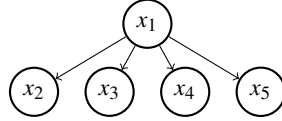


Fig. 2. Graph \mathcal{G} for $x_1 = \text{modal value}$

Interestingly, the upper bound given in Theorem 2 is the same as the number of extreme points of the credal set associated to a possibility measure, as showed in [14, Section 5]. Our intuition for this is that possibility measures also determine an order between the singletons, by means of their associated possibility distributions. On the other hand, an example where the lower bound is reached is the case where \mathcal{L} forms a complete ordering of singletons $\{x_1, \dots, x_n\}$ (this is the case considered in [5, P. 195]; note that the result there is now a particular case of Theorem 1).

4.2 Extraction Algorithm

Using the results of Section 3, we can propose a pseudo-algorithm to extract extreme points, summarised in Algorithm 1.

Implementing this algorithm mainly requires to be able, for a given set B , to check whether elements of B are strongly disconnected and to compute $H(B)$. An instrumental tool to do this is the matrix M corresponding to the transitive closure $\mathcal{C}(\mathcal{L}) \subseteq \mathcal{X} \times \mathcal{X}$ of \mathcal{L} , with $M(i, j) = 1$ iff $(i, j) \in \mathcal{C}(\mathcal{L})$. M can be efficiently computed by applying Warshall algorithm (see [19]) to matrix L with $L(i, j) = 1$ iff $(i, j) \in \mathcal{L}$.

Once this is done, checking whether two elements x_i, x_j are strongly disconnected can be done in linear time. Checking that B is made of strongly disconnected elements is equivalent to check whether all pairs of elements $x_i, x_j \in B$ are strongly disconnected, hence at most in quadratic time. As $H(B) = \cup_{x \in B} H(x)$, computing $H(B)$ is also linear. This means that the complexity of the loop going from Line 10 to 13 in Algorithm 1 is quadratic.

Algorithm 1 also tries to minimize the number of sets of nodes to check by reducing the search to sets that are not known to be sets containing connected nodes, rather than making a naïve search among all subsets $B \subseteq \mathcal{X}$. However, the algorithm would still have to check, at worst, an exponential number of sets.

4.3 n-monotonicity

Next, we investigate in more detail the set of probabilities $\mathcal{P}(\mathcal{L})$ from the point of view of the theory of coherent lower previsions developed in [5]. Since the set $\mathcal{P}(\mathcal{L})$

Algorithm 1: Extreme point search

Input: Set \mathcal{L} of comparisons
Output: Extreme points of $\mathcal{P}(\mathcal{L})$

```

1 List  $\leftarrow \emptyset$ ;
2 for  $i = 1, \dots, n$  do
3   Build extreme points corresponding to  $H(x_i)$ ;
4   List  $\leftarrow x_i$ ;
5 Candidate set  $\leftarrow$  List ;
6 for  $i = 2, \dots, n$  do
7   List  $\leftarrow \emptyset$  ;
8   foreach set  $B$  in Candidate set do
9     for  $j = 1, \dots, n$  do
10      if  $x_j$  is strongly disconnected from the elements  $B$  and  $H(B \cup \{x_j\})$  is a new
      extreme point then
11        Add  $H(B \cup \{x_j\})$  to extreme points ;
12        List  $\leftarrow B \cup \{x_j\}$  ;
13      Candidate set  $\leftarrow$  List ;
  
```

is a closed convex set of probabilities, its lower envelope \underline{P} , given by

$$\underline{P}(A) = \min\{P(A) : P \in \mathcal{P}(\mathcal{L})\} \forall A \subseteq \mathcal{X} \quad (2)$$

is a *coherent lower probability*. As such, it can be given a behavioural interpretation in terms of acceptable betting rates.

