Completely monotone outer approximations of lower probabilities on finite possibility spaces

Erik Quaeghebeur[1]

Abstract Drawing inferences from general lower probabilities on finite possibility spaces usually involves solving linear programming problems. For some applications this may be too computationally demanding. Some special classes of lower probabilities allow for using computationally less demanding techniques. One such class is formed by the completely monotone lower probabilities, for which inferences can be drawn efficiently once their Möbius transform has been calculated. One option is therefore to draw approximate inferences by using a completely monotone approximation to a general lower probability; this must be an outer approximation to avoid drawing inferences that are not implied by the approximated lower probability. In this paper, we discuss existing and new algorithms for performing this approximation, discuss their relative strengths and weaknesses, and illustrate how each one works and performs.

Key words: lower probabilities, outer approximation, complete monotonicity, belief functions, Möbius transform

1 Introduction

In the theory of coherent lower previsions—or, more colloquially, of imprecise probabilities—the procedure of natural extension is the basic technique for drawing inferences [11, §3.1]. In a finitary setting, i.e., one with a finite possibility space Ω and in which the lower prevision $\underline{P}$ is assessed for a finite collection of gambles (random variables) $\mathscr{K} \subseteq \mathbb{R}^\Omega$, calculating the natural extension $\underline{E}_{\underline{P}} f$ for a gamble f in $\mathbb{R}^\Omega$ corresponds to solving a linear programming (LP) problem:

$$\underline{E}_{\underline{P}} f := \max\left\{\alpha \in \mathbb{R} : f - \alpha \geq \textstyle\sum_{g \in \mathscr{K}} \lambda_g \cdot (g - \underline{P} g), \lambda \in \mathbb{R}^{\mathscr{K}}_{\geq 0}\right\}. \tag{1}$$

SYSTeMS Research Group, Ghent University, Technologiepark-Zwijnaarde 914, 9052 Gent, Erik.Quaeghebeur@UGent.be

In applications where $\mathscr{K}$ is large or where the natural extension needs to be calculated for a large number of gambles, such LP problems may be too computationally demanding. But natural extension preserves dominance: if $\underline{P}_* \leq \underline{P}$ then $\underline{E}_{\underline{P}_*} \leq \underline{E}_{\underline{P}}$.

So first we restrict $\mathscr{K}$ and only consider lower previsions $\underline{P}$ that are defined on the set of (indicators of) events; i.e., we only consider lower probabilities $\underline{P}$ defined on the power set 2^Ω. Now, for 2-monotone lower probabilities $\underline{P}_*$, which form a subclass of the coherent lower probabilities, the natural extension $\underline{E}_{\underline{P}_*}$ can be calculated more efficiently using Choquet integration [see, e.g., 11, §3.2.4]:

$$\underline{E}_{\underline{P}_*} f = (C) \int f \mathrm{d}\underline{P}_* := \min f + \int_{\min f}^{\max f} \underline{P}_* \{\omega \in \Omega : f\omega \geq t\} \mathrm{d}t, \tag{2}$$

So if we can find a 2-monotone outer approximation $\underline{P}_*$ to $\underline{P}$, i.e., such that $\underline{P}_* A \leq \underline{P} A$ for all events $A \subseteq \Omega$, we can efficiently calculate the outer approximation $\underline{E}_{\underline{P}_*}$ to $\underline{E}_{\underline{P}}$.

How can we go about this? Every coherent lower prevision $\underline{P}$ can be written as a convex combination of extreme coherent lower previsions [7] that is not-necessarily unique [e.g., 6, §2.3.3, ¶4]; in the finitary case, the set $\mathscr{E}^{\mathrm{c}}(\mathscr{K})$ of extreme coherent lower previsions on $\mathscr{K}$ is finite. So $\underline{P} = \sum_{\underline{Q} \in \mathscr{E}^{\mathrm{c}}(\mathscr{K})} \lambda \underline{Q} \cdot \underline{Q}$, where $\lambda : \mathscr{E}^{\mathrm{c}}(\mathscr{K}) \to [0,1]$ is a function that generates coefficients of a convex $\mathscr{E}^{\mathrm{c}}(\mathscr{K})$-decomposition of $\underline{P}$. The same holds for 2-monotone lower probabilities, but with a different set of extreme members $\mathscr{E}^2(\Omega)$ [8, 6]. The idea is to find a $\nu : \mathscr{E}^2(\Omega) \to [0,1]$ such that $\underline{P}_* := \sum_{\underline{Q} \in \mathscr{E}^2(\Omega)} \nu \underline{Q} \cdot \underline{Q}$ is an—in some sense—good outer approximation to $\underline{P}$.

