Information Sciences 638 (2023) 118995

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Imprecise probabilistic models based on hierarchical intervals

Serafín Moral *, Andrés Cano, Manuel Gómez-Olmedo

Computer Science and Artificial Intelligence, University of Granada, Spain

ARTICLE INFO

Keywords:
Probability intervals
Imprecise probability
Coherence
Natural extension
Order-2 capacities

ABSTRACT

This paper proposes a generalization of the imprecise probability model given by probability intervals on singleton sets. This enables us not only to represent probability intervals in a hierarchy of sets with a tree structure but also to represent other models such as possibility measures and generalized p-boxes. The paper also shows how the resulting model is always an order-2 capacity and that the basic operations of checking coherence, computing the natural extension or conditioning can be performed in an extremely efficient way.

1. Introduction

One of the basic problems of imprecise probabilities is to check whether a set of assessments (upper and lower previsions) avoids sure loss, and, if it does, to compute its natural extension [25]. Other interesting problems include computing the set of extreme probabilities or determining the conditional upper and lower previsions. Generally speaking, these are expensive computational problems for general previsions, but in some particular relevant cases, they can be efficiently solved. In the general case, checking coherence or computing the natural extension can be performed using linear programming [27].

Due to the computational complexity of using general intervals in real applications, less general and more efficient models are usually considered. The case of elementary probability intervals [2] is of special interest as there are explicit expressions for checking coherence and efficient algorithms to compute the natural extension. There are also algorithms to compute all the extreme probabilities of the associated credal set [2]. If X is a finite frame of discernment, an elementary interval probabilistic specification is a set of probability intervals for each element $x \in X$, i.e. a family of intervals $[\underline{P}(\{x\}), \overline{P}(\{x\})], \quad x \in X$.

A key issue of elementary intervals is that they always result in an order-2 capacity [4,14,15] and, therefore, certain operations such as computing extreme probabilities of the associated credal set or calculating the conditional upper and lower probabilities can be performed more efficiently and directly due to the special characterization of extreme probabilities [6] and conditional intervals [3] in order-2 capacities.

In this paper, we extend the results for elementary intervals to the more general setting of hierarchical interval specifications. A hierarchical interval specification is a family of intervals $[\underline{P}(A), \overline{P}(A)]$ in which events can be structured as a tree in such a way that the events associated with the children of a node are a partition of the event associated with the node. We also show how simple algorithms can be used to check coherence and compute the natural extension. Previous algorithms in [2] for elementary intervals are a particular case of the algorithms provided in this paper. We also demonstrate that these intervals always result in an order-2

* Corresponding author.
E-mail addresses: smc@decsai.ugr.es (S. Moral), acu@decsai.ugr.es (A. Cano), mgomez@decsai.ugr.es (M. Gómez-Olmedo).

https://doi.org/10.1016/j.ins.2023.118995
Received 22 June 2022; Received in revised form 21 February 2023; Accepted 17 April 2023

Available online 23 April 2023

capacity, providing fast procedures to compute not only extreme probabilities and conditional probabilities but also the upper and lower previsions of a general gamble.

Hierarchical intervals should not be confused with hierarchical Bayesian probabilistic models which have a family of variables that are dependent on a set of parameters related by means of a sequence of conditional probability distributions [12]. Although this paper considers imprecision in probabilities, there is another important difference: our model is applied to unconditional rather than conditional probabilistic specifications. The imprecise probability tree model [7] is more similar to the Bayesian hierarchical model that is also based on conditional information.

Our model also includes other well-known representations of uncertainty that are very common in literature such as possibility-necessity intervals [10] and p-boxes [8]. In this way, we unify some of the most common imprecise specifications used in practice into a single framework.

The paper is structured in the following way: Section 2 provides the basic framework to introduce the important concepts and results, and follows a logical approach to differentiate between the initial specification from the associated consequences (the natural extension). Section 3 introduces the concept of hierarchical interval specification and provides the basic algorithms for checking coherence and computing the natural extension. Furthermore, a procedure is given to compute the natural extension in an incremental way whenever a new element is added to a set. Section 4 proves that hierarchical intervals always produce order-2 capacities and provides an algorithm to compute the extreme probability associated with a specific order of the elements of the referential set. The comparison with existing models is presented in Section 5, where the differences with the imprecise probability tree model is discussed and the inclusion of other models as p-boxes is presented. Finally, Section 6 outlines our conclusions and discusses our future lines of work.

2. Basic concepts

Let us assume a referential set with a finite number of elements, $X = \{x_1, \dots, x_n\}$. A probabilistic interval specification [25,28] is a triplet $(\mathcal{A}, \underline{P}, \overline{P})$, where $\mathcal{A}$ is a family of subsets of X ($\mathcal{A} \subseteq 2^X$) and $\underline{P}, \overline{P}$ are mappings from $\mathcal{A}$ into $[0,1]$, such that $\underline{P}(A) \leq \overline{P}(A)$, $\forall A \in \mathcal{A}$.

Given a probabilistic interval specification $S = (\mathcal{A}, \underline{P}, \overline{P})$, its associated set of dominated probability measures [16,18,25] is the set

$$\mathcal{M}(S) = \{P \in \mathcal{P}(X) \,:\, \underline{P}(A) \leq P(A) \leq \overline{P}(A),\ \forall A \in \mathcal{A}\}, \tag{1}$$

where $\mathcal{P}(X)$ is the set of all the probability measures on X. $\mathcal{M}(S)$ is always closed and convex and is called the associated credal set with S. If P is a probability measure on X, we will denote its probability distribution as p. We will also use $m(S)$ to denote the set of probability distributions associated with the probability measures in $\mathcal{M}(S)$.

It is said that $S = (\mathcal{A}, \underline{P}, \overline{P})$ avoids sure loss if and only if $\mathcal{M}(S) \neq \emptyset$. Henceforth, all probabilistic specifications are assumed to avoid sure loss, even when this is not explicitly specified.

The natural extension of S is the interval specification $E(S) = (2^X, \underline{P}', \overline{P}')$ where

$$\underline{P}'(A) = \inf_{P \in \mathcal{M}(S)} P(A),\ \overline{P}'(A) = \sup_{P \in \mathcal{M}(S)} P(A),\ \forall A \in 2^X. \tag{2}$$

It is said that $(\mathcal{A}, \underline{P}, \overline{P})$ is coherent if and only if $\underline{P}(A) = \underline{P}'(A)$, $\overline{P}(A) = \overline{P}'(A)$, $\forall A \in \mathcal{A}$.

In the natural extension, only one interval limit is required since the following equality is always satisfied:

$$\underline{P}'(A) + \overline{P}'(A^c) = 1, \tag{3}$$

where A^c is the complementary set of A with respect to X.

It is said that $S_1 = (\mathcal{A}_1, \underline{P}_1, \overline{P}_1)$ and $S_2 = (\mathcal{A}_2, \underline{P}_2, \overline{P}_2)$ are equivalent if and only if they have the same natural extension $E(S_1) = E(S_2)$. Generally speaking, equivalent probabilistic intervals provide the same information about an unknown value on X. It is desirable that a system of probabilistic intervals be coherent. If S_1, S_2 are coherent and equivalent specifications and $\mathcal{A}_1 \subseteq \mathcal{A}_2$, then S_1 is shorter and requires less space to be stored, but S_2 is more complete as the upper and lower probabilities for events $A \in \mathcal{A}_2 \setminus \mathcal{A}_1$ are already computed and we do not need to use Expression (2) to obtain them. Therefore, there is a space/time trade-off between short/large specifications. The coherent specifications defined in $\mathcal{A} = 2^X$ are said to be complete. The natural extension is always complete.

An interval probabilistic specification $S = (\mathcal{A}, \underline{P}, \overline{P})$ is said to be elementary if and only if $\mathcal{A} = \{\{x\} \,:\, x \in X\}$, i.e. the intervals are given for singleton subsets of X. These elementary interval specifications can be equivalently determined by giving intervals in the associated probability distribution, $[\underline{p}(x), \overline{p}(x)], \forall x \in X$.

In [2], it was proved that an elementary specification avoids sure loss if and only if

$$\sum_{x \in X} \underline{p}(x) \leq 1 \leq \sum_{x \in X} \overline{p}(x), \tag{4}$$

and it is coherent if and only if

$$\sum_{\substack{x \in X \\ x \neq y}} \underline{p}(x) + \overline{p}(y) \leq 1 \leq \sum_{\substack{x \in X \\ x \neq y}} \overline{p}(x) + \underline{p}(y),\ \forall y \in X. \tag{5}$$

If an elementary interval specification S avoids sure loss, its natural extension $E(S) = (2^X, \underline{P}', \overline{P}')$ is given by [2],

$$\begin{aligned} \underline{P}'(A) &= \max\{\sum_{x\in A} \underline{p}(x), 1 - \sum_{x\notin A} \overline{p}(x)\}, \\ \overline{P}'(A) &= \min\{\sum_{x\in A} \overline{p}(x), 1 - \sum_{x\notin A} \underline{p}(x)\}, \end{aligned} \quad \forall A \subseteq X. \tag{6}$$

Example 1. Let us assume $X = \{x_1, x_2, x_3, x_4\}$ and the elementary specification given by the intervals $\underline{p}(x_i) = 0.2, \overline{p}(x_i) = 0.3$ for any x_i. We can easily verify that this specification avoids sure loss and is coherent. In this case, the lower intervals of the natural extension are given by

$$\underline{P}'(A) = 0.2 \min\{|A|, 2\} + 0.3 \max\{|A| - 2, 0\}, \forall A \subseteq X.$$

The natural extension of an interval specification $S = (\mathcal{A}, \underline{P}, \overline{P})$ is always complete, but it might be convenient to represent it by providing intervals in a more reduced set of intervals. The natural extension can, in fact, always be equivalently specified on the same set $\mathcal{A}$ as S, as it is proved by the following result.

Proposition 1. *Let $S = (\mathcal{A}, \underline{P}, \overline{P})$ be an interval specification that avoids sure loss, and $E(S) = (2^X, \underline{P}', \overline{P}')$ be its natural extension. If $S' = (\mathcal{A}, \underline{P}', \overline{P}')$, then S' and $E(S)$ are equivalent.*

Proof. S' and $E(S)$ are equivalent when they define the same set of probability measures, i.e. when $\mathcal{M}(S') = \mathcal{M}(E(S))$.

Since S' assigns the same values of lower and upper probabilities as $E(S)$ but in a smaller family of sets, then it is obvious that the associated sets computed by expression (1) satisfy $\mathcal{M}(E(S)) \subseteq \mathcal{M}(S')$.

Moreover, it is well known [25, pag. 136] that $\mathcal{M}(S) = \mathcal{M}(E(S))$, and that S is defined for the same sets as S' and that $\underline{P}'(A) \geq \underline{P}(A)$, $\overline{P}'(A) \leq \overline{P}(A)$. Consequently, the reverse inclusion $\mathcal{M}(S') \subseteq \mathcal{M}(S) = \mathcal{M}(E(S))$ is also satisfied.

Henceforth, $S' = (\mathcal{A}, \underline{P}', \overline{P}')$ will be called the partial natural extension of S and denoted by $E'(S)$.

An interval specification S is said to be an order-2 capacity [14,15] if and only if

$$\underline{P}'(A_1 \cup A_2) + \underline{P}'(A_1 \cap A_2) \geq \underline{P}'(A_1) + \underline{P}'(A_2), \ \forall A_1, A_2 \subseteq X, \tag{7}$$

or, equivalently,

$$\overline{P}'(A_1 \cup A_2) + \overline{P}'(A_1 \cap A_2) \leq \overline{P}'(A_1) + \overline{P}'(A_2), \ \forall A_1, A_2 \subseteq X, \tag{8}$$

where $(\underline{P}', \overline{P}')$ are the probability intervals of the natural extension of S.

It is well known that an elementary specification is always an order-2 capacity [2]. In order-2 capacities, the extreme points of their associated credal set can be computed in the following way: for each σ permutation in $\{1, \ldots, n\}$, we consider the probability distribution p_σ on X given by

$$p_\sigma(x_{\sigma(i)}) = \underline{P}'(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}) - \underline{P}'(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}), \ i = 1, \ldots, n, \tag{9}$$

where $\underline{P}'$ is the lower probability in the natural extension.

In the case of elementary probability intervals, the number of extreme probability distributions is usually much lower than $n!$ and in [2] an efficient algorithm is given to compute them. There is an implicit result underlying this algorithm that we will now proceed to formally state and prove.

Proposition 2. *If we have a coherent set of elementary probability intervals S given by lower and upper probability distributions $(\underline{p}, \overline{p})$ such that $\sum_{y\in X} \underline{p}(y) < 1$, then the set of the extreme probabilities of their associated credal set $m(S)$ is given by the probability distributions $p_{(A,x)}$, where (A, x) is the set of pairs in which A is a subset $A \subseteq X$ and x is an element $x \in X \setminus A$ such that*

$$\underline{P}'(A) = \sum_{y\in A} \underline{p}(y), \ \underline{P}'(A \cup \{x\}) > \sum_{y\in A\cup\{x\}} \underline{p}(y). \tag{10}$$

In this case, $p_{(A,x)}$ is given by

$$p_{(A,x)}(y) = \begin{cases} \underline{p}(y) & \text{if } y \in A \\ 1 - \sum_{z\notin(A\cup\{x\})} \overline{p}(z) - \sum_{z\in A} \underline{p}(z) & \text{if } y = x \\ \overline{p}(y) & \text{if } y \in X \setminus (A \cup \{x\}) \end{cases} \tag{11}$$

See proof on page 16.