Coherent lower probabilities include as particular cases most of the imprecise probability models that we can find in the literature, such as 2-monotone capacities, belief functions, or necessity measures; see [20] for more details. In particular, a coherent lower probability is 2-monotone when for any $A, B \subseteq \mathcal{X}$ we have

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B). \quad (3)$$

These are also called *convex* functions on Choquet capacities of order 2 [3,21]. When $|\mathcal{X}| \leq 3$, a coherent lower probability on $\mathcal{P}(\mathcal{X})$ is always 2-monotone [22], and as consequence this is also true for the comparative probability models we consider in this paper. On the other hand, when $|\mathcal{X}| \geq 4$, there exist coherent lower probabilities on $\mathcal{P}(\mathcal{X})$ which are not 2-monotone. We next show that, in general, the coherent lower probabilities induced by comparative probability models will not be 2-monotone.

Example 3. Consider $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and $\mathcal{L} = \{(1, 2), (1, 3), (2, 4), (3, 4)\}$. From Theorem 1, the extreme points of $\mathcal{P}(\mathcal{L})$ are associated to the mass functions

$$\left\{ \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right), \left(\frac{1}{2}, \frac{1}{2}, 0, 0 \right), \left(\frac{1}{2}, 0, \frac{1}{2}, 0 \right), \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0 \right), (1, 0, 0, 0) \right\};$$

as a consequence, if we consider the events $A = \{x_1, x_3\}$ and $B = \{x_1, x_4\}$, we see that

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) = 1/2 + 1/4 < \underline{P}(A) + \underline{P}(B) = 1/2 + 1/3.$$

Hence, \underline{P} violates the 2-monotonicity condition. ♦

From this, we can deduce that belief functions, that are in particular 2-monotone, are not expressive enough to represent comparative probability models.

On the other hand, from a convex set of probability measures we can also determine lower and upper expectation functionals. Similarly to Eq. (2), the real-valued functional \underline{P} given by

$$\underline{P}(f) = \min\{P(f) : P \in \mathcal{P}(\mathcal{L})\} \quad (4)$$

for any function $f : \mathcal{X} \rightarrow \mathbb{R}$ is called a *coherent lower prevision*. Here, we are also using P to denote the expectation functional associated to the probability measure P , given by $P(f) = \sum_{x \in \mathcal{X}} f(x)p(x)$.

Similarly to Eq. (3), a coherent lower prevision is called 2-monotone when

$$\underline{P}(f \vee g) + \underline{P}(f \wedge g) \leq \underline{P}(f) + \underline{P}(g)$$

for any $f, g : \mathcal{X} \rightarrow \mathbb{R}$, where \vee denotes the point-wise maximum and \wedge denotes the point-wise minimum. This type of lower previsions has been studied in detail in [22,23]. They are interesting, because, unlike coherent lower previsions, they can be calculated as the Choquet integral with respect to the lower probability that is their restriction to events. Moreover, 2-monotonicity has been showed to be equivalent to comonotone additivity [23, Theorem 15]. However, we can prove that the coherent lower prevision associated to a one-to-one comparison model is not 2-monotone as soon as \mathcal{X} has more than two elements:

Theorem 3. *Consider a space \mathcal{X} with $|\mathcal{X}| \geq 3$, and let \mathcal{L} be a number of probability comparisons on the elements of \mathcal{X} whose associated graph satisfies (G1) and (G2). Let \underline{P} be the coherent lower prevision determined by (4). Then \underline{P} is not 2-monotone.*

Although it is an open problem at this stage, we think that with similar arguments to those in the proof it can be showed that a comparative probability model on the singletons never determines a 2-monotone lower prevision even when the associated graph \mathcal{G} violates (G1) or (G2).

4.4 Conditioning

A classical operation when dealing with uncertainty is that of conditioning. Here we will study the problem of computing lower conditional probabilities $\underline{P}(A|B)$ from the credal set $\mathcal{P}(\mathcal{L})$. Out of the many possible notions we can consider in this case, we think that the most intuitive under the robust Bayesian interpretation we are considering in this paper is that of *regular extension* [5, Appendix J], that produces

$$\underline{P}(A|B) = \inf_{P \in \mathcal{P}(\mathcal{L})} \{P(A|B) : P(B) > 0\}, \quad (5)$$

where $P(A|B)$ is obtained from P through Bayes Rule of conditioning.