It is impractical to consider all elements of $\mathscr{E}^2(\Omega)$: finding this set is computationally very demanding and with increasing $|\Omega|$ it quickly becomes very large [8, §4]. In this paper, our strategy is to only retain the subclass $\mathscr{E}^\infty(\Omega)$ of vacuous lower probabilities: each such lower prevision essentially corresponds to an assessment that a given event A of Ω occurs; the corresponding natural extension is given by $\underline{E}_A f := \min_{\omega \in A} f\omega$. Lower probabilities $\underline{P}_*$ that can be written as a convex combination of vacuous lower probabilities are called completely monotone. The decomposition of such a lower probability, i.e., the coefficient function $\nu : 2^\Omega \to [0,1]$, is unique and determines it as follows:

$$\underline{P}_* A = \sum_{B \subseteq A} \nu B, \qquad \underline{E}_{\underline{P}_*} f = \sum_{B \subseteq \Omega} \nu B \cdot \underline{E}_B f = \sum_{B \subseteq \Omega} \nu B \cdot \min_{\omega \in B} f\omega. \tag{3}$$

The left-hand equation is called Möbius inversion; the right-hand one is an alternative to Choquet integration for calculating the natural extension.

Mathematically, completely monotone lower probabilities coincide with the belief functions of Dempster–Shafer theory [4, 9]. From this theory, we know that the coefficients of the decomposition—which we call basic belief mass assignments in this paper—can be obtained by using the Möbius transform of $\underline{P}_*$; i.e., the coefficient of $\underline{E}_A$ is

$$\nu A = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \cdot \underline{P}_* B = \underline{P}_* A - \sum_{B \subset A} \nu B, \tag{4}$$

where the last expression shows how these coefficients can be calculated recursively. Obviously, $\underline{P}_*$ must be defined for all events to calculate these coefficients; if necessary, one should extend it to all indicator functions first to obtain the lower probability on all events.

In this paper, we assume a lower probability $\underline{P}$ is given that is defined on all events. We discuss a number of algorithms that allow us to obtain a basic belief mass assignment function ν that determines—via Equation (3)—a completely monotone lower probability $\underline{P}_*$ that is an outer approximation to $\underline{P}$, i.e., $\underline{P}_* \leq \underline{P}$. The reason we focus on outer approximations is that they are conservative in the sense that they do not lead to inferences unwarranted by the approximated lower probability. For the algorithms to work, it is sufficient that $\underline{P}$ satisfies $\sum_{\omega\in\Omega} \underline{P}\{\omega\} \leq 1$, is nonnegative ($\underline{P} \geq 0$), monotone ($B \subseteq A \Rightarrow \underline{P}B \leq \underline{P}A$), and normed ($\underline{P}\emptyset = 0$ and $\underline{P}\Omega = 1$), all four of which we assume to be the case.

2 Completely monotone outer approximation algorithms

We are going to discuss four algorithms—one trivial new one, two from the literature, and one substantive new one—that fall into three classes: the first one creates a linear-vacuous mixture, the second one reduces the problem to an LP problem, and the last two are based on modifications of the Möbius transformation and are more heuristic in nature. (All algorithms have been implemented in Troffaes's `imprecise` software package/framework [10].) But before jumping into this material, we discuss a useful preprocessing step and introduce the lower probabilities that are used to illustrate (the results) of the techniques.

First the preprocessing step: we mentioned that the decomposition into extreme coherent lower previsions of a coherent lower prevision $\underline{P}$ is in general non-unique. However, the coefficients $\lambda\{\omega\}$ of the degenerate lower previsions—i.e., vacuous lower previsions relative to singletons $\{\omega\}$ of Ω—are unique [7, Prop. 1], so we can write any coherent lower probability as a linear-imprecise mixture:

$$\underline{P}A = \kappa\cdot\textstyle\sum_{\omega\in A} p\omega + (1-\kappa)\cdot\underline{R}A, \qquad \underline{E}_{\underline{P}}f = \kappa\cdot\textstyle\sum_{\omega\in\Omega} p\omega\cdot f\omega + (1-\kappa)\cdot\underline{E}_{\underline{R}}f, \quad (5)$$

where $\kappa := \sum_{\omega\in\Omega}\lambda\{\omega\}$ and $p\omega := \lambda\{\omega\}/\kappa$, which allows us to solve for the imprecise part $\underline{R}$. This is a coherent lower probability whose lower probability on singletons is zero. (The second equation then follows from [11, §3.4.1].) In case $\underline{P}$ is not coherent, $\underline{R}$ may end up with negative values, but we may infer from the zero values of $\underline{R}$ on singletons that these may be set to zero. Given that linear previsions are completely monotone, it makes sense to separate out the linear part—represented by the probability mass function p—and only approximate the imprecise part.