Although the case $\sum_{y\in X} \underline{p}(y) = 1$ is not covered by the previous proposition, it is a trivial situation since in this case under coherence, we must have $\underline{p}(x) = \overline{p}(x), \forall x \in X$ and, therefore, there is a single probability distribution in $m(S)$ given by $p(x) = \underline{p}(x), \forall x \in X$.

Example 2. With the intervals given in Example 1, the extreme probabilities are associated with pairs (A, x) where A has two elements. Furthermore, since $p_{(A,x)} = p_{(A,y)}$, for any $x, y \notin A$, the extreme probabilities only depend on A. There is one for each $x_i, x_j \in X$ denoted by p_{ij} given by

$$p_{ij}(x_i) = p_{ij}(x_j) = 0.2,\ p_{ij}(x_l) = 0.3\ (l \neq i, l \neq j) \tag{12}$$

A set of probability measures Q in 2^X always defines a coherent and complete set of probability intervals $S(Q)$, given by

$$\underline{P}(A) = \inf_{P \in Q} P(A),\ \overline{P}(A) = \sup_{P \in Q} P(A), \forall A \in 2^X. \tag{13}$$

It is important to note that several sets of probabilities can define the same set of probability intervals, with $\mathcal{M}(S)$ being the greatest one. In general, a set of probability intervals S is considered to be equivalent to $\mathcal{M}(S)$, while other smaller sets defining the same intervals are considered to be more informative [25]. If we start with an arbitrary set Q and compute its associated set of probability intervals, $S(Q)$ can be considered to be an approximate outer representation of Q (equivalent to a larger set of probability measures).

There are several possible definitions of conditioning in imprecise probability. In this paper, we consider regular conditioning [25]. This is defined in terms of credal sets. If Q is a credal set and $A \subseteq X$, then the conditioning of Q to A is the credal set

$$Q_A = \{P(.|A)\ :\ P \in Q,\ P(A) > 0\}, \tag{14}$$

where $P(.|A)$ denotes the conditional probability P given A, which is properly defined if and only if $\overline{P}(A) > 0$.

The conditional set of an interval specification $S = (\mathcal{A}, \underline{P}, \overline{P})$ to event $A \subseteq X$ can be defined as the complete interval specification $S_A = (2^X, \underline{P}(.|A), \overline{P}(.|A))$ obtained with the following procedure:

- Compute the credal set associated with S (Expression (1)): $\mathcal{M}(S)$
- Compute the conditioning of $\mathcal{M}(S)$ to A (Expression (14)): $\mathcal{M}(S)_A$
- Compute S_A as the interval specification associated with $\mathcal{M}(S)_A$ (Expression (13)): $S(\mathcal{M}(S)_A)$.

We can see how the conditional interval specification is always complete, except when it is not defined, i.e. when $\overline{P}'(A) = 0$, where $\overline{P}'(A)$ denotes the upper probability of A in the natural extension.

Example 3. In the case of the intervals given in Example 1, if the conditioning set is $A = \{x_1, x_2, x_3\}$, then the credal set will be generated by $p_{ij}(.|A)$, where p_{ij} are the probability distributions given in Example 2. These conditional probabilities are[1]:

$$\begin{array}{ll} p_{ij}(x_i|A) = p_{ij}(x_j|A) = 2/7,\ p_{ij}(x_l|A) = 3/7,\ l \neq i, l \neq j & \text{if } x_i, x_j \in A \\ p_{ij}(x_i|A) = 1/4,\ p_{ij}(x_l|A) = 3/8,\ l \neq i, & \text{if } x_i \in A,\ x_j \notin A. \end{array} \tag{15}$$

The intervals associated with these probability distributions are the intervals of $S(\mathcal{M}(S)_A)$ and are given for the subsets of A by

$$\underline{P}(\{x_i\}|A) = 1/4,\ \underline{P}(\{x_i, x_j\}|A) = 4/7,\ \underline{P}(\{x_1, x_2, x_3\}|A) = 1$$

In this case, it can be seen how this set of probability intervals can be represented by a set of elementary probability intervals $(\underline{p}_A, \overline{p}_A)$:

$$\underline{p}_A(x_i) = 1/4,\ \overline{p}_A(x_i) = 3/7,\ i = 1, 2, 3$$

One of the extreme probabilities of the credal set associated with this interval specification is given by

$$p(x_1) = 1/4,\ p(x_2) = 9/28,\ p(x_3) = 3/7$$

This probability distribution is not a convex combination of the probability distributions in Expression (15). In order to obtain $p(x_1) = 1/4$, it is only possible to select the probability in the second row but then $p(x_2)$ should be equal to $3/8$.

The conclusion is that the step from the credal set defined in Expression (15) to the probability intervals increases the associated credal set and there is an ensuing loss of information. In order to obtain the entire credal set exactly, it is necessary to add the upper and lower previsions (intervals on general gambles or functions defined on X) [16,17].

When the initial specification S is an order-2 capacity, the upper and lower conditional intervals can be computed directly without the use of the associated credal set $\mathcal{M}(S)$ using the following result [3]:

Proposition 3. *If the interval specification S is an order-2 capacity, then the conditional intervals can be computed using the following expressions:*

[1] In order to simplify, we only specify the conditional information of elements or sets in the conditioning set A.

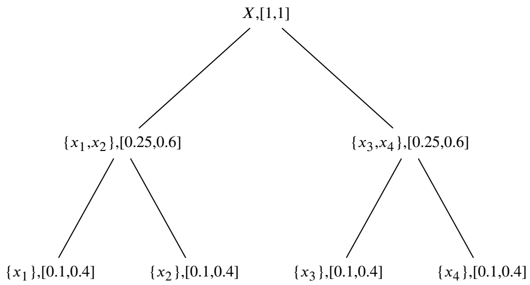


Fig. 1. Interval tree for Example 5.

$$\underline{P}(B|A) = \frac{\underline{P}'(A\cap B)}{\underline{P}'(A\cap B) + \overline{P}'(A^c\cap B)},\quad \overline{P}(B|A) = \frac{\overline{P}'(A\cap B)}{\overline{P}'(A\cap B) + \underline{P}'(A^c\cap B)}, \tag{16}$$

where $(\underline{P}', \overline{P}')$ are the intervals of the natural extension of S.

It is important to note that if S is defined for sets in $\mathcal{A}$, the conditional specification S_A should be complete, as it is generally not equivalent to the interval specification defined on the same set $\mathcal{A}$ with the intervals given by Expression (16) as the following example shows:

Example 4. Let us consider S on $X = \{x_1, x_2, x_3, x_4, x_5\}$ and the elementary interval specification given by $(\underline{p}, \overline{p})$ where $\underline{p}(x_i) = 0.1, \overline{p}(x_i) = 0.3$, $\forall x_i \in X$. If $A = \{x_1, x_2, x_3, x_4\}$, then

$$\underline{P}(\{x_i\}|A) = 1/9,\ \overline{P}(\{x_i\}|A) = 3/7,\ \forall i = 1,\dots,4.$$

Furthermore, $\underline{P}(\{x_1, x_2\}|A) = 1/4$. The natural extension of conditional elementary intervals to this set, however, would imply $\underline{P}'(\{x_1, x_2\}|A) = 2/9$ (using Expression (6)). In this way, the conditional intervals are not equivalent to these intervals that are restricted to the set $\mathcal{A}$ in which S is defined. If the conditional intervals are only computed for sets in $\mathcal{A}$, then an additional approximation is performed.

3. Hierarchical intervals

In this section, we generalize elementary probability intervals to specifications that can provide intervals for additional sets while retaining most of the properties of elementary interval specifications, and provide algorithms to check sure loss and compute natural extension.

It is said that $S = (\mathcal{A}, \underline{P}, \overline{P})$ is a hierarchical interval specification if and only if $\mathcal{A}$ can be structured as a tree T with a set of $\mathcal{A}$ labeling each of its nodes, in which the root is always associated with X, and that for each non-leaf node with set B, the sets associated with its children, $B_1 \dots, B_k$ are a partition of B. These sets are denoted by $Ch(B)$.

Example 5. Let us assume that $X = \{x_1, x_2, x_3, x_4\}$ and the specification given by the following sets of intervals is

$$\{x_i\} \to [0.1, 0.4],\ \ i = 1, 2, 3, 4$$

$$\{x_1, x_2\} \to [0.25, 0.6],\ \{x_3, x_4\} \to [0.25, 0.6]$$

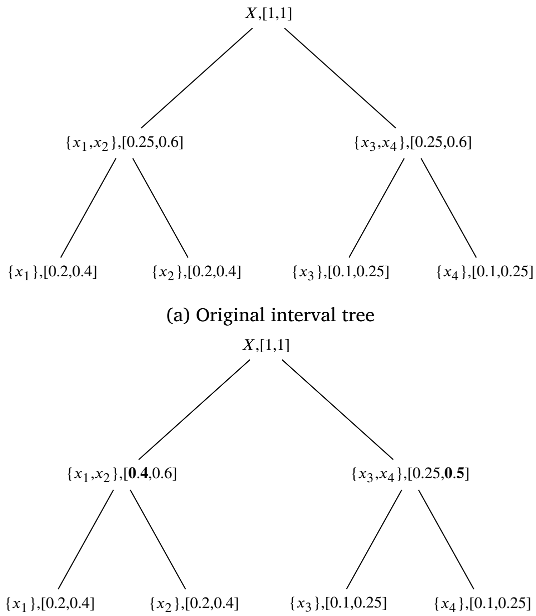
$$X \to [1, 1]$$

The associated interval tree is depicted in Fig. 1.

We will now provide an example of an interval specification which cannot be expressed as a hierarchical one.

Example 6. In order to provide an example of an interval specification which cannot be represented as a hierarchy, we must consider a frame with at least 4 elements. Any interval specification can be expressed as an elementary interval if $X = \{x_1, x_2, x_3\}$. Correspondingly, an interval $[\alpha, \beta]$ for $\{x_1, x_2\}$ is equivalent to an interval $[1-\beta, 1-\alpha]$ for its complementary set $\{x_3\}$ which is a unitary set.

Let us consider $X = \{x_1, x_2, x_3, x_4\}$ and the interval specification given by $[0, 0.5]$ for each unitary set, and $[0.2, 0.8]$ for any set with two elements. If we wish to reduce this to a hierarchical specification, we should include all the intervals for elementary events. The intervals for sets with 3 elements are not informative since they are the complementary set of singletons and these intervals can be obtained by duality. For sets of 2 elements, we can only keep 2 intervals for complementary events in order to obtain a hierarchical specification. Without loss of generality, let us assume that these sets are $\{x_1, x_2\}$ and $\{x_3, x_4\}$, both with intervals $[0.2, 0.8]$. With this



(a) Original interval tree

(b) Corrected interval tree with $(\underline{P}^*, \overline{P}^*)$ computed as in Proposition 4

Fig. 2. Checking sure loss in interval trees (Example 7).

specification, if we compute the natural extension for set $\{x_2, x_3\}$, we obtain the vacuous interval $[0,1]$, i.e. by using a hierarchy of intervals instead of the original ones, we lose information (wider intervals for some sets), and the original one cannot be specified by a hierarchy.

Since the root node always corresponds to the complete X set with lower and upper probabilities equal to 1, it can always be considered to be implicit in any hierarchical specification without the need to include it explicitly. On account of this, an elementary specification is always a hierarchical specification in which the root node has as children all the elementary events, $\{x_i\}$, for $x_i \in X$. There are, however, hierarchical specifications that are not equivalent to elementary specifications such as the one in Example 5 since the natural extension of the specification considering only the intervals on the elementary events would result in $\underline{P}'(\{x_1, x_2\}) = 0.2$, which is different from the value of 0.25 assigned by the hierarchical specification.

It is possible to check whether a hierarchical specification avoids sure loss in linear time, as the following result shows:

Proposition 4. *Let us consider S to be a hierarchical specification with an associated tree T. If the tree nodes are visited from the leaf nodes to the root and the following adjustment is performed on the inner nodes:*

$$\underline{P}^*(B) = \max\{\underline{P}(B), \sum_{B_i \in Ch(B)} \underline{P}^*(B_i)\}, \;\; \overline{P}^*(B) = \min\{\overline{P}(B), \sum_{B_i \in Ch(B)} \overline{P}^*(B_i)\},$$

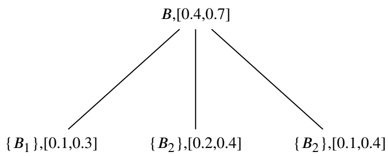
Then, S avoids sure loss if and only if for any B in tree T, $\underline{P}^*(B) \leq \overline{P}^*(B)$.

