Approximation of Belief Functions

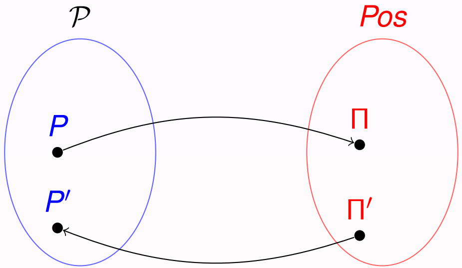
Serafín Moral
7th School on Belief Functions and their Applications

Granada

BFTA 2025, Granada, 19-23 October

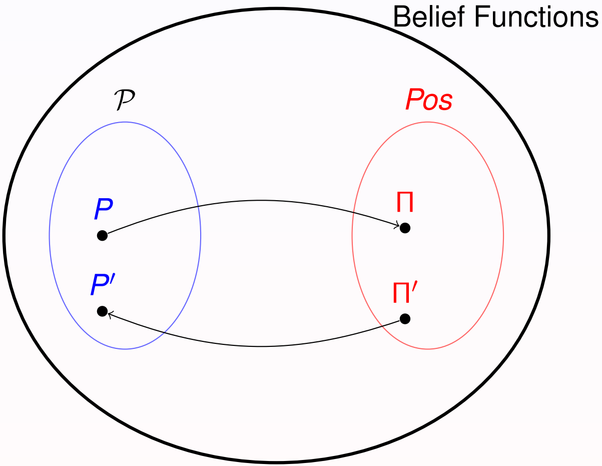
My PhD Thesis

"Información difusa, relaciones entre probabilidad y posibilidad".
University of Granada, 1985



My PhD Thesis

"Información difusa, relaciones entre probabilidad y posibilidad".
University of Granada, 1985

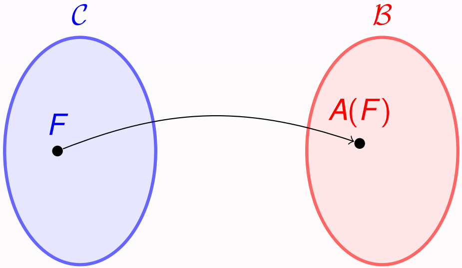


The Problem

We have two spaces $\mathcal{C}$ and $\mathcal{B}$ of uncertainty representations, we aim to produce a mapping:

$$A : \mathcal{C} \to \mathcal{B}$$

In such a way that if $F \in \mathcal{C}$, then $A(F)$ should be as 'similar' as possible to F.



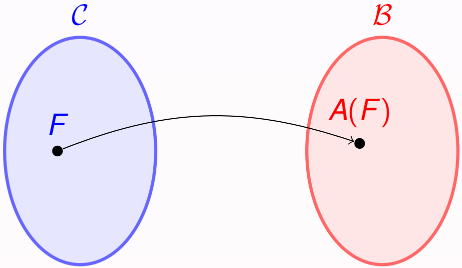
For example, if $F \in \mathcal{B}$ we should have $A(F) = F$.

The Problem

We have two spaces $\mathcal{C}$ and $\mathcal{B}$ of uncertainty representations, we aim to produce a mapping:

$$A : \mathcal{C} \rightarrow \mathcal{B}$$

In such a way that if $F \in \mathcal{C}$, then $A(F)$ should be as 'similar' as possible to F.



For example, if $F \in \mathcal{B}$ we should have $A(F) = F$.

Example

Fact

If a belief function is represented by its mass assignment m, then this representation is efficient if the number of focal elements is small.

So, we can consider that:

- $\mathcal{C}$ is the space of all belief functions on a frame.
- $\mathcal{B}$ is the space of all belief functions on a frame with a number of focal elements limited to K.

$$m(\{x_1\}) = 0.1, \quad m(\{x_1, x_2\} = 0.2, \quad m(\{x_2, x_4\}) = 0.1,$$
$$m(\{x_1, x_2, x_4\}) = 0.2, \quad m(\{x_2, x_3, x_4\} = 0.1, \quad m(\{x_1, x_2, x_3, x_4\} = 0.1$$

$A(m) = m'$ with 4 focal elements?

Example

Fact

If a belief function is represented by its mass assignment m, then this representation is efficient if the number of focal elements is small.

