A new contextual discounting rule for lower probabilities

Sebastien Destercke[1]

INRA/CIRAD, UMR1208, 2 place P. Viala, F-34060 Montpellier cedex 1, France
sdestercke@gmail.com

Abstract. Sources providing information about the value of a variable may not be totally reliable. In such a case, it is common in uncertainty theories to take account of this unreliability by a so-called discounting rule. A few discounting rules have been proposed in the framework of imprecise probability theory, but one of the drawback of those rules is that they do not preserve interesting properties (i.e. n-monotonicity) of lower probabilities. Another aspect that only a few of them consider is that source reliability is often dependent of the context, i.e. a source may be more reliable to identify some values than others. In such cases, it is useful to consider contextual discounting, where reliability information is dependent of the variable values. In this paper, we propose such a contextual discounting rule that also preserves some of the interesting mathematical properties a lower probability can have.
keywords: information fusion, reliability, discounting, probability sets

1 Introduction

When sources providing uncertain information about the value assumed by a variable X on the (finite) domain $\mathscr{X}$ are not fully reliable, it is necessary to integrate information about this reliability in uncertainty representations. In imprecise probability theories (i.e. possibility theory, evidence theory, transferable belief model, lower previsions), where imprecision in beliefs or information is explicitly modelled in uncertainty representations, it is usual to take account of this reliability through the operation commonly called discounting. Roughly speaking, the discounting operation consists in making the information all the more imprecise (i.e. less relevant) as it is unreliable.

Many authors have discussed discounting operations in uncertainty theories [1,2,3]. In most cases, authors consider that reliability is modelled by a single weight (possibly imprecise) λ whose value is in the unit interval, i.e. $\lambda \in [0,1]$. In a few other cases, they consider that different weights can be given to different elements of a partition of the referential $\mathscr{X}$, and in this case reliability information is given by a vector of weights $\lambda = (\lambda_1, \ldots, \lambda_L)$, with L the cardinality of the partition and $\lambda_i \in [0,1]$. The reason for considering such weights is that, in some cases, the ability of the source to recognise the true value of X may depend on this value. For example, a specialised physician will be very reliable when it comes to recognise diseases corresponding to its speciality, but less reliable when the patient has other diseases. A sensor may be very discriminative for some kinds of objects, while often confusing other objects between them.

Many rules handling more than precise single reliability weight have been proposed in the framework of imprecise probability theory [2,4,5], in which uncertain information is represented by bounds over expectation values or by associated convex probability sets, the two representations being formally equivalent. Both Karlsson *et al.* [4] and Benavoli and Antonucci [5] consider the case where a unique but possibly imprecise reliability weight is given for the whole referential $\mathscr{X}$, but start from different requirements, hence proposing different discounting rules. Karlsson *et al.* [4] require a discounted probability set to be insensitive to Bayesian combination (i.e. using the product) when the source is completely unreliable. It brings them to the requirement that the information provided by a completely unreliable source should be transformed into the precise uniform probability distribution. Benavoli and Antonucci [5] model reliability by the means of coherent conditional lower previsions [6] and directly integrates it to an aggregation process, assuming that the information provided by a completely unreliable source should be transformed into a so-called vacuous probability set (i.e. the probability set corresponding to all probabilities having $\mathscr{X}$ for support). Moral and Sagrado [2] start from constraints given on expectations value and assume that reliability weights are precise but can be contextual (i.e., one weight per element of $\mathscr{X}$) or can translate some (fuzzy) indistinguishability relations.

Each of these rules is justified in its own setting. However, a common defect of all these rules is that when reliability weights are not reduced to a single precise number, the discounted probability set is usually more complex and difficult to handle than the initial one. This is a major inconvenient to their practical use, since using generic probability sets often implies an heavy computational burden. In this paper, we propose a new discounting rule for lower and upper probabilities, inspired from the discounting rule proposed by Mercier et al. [3] in the framework of the transferable belief model [7]. We show that this rule preserves both the initial probability set complexity, as well as some of its interesting mathematical properties, provided the initial lower probability satisfies them.