See proof on page 17.

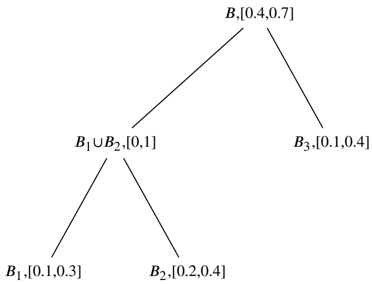
We can see that this computation is linear according to the size of the problem specification since it involves visiting and operating with the values of each node twice at most.

Example 7. Let us assume that $X = \{x_1, x_2, x_3, x_4\}$ and the specification by interval tree is depicted in Fig. 2a. The correction in Proposition 4 is applied in order to obtain the interval tree in Fig. 2b with P^*. Since $\underline{P}(\{x_1\}) = \underline{P}(\{x_2\}) = 0.2$, then the value $P(\{x_1, x_2\}) = 0.25$ has been changed to $\underline{P}^*(\{x_1, x_2\}) = 0.4$. In this case, as the new intervals always satisfy $\underline{P}^*(A) \leq \overline{P}^*(A), \forall A \in 2^X$, the initial assessment avoids sure loss.

The following result enables us to compute in linear time the partial natural extension of a hierarchical specification that avoids sure loss.



(a) Original node with 3 children



(b) After adding an intermediary node

Fig. 3. Example of the basic step in a binary tree transformation.

Proposition 5. *If we assume that S is a hierarchical specification with tree T that avoids sure loss, then the partial natural extension defined for nodes in T can be computed from the bounds $\underline{P}^*, \overline{P}^*$ defined in Proposition 4 with the following equations applied to every T node from root to leaves: if B is an inner node, then $\forall B_i \in Ch(B)$,*

$$\underline{P}'(B_i) = \max\{\underline{P}^*(B_i), \underline{P}'(B) - \sum_{\substack{B_j \in Ch(B)\\ j\neq i}} \overline{P}^*(B_j)\}$$

$$\overline{P}'(B_i) = \min\{\overline{P}^*(B_i), \overline{P}'(B) - \sum_{\substack{B_j \in Ch(B)\\ j\neq i}} \underline{P}^*(B_j)\}$$

See proof on page 17.

The complexity of this procedure is also linear to the size of the problem specification (number of nodes, m), but multiplying by the branching factor d of the tree $O(md)$, as this is the maximum number of times that a node is visited or that its values are used in operations. However, the complexity could be reduced to linear by considering the fact that a hierarchical specification can always be transformed into a binary specification (two children at most for each node), with a linear increase in the number of nodes. The transformation is based on the following basic step: if node B has children $B_1, \ldots, B_k$, then we can combine two child nodes, for example B_1 and B_2, to include an intermediary node (parent of both) with the set $B_1 \cup B_2$ and to add a trivial interval $[0,1]$ for this set. This set will now have two children, B_1, B_2, each one of them with their original specifications. A simple example can be seen in Fig. 3. The resulting tree is not coherent, but coherence will be achieved once the previous procedure has been applied. This reduces the branching by 1 by adding an additional node. As this can only be applied once per node, the number of new nodes is limited to the number of original nodes, and the new tree can be at most double the size of the original. We can now check coherence and its complexity is linear in the binary tree. We can see that checking coherence for elementary intervals using Expression (5) is quadratic in terms of the problem size (number of elements in X). However, if we express the representation as a hierarchical specification and transform it into a binary tree, then the complexity would be linear since the additional nodes (with trivial specifications) serve as a cache to avoid repetitions in the computations.

Example 8. Let us assume that $X = \{x_1, x_2, x_3, x_4\}$ and intervals $(\underline{P}^*, \overline{P}^*)$ in Fig. 4a. The partial natural extension is given by the intervals $(\underline{P}', \overline{P}')$ in Fig. 4b. We can see how $\underline{P}'(\{x_1\}) = 0.3$ since $\underline{P}^*(\{x_1, x_2\}) = 0.4$ and $\overline{P}^*(\{x_2\}) = 0.1$. At the same time, $\overline{P}'(\{x_3, x_4\} = 0.6$ since $\overline{P}^*(X) = 1$ and $\underline{P}^*(\{x_1, x_2\}) = 0.4$.

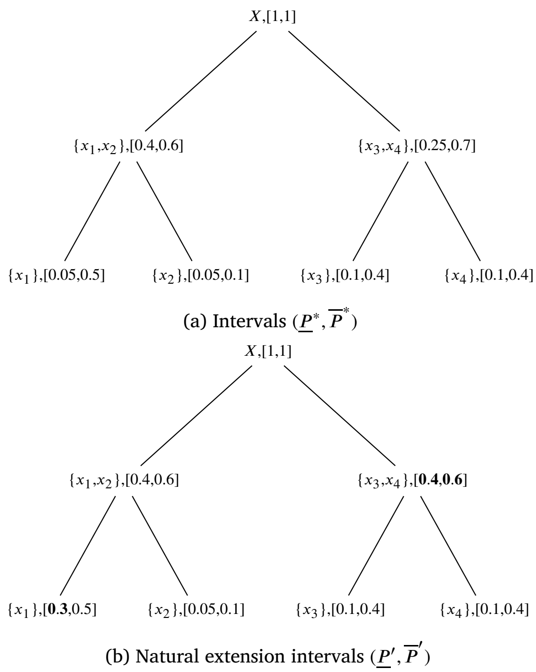
Henceforth, all hierarchical specifications are assumed to be coherent. If any specification is not coherent, then it can be always replaced by its partial natural extension which can be efficiently computed with the previous propositions. We will now prove a result to show that the natural extension can also be efficiently computed for any set $A \in 2^X$ using Algorithm 1 which will return two values $(\overline{P}'(A), 1)$. The second value is always 1 but since this is a recursive algorithm, this value will be meaningful in intermediate steps.

Proposition 6. *If S is a coherent hierarchical specification with the associated tree T, then Algorithm 1 returns the pair of values (α, β) where α is the upper natural extension, $\overline{P}'(A)$, for any $A \in 2^X$ and $\beta = 1$.*

See proof on page 18.

The computational complexity of this algorithm is linear in relation to the number of nodes m since each node is visited at most once and the values (α, β) of each node are used once in the parent node computations.

In order to compute $\underline{P}'(A)$, we can apply the same algorithm by considering that $\underline{P}'(A) = 1 - \overline{P}'(A^c)$.



(a) Intervals $(\underline{P}^*, \overline{P}^*)$

(b) Natural extension intervals $(\underline{P}', \overline{P}')$

Fig. 4. Checking partial natural extension (Example 8).

Algorithm 1 Computing the Natural Extension.

Require: A, a set for computing its natural extension
Require: T, the tree of a hierarchical specification, such that A is included in the root of T
Ensure: (α, β), where α is the upper natural extension of A, $\overline{P}'(A)$, and β is 1
1: **procedure** NATURALEXTENSION(A,T)
2: $H \leftarrow Root(T)$
3: **if** $H \cap A = \emptyset$ **then**
4: **return** $(0, \underline{P}(H))$
5: **end if**
6: **if** H is a leaf **then**
7: **return** $(\overline{P}(H), \overline{P}(H))$
8: **end if**
9: $\{C_i\}_{i=1}^{l} = Ch(H)$
10: **for** i = 1,…,l **do**
11: $(\alpha_i, \beta_i) \leftarrow$ NATURALEXTENSION$(A \cap C_i, C_i)$
12: **end for**
13: $\beta \leftarrow \min\{\max\{\sum_{i=1}^{l} \beta_i, \underline{P}(H)\}, \overline{P}(H)\}$
14: $\alpha \leftarrow \sum_{i=1}^{l} \alpha_i + \beta - \max\{\sum_{i=1}^{l} \beta_i, \underline{P}(H)\}$
15: **return** (α, β)
16: **end procedure**

Example 9. In Fig. 5, we have a hierarchical specification with the intervals (α, β) computed in each node of the tree for $\overline{P}'(\{x_1, x_3\})$ (lower part of each node in the figure). The final obtained value is $\overline{P}'(\{x_1, x_3\}) = 0.85$.

Let us now assume that we have computed $\overline{P}'(A)$ using Algorithm 1 and that we want to compute $\overline{P}'(A \cup \{x\})$ where $x \notin A$, i.e. the natural extension after adding a value x to A. This computation is relevant since it will be instrumental in showing that a hierarchical specification is always an order-2 capacity in Proposition 8.

Let us call $\alpha(H), \beta(H)$ the values computed in order to obtain the upper natural extension of A, assuming that $\alpha(H) = 0$ and $\beta(H) = \underline{P}(H)$ if $H \cap A = \emptyset$. We can see that these would be the values computed by Algorithm 1 if we apply the equations shown in lines 13 and 14 until the leaf nodes in which $H \cap A = \emptyset$ is fulfilled. We also consider that $\alpha^x(H), \beta^x(H)$ are the values computed after adding x and applying the same algorithm to obtain the natural extension of $A \cup \{x\}$. Since it is evident that $\alpha(H) = \alpha^x(H)$ and $\beta(H) = \beta^x(H)$ if $x \notin H$, it is only necessary to compute the new values for the events in the path from the leaf node $\{x\}$ to the root. Let us assume that these sets are $H_0, \ldots, H_m$, where $H_0 = \{x\}$ and $H_m = X$.

If we call $\delta(H_j) = \alpha^x(H_j) - \alpha(H_j)$ and $\phi(H_j) = \beta^x(H_j) - \beta(H_j)$, then $\delta(H_0) = \overline{P}(\{x\})$ and $\phi(H_0) = \overline{P}(\{x\}) - \underline{P}(\{x\})$.

It can be seen how if for an index j, $\phi(H_j) = 0$, then we have $\beta^x(H_k) = \beta(H_k)$ and $\delta(H_k)$, is constant for $k \geq j$. Therefore, $\overline{P}'(A \cup \{x\}) = \overline{P}(A) + \delta(H_j)$.

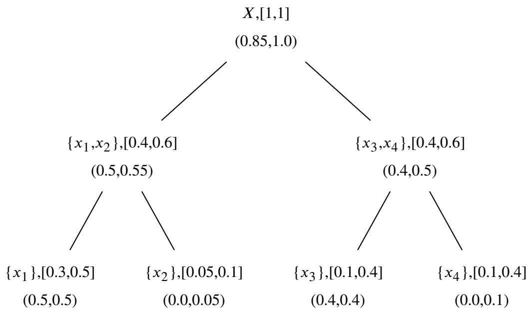


Fig. 5. Natural extension $\overline{P}'(\{x_1,x_3\})$ according to Algorithm 1.

The values of δ and ϕ can be computed recursively for $H_0,\ldots,H_m$ starting at H_0. In order to compute the values for H_j, we consider the following two values:

$$\epsilon_1 = \min\{\max\{B(H_j)+\phi(H_{j-1})-\overline{P}(H_j),0\},\phi(H_{j-1})\}$$
$$\epsilon_2 = \min\{\max\{\underline{P}(H_j)-B(H_j),0\},\phi(H_{j-1})\}$$

where $B(H_j)=\sum_{C_i\in Ch(H_j)}\beta(C_i)$ and $\beta(C_i)$ is the value computed in Algorithm 1 when computing $\overline{P}'(A)$.

The updating can then be performed using the following expressions:

$$\delta(H_j)=\delta(H_{j-1})-\epsilon_1,\quad \phi(H_j)=\phi(H_{j-1})-\epsilon_1-\epsilon_2.$$

A detailed description of the process is provided in Algorithm 2. Although it is not specified in the algorithm, if after computing $\overline{P}'(A\cup\{x\})$ we want to compute $\overline{P}'(A\cup\{x,y\})$, we can apply the same algorithm to y but now the $B(H)$ values should be updated in nodes $(H_0,\ldots,H_m)$ using the expression $B(H_j)\leftarrow B(H_j)+\phi(H_j)$, when $\phi(H_j)>0$.

Algorithm 2 Extending the Natural Extension.

Require: T, the tree of a hierarchical specification, such that A is included in the root of T
Require: $\overline{P}'(A)$, the natural extension of A computed with Algorithm 1
Require: $B(H)$, the value $B(H)=\sum_{C_i\in C(H)}\beta(C_i)$ of $\overline{P}'(A)$ computations with Algorithm 1
Require: x, a value in $X\setminus A$
Ensure: $\overline{P}'(A\cup\{x\})$, the natural extension of $A\cup\{x\}$.

procedure EXTENDNATURALEXTENSION(T,$\overline{P}'(A)$,x)
  Compute the nodes $(H_0,\ldots,H_m)$ in the path from leaf node $\{x\}$ to the root.
  $\delta(H_0)\leftarrow\overline{P}(\{x\})$
  $\phi(H_0)\leftarrow\overline{P}(\{x\})-\underline{P}(\{x\})$
  for $j=1$ to m **do**
    $\epsilon_1\leftarrow\min\{\max\{B(H_j)+\phi(H_{j-1})-\overline{P}(H_j),0\},\phi(H_{j-1})\}$
    $\epsilon_2\leftarrow\min\{\max\{\underline{P}(H_j)-B(H_j),0\},\phi(H_{j-1})\}$,
    $\delta(H_j)=\delta(H_{j-1})-\epsilon_1$
    $\phi(H_j)\leftarrow\phi(H_{j-1})-\epsilon_1-\epsilon_2$
    if $\phi(H_j)=0$ **then**
      Break
    end if
  end for
  return $\overline{P}'(A)+\delta(H_j)$
end procedure

Proposition 7. *If in Algorithm 2 $B(H)=\sum_{C_i\in Ch(H)}\beta(C_i)$, where $\beta(C_i)$ are the values computed in Algorithm 1 when computing $\overline{P}'(A)$, then it returns $\overline{P}'(A\cup\{x\})$.*

See proof on page 19.