So, we can consider that:

- $\mathcal{C}$ is the space of all belief functions on a frame.
- $\mathcal{B}$ is the space of all belief functions on a frame with a number of focal elements limited to K.

$$m(\{x_1\}) = 0.1, \quad m(\{x_1, x_2\} = 0.2, \quad m(\{x_2, x_4\}) = 0.1,$$
$$m(\{x_1, x_2, x_4\}) = 0.2, \quad m(\{x_2, x_3, x_4\} = 0.1, \quad m(\{x_1, x_2, x_3, x_4\} = 0.1$$

$A(m) = m'$ with 4 focal elements?

Outline

- The relation between and information and its approximation
- Measures of divergence
- Putting a limit to the number of focal elements
- Outer approximating interval probabilities by a belief function
- Inner approximating interval probabilities by a belief function
- Partially specified belief functions
- Combination as approximation
- Approximating belief functions by consonant beliefs (possibilities)?

Outline

- The relation between and information and its approximation
- Measures of divergence
- Putting a limit to the number of focal elements
- Outer approximating interval probabilities by a belief function
- Inner approximating interval probabilities by a belief function
- Partially specified belief functions
- Combination as approximation
- Approximating belief functions by consonant beliefs (possibilities)?

Outline

- The relation between and information and its approximation
- Measures of divergence
- Putting a limit to the number of focal elements
- Outer approximating interval probabilities by a belief function
- Inner approximating interval probabilities by a belief function
- Partially specified belief functions
- Combination as approximation
- Approximating belief functions by consonant beliefs (possibilities)?

Outline

- The relation between and information and its approximation
- Measures of divergence
- Putting a limit to the number of focal elements
- Outer approximating interval probabilities by a belief function
- Inner approximating interval probabilities by a belief function
- Partially specified belief functions
- Combination as approximation
- Approximating belief functions by consonant beliefs (possibilities)?

Outline

- The relation between and information and its approximation
- Measures of divergence
- Putting a limit to the number of focal elements
- Outer approximating interval probabilities by a belief function
- Inner approximating interval probabilities by a belief function
- Partially specified belief functions
- Combination as approximation
- Approximating belief functions by consonant beliefs (possibilities)?

Outline

- The relation between and information and its approximation
- Measures of divergence
- Putting a limit to the number of focal elements
- Outer approximating interval probabilities by a belief function
- Inner approximating interval probabilities by a belief function
- Partially specified belief functions
- Combination as approximation
- Approximating belief functions by consonant beliefs (possibilities)?

Outline

- The relation between and information and its approximation
- Measures of divergence
- Putting a limit to the number of focal elements
- Outer approximating interval probabilities by a belief function
- Inner approximating interval probabilities by a belief function
- Partially specified belief functions
- Combination as approximation
- Approximating belief functions by consonant beliefs (possibilities)?

Outline

- The relation between and information and its approximation
- Measures of divergence
- Putting a limit to the number of focal elements
- Outer approximating interval probabilities by a belief function
- Inner approximating interval probabilities by a belief function
- Partially specified belief functions
- Combination as approximation
- Approximating belief functions by consonant beliefs (possibilities)?

Relation between F and $A(F)$

Relation between F and $A(F)$

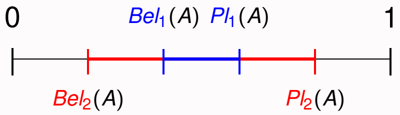
Basic Question

Should $A(F)$ be less or more informative than F?

We are going to consider two definitions of 'more informative'.

Belief Based Ordering

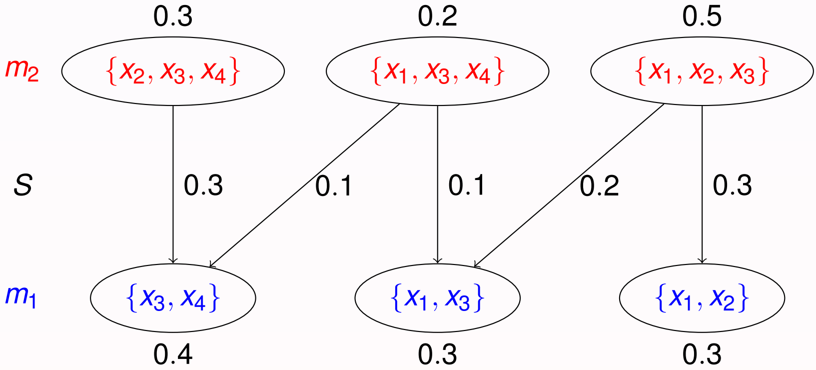
$$m_1 \sqsubseteq_{Bel} m_2 \Leftrightarrow Bel_1(A) \geq Bel_2(A), \forall A \subseteq \Omega.$$