Section 2 recalls the basics of lower/upper probabilities needed here, as well as some considerations about the properties discounting rules can satisfy. Section 3 then presents our rule, discusses its properties and possible interpretation, and compares its properties with those of other discounting rules.

2 Preliminary notions

This section recalls both the notion of lower probabilities and of associated sets of probabilities. It then details some properties that may or may not have a given discounting rule.

2.1 Probability sets and lower probabilities

In this paper, we consider that our uncertainty about the value assumed by a variable X on a finite space $\mathscr{X} = \{x_1, \ldots, x_N\}$ is modelled by a lower probability $\underline{P} : \wp(\mathscr{X}) \rightarrow [0,1]$, i.e. a mapping from the power set of $\mathscr{X}$ to the unit interval, satisfying the boundary constraints $\underline{P}(\emptyset) = 0$, $\underline{P}(\mathscr{X}) = 1$ and monotonic with respect to inclusion, i.e. for

any $A,B \subseteq \mathscr{X}$ such that $A \subseteq B$, $\underline{P}(A) \leq \underline{P}(B)$. To a lower probability can be associated an upper probability $\overline{P}$ such that, for any $A \subseteq \mathscr{X}, \underline{P}(A) = 1 - \overline{P}(A^c)$, with A^c the complement of A. A lower probability induce a probability set $\mathscr{P}_{\underline{P}}$ such that

$$\mathscr{P}_{\underline{P}} := \{p \in \Sigma_{\mathscr{X}} | (\forall A \subseteq \mathscr{X})(P(A) \geq \underline{P}(A)\},$$

with p a probability mass, P the induced probability measure and $\Sigma_{\mathscr{X}}$ the set (simplex) of all probability mass functions on $\mathscr{X}$. A lower probability is said to be *coherent* if and only if $\mathscr{P}_{\underline{P}} \neq \emptyset$ and $\underline{P}(A) = \min\{P(A) | p \in \mathscr{P}_{\underline{P}}\}$ for all $A \subseteq \mathscr{X}$, i.e., if $\underline{P}$ is the *lower envelope* of $\mathscr{P}_{\underline{P}}$ on events.

Inversely, from any probability set $\mathscr{P}$, one can extract a lower probability measure defined, for any $A \subseteq \mathscr{X}$, as $\underline{P}(A) = \min\{P(A) | p \in \mathscr{P}\}$. Note that lower probabilities alone are not sufficient to describe any probability set. Let $\mathscr{P}$ be a probability set and $\underline{P}$ its lower probability, then the probability set $\mathscr{P}_{\underline{P}}$ induced by this lower probability is such that $\mathscr{P} \subseteq \mathscr{P}_{\underline{P}}$ with the inclusion being usually strict. In general, one needs the richer language of expectation bounds to describe any probability set [8].

In this paper, we will restrict ourselves to credal sets induced by lower probabilities alone. Note that such lower probabilities already encompass an important number of practical uncertainty representations, such as necessity measures [9], belief functions [10] or so-called p-boxes [11]. An important classes of probability sets induced by lower probabilities alone and encompassing these representations are the one for which lower probabilities satisfy the property of n-monotonicity for $n \geq 2$. n-monotonicity is defined as follows:

Definition 1. *A lower probability $\underline{P}$ is n-monotone, where $n > 0$ and $n \in \mathbb{N}$, if and only if for any set $\mathscr{A} = \{A_i | i \in \mathbb{N}, 0 < i \leq n\}$ of events $A_i \subseteq \mathscr{X}$, it holds that*

$$\underline{P}(\bigcup_{A_i \in \mathscr{A}} A_i) \geq \sum_{I \subseteq \mathscr{A}} (-1)^{|I|+1} \underline{P}(\bigcap_{A_i \in I} A_i).$$

An ∞-montone lower probability (i.e., a belief function) is a lower probability n-monotone for every n. Both 2-monotonicity and ∞-monotonicity have been studied with particular attention in the literature [12,10,13,14], for they have interesting mathematical properties that facilitate their practical handling. When processing lower probabilities, it is therefore desirable to preserve such properties, if possible.