We use some example lower probabilities to illustrate the algorithms. All have $\Omega = \{a,b,c,d\}$: a cardinality of four allows the lower probabilities to be complex enough to be interesting without resulting in unending lists of numbers. (Also, for $|\Omega| < 4$, all lower probabilities are probability intervals and therefore 2-monotone [2, Prop. 5].) The first one in Table 1 is from the literature [5, Ex. 2], for which we also give the linear-imprecise decomposition; the second and third ones in Table 2 are especially chosen extreme coherent lower probabilities [8, 6, App. A] that highlight some of the algorithms' features and consist of an imprecise part only.

Table 1 A lower probability $\underline{P}$, its linear-precise decomposition $(p, \underline{R}, \kappa = 0.737)$, the 2-monotone probability interval outer approximation $\underline{R}_{\mathrm{PI}}$ of $\underline{R}$ (generated by efficient natural extension from $[\underline{R}\{\omega\}, \overline{R}\{\omega\} = 1 - \underline{R}(\Omega \setminus \{\omega\})]_{\omega\in\Omega}$ [cf. 2, particularly Prop. 4]), the Möbius transform ρ of $\underline{R}$, and six completely monotone outer approximations of $\underline{R}$: the linear-vacuous one $\underline{R}_{\mathrm{LV}}$; two optimal ones, $\underline{R}_{\mathrm{LP}^{\mathrm{DS}}}$ using a dual simplex solver and $\underline{R}_{\mathrm{LP}^{\mathrm{CC}}}$ using a criss-cross solver; two IRM-approximations, $\underline{R}_{\mathrm{IRM}}$ using the lexicographic order and $\pi\underline{R}_{\mathrm{IRM}}$ the inverse order; the IMRM-approximation $\underline{R}_{\mathrm{IMRM}}$. Negative values in the Möbius transform ρ of $\underline{R}$ are highlighted with a gray background. Approximation values that differ from the approximated values of $\underline{R}$ are in boldface. The last row contains the sum-norm-differences between $\underline{R}$ and the approximations. The number of significant digits used has been chosen to facilitate comparisons and verification.

Event	$\underline{P}$	p	$\underline{R}$	ρ	$\underline{R}_{\mathrm{LV}}$	$\underline{R}_{\mathrm{LP}^{\mathrm{DS}}}$	$\underline{R}_{\mathrm{LP}^{\mathrm{CC}}}$	$\underline{R}_{\mathrm{IRM}}$	$\pi\underline{R}_{\mathrm{IRM}}$	$\underline{R}_{\mathrm{IMRM}}$	$\underline{R}_{\mathrm{PI}}$
a	0.0895	0.122									
b	0.2743	0.372									
c	0.2668	0.362									
d	0.1063	0.144									
$a\ b$	0.3947		0.117	0.117	**0**	**0**	0.117	**0.046**	**0.091**	**0.066**	**0**
$a\ c$	0.4506		0.358	0.358	**0**	**0.196**	**0.079**	**0.185**	**0.193**	**0.211**	**0**
$a\ d$	0.2959		0.381	0.381	**0**	**0.352**	**0.352**	**0.242**	**0.249**	**0.244**	**0.129**
$b\ c$	0.5837		0.162	0.162	**0**	0.162	0.162	**0.074**	**0.074**	**0.082**	**0**
$b\ d$	0.4835		0.391	0.391	**0**	**0.227**	**0.110**	**0.219**	**0.216**	**0.227**	**0**
$c\ d$	0.4079		0.132	0.132	**0**	**0.002**	**0.119**	**0.099**	**0.051**	**0.081**	**0**
$a\ b\ c$	0.7248		0.358	−0.280	**0**	0.358	0.358	**0.305**	0.358	0.358	0.358
$a\ b\ d$	0.6224		0.579	−0.310	**0**	0.579	0.579	**0.507**	**0.556**	0.579	0.579
$a\ c\ d$	0.6072		0.550	−0.322	**0**	0.550	0.550	**0.526**	**0.493**	0.550	0.550
$b\ c\ d$	0.7502		0.391	−0.295	**0**	0.391	0.391	0.391	**0.341**	0.391	0.391
$a\ b\ c\ d$	1		1	0.664	1	1	1	1	1	1	1
$\|\underline{R} - \underline{R}_*\|_1$					3.419	0.603	0.603	0.827	0.797	0.631	1.413