Strong Ordering (Moral, 1985; Yager, 1986)

$m_1 \sqsubseteq_s m_2$ if and only if, there is a $S : 2^\Omega \times 2^\Omega \rightarrow [0, 1]$ such that

- $S(A, B) > 0$ only if $A \subseteq B$,
- $\sum_B S(A, B) = m_1(A)$,
- $\sum_A S(A, B) = m_2(B)$.



Which concept?

- There are more concepts of 'more informative'.
- We know that $m_1 \sqsubseteq_s m_2 \Rightarrow m_1 \sqsubseteq_{Bel} m_2$.
- We will consider both concepts.
- Strong ordering is more according to the belief functions fundamentals.
- Belief based ordering is sometimes more appropriate from a computational point of view and it is useful to compare with other formalisms as interval probabilities: $Bel(A) \geq \underline{P}(A)$.

Which concept?

- There are more concepts of 'more informative'.
- We know that $m_1 \sqsubseteq_s m_2 \Rightarrow m_1 \sqsubseteq_{Bel} m_2$.
- We will consider both concepts.
- Strong ordering is more according to the belief functions fundamentals.
- Belief based ordering is sometimes more appropriate from a computational point of view and it is useful to compare with other formalisms as interval probabilities: $Bel(A) \geq \underline{P}(A)$.

Which concept?

- There are more concepts of 'more informative'.
- We know that $m_1 \sqsubseteq_s m_2 \Rightarrow m_1 \sqsubseteq_{Bel} m_2$.
- We will consider both concepts.
- Strong ordering is more according to the belief functions fundamentals.
- Belief based ordering is sometimes more appropriate from a computational point of view and it is useful to compare with other formalisms as interval probabilities: $Bel(A) \geq \underline{P}(A)$.

Which concept?

- There are more concepts of 'more informative'.
- We know that $m_1 \sqsubseteq_s m_2 \Rightarrow m_1 \sqsubseteq_{Bel} m_2$.
- We will consider both concepts.
- Strong ordering is more according to the belief functions fundamentals.
- Belief based ordering is sometimes more appropriate from a computational point of view and it is useful to compare with other formalisms as interval probabilities: $Bel(A) \geq \underline{P}(A)$.

Which concept?

- There are more concepts of 'more informative'.
- We know that $m_1 \sqsubseteq_s m_2 \Rightarrow m_1 \sqsubseteq_{Bel} m_2$.
- We will consider both concepts.
- Strong ordering is more according to the belief functions fundamentals.
- Belief based ordering is sometimes more appropriate from a computational point of view and it is useful to compare with other formalisms as interval probabilities: $Bel(A) \geq \underline{P}(A)$.

Relation between F and $A(F)$

Basic Question

Should $A(F)$ be less or more informative than F?

In general, it is assumed that $A(F)$ should be less informative than F: **No new information is created by the approximation.** Outer Approximation.

Additional question?

It is always the case?

Not always: Sometimes we want to approximate a belief function by an additive probability (pignistic probability) or we want something similar to natural extension in imprecise probability,

Initial Specification + Belief function (hypothesis) $\Longrightarrow$ More informative

More informative $\rightarrow$ Inner Approximation

Relation between F and $A(F)$

Basic Question

Should $A(F)$ be less or more informative than F?

In general, it is assumed that $A(F)$ should be less informative than F: **No new information is created by the approximation.** Outer Approximation.

Additional question?

It is always the case?

Not always: Sometimes we want to approximate a belief function by an additive probability (pignistic probability) or we want something similar to natural extension in imprecise probability,

$$\text{Initial Specification + Belief function (hypothesis)} \Longrightarrow \text{More informative}$$

$$\text{More informative} \rightarrow \text{Inner Approximation}$$

Relation between F and $A(F)$

Basic Question

Should $A(F)$ be less or more informative than F?

In general, it is assumed that $A(F)$ should be less informative than F: **No new information is created by the approximation.** Outer Approximation.

Additional question?

It is always the case?