2.2 Discounting operation: definition and properties

The discounting operation consists in using the reliability information λ to transform an initial lower probability $\underline{P}$ into another lower probability $\underline{P}^{\lambda}$. λ can take different forms, ranging from a single precise number to a vector of imprecise numbers. In order to discriminate between different discounting rules, we think it is useful to list some of the properties that they can satisfy.

*Property 1 (coherence preservation, **CP**).* A discounting rule satisfies coherence preservation **CP** when $\underline{P}^{\lambda}$ is coherent whenever $\underline{P}$ is coherent.

This property ensures some consistency to the discounting rule.

*Property 2 (Imprecision monotony, **IM**).* A discounting rule satisfies Imprecision monotony **IM** if and only if $\underline{P}^{\lambda} \leq \underline{P}$, that is if the discounted information is less precise than the original one.[1]

This property simply means that imprecision should increase when a source is partially unreliable. This may seem a reasonable request, however for some particular cases [4], there may exist arguments against such a property.

*Property 3 (n-monotonicity preservation, **MP**).* A discounting rule satisfies n-monotonicity preservation **MP** when $\underline{P}^{\lambda}$ is n-monotone whenever $\underline{P}$ is n-monotone.

Such a property ensures that interesting mathematical properties of a lower probabilities will be preserved by the discounting operation.

*Property 4 (lower probability preservation, **LP**).* A discounting rule satisfies lower probability preservation, **LP** when the discounted probability set $\mathscr{P}^{\lambda}$ resulting from discounting is such that $\mathscr{P}^{\lambda} = \mathscr{P}_{\underline{P}^{\lambda}}$, provided initial information was given as a lower probability.

This property ensures that if the initial information is entirely captured by a lower probability, so will be the discounted information. It ensures to some extent that the uncertainty representation structure will keep a bounded complexity.

*Property 5 (Reversibility, **R**).* A discounting rule satisfies reversibility **R** if the initial information $\underline{P}$ can be recovered from the knowledge of the discounted information $\underline{P}^{\lambda}$ and λ alone, when $\lambda > 0$.

This property, similar to the de-discounting discussed by Denoeux and Smets [15], ensures that, if one receives as information the discounted information together with the source reliability information, he can still come back to the original information provided by the source. This can be useful if reliability information is revised. This requires the discounting operation to be an injection.

3 The discounting rule

We now propose our contextual discounting rule, inspired from the contextual discounting rule proposed by Mercier at al. [3] in the context of the transferable belief model. We show that, from a practical viewpoint, this discounting rule has interesting properties, and briefly discuss its interpretation.

3.1 Definition

We consider that source reliability comes into the form of a vector of weights $\lambda = (\lambda_1, \ldots, \lambda_L)$ associated to elements of a partition $\Theta = \{\theta_1, \ldots, \theta_L\}$ of $\mathscr{X}$ (i.e. $\theta_i \subseteq \mathscr{X}$, $\cup_{i=1}^{L} \theta_i = \mathscr{X}$ and $\theta_i \cap \theta_j = \emptyset$ if $i \neq j$). We denote by $\mathscr{H}$ the field induced by Θ. Value

[1] This is equivalent to ask for $\mathscr{P}_{\underline{P}} \subseteq \mathscr{P}_{\underline{P}^{\lambda}}$

one is given to λ_i when the source is judged completely reliable for detecting element x_i, and zero if it is judged completely unreliable. We do not consider imprecise weights, simply because in such a case one can still consider the pessimistic case where the lowest weights are retained.