Table 2 Two lower probabilities $\underline{P}$—the one on the left is permutation invariant [cf., e.g., 6, §2.2.6]—with their Möbius transforms μ and IMRM-approximations $\underline{P}_{\mathrm{IMRM}}$. On the left moreover two other completely monotone outer approximations, an optimal one $\underline{P}_{\mathrm{LP}}$ and an IRM-approximation $\underline{P}_{\mathrm{IRM}}$. On the right moreover the Möbius transform ν_6 of the IMRM-approximation and two intermediate basic belief mass assignments (ν_3 and ν_5) used in its construction. (N.B.: for this lower probability, we have $\|\underline{P} - \underline{P}_{\mathrm{LP}}\|_1 = 1/2$, $\|\underline{P} - \underline{P}_{\mathrm{IRM}}\|_1 = 3/4$.) Other table elements and stylings have the same meaning as in Table 1.

Event	$\underline{P}$	μ	$\underline{P}_{\mathrm{LP}}$	$\underline{P}_{\mathrm{IRM}}$	$\underline{P}_{\mathrm{IMRM}}$	$\underline{P}$	μ	ν_3	ν_5	ν_6	$\underline{P}_{\mathrm{IMRM}}$
$a\ b$	1/3	1/3	**0**	**1/10**	**1/6**	1/2	1/2	1/2	1/4	1/4	**1/4**
$a\ c$	1/3	1/3	**1/6**	**0.95/8**	**1/6**	1/4	1/4	1/4	1/8	1/8	**1/8**
$a\ d$	1/3	1/3	1/3	**1/7**	**1/6**	1/4	1/4	1/4	1/8	1/8	**1/8**
$b\ c$	1/3	1/3	1/3	**0.96/7**	**1/6**	1/4	1/4	1/4	1/8	1/8	**1/8**
$b\ d$	1/3	1/3	**1/6**	**0.99/6**	**1/6**	0	0	0	0	0	0
$c\ d$	1/3	1/3	**0**	**0.98/5**	**1/6**	0	0	0	0	0	0
$a\ b\ c$	1/2	−1/2	1/2	**1.07/3**	1/2	1/2	−1/2	−1/2	0	0	1/2
$a\ b\ d$	1/2	−1/2	1/2	**0.82/2**	1/2	3/4	0	0	3/8	9/32	**21/32**
$a\ c\ d$	1/2	−1/2	1/2	**0.91/2**	1/2	1/4	−1/4	−1/4	0	0	1/4
$b\ c\ d$	1/2	−1/2	1/2	1/2	1/2	1/4	0	0	1/8	3/32	**7/32**
$a\ b\ c\ d$	1	0	1	1	1	1	1/2		−1/8	0	1
$\|\underline{P} - \underline{P}_*\|_1$			1	99/70	1						3/4

In the tables, we have also given the Möbius transform μ of each of these lower probabilities—or the transform ρ of their imprecise part—by applying Equation (4). As is the case for all lower probabilities that are not completely monotone, some of the basic belief mass assignments so obtained are negative, but they still sum up to one [3, 9]. The algorithms we discuss all essentially construct a nonnegative basic belief mass assignment function which can be seen as resulting from shifting positive mass up in the poset of events ordered by inclusion (or shifting negative mass down) to compensate the negative mass assignments. This mass shifting is also the basic idea behind the last two algorithms.

Linear-vacuous approximation The first algorithm is trivial: it consists in replacing a lower probability's imprecise part by the vacuous lower probability, which is identically zero except in Ω, where it is 1. In terms of mass shifts, all mass of non-singletons is shifted up to the event poset's top Ω.

Table 1 contains a—due to the triviality—not very interesting illustration.