The complexity of this algorithm is linear in the size of the path from the root to node $\{x\}$. If the tree is balanced, this length is logarithmic in the number of tree nodes.

Example 10. Let us consider the case of Example 9 in which the natural extension of $A=\{x_1,x_3\}$ was computed and that we now want to compute the natural extension of $A\cup\{x_4\}$. The sequence of sets for which δ and ϕ must be computed is $(\{x_4\},\{x_3,x_4\},\{x_1,x_2,x_3,x_4\})$. These values are computed in Fig. 6. The values are in the lower part of each node: $(\delta(H_j),\phi(H_j))$. As $\delta(X)=0.1$, the value $\overline{P}'(\{x_1,x_3,x_4\})=\overline{P}'(\{x_1,x_3\})+0.1=0.95$. This value is correct since $\underline{P}'(\{x_2\})=0.05$.

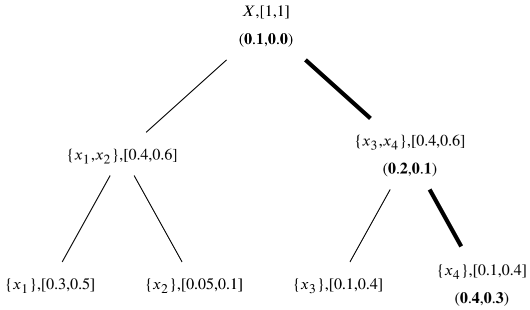


Fig. 6. (δ,ϕ) values to compute $\overline{P}'(A\cup\{x_4\})$ where $A=\{x_1,x_3\}$.

4. Hierarchical intervals and order-2 capacities

Proposition 8. *If S is a hierarchical specification, then it is always an order-2 capacity.*

See proof on page 20.

Since a hierarchical specification is an order-2 capacity, it is easy to compute the conditional intervals using Expression (16). The necessary values can be computed with Algorithm 1. When a lower probability is needed, this is computed by duality: $\underline{P}'(C)=1-\overline{P}'(C^c)$.

The extreme probabilities are also associated with permutations σ on $\{1,2,\dots,n\}$ and can be computed using Expression (9).

Algorithm 3 Computing the Extreme Probability.

Require: T, the tree of a coherent hierarchical specification on X
Require: σ, a permutation on X
Ensure: p_σ, the extreme probability associated with this permutation
1: **procedure** EXTREMEPROBABILITY(T,σ)
2: Let p_σ be a mapping from X to $[0,1]$
3: **for** $j=1$ to n **do**
4: $p_\sigma(x_{\sigma(i)})\leftarrow \underline{P}(x_{\sigma(i)})$ in tree specification T
5: $\overline{P}(x_{\sigma(i)})\leftarrow \underline{P}(x_{\sigma(i)})$ and add this restriction to T
6: Compute intervals $(\underline{P}^*,\overline{P}^*)$ ▷ Proposition 4
7: Compute natural extension intervals $(\underline{P}',\overline{P}')$ ▷ Proposition 5
8: $(\underline{P},\overline{P})\leftarrow(\underline{P}',\overline{P}')$
9: **end for**
10: **end procedure**

In order to compute these extreme probabilities, we can follow Algorithm 3. Since $\underline{P}'$ is an order-2 capacity, this procedure computes the values of the extreme probability given in Expression (9). In Step 4, the probability value of $x_{\sigma(i)}$ is fixed to its lower value and the upper probability is then set to this lower value. In Steps 6 and 7, this new specification is propagated to the rest of the tree, making it coherent.

The complexity of this algorithm is quadratic in relation to the number of nodes of the tree (m) since the number of elements in X is proportional to the number of nodes (at most, the number of leaves) and the complexity of each iteration of the loop in lines 3-9 of the algorithm is linear.

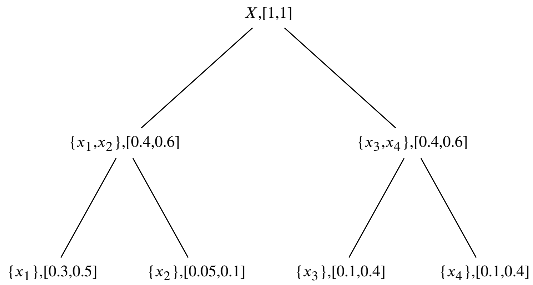
Example 11. Let us assume the coherent hierarchical specification of Fig. 7a and the permutation given by $\sigma(1)=3,\sigma(2)=2,\sigma(3)=4,\sigma(4)=1$.

First, x_3 must take its lower value $p_\sigma(x_3)=0.1$. The upper interval of x_3 is then set to this value 0.1, and new intervals are computed for the other values according to Propositions 4 and 5. The result is shown in Fig. 7b. It can be observed how the upper probability of $\{x_3,x_4\}$ has changed to 0.5 (upper propagation when computing $(\underline{P}^*,\overline{P}^*)$ intervals) and that the lower probability of $\{x_1,x_2\}$ has increased to 0.5 (lower propagation for intervals $(\underline{P}',\overline{P}')$). This is a consequence of the fact that the upper probability of its complementary set is 0.5. Similarly, the lower probability of $\{x_1\}$ has increased to 0.4 and the lower probability of $\{x_4\}$ to 0.3.

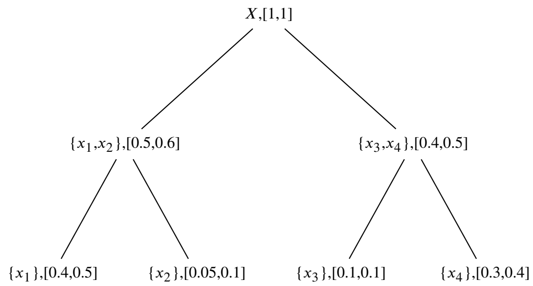
We then set $p_\sigma(x_2)$ to its lower interval value $p_\sigma(x_2)=0.05$, the upper probability is set to its lower probability and the new interval is propagated again to compute the new coherent intervals. The result is shown in Fig. 7c.

The probability of X_4 is then fixed to its lower probability $p(x_4)=0.35$ and after propagating, the probability of x_1 is fixed to 0.5.

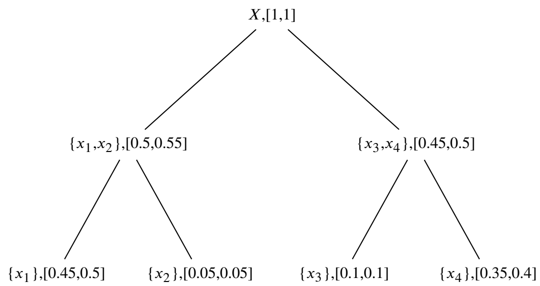
Finally, if we have a coherent hierarchical specification S on X and f is a gamble, i.e. a mapping $f:X\rightarrow\mathbb{R}$, the lower prevision of f, also denoted by $\underline{P}(f)$, is the minimum of $P(f)=\sum_{x\in X}p(x)f(x)$ for every probability $P\in\mathcal{M}(S)$. Since S is an order-2 capacity,



(a) Original Coherent Specification $(\underline{P}, \overline{P})$



(b) Probability intervals after setting $p_\sigma(x_3) = 0.1$



(c) Probability intervals after setting $p_\sigma(x_2) = 0.05$

Fig. 7. Computing the Extreme Probabilities (Example 11).

this computation is very simple and we only need to compute a permutation σ such that if $f(x_i) \leq f(x_j)$, then $\sigma(i) \leq \sigma(j)$ and to make $\underline{P}(f) = p_\sigma(f)$. This is due to the fact that the lower prevision in an order-2 capacity is the mathematical expectation with respect to the extreme probability which is obtained with a permutation that is co-monotone with the gamble [6]. With some additional detail, the procedure is as follows:

- Determine a permutation σ over X so that $\sigma(x_i) < \sigma(x_j)$ if $f(x_i) < f(x_j)$. If $f(x_i) = f(x_j)$, the order of $\sigma(x_i)$ and $\sigma(x_j)$ is chosen in an arbitrary way.
- Compute p_σ with Algorithm 3. Make $\underline{P}(f) = p_\sigma(f) = \sum_{x \in X} f(x) p_\sigma(x)$.

$\overline{P}(f)$ can be computed using duality $\overline{P}(f) = -\underline{P}(-f)$ or directly by computing a permutation σ such that $f(x_i) \geq f(x_j)$ then $\sigma(i) \leq \sigma(j)$ and making $\overline{P}(f) = p_\sigma(f)$.

5. Relationships with existing approaches

The most similar approach to the one presented in this paper is the elementary interval model [2]. This corresponds to a hierarchical specification with only one level below the root, and on this, intervals are provided for the singletons. One example of this is shown in Fig. 8.

If we compare this specification with the one corresponding to Fig. 1, we can see that the lower probability $\overline{P}'(\{x_1, x_2\}) =$ 0.2, whereas in Fig. 1, with the same initial intervals for elementary events, the lower bound is 0.4, which is more informative. Hierarchical specifications are, therefore, a proper generalization of elementary intervals.

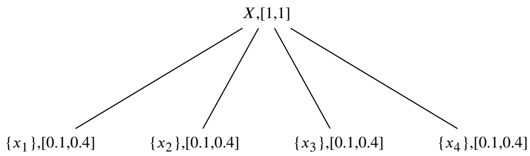


Fig. 8. Interval tree for elementary intervals.

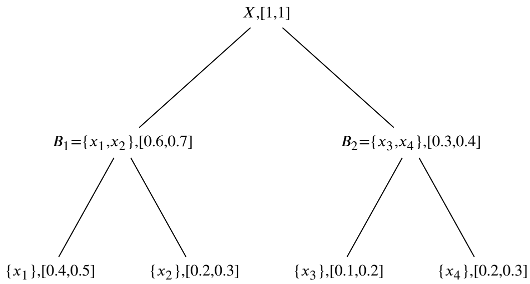


Fig. 9. Interval hierarchical tree for Example 12.

Specifications with only one level on the tree in which leaves are not necessarily singletons but a general partition of X: $B_1, \ldots, B_k$ have been considered by Walley [25, Page 198]. The conditions for avoiding sure loss and checking coherence are given for this case and are a direct generalization of the ones for elementary intervals but are a particular case of the algorithms in this paper.

Much has been written and published on the topic of avoiding sure loss and checking coherence for general specifications and also conditional probability intervals [1,13,20,27]. However, they are based on solving linear programming problems which are less efficient. The approach in [21] is efficient but it is limited to syllogistic deductions involving only three probability intervals.

Although hierarchical specifications of imprecise probabilities are discussed in [7,29], both these referenced models are based on conditional previsions [25, Ch. 6]. The difference is that if we have a node B with children B_1, B_2, (B_1, B_2 are a partition of B), then in the conditional approach, the information in the children is about conditional events $P(B_i|B)$, whereas in our approach, we provide intervals for unconditional events $P(B_i)$. In the conditional approach, coherence is simpler as there is no interaction between the coherence of different levels. If we have interval probabilities for $P(B)$ and $P(B_1|B)$ and $P(B_2|B)$, we only need to check the separated coherence of the conditional specifications, but in the case of our hierarchical specifications, if we have the interval $[0.1, 0.2]$ for both B_1 and B_2, and we include the interval $[0.15, 0.5]$ for B, the resulting specification is not coherent. The following example shows a more complex situation:

Example 12. Let us assume that $X = \{x_1, x_2, x_3, x_4\}$ and that we have a set of observations for these values from which we want to estimate the probabilities $P(x_i)$. If we have observed that $n_1 = 4, n_2 = 2, n_3 = 1, n_4 = 2$, where n_i is the number of observations for x_i, and the imprecise Dirichlet model (IDM) [26] is applied with $s = 1$, then the following intervals are obtained for the different values:

$$\{x_1\} \to [0.4, 0.5],\ \{x_2\} \to [0.2, 0.3],\ \{x_3\} \to [0.1, 0.2],\ \{x_4\} \to [0.2, 0.3]$$

Once again, we can apply the IDM at a different level: for example, to the set $\{B_1, B_2\}$ where $B_1 = \{x_1, x_2\}, B_2 = \{x_3, x_4\}$, with frequencies $n(B_1) = 6, n(B_2) = 3$, and the resulting intervals for B_1 and B_2 are

$$B_1 \to [0.6, 0.7],\ B_2 \to [0.3, 0.4]$$

We can see how this application does not contain any new information as these intervals could be obtained by natural extension from the original ones. In any case, in Fig. 9 we can see the associated hierarchical specification containing these two intervals.