Given a set $A \subseteq \mathscr{X}$, its inner and outer approximations in $\mathscr{H}$, respectively denoted A_* and A^*, are:

$$A_* = \bigcup_{\substack{\theta \in \Theta \\ \theta \subseteq A}} \theta \quad \text{and} \quad A^* = \bigcup_{\substack{\theta \in \Theta \\ \theta \cap A \neq \emptyset}} \theta.$$

We then propose the following discounting rule that transforms an initial information $\underline{P}$ into $\underline{P}^\lambda$ such that, for every event $A \subseteq \mathscr{X}$, we have

$$\underline{P}^\lambda(A) = \underline{P}(A) \prod_{\theta_i \subseteq (A^c)^*} \lambda_i, \tag{1}$$

with the convention $\prod_{\theta_i \subseteq \emptyset} \lambda_i = 1$, ensuring that $\underline{P}(\mathscr{X}) = \underline{P}^\lambda(\mathscr{X}) = 1$ and $\overline{P}(\emptyset) = \overline{P}^\lambda(\emptyset) = 0$.

Example 1. Let us illustrate our proposition on a 3-dimensional space $\mathscr{X} = \{x_1, x_2, x_3\}$. Assume the lower probability is given by the following constraints:

$$0.1 \leq p(x_1) \leq 0.3; \quad 0.4 \leq p(x_2) \leq 0.5; \quad 0.3 \leq p(x_3) \leq 0.5.$$

Lower probabilities induced by these constraints (through natural extension [8]) can be easily computed, as they are probability intervals [16]. They are summarised in the next table:

	x_1	x_2	x_3	$\{x_1,x_2\}$	$\{x_1,x_3\}$	$\{x_2,x_3\}$
$\underline{P}$	0.1	0.4	0.3	0.5	0.5	0.7

Let us now assume that $\Theta = \{\{x_1,x_2\} = \theta_1, \{x_3\} = \theta_2\}$ and that $\lambda_1 = 0.5$, $\lambda_2 = 1$. The discounted lower probability $\underline{P}^\lambda$ is given in the following table

	x_1	x_2	x_3	$\{x_1,x_2\}$	$\{x_1,x_3\}$	$\{x_2,x_3\}$
$\underline{P}^\lambda$	0.05	0.2	0.15	0.5	0.25	0.35

Figure 1 pictures, in barycentric coordinates (i.e. each point in the triangle is a probability mass function over $\mathscr{X}$, with the probability of x_i equals to the distance of the point to the side opposed to vertex x_i), both the initial probability set and the discounted probability set resulting from the application or the proposed rule. As we can see, only the upper probability of $\{x_3\}$ (the element we are certain the source can recognise with full reliability) is kept at its initial value.

3.2 Properties of the discounting rule

Let us now discuss the properties of this discounting rule. First, by Equation (1), we have that the results of the discounting rule is still a lower probability, and since $\lambda \in [0,1]$, $\underline{P}^\lambda \leq \underline{P}$, hence the property of imprecision monotony is satisfied. We can also show the following proposition:

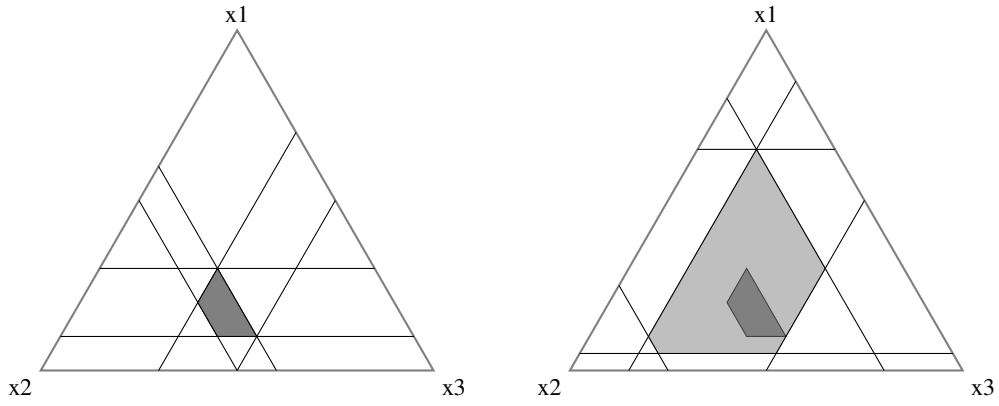

Fig. 1. Initial (right) and discounted (left) probability sets of Example 1.