Approximation via optimization The second algorithm is based on the formulation of the problem as an optimization problem: we wish to find a nonnegative basic belief mass assignment function ν such that its Möbius inverse $\underline{P}_{\mathrm{LP}}$ minimizes some distance to the approximated lower probability $\underline{P}$. We can force $\underline{P}_{\mathrm{LP}}$ to be an outer approximation by adding constraints that express its dominance by $\underline{P}$. By choosing the distance to be a linear function of ν's components, the optimization problem becomes an LP problem [1, §7]; we choose the sum-norm-distance:

$$\begin{aligned}
\nu &= \operatorname{argmin}\left\{ \|\underline{P} - \underline{P}_{\mathrm{LP}}\|_1 : \underline{P}_{\mathrm{LP}} \leq \underline{P} \text{ and } \underline{P}_{\mathrm{LP}} \text{ is completely monotone} \right\} \qquad (6)\\
&= \operatorname{argmin}\left\{ \textstyle\sum_{A\subseteq\Omega} |\underline{P}A - \underline{P}_{\mathrm{LP}}A| : \underline{P}_{\mathrm{LP}} \leq \underline{P} \text{ and } \underline{P}_{\mathrm{LP}} \text{ is completely monotone} \right\}\\
&= \operatorname{argmax}\left\{ \textstyle\sum_{A\subseteq\Omega} \underline{P}_{\mathrm{LP}}A : \underline{P}_{\mathrm{LP}} \leq \underline{P} \text{ and } \underline{P}_{\mathrm{LP}} \text{ is completely monotone} \right\}\\
&= \operatorname{argmax}\left\{ \textstyle\sum_{A\subseteq\Omega} \sum_{B\subseteq A} \nu B : \forall_{A\subseteq\Omega} (\sum_{B\subseteq A} \nu B \leq \underline{P}A) \text{ and } \nu \geq 0, \sum_{B\subseteq\Omega} \nu B = 1 \right\}\\
&= \operatorname{argmax}\left\{ \textstyle\sum_{B\subseteq\Omega} 2^{|\Omega\setminus B|} \nu B : \forall_{A\subseteq\Omega} (\sum_{B\subseteq A} \nu B \leq \underline{P}A) \text{ and } \nu \geq 0, \sum_{B\subseteq\Omega} \nu B = 1 \right\}, \qquad (7)
\end{aligned}$$

where $2^{|\Omega\setminus B|}$ is the number of events A that contain B. The third equality follows from taking into account the dominance constraints; the fourth from making the dependence on ν explicit using Equation (4). The linear-vacuous approximation shows that this linear program is feasible.

The results of this optimization approach are given for the lower probability in Table 1 and the one on the left in Table 2. In Table 1, two *differing* 'optimal' outer approximations are given, resulting from using different LP solvers. The optimal outer approximation given for the permutation invariant lower probability in Table 2 on the left is not permutation invariant itself. Both are due to the fact that in general there is no unique optimal solution and that solvers return the first one reached, which for the typical (non-interior-point) methods used lies on the border of the convex set of solutions.

The sum-norm distance can also be used as a quality criterion—one that obviously does not take symmetry aspects into account—for other approximation tech-

niques. It has therefore also been calculated and included for the other approximations in Table 1 and 2; $\|\underline{P} - \underline{P}_{\mathrm{LP}}\|_1$ and $\|\underline{P} - \underline{P}_{\mathrm{LV}}\|_1$ provide lower and upper bounds.

Iterative rescaling method The Iterative Rescaling Method or IRM [5] builds on the recursive Möbius transform formula in Equation (4), interrupting it to shift mass whenever negative mass assignments are encountered for some event.

Algorithm 1: $\mathrm{IRM}(\Omega, \underline{P}) := \underline{P}_{\mathrm{IRM}}$

Form a sequence A of length $2^{|\Omega|}$ by ordering all events in 2^Ω by increasing cardinality and arbitrarily for events of equal cardinality and set $\nu\emptyset := \underline{P}\emptyset = 0$
for $i := 1$ **to** $2^{|\Omega|} - 1$ **do** $\nu A_i := \underline{P}A_i - \sum_{B \subset A_i} \nu B$
  if $\nu A_i < 0$ **then** $(\ell, \alpha) := \mathrm{MassBasin}(A_i, \nu)$
    foreach $B \subset A_i : |B| \geq \ell$ **do** $\nu B := \frac{\alpha + \nu A_i}{\alpha} \cdot \nu B$
    $\nu A_i := 0$
return *the Möbius inverse $\underline{P}_{\mathrm{IRM}}$ of ν*