If we repeat the two-level application of the IDM but using a conditional approach as in [7,29], then the result can be seen in Fig. 10. We can see how the first level is identical but the second level is different: for B_1 children, the conditional probabilities $\{x_1\}|B_1$ and $\{x_2\}|B_1$ are computed. As the frequencies are $n_1 = 4, n_2 = 2$ given B_1, the following intervals are obtained when the IDM is applied:

$$\{x_1\}|B_1 \to [4/7, 5/7],\ \{x_2\}|B_1 \to [2/7, 3/7]$$

If once again, we apply the conditioned IDM model to B_2, we obtain

$$\{x_3\}|B_2 \to [1/4, 2/4],\ \{x_4\}|B_2 \to [2/4, 3/4]$$

We can observe how the interval hierarchical specification of Fig. 9 is more informative than the hierarchical conditional specification of Fig. 10. If the conditional probabilities $\{x_i\}|B_j$ are computed in the tree of Fig. 9, we obtain the same conditional

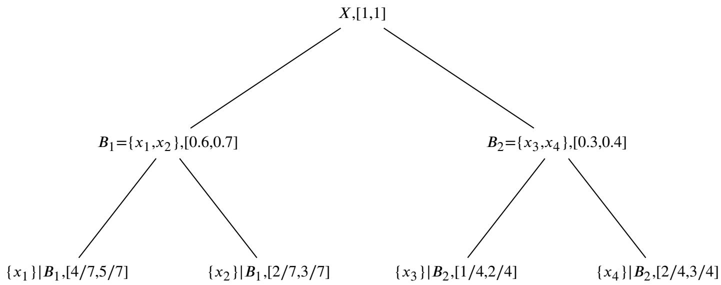


Fig. 10. Interval conditional hierarchical tree for Example 12.

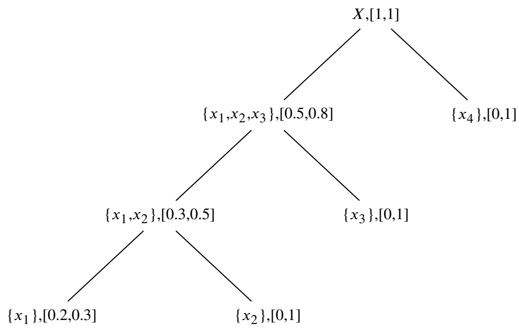


Fig. 11. Interval hierarchical tree for a p-box.

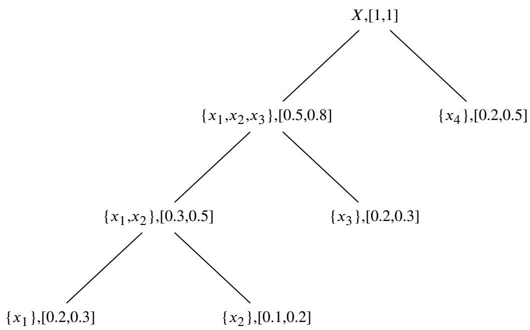


Fig. 12. Interval hierarchical tree for a p-box after natural extension.

probabilities as in Fig. 10. However, the lower probability of $\{x_1\}$ in the conditional tree is $0.6 \times 4/7 = 24/7$, which is strictly lower than the lower probability of $\{x_1\}$ in the hierarchical specification (0.4).

Former example does not claim that interval hierarchical specifications can always represent hierarchical conditional specifications without losing information. They are different and complementary models, and the most appropriate one will depend on the particular situation we are representing. In fact, in the example, we start with a sample that provides a joint information about X and therefore, in this case the interval specification is more appropriate than the conditional one.

Another related approach is given in [5], but in that paper, it is studied when a partial assignment can be extended to different types of measures, but not always considering the natural extension as in our case.

Another model that is included in the hierarchical specification is generalized p-boxes [8] which are receiving increasing interest in engineering applications [11]. A generalized p-box is given by an interval specification for a set of nested sets. For example, if $X = \{x_1, x_2, x_3, x_4\}$, we may have:

$$\{x_1\} \to [0.2, 0.3],\ \{x_1, x_2\} \to [0.3, 0.5],\ \{x_1, x_2, x_3\} \to [0.5, 0.8],\ X \to [1, 1]$$

This is a hierarchical specification, where the initial intervals are given in Fig. 11 and the intervals after computing natural extension are given in Fig. 12.

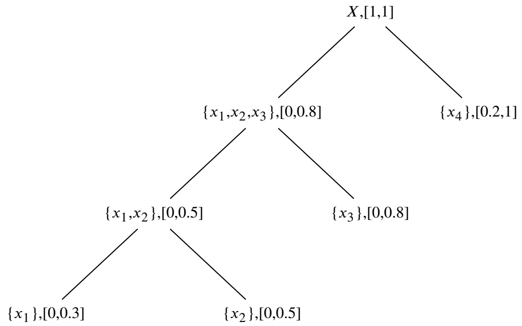


Fig. 13. Interval hierarchical tree for a possibility measure.

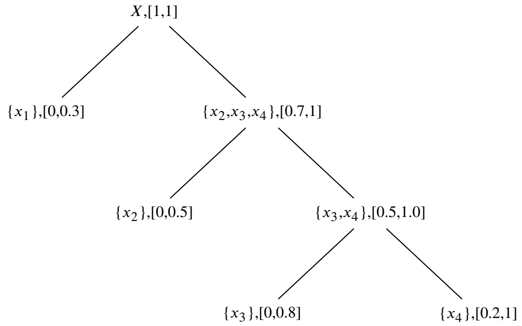


Fig. 14. Interval hierarchical tree for a necessity measure.

Possibility and necessity measures [10] are also particular cases of p-boxes in the finite case [24]. They correspond to nested specifications where either the lower values are always 0 (possibility specification) or the upper values are always 1 (necessity specification). By way of example, we may consider

$$\{x_1\} \to [0, 0.3],\ \{x_1, x_2\} \to [0, 0.5],\ \{x_1, x_2, x_3\} \to [0, 0.8],\ X \to [1, 1]$$

Its natural extension can be seen in Fig. 13.

Equivalently, the same information (identical credal set) can be represented by the necessity values

$$\{x_4\} \to [0.2, 1],\ \{x_3, x_4\} \to [0.5, 1],\ \{x_2, x_3, x_4\} \to [0.7, 1],\ X \to [1, 1]$$

The associated hierarchical specification is depicted in Fig. 14.

This example also shows how an interval hierarchical specification is not unique for a given credal set since we have provided two different representations for the same credal set.

However, hierarchical specifications offer new possibilities such as mixing interval and possibilistic specifications, as the following example shows.

Example 13. Let us assume that $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, that $B_1 = \{x_1, x_2, x_3\}, B_2 = \{x_4, x_5, x_6\}$ and that we have a set of observations about the membership to B_1, B_2, with these observations for the two sets being $n_1 = 3, n_2 = 6$. By applying the IDM, we could obtain the following intervals:

$$\{x_1, x_2, x_3\} \to [0.3, 0.4],\ \{x_4, x_5, x_6\} \to [0.6, 0.7]$$

At the same time, we have two experts and each one knows about each set B_1 and B_2 and they provide the following nested specifications for both sets:

$$\{x_1\} \to [0, 0.2],\ \{x_1, x_2\} \to [0, 0.3],\ \{x_1, x_2, x_3\} \to [0, 0.6]$$

$$\{x_4\} \to [0, 0.5], \{x_4, x_5\} \to [0, 0.8],\ \{x_4, x_5, x_6\} \to [0, 0.9]$$

The resulting specification is shown in Fig. 15.

Although the specification is not coherent, it can be efficiently extended to the coherent specification of Fig. 16.

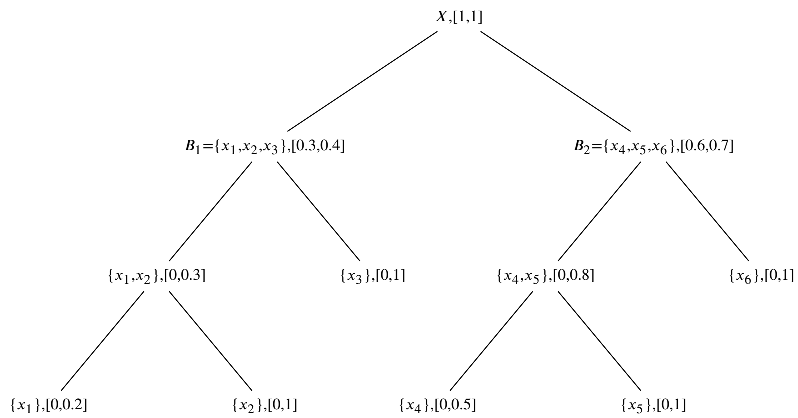


Fig. 15. Hierarchical tree combining interval and possibilistic information for Example 13.

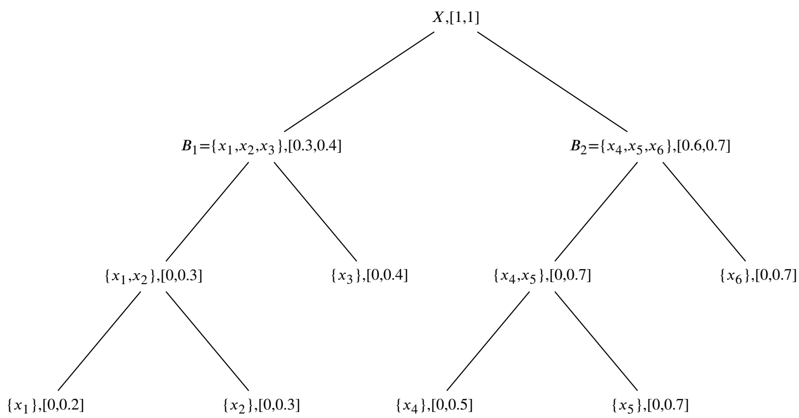


Fig. 16. Natural extension of a hierarchical tree combining interval and possibilistic information for Example 13.

6. Conclusions and future work

In this work, we have generalized the computations associated with elementary intervals to the more general case of hierarchical intervals. We have provided efficient algorithms to check coherence and compute natural extension. Previous algorithms for elementary intervals are a particular case of the algorithms provided in this paper. Hierarchical specifications also include other well-known models in literature, but there are limitations in the type of interval that can be provided (intervals for disjoint or nested sets), and therefore models such as belief functions cannot always be represented by hierarchical specifications. A procedure to overcome this limitation will be to study convex combinations of hierarchical specifications: if $\{(\underline{P}'_i, \overline{P'}_i)\}_{i=1}^k$ are such that each $(\underline{P}'_i, \overline{P'}_i)$ is the natural extension of a hierarchical specification and $\{\alpha_i\}_{i=1}^k$ are real non-negative numbers such that $\sum_{i=1}^k \alpha_i = 1$, the convex combination is given by lower and upper intervals $(\sum_{i=1}^k \alpha_i \underline{P}'_i, \sum_{i=1}^k \alpha_i \overline{P'}_i)$. Our objective will be to extend algorithms to these convex combinations. The convex combinations of hierarchical specifications include the particular case of belief functions. If we have a belief function Bel with focal elements $A_1, \ldots, A_l$, then for each A_i we can define the interval specification given by the interval $[1,1]$ for A and the interval $[0,0]$ for the complementary set (the belief function focused on A_i). If we use (Bel_i, Pl_i) to identify its associated pair of belief-plausibility functions ($Bel_i(B) = 1$, if $A_i \subseteq B$ and 0 otherwise), then it is clear that $(Bel, Pl) = \sum_{i=1}^l m(A_i)(Bel_i, Pl_i)$ and so the convex combinations of hierarchical specifications include belief functions as a particular case. We can easily verify that a convex combination of hierarchical specifications is always an order-2 capacity. An important question to answer is whether there is an example of an order-2 capacity which cannot be expressed as a convex combination of hierarchical interval specifications. If this is not the case, then hierarchical specifications would be the building blocks of order-2 capacities and play a similar role to the focal elements for belief functions.[2] This represents a very important research question for us as it would provide a better and deeper understanding of the nature of order-2 capacities and would provide a tool to represent them efficiently.

[2] As one of the reviewers of this paper mentioned, the paper by Quaeghebeur and De Cooman [22], which studies extreme lower previsions, can provide valuable tools.

The number of extreme probabilities of a hierarchical specification is limited by $n!$ as they are particular cases of order-2 capacities, but in most cases, as with elementary probability intervals, this number is lower. Theoretical and experimental studies for determining this number of extreme probabilities represent an extremely interesting area of research for the future.

Another possible use of hierarchical specifications is as an approximation tool. There are many situations in which general interval specifications are transformed into elementary probability intervals with loss of information but improved efficiency [23]. Hierarchical specifications can serve for this purpose, but with less information loss. An important optimization task would be to determine the hierarchical structure with a minimum loss of information for a general set of interval probabilities.