Proposition 1. *Let $\underline{P}$ be a lower probability and λ a strictly positive weight vector. The contextual discounting rule preserves the following properties:*

1. *Coherence*
2. *2-monotonicity*
3. *∞-monotonicity*

See Appendix A for the proof. These properties ensure us that the discounting rule preserves the desirable properties of lower probabilities that are coherence, as well as other more "practical" properties that keep computational complexity low, such as 2-monotonicity. The discounting operator is also reversible.

Property 6 (Reversibility). Let $\underline{P}^{\lambda}$ and λ be the provided information. Then, $\underline{P}$ can be retrieved by computing, for any $A \subseteq \mathscr{X}$,

$$\underline{P}(A) = \frac{\underline{P}^{\lambda}(A)}{\prod_{\theta_i \subseteq (A^c)^*} \lambda_i}.$$

Table 1 summarises the properties of the discounting rule proposed here, together with the properties of other discounting rules proposed in the literature. It considers the following properties and features: whether a discounting can cope with generic probability sets, with imprecise weights and with contextual weights, and if it satisfies or not the properties proposed in Section 2.2. This table displays some of the motivations that have led to the rule proposed in this paper. Indeed, while most rules presented in the literature have been justified and have the advantages that they can be applied to any probability set (not just the ones induced by lower probabilities), applying them also implies losing properties that have a practical interest and importance, especially the properties of $2-$ and $\infty-$monotonicity. When dealing with lower probabilities, our rule offers a convenient alternative, as it preserves important properties.

Paper	Any $\mathscr{P}$	Imp. weights	contextual	CP	IM	MP	LP	R
This paper	×	×	✓	✓	✓	✓	✓	✓
Moral et al. [2]	✓	×	✓	✓	✓	×	×	×
Karlsson et al. [4]	✓	✓	×	✓	×	×	×	✓
Benavoli et al. [5]	✓	✓	×	✓	✓	×	×	×

Table 1. Discounting rules properties.

3.3 Interpretation of the discounting rule

In order to give an intuitive interpretation of the proposed discounting rule, let us consider the case where $\Theta = \{x_1, \ldots, x_N\}$ and $\mathscr{H}$ is the power set of $\mathscr{X}$, that is one weight is given to each element of $\mathscr{X}$. In this case, Eq. (1) becomes, for an event $A \subseteq \mathscr{X}$,

$$\underline{P}^{\lambda}(A) = \underline{P}(A) \prod_{x_i \in A^c} \lambda_i,$$

and the upper discounted probability $\overline{P}^{\lambda}$ of an event A becomes

$$\overline{P}^{\lambda}(A) = 1 - \underline{P}^{\lambda}(A^c) = 1 - (\underline{P}(A^c) \prod_{x_i \in A} \lambda_i) = 1 - \prod_{x_i \in A} \lambda_i + \overline{P}^{\lambda}(A) \prod_{x_i \in A} \lambda_i.$$

Hence, in this particular case, we have the following lemma:

Lemma 1. *For any event $A \subseteq \mathscr{X}$, we have*

- $\underline{P}^{\lambda}(A) = \underline{P}(A)$ *iff* $\lambda_i = 1$ *for any* $x_i \in A^c$,
- $\overline{P}^{\lambda}(A) = \overline{P}(A)$ *iff* $\lambda_i = 1$ *for any* $x_i \in A$.

This means that our certainty in the fact that the true answer lies in A (modeled by $\underline{P}(A)$) does not change, provided that we are certain that the source is able to eliminate all possible values outside of A. Consider for instance the case $\underline{P}(A) = 1$, meaning that we are sure that the true answer is in A. It seems rational to require, in order to fully trust this judgement, that the source can eliminate with certainty all possibilities outside A. Conversely, the plausibility that the true value lies in A ($\overline{P}(A)$) does not change when the source is totally able to recognise elements of A. Consider again the extreme case $\overline{P}(A) = 0$, then it is again rational to ask for $\overline{P}(A)$ to increase if the source is not fully able to recognise elements of A, and for it to remain the same otherwise, as in this case the source would have recognised an element of A for sure.