Not always: Sometimes we want to approximate a belief function by an additive probability (pignistic probability) or we want something similar to natural extension in imprecise probability,

$$\text{Initial Specification} + \text{Belief function (hypothesis)} \Longrightarrow \text{More informative}$$

$$\text{More informative} \rightarrow \text{Inner Approximation}$$

Relation between F and $A(F)$

Basic Question

Should $A(F)$ be less or more informative than F?

In general, it is assumed that $A(F)$ should be less informative than F: **No new information is created by the approximation.** Outer Approximation.

Additional question?

It is always the case?

Not always: Sometimes we want to approximate a belief function by an additive probability (pignistic probability) or we want something similar to natural extension in imprecise probability,

$$\text{Initial Specification} + \text{Belief function (hypothesis)} \Longrightarrow \text{More informative}$$

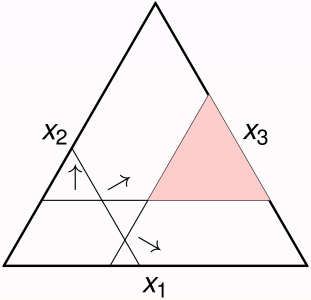
$$\text{More informative} \rightarrow \text{Inner Approximation}$$

Natural Extension: Imprecise Probability

$$P_*(\{x_1\}) = 0.2; P_*(\{x_2\}) = 0.3, P_*(\{x_1, x_2\}) = 0.4$$

Natural extension allows to correct it to more informative bounds

$$P'_*(\{x_1\}) = 0.2; P'_*(\{x_2\}) = 0.3, P'_*(\{x_1, x_2\}) = 0.5$$

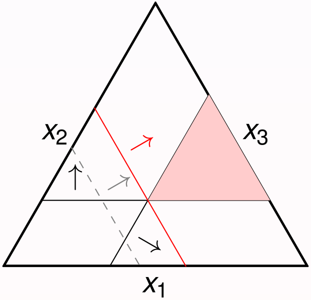


Natural Extension: Imprecise Probability

$$P_*(\{x_1\}) = 0.2; P_*(\{x_2\}) = 0.3, P_*(\{x_1, x_2\}) = 0.4$$

Natural extension allows to correct it to more informative bounds

$$P'_*(\{x_1\}) = 0.2; P'_*(\{x_2\}) = 0.3, P'_*(\{x_1, x_2\}) = 0.5$$



Combination as Approximation

- Imagine that we have two mass assignments: m_1 and m_2,
- and we want to 'combine' them, but we can not apply Dempster's rule (a catious rule),
- we can look for a mass m' which is more informative than m_1 and m_2: $m' \sqsubseteq m_1$ and $m' \sqsubseteq m_2$.
- We could compute a mass m' which is an approximation of m_1 and m_2: we look for an inner approximation of the two masses.

Combination as Approximation

- Imagine that we have two mass assignments: m_1 and m_2,
- and we want to 'combine' them, but we can not apply Dempster's rule (a catious rule),
- we can look for a mass m' which is more informative than m_1 and m_2: $m' \sqsubseteq m_1$ and $m' \sqsubseteq m_2$.
- We could compute a mass m' which is an approximation of m_1 and m_2: we look for an inner approximation of the two masses.

Combination as Approximation

- Imagine that we have two mass assignments: m_1 and m_2,
- and we want to 'combine' them, but we can not apply Dempster's rule (a catious rule),
- we can look for a mass m' which is more informative than m_1 and m_2: $m' \sqsubseteq m_1$ and $m' \sqsubseteq m_2$.
- We could compute a mass m' which is an approximation of m_1 and m_2: we look for an inner approximation of the two masses.

Combination as Approximation

- Imagine that we have two mass assignments: m_1 and m_2,
- and we want to 'combine' them, but we can not apply Dempster's rule (a catious rule),
- we can look for a mass m' which is more informative than m_1 and m_2: $m' \sqsubseteq m_1$ and $m' \sqsubseteq m_2$.
- We could compute a mass m' which is an approximation of m_1 and m_2: we look for an inner approximation of the two masses.

Least Commitment Principle

Intuitive Basis

If we have information m and look for a more informative (inner approximation) m', then m' should not add too much information: it should be "as least informative as possible".