Finally, we can consider the extension of marginal hierarchical specifications to a joint representation and under which conditions (dependence structure) it is possible to obtain again a hierarchical specification, similarly to [19] with p-boxes and other models.

CRediT authorship contribution statement

Serafín Moral: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Andrés Cano:** Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Manuel Gómez-Olmedo:** Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Serafin Moral reports financial support was provided by Spanish Ministry of Education and Science under project PID2019-106758GB-C31.

Data availability

No data was used for the research described in the article.

Acknowledgements

This research was supported by the Spanish Ministry of Education and Science under project PID2019-106758GB-C31. We are very grateful to the anonymous reviewers for their useful comments and suggestions that helped improve the first version of this paper.

Appendix A. Proofs of main results

Proposition 2. *If we have a coherent set of elementary probability intervals S given by lower and upper probability distributions $(\underline{p}, \overline{p})$ such that $\sum_{y\in X} \underline{p}(y) < 1$, then the set of the extreme probabilities of their associated credal set $m(S)$ is given by the probability distributions $p_{(A,x)}$, where (A, x) is the set of pairs in which A is a subset $A \subseteq X$ and x is an element $x \in X \setminus A$ such that*

$$\underline{P}'(A) = \sum_{y\in A} \underline{p}(y),\ \underline{P}'(A \cup \{x\}) > \sum_{y\in A\cup\{x\}} \underline{p}(y). \tag{10}$$

In this case, $p_{(A,x)}$ is given by

$$p_{(A,x)}(y) = \begin{cases} \underline{p}(y) & \text{if } y \in A \\ 1 - \sum_{z\notin(A\cup\{x\})} \overline{p}(z) - \sum_{z\in A} \underline{p}(z) & \text{if } y = x \\ \overline{p}(y) & \text{if } y \in X \setminus (A \cup \{x\}) \end{cases} \tag{11}$$

Proof of Proposition 2. The proof is based on demonstrating that the probability distribution $p_{(A,x)}$ is the same as p_σ for any σ such that

$$x_{\sigma(i)} \in A, \text{ if } i < k, \text{ and } x_{\sigma(l)} \in X \setminus (A \cup \{x\}), \text{ if } i > k \tag{A.1}$$

where k is the number of elements of $A \cup \{x\}$.

According to Expression (6) for computing $\underline{P}'(B)$ and condition (10), we have that $\underline{P}'(A) = \sum_{y\in A} \underline{p}(y)$ and $\underline{P}'(A \cup \{x\}) = 1 - \sum_{y\notin A\cup\{x\}} \overline{p}(y)$.

It can be seen that the set function $f(B) = \sum_{y\in B} \underline{p}(y) - (1 - \sum_{x\notin B} \overline{p}(y))$ is not increasing in B. In fact, $f(B \cup \{z\}) - f(B) = \underline{p}(z) - \overline{p}(z) \leq 0$, as the set of probability intervals is coherent. Therefore, $\underline{P}'(C) = \sum_{y\in C} \underline{p}(y)$ for any $C \subseteq A$ and $\underline{P}'(D) = 1 - \sum_{y\notin D} \overline{p}(y)$ for any D, such that $A \cup \{x\} \subseteq D$.

For any σ satisfying (A.1), therefore, $\underline{P}'(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}) = \sum_{j=1}^{i} \underline{p}(x_{\sigma(j)})$ if $x_{\sigma(i)} \in A$ and $\underline{P}'(\{x_{\sigma(1)}, \ldots, x_{\sigma(l)}\}) = 1 - \sum_{j=l+1}^{n} \overline{p}(\sigma(x_{\sigma(j)})$ if $x_{\sigma(l)} \notin A \cup \{x\}$.

If we apply Expression (9), we obtain, $p_\sigma(x_{\sigma(i)}) = \underline{p}(x_{\sigma(i)})$ if $x_{\sigma(i)} \in A$ and $p_\sigma(x_{\sigma(l)}) = \overline{p}(x_{\sigma(l)})$ if $x_{\sigma(l)} \notin A \cup \{x\}$, i.e. the same expression as in (11) for $y \in X \setminus \{x\}$. Since the two probability distributions that assign the same value for $n-1$ elements in a set of n elements are equal, $p_\sigma = P_{(A,x)}$.

In order to conclude the proof, we must show that for each permutation p_σ, there is a pair (A, x) that satisfies the conditions in (10). It is necessary for one of the pairs $(\{x_{\sigma(1)}, \dots, x_{\sigma(i)}\}, x_{\sigma(i+1)})$ $(i = 1, \dots, n-1)$ to satisfy this since we have $\underline{P}'(\{x_{\sigma(1)}\}) = \underline{p}(x_{\sigma(1)})$ and $\underline{P}'(X) = 1 > \sum_{y \in X} \underline{p}(y)$. More precisely, (A, x) is the pair $(\{x_{\sigma(1)}, \dots, x_{\sigma(i)}\}, x_{\sigma(i+1)})$ given by the minimum value of i satisfying $\underline{P}'(\{x_{\sigma(1)}, \dots, x_{\sigma(i)}, x_{\sigma(i+1)}\}) > \sum_{l=1}^{i+1} \underline{p}(x_{\sigma(l)})$.

Proposition 4. *Let us consider S to be a hierarchical specification with an associated tree T. If the tree nodes are visited from the leaf nodes to the root and the following adjustment is performed on the inner nodes:*

$$\underline{P}^*(B) = \max\{\underline{P}(B), \sum_{B_i \in Ch(B)} \underline{P}^*(B_i)\}, \quad \overline{P}^*(B) = \min\{\overline{P}(B), \sum_{B_i \in Ch(B)} \overline{P}^*(B_i)\},$$

Then, S avoids sure loss if and only if for any B in tree T, $\underline{P}^*(B) \leq \overline{P}^*(B)$.

Proof of Proposition 4. We first prove that if for any B in tree T, $\underline{P}^*(B) \leq \overline{P}^*(B)$, then the specification avoids sure loss. It can be seen how after adjustment, intervals $[\underline{P}^*(B), \overline{P}^*(B)]$ are narrower than the original ones $[\underline{P}(B), \overline{P}(B)]$. Therefore, if the new intervals avoid sure loss, then so will the original ones.

We can also observe that if for any leaf B on T where B is not an elementary set, if we add a child to this node for each $x \in B$ with set $\{x\}$ with interval $[0, 1]$, the tree is extended so that every leaf is elementary and with no effect on the adjustment. We can, therefore, assume that all the leaves are elementary. In such conditions, we can select a probability measure starting at the root node, i.e. by assigning $P(X) = 1$ and then assigning probabilities to every node from root to leaves, taking into account that for every inner node B the new intervals satisfy

$$\sum_{B_i \in Ch(B)} \underline{P}^*(B_i) \leq \underline{P}^*(B) \leq \overline{P}^*(B) \leq \sum_{B_i \in Ch(B)} \overline{P}^*(B_i).$$

Since the assigned value $P(B)$ belongs to the interval $[\underline{P}(B), \overline{P}(B)]$, we therefore find that

$$\sum_{B_i \in Ch(B)} \overline{P}^*(B_i) \leq P(B) \leq \sum_{B_i \in Ch(B)} \overline{P}^*(B_i).$$

If this is satisfied, then there are values $P(B_i)$ such that $P(B_i) \in [\underline{P}^*(B_i), \overline{P}^*(B_i)]$ such that $\sum_{B_i \in \in Ch(B)} P(B_i) = P(B)$. For example, if $P(B) = \alpha \sum_{B_i \in Ch(B)} \underline{P}^*(B_i) + (1-\alpha) \sum_{B_i \in Ch(B)} \overline{P}^*(B_i)$, one solution is obtained by setting $P(B_i) = \alpha.\underline{P}^*(B_i) + (1-\alpha).\overline{P}^*(B_i)$ for any $B_i \in Ch(B)$. Since the leaves are elementary sets, we obtain a fully specified probability measure that satisfies all the intervals and thereby avoids sure loss.

For the reverse implication, assume that the original assignment avoids sure loss, there is a probability measure P that satisfies every inequality. We will then prove that P must also satisfy the adjusted intervals. This is true for leaves that do not change and the inner node result is proved on account of the fact that if $P(B_i) \geq \underline{P}^*(B_i)$ for any $B_i \in Ch(B_i)$ and the child nodes of node B are a partition of B, then it also follows that $P(B) = \sum_{B_i \in Ch(B)} P(B_i) \geq \sum_{B_i \in Ch(B)} \underline{P}^*(B_i)$. Since $P(B) \geq \underline{P}(B)$, it follows that
$P(B) \geq \max\{\underline{P}(B), \sum_{B_i \in Ch(B)} \underline{P}^*(B_i)\} = \underline{P}^*(B)$.

The new lower bounds for B are satisfied and this can also be proved for the new upper bounds. As there is a probability satisfying $\underline{P}^*(B) \leq P(B) \leq \overline{P}^*(B)$, it follows that $\underline{P}^*(B) \leq \overline{P}^*(B)$ for any node B.

Proposition 5. *If we assume that S is a hierarchical specification with tree T that avoids sure loss, then the partial natural extension defined for nodes in T can be computed from the bounds $\underline{P}^*, \overline{P}^*$ defined in Proposition 4 with the following equations applied to every T node from root to leaves: if B is an inner node, then $\forall B_i \in Ch(B)$,*

$$\underline{P}'(B_i) = \max\{\underline{P}^*(B_i), \underline{P}'(B) - \sum_{\substack{B_j \in Ch(B) \\ j \neq i}} \overline{P}^*(B_j)\}$$

$$\overline{P}'(B_i) = \min\{\overline{P}^*(B_i), \overline{P}'(B) - \sum_{\substack{B_j \in Ch(B) \\ j \neq i}} \underline{P}^*(B_j)\}$$

Proof of Proposition 5. We begin by proving that any probability measure P that satisfies $\underline{P}^*(B_i) \leq P(B_i) \leq \underline{P}^*(B_i)$ also satisfies $\underline{P}'(B_i) \leq P(B_i) \leq \underline{P}'(B_i)$. This is true for the root node since the bounds for this node do not change. We can then prove that if this is true for an inner node, then it is also true for its leaves. We have a probability measure P such that $\underline{P}'(B) \leq P(B) \leq \overline{P}'(B)$ and $\underline{P}^*(B_i) \leq P(B_i) \leq \overline{P}^*(B_i)$ for any child B_i of B. Since P is a probability measure and the children of B are a partition of the set, it follows that

$$\underline{P}'(B) \le P(B) = \sum_{B_i \in Ch(B)} P(B_i).$$

Therefore,

$$P(B_i) \ge \underline{P}'(B) - \sum_{\substack{B_j \in Ch(B)\\ j \ne i}} P(B_j) \ge \underline{P}'(B) - \sum_{\substack{B_j \in Ch(B)\\ j \ne i}} \overline{P}^*(B_j).$$

Since we also know that $P(B_i) \ge \underline{P}^*(B_i)$, then $P(B_i) \ge \max\{\underline{P}^*(B_i), \underline{P}'(B) - \sum_{\substack{B_j \in Ch(B)\\ j\ne i}} \overline{P}^*(B_j)\} = \underline{P}'(B_i)$. Similarly, this can be proved for the upper bounds.

This result also proves that the specifications with $(\underline{P}^*, \overline{P}^*)$ and $(\underline{P}', \overline{P}')$ are equivalent, since they define the same credal set, and that $\underline{P}'(B) \le \overline{P}'(B)$ for any set B in the tree. In order to prove that $(\underline{P}', \overline{P}')$ is the partial natural extension, we only need to prove that for every C in T there is a probability in the credal set associated with $(\underline{P}', \overline{P}')$ such that $P(C) = \underline{P}'(C)$ and another probability Q such that $Q(C) = \overline{P}'(C)$, then the intervals given by $(\underline{P}', \overline{P}')$ are tight and cannot be made smaller.

We will only prove the case of the lower bound. If C is the root node, the result can easily be verified since $C = X$ and $\underline{P}(X) = 1$, and any probability in the associated credal set will satisfy the equality.

If C is not a root node, then we can proceed as in the proof of Proposition 4 by assigning probability values for every B node from root to leaves using bounds $\underline{P}'(B), \overline{P}'(B)$, except for C and its siblings. Let us assume that B is its parent node, i.e. $C = B_i \in Ch(B)$. We have an assignation $P(B)$ such that $\underline{P}'(B) \le P(B) \le \overline{P}'(B)$. The probabilities $P(B_j)$ are then assigned as follows:

$$P(B_j) = \left\{ \begin{array}{ll} \underline{P}'(B_j), & \text{if } i = j \\ \alpha \underline{P}^*(B_i) + (1-\alpha)\overline{P}^*(B_i), & \text{if } i \ne j \end{array} \right\}$$

where $\alpha \in [0,1]$ is such that $P(B) - \underline{P}'(B_i) = \alpha \sum_{\substack{B_j \in Ch(B)\\ j\ne i}} \underline{P}^*(B_j) + (1-\alpha) \sum_{\substack{B_j \in Ch(B)\\ j\ne i}} \overline{P}^*(B_i)$. This value α always exists as

$$P(B) - \underline{P}'(B_i) \le P(B) - \underline{P}'(B) + \sum_{\substack{B_j \in Ch(B)\\ j\ne i}} \underline{P}^*(B_j) \le \sum_{\substack{B_j \in Ch(B)\\ j\ne i}} \underline{P}^*(B_j),$$

and similarly, it can be proved that $P(B) - \underline{P}'(B_i) \ge \sum_{\substack{B_j \in Ch(B)\\ j\ne i}} \overline{P}^*(B_i)$.