Now, consider the case where $\Theta = \{\mathscr{X}\}$ and $\mathscr{H} = \{\emptyset, \mathscr{X}\}$, with λ the associated unique weight. We retrieve the classical discounting rule consisting in mixing the initial probability set with the vacuous one, that is $\underline{P}^{\lambda}(A) = \lambda \underline{P}(A)$ for any $A \subseteq \mathscr{X}$ and we have $\mathscr{P}_{\underline{P}^{\lambda}} = \{\lambda \cdot p + (1-\lambda) \cdot q | p \in \mathscr{P}_{\underline{P}}, q \in \Sigma_{\mathscr{X}}\}$. Note that when $\Theta = \{\theta_1, \ldots, \theta_L\}$ with $L > 1$ and $\lambda := \lambda_1 = \ldots = \lambda_L$, the lower probability $\underline{P}^{\lambda}$ obtained from $\underline{P}$ is not equivalent to the one obtained by considering $\Theta = \{\mathscr{X}\}$ with λ, contrary to the rule of Moral and Sagrado [2]. However, if one thinks that reliability scores have to be distinguished for some different parts of the domain $\mathscr{X}$, there is no reason that the rule should act like if there was only one weight when the different weights are equal.

4 Conclusion

In this paper, we have proposed a contextual discounting rule for lower probabilities that can be defined on general partitions of the domain $\mathscr{X}$ on which a variable X assumes its values. Compared to previously defined rules for lower probabilities, the present rule have the advantage that its result is still a lower probability (one does not need to use general lower expectation bounds). It also preserves interesting mathematical properties, such as $2-$ and ∞-monotonicity, which are useful to compute the so-called natural extension.

Next moves include the use of this discounting rule and of others in practical applications (e.g. merging of classifier results, of expert opinions, ...), in order to empirically compare their practical results. From a theoretical point of view, the rule presented here should be extended to the more general case of lower previsions, so as to ensure that extensions of n-monotonicity [17] are preserved. Although preserving n-monotonicity for values others than $n = 2$ and $n = \infty$ has less practical interest, it would also be interesting to check whether it is preserved by the proposed rule (we can expect that it is, given results for 2-monotonicity and ∞-monotonicity). Another important issue is to provide a stronger and proper interpretation (e.g. in terms of betting behaviour) to this rule, as the interpretation given in the framework of the TBM [3] cannot be applied to generic lower probabilities.

A Proof of proposition 1

Proof. Let $\underline{P}$ be the lower probability given by the source

- Let us start with property 3, as we will use it to prove the other properties. This property has been proved by Mercier et al. [3] in the case of the transferable belief models, in which are included normalized belief functions (i.e. ∞-monotone lower probabilities).
- Let us now show that property 1 of coherence is preserved. First, note that if $\underline{P}$ is coherent, it means that $\mathscr{P}_{\underline{P}} \neq \emptyset$, and since $\underline{P}^{\lambda} \leq \underline{P}$, $\mathscr{P}_{\underline{P}^{\lambda}} \neq \emptyset$ too. Now, consider a particular event A. If $\underline{P}$ is coherent, it means that there exists a probability measure $P \in \mathscr{P}_{\underline{P}}$ such that it dominates $\underline{P}$ (i.e., $\underline{P} \leq P$) and moreover $\underline{P}(A) = P(A)$. P being a special kind of ∞-monotone lower probability, we can also apply the discounting rule to P and obtain a lower probability P^{λ} which remains ∞-monotone (property (3)) and is such that $P^{\lambda}(A) = \underline{P}^{\lambda}(A)$. The fact that $\prod_{\theta_i \subseteq \emptyset} \lambda_i = 1$ ensures us that $P^{\lambda}(\emptyset) = 0$ and $P^{\lambda}(\mathscr{X}) = 1$, hence P^{λ} is coherent. Also note that P^{λ} still dominates $\underline{P}^{\lambda}$, since both P and $\underline{P}$ are multiplied by the same numbers on every event to obtain P^{λ} and $\underline{P}^{\lambda}$. Therefore, $\exists P'$ such that $\underline{P}^{\lambda} \leq P^{\lambda} \leq P'$ and $P'(A) = P^{\lambda}(A) = \underline{P}^{\lambda}(A)$. As this is true for every event A, this means that $\underline{P}^{\lambda}$ is coherent.
- We can now show property 2. If $\underline{P}$ is 2-monotone, it means that $\forall A, B \subseteq \mathscr{X}$, the inequality