In the **if**-block, the negative mass νA_i is distributed proportionally to its subevents of a for compensation lowest needed cardinality and up; i.e.,

$$\ell := \max\{k < |A_i| : \textstyle\sum_{B \subset A : |B| \geq k} \nu B =: \alpha_k > \nu A_i\}, \qquad \alpha := \alpha_\ell.$$

For clarity, we have separated out the algorithm that calculates these parameters:

Algorithm 2: $\mathrm{MassBasin}(A, \nu) := (\ell, \alpha)$

Set $\ell := |A|$ and $\alpha := 0$
while $\alpha < -\nu A$ **do** $\ell := \ell - 1$ *and* $\alpha := \alpha + \sum_{B \subset A : |B| = \ell} \nu B$
return *the (lowest needed) cardinality ℓ and the compensation mass α*

The results of the IRM-algorithm are given for the lower probability in Table 1 and the one on the left in Table 2. In Table 1, two *differing* outer approximations are given, resulting from using different 'arbitrary' orderings of the events of equal cardinality. Using the sum-norm criterion, we see that the quality of the approximation depends on the order chosen. Also, the outer approximation given for the permutation invariant lower probability in Table 2 on the left is not permutation invariant itself, reflecting the impact of the arbitrary order.

Furthermore, it can be seen in the Tables that for events of a cardinality for which the optimization approximation is always exact, this is not so for the IRM-approximation; there only the last such event of the arbitrary order is exact. This is due to the fact that the IRM-algorithm does not backtrack to recalculate the mass assignments for an event after rescaling some of its subevents due to negative masses encountered for *subsequent* events.

Iterative minimal rescaling method Inspired by the IRM, we have designed an approximation algorithm that avoids its defects mentioned above, at the cost of increased complexity. Furthermore, our algorithm is permutation-invariant, so it improves on the LP approach as well, in that regard.

The algorithm is still based on the recursive Möbius transform formula in Equation (4), but the rescaling approach is a bit more involved than with the IRM and gives rise to a higher number of these recursion calculations:

Algorithm 3: IMRM$(\Omega, \underline{P}) := \underline{P}_{\text{IMRM}}$

Set $\nu\emptyset := \underline{P}\emptyset = 0$ *and* $k := 1$
while $k \leq |\Omega|$ **do** $\mathscr{A} := \emptyset$
  foreach $A \subseteq \Omega : |A| = k$ **do** $\nu A := \underline{P}A - \sum_{B \subset A} \nu B$
  $\mathscr{A} := \{A \subseteq \Omega : |A| = k \wedge \nu A < 0\}$
  if $\mathscr{A} = \emptyset$ **then** $k := k + 1$
  else foreach $A \in \mathscr{A}$ **do** $(\ell_A, \alpha_A) := \text{MassBasin}(A, \nu)$
    $\ell := \min_{A \in \mathscr{A}} \ell_A$ *and* $\mathscr{B} := \{A \in \mathscr{A} : \ell_A = \ell\}$
    foreach $A \in \mathscr{B}$ **do** $\beta_A := \sum_{B \subset A : |B| = \ell} \nu B$
    foreach $B \in \bigcup_{A \in \mathscr{B}} 2^A : |B| = \ell$ **do** $\nu B := \max_{A \in \mathscr{B} : B \subset A} \frac{\alpha_A + \nu A}{\beta_A} \cdot \nu B$
    $k := \ell + 1$
return *the Möbius inverse* $\underline{P}_{\text{IMRM}}$ *of* ν

Per cardinality k, all basic belief mass assessments are calculated before doing any rescaling due to negative masses encountered for the events in $\mathscr{A}$. To limit the mass loss for events of lowest needed cardinality ℓ—i.e., those most heavily penalized by the sum-norm criterion—, only their masses are rescaled during that iteration of the **while**-loop, which is the reason to restrict attention to $\mathscr{B}$. For a single event A of $\mathscr{B}$, the mass loss for its cardinality-ℓ subevents is limited by only shifting that mass down which cannot be compensated higher up, which explains the scaling factor β_A used. We avoid overcompensation of negative mass in one element of $\mathscr{B}$ due to a bigger deficit in another by using the largest scaling factor available, which leads to a minimal rescaling. This last point is what lead us to name the algorithm the Iterative Minimal Rescaling Method or IMRM. In general, the IMRM will not be as good as the optimization approach in terms of the sum-norm criterion: the mass is still shifted proportionally, which is not necessarily optimal.