In order to finish the probability specification, if a leaf B is not elementary, we can assign $P(\{x_i\}) = \frac{1}{|B|}$, for any $x_i \in B$, with which the desired probability is obtained.

Proposition 6. *If S is a coherent hierarchical specification with the associated tree T, then Algorithm 1 returns the pair of values (α, β) where α is the upper natural extension, $\overline{P}'(A)$, for any $A \in 2^X$ and $\beta = 1$.*

Proof of Proposition 6. We will use $\alpha(H)$ and $\beta(H)$ to denote the couple of values (α, β) computed by the algorithm for node H.

Given the expression for computing β, we can easily verify that $\underline{P}(H) \le \beta(H) \le \overline{P}(H)$.

As in the root node, we have the full set X and $\overline{P}(X) = \underline{P}(X) = 1$, so in this node $\beta = 1$.

We will prove by induction on the number of nodes which are descendants of H that for any probability P satisfying all the restrictions in the tree, it follows that $P(H \cap A) \le \alpha(H)$ and $P(H \cap A^c) \ge \beta(H) - \alpha(H)$.

If H is a leaf node or $H \cap A = \emptyset$, verification is easy given lines 4 and 7 of the algorithm and on account of the fact that the current assessment is coherent.

If H is not a leaf node and $H \cap A \ne \emptyset$, then for any of its children C_i, it follows that for any probability that satisfies the restrictions in the subtree nodes given by the nodes C_i and their descendents, it holds that $P(C_i \cap A) \le \alpha(C_i) = \alpha_i$.

We now consider two situations:

- When $\beta(H) = \max\{\sum_{i=1}^l \beta_i, \underline{P}(H)\}$, then $\overline{P}(H) \ge \sum_{i=1}^l \beta_i$,
  $\alpha = \sum_{i=1}^l \alpha_i$, and $P(H \cap A) = \sum_{i=1}^l P(C_i \cap A) \le \sum_{i=1}^l \alpha_i = \alpha(H)$.
  Furthermore, $\beta - \alpha = \max\{\sum_{i=1}^l \beta_i, \underline{P}(H)\} - \sum_{i=1}^l \alpha_i$.
  This implies that $P(H \cap A^c) = \sum_{i=1}^l P(C_i \cap A^c) \ge \sum_{i=1}^l (\beta_i - \alpha_i)$ (by the induction hypothesis) and that $P(H \cap A^c) = P(H) - P(H \cap A) \ge \underline{P}(H) - \sum_{i=1}^l \alpha_i$.
  Therefore, $P(H \cap A^c) \ge \max\{\sum_{i=1}^l \beta_i, \underline{P}(H)\} - \sum_{i=1}^l \alpha_i = \beta - \alpha$.
- When $\beta(H) < \max\{\sum_{i=1}^l \beta_i, \underline{P}(H)\}$, then $\beta(H) = \overline{P}(H)$ and $\sum_{i=1}^l \beta_i > \underline{P}(H)$.
  In these conditions, $\alpha = \overline{P}(H) + \sum_{i=1}^l \alpha_i - \sum_{i=1}^l \beta_i$.
  This follows that $P(H \cap A) = P(H) - \sum_{i=1}^l P(C_i \cap A^c) \le \overline{P}(H) - \sum_{i=1}^l (\beta_i - \alpha_i) = \alpha$, on account of the fact that we have applied $P(C_i \cap A^c) \ge \beta_i - \alpha_i$ by induction.
  Moreover, $\beta - \alpha = \sum_{i=1}^l (\beta_i - \alpha_i) \le \sum_{i=1}^l P(C_i \cap A^c) = P(H \cap A^c)$.

If we apply this to the root node, then any probability satisfying every restriction represented by the tree also satisfies $P(A) \le \alpha(X)$, which is the value provided by Algorithm 1.

We will now prove that the bounds are tight by virtue of the following facts which can be proved by induction:

a) For any γ such that $\beta < \gamma \le \overline{P}(H)$, there is a probability that satisfies all the constraints below node H such that $P(H) = \gamma$, $P(H \cap A) = \alpha$.
b) For any γ such that $\underline{P}(H) \le \gamma \le \beta$, there is a probability satisfying all the constraints below node H such that $P(H) = \gamma$, $P(H \cap A^c) = \beta - \alpha$.

For nodes with $H \cap A = \emptyset$, we have that $\beta = \underline{P}(H)$, $\alpha = 0$. The first fact a) is satisfied since for any value γ satisfying $\underline{P}(H) < \gamma \le \overline{P}(H)$, there is a probability P, with $P(H) = \gamma$. The second fact b) is also satisfied since for any $\gamma = \underline{P}(H)$, there is a probability P, with $P(H) = \gamma = \underline{P}(H) = \beta - \alpha$, because all the intervals are reachable by coherence. For nodes with $H \cap A = \emptyset$, we have that $\beta = \underline{P}(H)$, $\alpha = 0$. The first fact a) is satisfied since for any value γ satisfying $\underline{P}(H) < \gamma \le \overline{P}(H)$, there is a probability P, with $P(H) = \gamma$. The second fact b) is also satisfied since for any $\gamma = \underline{P}(H)$, there is a probability P, with $P(H) = \gamma = \underline{P}(H) = \beta - \alpha$, because all the intervals are reachable by coherence.

For other leaf nodes these conditions are easily verified.

For any non-leaf node H, such that $H \cap A \neq \emptyset$, we assume by induction that this is also fulfilled for any of its children $C_i, i = 1, \dots, l$. We consider two cases:

- $\gamma \le \sum_{i=1}^{l} \beta_i$: In this situation, we have γ_i, $i = 1, \dots, l$ such that $\underline{P}(C_i) \le \gamma_i \le \beta_i$, $i = 1, \dots, l$ and such that $\sum_{i=1}^{l} \gamma_i = \gamma$. We then apply the induction hypothesis to this case and there is a value P_i such that $P_i(C_i) = \gamma_i$, $P_i(A^c \cap C_i) = \beta_i - \alpha_i$. The probability P defined on H by $P(x) = P_i(x)$, for $x \in C_i$, then satisfies the given requirements. We must take into account the following facts:
  - Since $\underline{P}(H) \le \gamma \le \sum_{i=1}^{l} \beta_i$, this means that $\beta = \min\{\sum_{i=1}^{l} \beta_i, \overline{P}(H)\}$ and $\beta - \alpha = \sum_{i=1}^{l} (\beta_i - \alpha_i)$.
  - The situation in point a) is not possible:
    * if $\sum_{i=1}^{l} \beta_i \le \overline{P}(H)$, then $\beta = \sum_{i=1}^{l} \beta_i$ and $\beta = \sum_{i=1}^{l} \beta_i < \gamma \le \sum_{i=1}^{l} \beta_i = \beta$
    * if $\sum_{i=1}^{l} \beta_i > \overline{P}(H)$, then $\beta = \overline{P}(H)$ and we cannot have $\beta < \gamma \le \overline{P}(H) = \beta$
  - Therefore, only *b*) is possible, and in this case it can easily be verified that $P(H) = \sum_{i=1}^{l} \gamma_i = \gamma$ and that $P(H \cap A^c) = \sum_{i=1}^{l} (\beta_i - \alpha_i) = \beta - \alpha$.
- $\gamma > \sum_{i=1}^{l} \beta_i$: In this situation, we have γ_i, $i = 1, \dots, l$ $\beta_i < \gamma_i \le \overline{P}(C_i)$, $i = 1, \dots, l$ such that $\sum_{i=1}^{l} \gamma_i = \gamma$. We then apply the induction hypothesis to this case and there is a value P_i such that $P_i(C_i) = \gamma_i$, $P_i(A \cap C_i) = \alpha_i$. The probability P defined on H by $P(x) = P_i(x)$, for $x \in C_i$, satisfies the given requirements. It is necessary to consider the following facts:
  - Since $\sum_{i=1}^{l} \beta_i < \gamma \le \overline{P}(H)$, $\beta = \max\{\sum_{i=1}^{l} \beta_i, \underline{P}(H)\}$ and $\alpha = \sum_{i=1}^{l} \alpha_i$.
  - In situation *a*), i.e. $\beta < \gamma \le \overline{P}(H)$, it can easily be verified that $P(H) = \sum_{i=1}^{l} \gamma_i = \gamma$, and that $P(H \cap A) = \sum_{i=1}^{l} \alpha_i = \alpha$.
  - The situation in point *b*), i.e. $\underline{P}(H) \le \gamma \le \beta$, is only possible when $\underline{P}(H) = \gamma = \beta$. In this situation, it follows that $P(H) = \gamma$ and $P(H \cap A^c) = \gamma - \alpha = \beta - \alpha$.

At the root node, $\underline{P}(X) = \overline{P}(X) = \beta = 1$, and the value α is a tight bound.

Proposition 7. *If in Algorithm 2 $B(H) = \sum_{C_i \in Ch(H)} \beta(C_i)$, where $\beta(C_i)$ are the values computed in Algorithm 1 when computing $\overline{P}'(A)$, then it returns $\overline{P}'(A \cup \{x\})$.*

Proof of Proposition 7. The proof is based on showing that $\delta(H_j) = \alpha^x(H_j) - \alpha(H_j)$ and $\phi(H_j) = \beta^x(H_j) - \beta(H_j)$ in a recursive way. The values $\alpha(H), \beta(H)$ for nodes with $A \cap H = \emptyset$ also satisfy the equations in lines 13 and 14 of Algorithm 1, and we can proceed from the leaf to the root assuming that these equations are always satisfied.

For the leaf node H_0, this is the case because of the way in which it is defined.

For each inner node H_j, we can see that if this is true for H_{j-1} then it is true for H_j. Let us consider $A(H) = \sum_{C_i \in Ch(H)} \alpha(C_i)$. When superscript x appears in a parameter, then it is considered to correspond to the values of Algorithm 1 when applied to set $A \cup \{x\}$ and when the superscript does not appear then it corresponds to the values of the same algorithm applied to set A.

This follows that

$$\beta(H_j) = \min\{\max\{B(H_j), \underline{P}(H_j)\}, \overline{P}(H_j)\} \tag{A.2}$$

$$\alpha(H_j) = A(H_j) + \beta(H_j) - \max\{B(H_j), \underline{P}(H_j)\} \tag{A.3}$$

$$\beta^x(H_j) = \min\{\max\{\sum_{C_i \in Ch(H_j)} \beta^x(C_i), \underline{P}(H_j)\}, \overline{P}(H_j)\} \tag{A.4}$$

$$\alpha^x(H_j) = \sum_{C_i \in Ch(H_j)} \alpha^x(C_i) + \beta^x(H_j) - \max\{\sum_{i=1}^{l} \beta_i^x, \underline{P}(H_j)\} \tag{A.5}$$

Under the hypothesis of induction, it follows that $\beta^x(H_{j-1})) = \beta(H_{j-1}) + \phi(H_{j-1})$ and $\alpha^x(H_{j-1}) = \alpha(H_{j-1}) + \delta(H_{j-1})$ and for any child C_i different from H_{j-1} it follows that $\beta^x(C_i)) = \beta(C_i)$ and $\alpha^x(C_i) = \alpha(C_i)$. Therefore, $\sum_{C_i \in Ch(H_j)} \beta^x(C_i) = B(H_j) + \phi(H_{j-1})$ and $\sum_{C_i \in Ch(H_j)} \alpha^x(C_i) = A(H_j) + \delta(H_{j-1})$. The expressions for $\beta^x(H_j)$ and $\alpha^x(H_j)$ can be rewritten as

$$\beta^x(H_j) = \min\{\max\{B(H_j) + \phi(H_{j-1}), \underline{P}(H_j)\}, \overline{P}(H_j)\} \tag{A.6}$$

$$\alpha^x(H_j) = A(H_j) + \delta(H_{j-1}) + \beta^x(H_j) - \max\{B(H_j) + \phi(H_{j-1}), \underline{P}(H_j)\} \tag{A.7}$$