$$\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B) - \underline{P}(A \cap B)$$

holds. Now, considering $\underline{P}^{\lambda}$, we have to show that $\forall A,B \subseteq \mathscr{X}$, the following inequality holds

$$\underline{P}(A\cup B)\prod_{\theta_i\subseteq((A\cup B)^c)^*}\lambda_i \geq \underline{P}(A)\prod_{\theta_i\subseteq(A^c)^*}\lambda_i+\underline{P}(B)\prod_{\theta_i\subseteq(B^c)^*}\lambda_i-\underline{P}(A\cap B)\prod_{\theta_i\subseteq((A\cap B)^c)^*}\lambda_i. \quad (2)$$

Let us consider the three following partitions:

$$\begin{aligned}(A^c)^* &= ((A^c)^*\setminus((A\cup B)^c)^*)\cup((A\cup B)^c)^*,\\ (B^c)^* &= ((B^c)^*\setminus((A\cup B)^c)^*)\cup((A\cup B)^c)^*,\\ ((A\cap B)^c)^* &= ((A^c)^*\setminus((A\cup B)^c)^*)\cup((B^c)^*\setminus((A\cup B)^c)^*)\cup((A\cup B)^c)^*.\end{aligned}$$

To simplify notation, we denote by $\mathscr{S}=(A\cup B)^c$. We can reformulate Eq (2) as

$$\begin{aligned}\underline{P}(A\cup B)\prod_{\theta_i\subseteq\mathscr{S}}\lambda_i \geq \quad & \underline{P}(A)\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i\prod_{\theta_i\subseteq\mathscr{S}}\lambda_i+\underline{P}(B)\prod_{\theta_i\subseteq((B^c)^*\setminus\mathscr{S})}\lambda_i\prod_{\theta_i\subseteq\mathscr{S}}\lambda_i\\ & -\underline{P}(A\cap B)\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i\prod_{\theta_i\subseteq((B^c)^*\setminus\mathscr{S})}\lambda_i\prod_{\theta_i\subseteq\mathscr{S}}\lambda_i.\end{aligned}$$

Dividing by $\prod_{\theta_i\subseteq\mathscr{S}}$, we obtain

$$\begin{aligned}\underline{P}(A\cup B)\geq \quad & \underline{P}(A)\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i+\underline{P}(B)\prod_{\theta_i\subseteq((B^c)^*\setminus\mathscr{S})}\lambda_i\\ & -\underline{P}(A\cap B)\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i\prod_{\theta_i\subseteq((B^c)^*\setminus\mathscr{S})}\lambda_i.\end{aligned}$$

Now, using the fact that $\underline{P}$ is 2-monotone and replacing $\underline{P}(A\cup B)$ by the lower bound $\underline{P}(A)+\underline{P}(B)-\underline{P}(A\cap B)$ in the above equation, we must show

$$\begin{aligned}&\underline{P}(A)(1-\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i)+\underline{P}(B)(1-\prod_{\theta_i\subseteq((B^c)^*\setminus\mathscr{S})}\lambda_i)\\ &-\underline{P}(A\cap B)(1-\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i\prod_{\theta_i\subseteq((B^c)^*\setminus\mathscr{S})}\lambda_i)\geq 0.\end{aligned}$$