Note that in order to apply this rule, we need that there is some probability measure P in $\mathcal{P}(\mathcal{L})$ such that $P(B) > 0$ (or, in other words, that the *upper* probability $\bar{P}(B)$

is positive); but this is no restriction in the case of comparative probabilities, because there will always be an extreme point P of $\mathcal{P}(\mathcal{L})$ for which $P(B) > 0$: it suffices to consider $P_{H(x_i)}$ with $x_i \in B$. On the contrary, the lower probability $\underline{P}(B)$ will be positive if and only if for any $x_i \in \mathcal{X}$ it holds that $B \cap H(x_i) \neq \emptyset$, i.e., if and only if B contains all the nodes without a predecessor.

To attain the conditional lower probability $\underline{P}(A|B)$ given by Eq. (5), we need to find the extreme point for which $P(B)$ is positive and the fraction $P(A \cap B)/P(B)$ minimal. This can be done easily by the procedure described in Algorithm 2. Note that it is sufficient to concentrate on extreme points generated by subsets C of $B \setminus A$, as we want to minimise the ratio $|H(C) \cap B \cap A|/|H(C) \cap B|$. From this, we easily derive the following algorithm:

Algorithm 2: Conditional Probability computation

Input: Set \mathcal{L} of comparisons

Output: Lower conditional probability $P(A|B)$ with $A \subset B$

```

1 Cond ← 1 ;
2 foreach Set  $C \subseteq B \setminus A$  do
3   Value ←  $|H(C) \cap (A \cap B)|/|H(C) \cap B|$  ;
4   if Value < Cond then Cond ← Value
5 Return Cond ;

```

4.5 Multiple source merging

When multiple sources provide different comparisons, for instance when two different experts provide assessments \mathcal{L}_1 and \mathcal{L}_2 , it becomes necessary to merge them in a single representation. The two most common rules to do so are the conjunction and disjunction, that respectively come down to computing $\mathcal{P}(\mathcal{L}_1) \cap \mathcal{P}(\mathcal{L}_2)$ and $CH(\mathcal{P}(\mathcal{L}_1) \cup \mathcal{P}(\mathcal{L}_2))$, where CH denotes the convex hull (the disjunction usually producing non-convex probability sets). Our next result shows that simple operations on \mathcal{L}_1 and \mathcal{L}_2 can provide exact or approximated results of these operations.

Proposition 2. *1. The disjunctively merged set $CH(\mathcal{P}(\mathcal{L}_1) \cup \mathcal{P}(\mathcal{L}_2))$ is such that $CH(\mathcal{P}(\mathcal{L}_1) \cup \mathcal{P}(\mathcal{L}_2)) \subseteq \mathcal{P}(\mathcal{L}_1 \cup \mathcal{L}_2)$, and the inclusion can be strict.*
2. The conjunctively merged set satisfies $\mathcal{P}(\mathcal{L}_1) \cap \mathcal{P}(\mathcal{L}_2) = \mathcal{P}(\mathcal{L}_1 \cap \mathcal{L}_2)$.

5 Practical examples and extensions

In this section, we propose some particular examples of situations where elementary comparative probability models can be used, and discuss some possible extensions.

5.1 Imprecise mass functions

Elementary comparative probability models can be related to the work on imprecise mass functions discussed by Augustin [24] and Denoeux [25]. Recall that a belief function \underline{P} on the power set of \mathcal{X} is uniquely determined by its associated *basic probability assignment* m , by means of the formula [1]

$$\underline{P}(A) = \sum_{E \subseteq A} m(E). \quad (6)$$

The basic probability assignment $m(E)$ of a set E represents the weight of the available evidence supporting that the outcome of the experiment belongs to E . It holds that $\sum_{E \subseteq \mathcal{X}} m(E) = 1$, so we may regard m as the probability mass function of some probability measure on $\mathcal{P}(\mathcal{X})$. This arises for instance in the context of finite random sets.