The lower probability $\underline{P}_{\text{IMRM}}$ obtained is indeed completely monotone, because for the basic belief mass assignment function we have $\nu\emptyset = 0$ from the start, $\nu A \geq 0$ for all events such that $|A| < k$ at the end of each iteration of the **while**-loop, and $\nu\Omega = \underline{P}\Omega - \sum_{B \subset \Omega} \nu B = 1 - \sum_{B \subset \Omega} \nu B$ at the end of the last iteration. It is an outer approximation because the recursion formula used tries to make $\underline{P}_{\text{IMRM}}A$ equal to $\underline{P}A$ for all $A \subseteq \Omega$; subsequent rescalings can only lower this value.

The results of the IMRM-algorithm are given for the lower probability in Table 1 and both in Table 2. The positive impact of the algorithm's permutation invariance is especially clear for the left lower probability of Table 2. The algorithm itself is illustrated on Table 2's right side; there ν is given as it exists at the moment $\mathscr{A} = \emptyset$ is checked for the third, fifth, and sixth—final—iteration of the while loop; the impact of negative mass values on subsequent iterates is the prime point of interest here.

3 Conclusion

We have introduced and illustrated the linear-imprecise decomposition of lower probabilities, and the linear-vacuous and IMRM-algorithms for generating completely monotone outer approximations to lower probabilities. We have compared these to algorithms in the literature; permutation invariance is their main advantage.

One thing that still needs to be done is a complexity and parallelizability analysis to get a view of the relative computational burden of each of the algorithms discussed. Also interesting to investigate is the use of other objective functions in the optimization approach—e.g., using other norms, engendering nonlinear convex optimization problems—; this could lead to uniqueness of the solution and therewith permutation invariance. Both would allow for a more informed choice between the different possible completely monotone outer approximation algorithms.

Acknowledgements Many thanks to Matthias Troffaes for making his software package `improb` publicly available, for assisting me in getting familiar with the code—so I could implement the algorithms described in this paper—, and for his very useful feedback. Thanks to Jim Hall and Jonathan Lawry for their helpful responses to my questions pertaining to their IRM-algorithm paper. Thanks to Enrique Miranda and a reviewer for their useful comments and suggestions.

References

1. Baroni, P., Vicig, P.: An uncertainty interchange format with imprecise probabilities. Internat. J. Approx. Reason. 40, 147–180 (2005) doi: 10.1016/j.ijar.2005.03.001
2. de Campos, L.M., Huete, J.F., Moral, S.: Probability intervals: A tool for uncertain reasoning. Internat. J. Uncertain., Fuzziness Knowledge-Based Systems 2, 167–196 (1994) doi: 10.1142/S0218488594000146
3. Chateauneuf, A., Jaffray, J.Y.: Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. Math. Social Sci. 17, 263–283 (1989) doi: 10.1016/0165-4896(89)90056-5
4. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. 38, 325–339 (1967) doi: 10.1214/aoms/1177698950
5. Hall, J.W., Lawry, J.: Generation, combination and extension of random set approximations to coherent lower and upper probabilities. Reliab. Eng. Syst. Safety 85, 89–101 (2004) doi: 10.1016/j.ress.2004.03.005
6. Quaeghebeur, E.: Learning from samples using coherent lower previsions. PhD thesis, Ghent University (2009) url: http://hdl.handle.net/1854/LU-495650
7. Quaeghebeur, E.: Characterizing the set of coherent lower previsions with a finite number of constraints or vertices. In: Spirtes, P., Grünwald, P. (eds) Proceedings of UAI 2010, 466–473, AUAI Press (2010) url: http://hdl.handle.net/1854/LU-984156
8. Quaeghebeur, E., De Cooman, G.: Extreme lower probabilities. Fuzzy Sets and Systems 159, 2163–2175 (2008) doi: 10.1016/j.fss.2007.11.020 url: http://hdl.handle.net/1854/11713
9. Shafer, G.: A mathematical theory of evidence. Princeton University Press (1976)
10. Troffaes, M., Quaeghebeur, E.: `improb`: A python module for working with imprecise probabilities (2011) url: https://github.com/equaeghe/improb, fork of https://github.com/mcmtroffaes/improb. Latest public release at http://packages.python.org/improb/
11. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman & Hall (1991)