Now, we can replace $\underline{P}(A\cap B)$ by $\min(\underline{P}(A),\underline{P}(B))$, considering that $\min(\underline{P}(A),\underline{P}(B))\geq \underline{P}(A\cap B)$. Without loss of generality, assume that $\underline{P}(A)\leq\underline{P}(B)$, then we have

$$\begin{aligned}&\underline{P}(A)(\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i\prod_{\theta_i\subseteq((B^c)^*\setminus\mathscr{S})}\lambda_i-\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i)+\underline{P}(B)(1-\prod_{\theta_i\subseteq((B^c)^*\setminus\mathscr{S})}\lambda_i)\geq 0\\ &-\underline{P}(A)(1-\prod_{\theta_i\subseteq((B^c)^*\setminus\mathscr{S})}\lambda_i)(\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i)+\underline{P}(B)(1-\prod_{\theta_i\subseteq((B^c)^*\setminus\mathscr{S})}\lambda_i)\geq 0\\ &-\underline{P}(A)(\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i)+\underline{P}(B)\geq 0\end{aligned}$$

and, since $\underline{P}(A)(\prod_{\theta_i\subseteq((A^c)^*\setminus\mathscr{S})}\lambda_i)\leq\underline{P}(A)\leq\underline{P}(B)$, this finishes the proof.

References

1. Dubois, D., Prade, H.: Possibility theory and data fusion in poorly informed environments. Control Engineering Practice **2** (1994) 811–823
2. Moral, S., Sagrado, J.: Aggregation of imprecise probabilities. In BouchonMeunier, B., ed.: Aggregation and Fusion of Imperfect Information. Physica-Verlag, Heidelberg (1997) 162–188
3. Mercier, D., Quost, B., Denoeux, T.: Refined modeling of sensor reliability in the bellief function framework using contextual discounting. Information Fusion **9** (2008) 246–258
4. Karlsson, A., Johansson, R., Andler, S.F.: On the behavior of the robust bayesian combination operator and the significance of discounting. In: ISIPTA'09: Proc. of the Sixth Int. Symp. on Imprecise Probability: Theories and Applications. (2009) 259–268
5. Benavoli, A., Antonucci, A.: Aggregating imprecise probabilistic knowledge. In: ISIPTA'09: Proc. of the Sixth Int. Symp. on Imprecise Probability: Theories and Applications. (2009) 31–40
6. Miranda, E.: A survey of the theory of coherent lower previsions. Int. J. of Approximate Reasoning **48** (2008) 628–658
7. Smets, P., Kennes, R.: The transferable belief model. Artificial Intelligence **66** (1994) 191–234
8. Walley, P.: Statistical reasoning with imprecise Probabilities. Chapman and Hall, New York (1991)
9. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
10. Shafer, G.: A mathematical Theory of Evidence. Princeton University Press, New Jersey (1976)
11. Ferson, S., Ginzburg, L., Kreinovich, V., Myers, D., Sentz, K.: Constructing probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories (2003)
12. Chateauneuf, A., Jaffray, J.Y.: Some characterizations of lower probabilities and other monotone capacities through the use of Mobius inversion. Mathematical Social Sciences **17**(3) (1989) 263–283
13. Miranda, E., Couso, I., Gil, P.: Extreme points of credal sets generated by 2-alternating capacities. I. J. of Approximate Reasoning **33** (2003) 95–115
14. Bronevich, A., Augustin, T.: Approximation of coherent lower probabilities by 2-monotone measures. In: ISIPTA'09: Proc. of the Sixth Int. Symp. on Imprecise Probability: Theories and Applications, SIPTA (2009) 61–70
15. Denoeux, T., Smets, P.: Classification using belief functions: the relationship between the case-based and model-based approaches. IEEE Trans. on Syst., Man and Cybern. B **36**(6) (2006) 1395–1406
16. de Campos, L., Huete, J., Moral, S.: Probability intervals: a tool for uncertain reasoning. I. J. of Uncertainty, Fuzziness and Knowledge-Based Systems **2** (1994) 167–196
17. de Cooman, G., Troffaes, M., Miranda, E.: n-monotone lower previsions and lower integrals. In Cozman, F., Nau, R., Seidenfeld, T., eds.: Proc. 4th International Symposium on Imprecise Probabilities and Their Applications. (2005)