Requirements:

- A minimum is perfect, but at least it should be 'minimal' between all the more informative approximations:

$$m'' \in B, \; m'' \sqsubseteq m, \; m' \sqsubseteq m'' \Rightarrow m'' = m'$$

- A way of obtaining a minimal inner approximation (but not all the minimal ones) is to compute the approximation maximizing a non-specificity measure:

$$NS(m') = \sum_{A \subseteq \Omega} m'(A) \log(|A|)$$

Least Commitment Principle

Intuitive Basis

If we have information m and look for a more informative (inner approximation) m', then m' should not add too much information: it should be "as least informative as possible".

Requirements:

- A minimum is perfect, but at least it should be 'minimal' between all the more informative approximations:

$$m'' \in \mathcal{B},\ m'' \sqsubseteq m,\ m' \sqsubseteq m'' \Rightarrow m'' = m'$$

- A way of obtaining a minimal inner approximation (but not all the minimal ones) is to compute the approximation maximizing a non-specificity measure:

$$NS(m') = \sum_{A \subseteq \Omega} m'(A) \log(|A|)$$

Least Commitment Principle

Intuitive Basis

If we have information m and look for a more informative (inner approximation) m', then m' should not add too much information: it should be "as least informative as possible".

Requirements:

- A minimum is perfect, but at least it should be 'minimal' between all the more informative approximations:

$$m'' \in \mathcal{B},\ m'' \sqsubseteq m,\ m' \sqsubseteq m'' \Rightarrow m'' = m'$$

- A way of obtaining a minimal inner approximation (but not all the minimal ones) is to compute the approximation maximizing a non-specificity measure:

$$NS(m') = \sum_{A \subseteq \Omega} m'(A) \log(|A|)$$

Most Commitment Principle

If we are looking for an outer approximation of m, the least commitment principle does not make sense.
In this case we should look for an approximation that is 'as informative as possible'

Requirements:

- It should be maximal between all the less informative approximations:

$$m'' \in \mathcal{B},\ m \sqsubseteq m'',\ m'' \sqsubseteq m' \Rightarrow m'' = m'$$

- A way of obtaining a maximal inner approximation (but not all the maximal ones) is compute an approximation minimizing a non-specificity measure:

$$NS(m') = \sum_{A \subseteq \Omega} m'(A) \log(|A|)$$

Most Commitment Principle

If we are looking for an outer approximation of m, the least commitment principle does not make sense.
In this case we should look for an approximation that is 'as informative as possible'
Requirements:

- It should be maximal between all the less informative approximations:

$$m'' \in \mathcal{B},\ m \sqsubseteq m'',\ m'' \sqsubseteq m' \Rightarrow m'' = m'$$

- A way of obtaining a maximal inner approximation (but not all the maximal ones) is compute an approximation minimizing a non-specificity measure:

$$NS(m') = \sum_{A \subseteq \Omega} m'(A) \log(|A|)$$

Most Commitment Principle

If we are looking for an outer approximation of m, the least commitment principle does not make sense.

In this case we should look for an approximation that is 'as informative as possible'

Requirements:

- It should be maximal between all the less informative approximations:

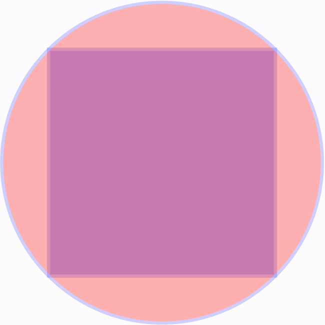
$$m'' \in \mathcal{B},\ m \sqsubseteq m'',\ m'' \sqsubseteq m' \Rightarrow m'' = m'$$

- A way of obtaining a maximal inner approximation (but not all the maximal ones) is compute an approximation minimizing a non-specificity measure:

$$NS(m') = \sum_{A \subseteq \Omega} m'(A) \log(|A|)$$

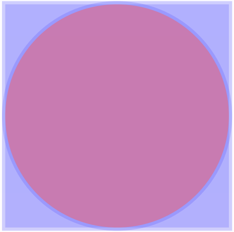
Basic Approximations

Outer Approximations



Maximal
Minimize non-specificity

Inner Approximations



Minimal:
Maximize non-specificity

General Approximations

Basic Question

Should we always look for inner or outer approximations or is it appropriate to look for a belief measure m' with no information relationship with the original information m?

For example, in probability theory we want to approximate a probability measure P which can be difficult to compute, using another probability measure P' which is esasier to compute: this is the case of variational probabilistic inference.

Formulation

Minimize some measure of error: $E(P, P')$.