We can then use our results to build imprecise mass functions. If we have assessments of the type $m(A_i) \geq m(A_j)$, we may consider the set of the mass functions compatible with these assessments. This is a convex set of probability measures whose extreme points can be determined by means of Theorem 1. Note that, by means of Eq. (6), each of these mass functions determines a belief function, that in turn is equivalent to a convex set of probability measures on $\mathcal{P}(\mathcal{X})$. Hence, a convex set of mass functions also induces a convex set of probabilities on $\mathcal{P}(\mathcal{X})$ [24]; however, its lower probability will not be, in general, a belief function (nor, as we can deduce from Example 3, 2-monotone).

This can be useful for instance in the context of inner/outer measures [26]. We may think of an infinite space \mathcal{X} that is partitioned into n sets A_1, \dots, A_n , and where a probability measure $P(A_i)$ is associated to each set A_i . Such an assessment induces on the power set of \mathcal{X} a set of probabilities that can be described by $m(A_i) = P(A_i)$. In this situation, comparative statements between the probabilities $P(A_i)$ are equivalent to comparative statements between the masses $m(A_i)$, and the set of extreme masses can then be derived using our results.

5.2 Extension to general comparative probability models: some comments

The most important extension of our work would be to consider arbitrary comparative probability models, where we allow for comparisons between any pair of events (the case of partitions is treated in Section 5.1), that is to allow any comparison $P(A) \geq P(B)$ with $A, B \subseteq \mathcal{X}$. These are the models studied extensively in [8,10,11], amongst others.

Note that, when considering comparative probability models, we can assume that the sets A, B we compare are disjoint, since the assessments $P(A) \geq P(B)$ and $P(A \setminus B) \geq P(B \setminus A)$ are equivalent. However, the existence of a probability compatible with the assessments is no longer trivial, and therefore the associated set $\mathcal{P}(\mathcal{L})$ may be empty: think for instance of the case of $\mathcal{X} = \{x_1, x_2, x_3\}$ and the assessments $P(\{x_1\}) \geq P(\{x_2, x_3\})$, $P(\{x_2\}) \geq P(\{x_1, x_3\})$ and $P(\{x_3\}) \geq P(\{x_1, x_2\})$. These are equivalent to $P(\{x_1\}) \geq 0.5$, $P(\{x_2\}) \geq 0.5$ and $P(\{x_3\}) \geq 0.5$, and there is no probability measure satisfying all of them simultaneously.

When $\mathcal{P}(\mathcal{L})$ is non-empty, then it is a closed convex set which is characterized by its finite number of extreme points. However, as the next example shows, we cannot expect the extreme points of such assessments to be as simple as the extreme points generated by the comparison of the probabilities of the states. In particular, the extreme points of the associated credal sets will not be necessarily associated with uniform probability distributions over some subsets, and finding an easy graphical representation from which they could be extracted seems hard.

Example 4. Consider $\mathcal{X} = \{x_1, x_2, x_3\}$ and the assessments $P(\{x_2\}) \geq P(\{x_1\})$ and $P(\{x_1, x_2\}) \geq P(\{x_3\})$, and let \mathcal{P} be the credal set determined by these assessments. The extreme points of \mathcal{P} are given by the mass functions

$$\{(0, 1, 0), (1/2, 1/2, 0), (1/4, 1/4, 1/2), (0, 1/2, 1/2)\}. \blacklozenge$$

6 Conclusions

Comparative probability models constitute a useful approach to modelling uncertain information about a probability model, especially when the available information is of a qualitative nature. However, most of the works in the literature about these models have focused on axiomatizing those comparative probability models that can be associated to a set of probability measures. In this paper, we have deepened on the link between elementary comparative probability models and imprecise probabilities, by: (a) characterizing the structure of the set of probability measures associated to a comparative probability model, and (b) studying the properties of the lower probability induced by this set. Interestingly, we have showed that this lower probability may not be 2-monotone, from which it follows that 2-monotone capacities (and in particular belief functions, or possibility measures) are not expressive enough to be able to deal with this type of qualitative information. Moreover, we have showed that the maximum number of extreme points is similar to the maximal number of extreme points of credal sets induced by possibility measures, and smaller than those induced by 2-monotone capacities or belief functions.