This is obtained by considering the following situations:

a) $B(H_j) + \phi(H_{j-1}) \le \underline{P}(H_j)$. In this case, $\epsilon_2 = \phi(H_{j-1})$, $\epsilon_1 = 0$, $\delta(H_j) = \delta(H_{j-1}), \phi(H_j) = 0$ and $\beta^x(H_j) = \beta(H_j)$, $\alpha^x(H_j) = \alpha(H_j) + \delta(H_{j-1})$. The desired expressions, therefore, are also valid for H_j.
b) $\underline{P}(H_j) \le B(H_j) + \phi(H_{j-1}) < \overline{P}(H_j)$. In this case, $\epsilon_1 = 0$, $\epsilon_2 = \max\{0, \underline{P}(H_j) - B(H_j)\}$, $\delta(H_j) = \delta(H_{j-1})$, $\phi(H_j) = \phi(H_{j-1}) - \max\{0, \underline{P}(H_j) - B(H_j)\}$.
Furthermore, $\beta^x(H_j) = B(H_j) + \phi(H_{j-1}), \beta(H_j) = \max\{B(H_j), \underline{P}(H_j)\}$ and $\alpha^x(H_j) = A(H_j) + \delta(H_{j-1}), \alpha(H_j) = A(H_j)$.
We can then easily verify that $\beta^x(H_j) - \beta(H_j) = \phi(H_j)$ and $\alpha^x(H_j) - \alpha(H_j) = \delta(H_j)$.
c) $B(H_j) > \overline{P}(H_j)$. In this case, we will also have $B(H_j) + \phi(H_{j-1}) > \overline{P}(H_j)$, $\epsilon_1 = \phi(H_{j-1}), \epsilon_2 = 0$. Therefore, $\phi(H_j) = 0$, $\delta(H_j) = \delta(H_{j-1}) - \phi(H_{j-1})$.
We also have $\beta^x(H_j) = \overline{P}(H_j), \beta(H_j) = \overline{P}(H_j)$ and $\alpha^x(H_j) = A(H_j) + \delta(H_{j-1}) + \overline{P}(H_j) - B(H_j) - \phi(H_{j-1})$, $\alpha(H_j) = A(H_j) + \overline{P}(H_j) - B(H_j)$.
With this, we can easily verify that $\beta^x(H_j) - \beta(H_j) = \phi(H_j)$ and $\alpha^x(H_j) - \alpha(H_j) = \delta(H_j)$.
d) $B(H_j) \le \overline{P}(H_j) \le B(H_j) + \phi(H_{j-1})$. In this situation, it follows that $\epsilon_1 = B(H_j) + \phi(H_{j-1}) - \overline{P}(H_j)$, $\epsilon_2 = \max\{\underline{P}(H_j) - B(H_j), 0\}$.
Therefore, $\phi(H_j) = \overline{P}(H_j) - B(H_j) - \max\{\underline{P}(H_j) - B(H_j), 0\}$, $\delta(H_j) = \delta(H_{j-1}) + \overline{P}(H_j) - B(H_j) - \phi(H_{j-1})$.
Furthermore, $\beta^x(H_j) = \overline{P}(H_j), \beta(H_j) = \max\{B(H_j), \underline{P}(H_j)\}$ and $\alpha^x(H_j) = A(H_j) + \delta(H_{j-1}) + \overline{P}(H_j) - B(H_j) - \phi(H_{j-1}), \alpha(H_j) = A(H_j) + \beta(H_j) - \max\{B(H_j), \underline{P}(H_j)\} = A(H_j)$.
Therefore, $\beta^x(H_j) - \beta(H_j) = \overline{P}(H_j) - \max\{B(H_j), \underline{P}(H_j)\} = \overline{P}(H_j) - B(H_j) - \max\{\underline{P}(H_j) - B(H_j), 0\} = \phi(H_j)$ and $\alpha^x(H_j) - \alpha(H_j) = A(H_j) + \delta(H_{j-1}) + \overline{P}(H_j) - B(H_j) - \phi(H_{j-1}) - A(H_j) = \delta(H_j)$.

Finally, when $\phi(H_j) = 0$, there is immediate verification that $\epsilon_1 = \epsilon_2 = 0$ if these values were computed for H_k with $k > j$ and the values $\delta(H_k)$ are all equal to $\delta(H_j)$ so the loop can be exited at this point.

Proposition 8. *If S is a hierarchical specification, then it is always an order-2 capacity.*

Proof of Proposition 8. We will prove that if $A \subseteq B$, then $\overline{P}'(A \cup \{x\}) - \overline{P}'(A) \ge \overline{P}'(B \cup \{x\}) - \overline{P}'(B)$. This property is known as the *law of diminishing returns* and it is an equivalent definition of an order-2 capacity [9].

In order to prove this, we will consider Algorithms 1 and 2 that compute $\overline{P}'(A)$ and the difference $\overline{P}'(A \cup \{x\}) - \overline{P}'(A)$, respectively.

We will also consider the same algorithms to compute $\overline{P}'(B)$ and the difference $\overline{P}'(B \cup \{x\}) - \overline{P}'(B)$. In order to discriminate the parameters computed in this case for A, let us add a prime symbol to the parameters of B.

In this situation, it follows that

- $B'(H) \ge B(H)$ as the beta values always increase for larger sets
- Since $B(H_j) + \phi(H_{j-1}) = B^x(H_j)$, this means that $B'(H_j) + \phi'(H_{j-1}) \ge B(H_j) + \phi(H_{j-1})$.

We will prove by induction that for any $j = 0, \dots, m$ it follows that

$$\delta(H_j) - \delta'(H_j) \ge 0, \quad \delta(H_j) - \phi(H_j) - (\delta'(H_j) - \phi'(H_j)) \ge 0$$

We have that

$$\delta(H_j) - \phi(H_j) - (\delta'(H_j) - \phi'(H_j)) = \delta(H_{j-1}) - \phi(H_{j-1}) - (\delta'(H_{j-1}) - \phi'(H_{j-1})) + \epsilon_2 - \epsilon_2' \tag{A.8}$$

We also have

$$\epsilon_2 = \min\{\max\{\underline{P}(H_j) - B(H_j), 0\}, \phi(H_{j-1})\},$$

$$\epsilon_2' = \min\{\max\{\underline{P}(H_j) - B'(H_j), 0\}, \phi'(H_{j-1})\},$$

If $\epsilon_2 = \max\{\underline{P}(H_j) - B(H_j), 0\}$, this is because

$$\underline{P}(H_j) \le B(H_j) + \phi(H_{j-1}),$$

then

$$\underline{P}(H_j) \le B'(H_j) + \phi'(H_{j-1}),$$

and therefore

$$\epsilon_2 = \max\{\underline{P}(H_j) - B(H_j), 0\} \ge \max\{\underline{P}(H_j) - B'(H_j), 0\} = \epsilon'_2.$$

Since $\epsilon_2 \ge \epsilon'_2$, if we consider induction and Expression (A.8), we obtain

$$\delta(H_j) - \phi(H_j) - (\delta'(H_j) - \phi'(H_j)) \ge 0.$$

Furthermore, if $\epsilon_2 = \phi(H_{j-1})$, using Expression (A.8), it follows by induction that

$$\delta(H_j) - \phi(H_j) - (\delta'(H_j) - \phi'(H_j)) = \delta(H_{j-1}) - (\delta'(H_{j-1}) - \phi'(H_{j-1})) - \epsilon'_2.$$

Since $\phi'(H_{j-1}) \ge \epsilon'_2$, it follows that

$$\delta(H_j) - \phi(H_j) - (\delta'(H_j) - \phi'(H_j)) \ge \delta(H_{j-1}) - \delta'(H_{j-1}) \ge 0.$$

Let us now prove by induction that we also have that $\delta(H_j) - \delta'(H_j) \ge 0$.

$$\delta(H_j) - \delta'(H_j) = \delta(H_{j-1}) - \delta'(H_{j-1}) - \epsilon_1 + \epsilon'_1.$$

If $\epsilon'_1 = \phi'(H_{j-1})$, as $\epsilon_1 \le \phi(H_{j-1})$, it follows that

$$\delta(H_j) - \delta'(H_j) \ge \delta(H_{j-1}) - \delta'(H_{j-1}) - \phi(H_{j-1}) + \phi'(H_{j-1}) \ge 0,$$

where the last inequality is true due to the induction hypothesis.

Furthermore, if $\epsilon'_1 < \phi'(H_{j-1})$, this is because

$$\epsilon'_1 = \max\{B'(H_j) + \phi'(H_{j-1}) - \overline{P}(H_j)\}, 0\} \ge \max\{B(H_j) + \phi(H_{j-1}) - \overline{P}(H_j)\}, 0\} \ge \epsilon_1.$$

Therefore

$$\delta(H_j) - \delta'(H_j) = \delta(H_{j-1}) - \delta'(H_{j-1}) - \epsilon_1 + \epsilon'_1 \ge \delta(H_{j-1}) - \delta'(H_{j-1}) \ge 0,$$

where the last inequality is also true as a result of the induction hypothesis.

Since the difference $\overline{P}'(A \cup \{x\}) - \overline{P}'(A) = \delta(H_m)$ and $\overline{P}'(B \cup \{x\}) - \overline{P}'(B) = \delta'(H_m)$, where H_m is the root node, we thereby obtain the desired result.

References

[1] V. Biazzo, A. Gilio, A generalization of the fundamental theorem of de Finetti for imprecise conditional probability assessments, Int. J. Approx. Reason. 24 (2000) 251–272.
[2] L. de Campos, J. Huete, S. Moral, Probability intervals: a tool for uncertain reasoning, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 2 (1994) 167–196.
[3] L. de Campos, M. Lamata, S. Moral, The concept of conditional fuzzy measure, Int. J. Intell. Syst. 5 (1990) 237–246.
[4] A. Chateauneuf, J.Y. Jaffray, Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, Math. Soc. Sci. 17 (1989) 263–283.
[5] E.A. Corsi, T. Flaminio, H. Hosni, When belief functions and lower probabilities are indistinguishable, in: International Symposium on Imprecise Probability: Theories and Applications, PMLR, 2021, pp. 83–89.
[6] L.M. de Campos Ibañez, J. Bolaños, Representation of fuzzy measures through probabilities, Fuzzy Sets Syst. 31 (1989) 23–36, https://doi.org/10.1016/0165-0114(89)90064-X.
[7] G. De Cooman, F. Hermans, Imprecise probability trees: bridging two theories of imprecise probability, Artif. Intell. 172 (2008) 1400–1427.
[8] S. Destercke, D. Dubois, E. Chojnacki, Unifying practical uncertainty representations–I: generalized p-boxes, Int. J. Approx. Reason. 49 (2008) 649–663.
[9] S. Destercke, A. Rico, O. Strauss, Using atomic bounds to get sub-modular approximations, in: International Conference on Scalable Uncertainty Management, Springer, 2022, pp. 64–78.
[10] D. Dubois, H. Prade, When upper probabilities are possibility measures, Fuzzy Sets Syst. 49 (1992) 65–74.
[11] M.G. Faes, M. Daub, S. Marelli, E. Patelli, M. Beer, Engineering analysis with probability boxes: a review on computational methods, Struct. Saf. 93 (2021) 102092.
[12] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, Chapman and Hall/CRC, 1995.
[13] P. Hansen, B. Jaumard, M.P. De Aragao, F. Chauny, S. Perron, Probabilistic satisfiability with imprecise probabilities, Int. J. Approx. Reason. 24 (2000) 171–189.
[14] P. Huber, The use of Choquet capacities in statistics, Bull. Int. Stat. Inst. 45 (1973) 181–188.
[15] P. Huber, Robust Statistics, Wiley, New York, 1981.
[16] E. Miranda, A survey of the theory of coherent lower previsions, Int. J. Approx. Reason. 48 (2008) 628–658.
[17] E. Miranda, G. de Cooman, Marginal extension in the theory of coherent lower previsions, Int. J. Approx. Reason. 46 (2007) 188–225.
[18] E. Miranda, G.d. Cooman, Lower previsions, in: Introduction to Imprecise Probabilities, 2014, pp. 28–55.
[19] I. Montes, Comonotone lower probabilities with robust marginal distributions functions, Rev. R. Acad. Cienc. Exactas Fís. Nat., Ser. A Mat. 116 (2022) 130.
[20] R. Pelessoni, P. Vicig, et al., A consistency problem for imprecise conditional probability assessments, in: Proceedings IPMU, 1998, pp. 1478–1485.
[21] N. Pfeifer, G. Sanfilippo, Probability propagation in selected Aristotelian syllogisms, in: European Conference on Symbolic and Quantitative Approaches with Uncertainty, Springer, 2019, pp. 419–431.
[22] E. Quaeghebeur, G. De Cooman, Extreme lower probabilities, Fuzzy Sets Syst. 159 (2008) 2163–2175.
[23] J. Su, Y. Deng, An interval method to measure the uncertainty of basic probability assignment, Soft Comput. (2022) 1–10.
[24] M. Troffaes, S. Destercke, Probability boxes on totally preordered spaces for multivariate modelling, Int. J. Approx. Reason. 52 (2011) 767–791.

[25] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.

[26] P. Walley, Inferences from multinomial data: learning about a bag of marbles (with discussion), J. R. Stat. Soc., Ser. B 58 (1996) 3–57.

[27] P. Walley, R. Pelessoni, P. Vicig, Direct algorithms for checking consistency and making inferences from conditional probability assessments, J. Stat. Plan. Inference 126 (2004) 119–151.

[28] K. Weichselberger, The theory of interval-probability as a unifying concept for uncertainty, Int. J. Approx. Reason. 24 (2000) 149–170.

[29] G. Yang, S. Destercke, M.H. Masson, Cautious classification with nested dichotomies and imprecise probabilities, Soft Comput. 21 (2017) 7447–7462.