We have also suggested some practical situations where this model can be useful, such as the elicitation of modal or least probable values or imprecise mass functions. However, this model remains quite simple and of limited expressiveness; it would be desirable to determine to which extent the results presented in this paper can be extended to the case of general comparisons between disjoint events, discussed in Section 5.2. Another important open problem would be to provide algorithms for the computation of the lower prevision induced by a comparative probability model, and to study in detail the applications of these results in fields such as qualitative decision making.

Acknowledgements

The research in this paper has been supported by project MTM2010-17844 and by an invited professorship provided by Universidad de Oviedo. We would like to thank Marco Zaffalon for making us aware of the work carried out in [17], as well as for other useful suggestions. We also thank Erik Quaeghebeur for stimulating discussion.

References

1. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton, NJ (1976)
2. Dubois, D., Prade, H.: Possibility Theory. Plenum Press, New York (1988)
3. Choquet, G.: Theory of capacities. *Annales de l'Institut Fourier* **5** (1953–1954) 131–295
4. Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D.S., Sentz, K.: Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories (January 2003)
5. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
6. Berger, J.O.: An overview of robust Bayesian analysis. *Test* **3** (1994) 5–124 With discussion.
7. Levi, I.: The enterprise of knowledge. MIT Press, Cambridge (1980)
8. Koopman, B.: The axioms and algebra of intuitive probability. *Annals of Mathematics* **41** (1940) 269–292
9. Fine, T.: Theories of Probability. Academic Press, New York (1973)
10. Suppes, P.: The measurement of belief. *Journal of the Royal Statistical Society. Series B (Methodological)* **36** (Jan 1974) 160–191
11. Walley, P., Fine, T.L.: Varieties of modal (classificatory) and comparative probability. *Synthese* **41** (1979) 321–374
12. Regoli, G.: Comparative probability and robustness. *Lecture Notes-Monograph Series* (1996) 343–352
13. Chateauneuf, A., Jaffray, J.Y.: Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences* **17**(3) (1989) 263–283
14. Miranda, E., Couso, I., Gil, P.: Extreme points of credal sets generated by 2-alternating capacities. *International Journal of Approximate Reasoning* **33**(1) (2003) 95–115
15. de Campos, L.M., Huete, J.F., Moral, S.: Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **2** (1994) 167–196
16. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics* **38** (1967) 325–339
17. Gulordava, K.: Empirical evaluation of credal classifiers. Master's thesis, University of Lugano (June 2010) Supervised by M. Zaffalon and G. Corani.
18. Wallner, A.: Extreme points of coherent probabilities in finite spaces. *International Journal of Approximate Reasoning* **44**(3) (2007) 339–357
19. Warshall, S.: A theorem on Boolean matrices. *Journal of the ACM* **9**(1) (1962) 11–12
20. Walley, P.: Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning* **24** (2000) 125–148
21. Denneberg, D.: Non-Additive Measure and Integral. Kluwer Academic, Dordrecht (1994)
22. Walley, P.: Coherent lower (and upper) probabilities. Technical Report Statistics Research Report 22, University of Warwick, Coventry (1981)
23. de Cooman, G., Troffaes, M.C.M., Miranda, E.: n -Monotone exact functionals. *Journal of Mathematical Analysis and Applications* **347** (2008) 143–156
24. Augustin, T.: Generalized basic probability assignments. *International Journal of General Systems* **4** (2005) 451–463
25. Denoeux, T.: Reasoning with imprecise belief structures. *International Journal of Approximate Reasoning* **20** (1999) 79–111
26. Fagin, R., Halpern, J.Y.: Uncertainty, belief, and probability. In: Computational Intelligence, Morgan Kaufmann (1989) 1161–1167