General Approximations

Basic Question

Should we always look for inner or outer approximations or is it appropriate to look for a belief measure m' with no information relationship with the original information m?

For example, in probability theory we want to approximate a probability measure P which can be difficult to compute, using another probability measure P' which is esasier to compute: this is the case of variational probabilistic inference.

Formulation

Minimize some measure of error: $E(P, P')$.

General Approximations

Basic Question

Should we always look for inner or outer approximations or is it appropriate to look for a belief measure m' with no information relationship with the original information m?

For example, in probability theory we want to approximate a probability measure P which can be difficult to compute, using another probability measure P' which is esasier to compute: this is the case of variational probabilistic inference.

Formulation

Minimize some measure of error: $E(P, P')$.

Measuring the error: distances

$$d_I^2(m, m') = \sqrt{\sum_{A \subseteq \Omega} (m(A) - m'(A))^2}$$

Jousselme's Distance: Jaccard L_2 Distance

$$d_J^2(m, m') = \sqrt{(m - m')Jac(m - m')}$$

$$\begin{array}{c c}
 & \begin{array}{ccccccc} \{x_1\} & \{x_2\} & \{x_3\} & \{x_1,x_2\} & \{x_1,x_3\} & \{x_2,x_3\} & \{x_1,x_2,x_3\} \end{array} \\
\begin{array}{c} \{x_1\} \\ \{x_2\} \\ \{x_3\} \\ \{x_1, x_2\} \\ \{x_1, x_3\} \\ \{x_2, x_3\} \\ \{x_1, x_2, x_3\} \end{array} &
\left(\begin{array}{ccccccc}
1 & 0 & 0 & 1/2 & 1/2 & 0 & 1/3 \\
0 & 1 & 0 & 1/2 & 0 & 1/2 & 1/3 \\
0 & 0 & 1 & 0 & 1/2 & 1/2 & 1/3 \\
1/2 & 1/2 & 0 & 1 & 1/3 & 1/3 & 2/3 \\
1/2 & 0 & 1/2 & 1/3 & 1 & 1/3 & 2/3 \\
0 & 1/2 & 1/2 & 1/3 & 1/3 & 1 & 2/3 \\
1/3 & 1/3 & 1/3 & 2/3 & 2/3 & 2/3 & 1
\end{array} \right)
\end{array}$$

Other Distances

- L_2 Belief Functions Distance:

$$d^2_{Bel}(m, m') = \sqrt{\sum_{A \subseteq \Omega} (Bel(A) - Bel'(A))^2}$$

- L_1 Belief Functions Distance (Imprec. Prob. Baroni, Vicig):

$$d^1_{Bel}(m, m') = \sum_{A \subseteq \Omega} |Bel(A) - Bel'(A))|$$

- L_2 Betting Probability Distance:

$$d^2_{BetP} = \sqrt{\sum_{x \in \Omega} \left(BetP(x) - BetP'(x))^2\right)}$$

Some properties in relation with approximation

- All the distances, except d^2_{BetP} satisfy

$$d(m, m') = 0 \Leftrightarrow m = m'$$

- In general, with these distances, we have to solve convex optimization problems, which are not too hard.
- d^2_{Bel} and d^2_J are strictly convex and under suitable conditions the solution is unique.

Generalized Kullback Leibler Divergence

Kullback Leibler Divergence

- In probabilitic reasoning is more common to use Kullback Leibler Divergence as a measure of error.
- The definition is:

$$KL(P, P') = \sum_{x \in \Omega} P(x) \log \left(\frac{P(x)}{P'(x)} \right)$$

- $KL(P, P') = 0$ if and only if $P = P'$
- It is not symmetrical:

$$P \longleftarrow P'$$

Reference Estimation

- If $P(x) > 0$ an $P'(x) = 0$, then $KL(P, P') = +\infty$.
- No special problem with $P(x) = 0$ and $P'(x) > 0$.

Kullback Leibler Divergence

- In probabilitic reasoning is more common to use Kullback Leibler Divergence as a measure of error.
- The definition is:

$$KL(P, P') = \sum_{x \in \Omega} P(x) \log \left(\frac{P(x)}{P'(x)} \right)$$

- $KL(P, P') = 0$ if and only if $P = P'$
- It is not symmetrical:

$$\underset{\text{Reference}}{P} \longleftarrow\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!- \underset{\text{Estimation}}{P'}$$

- If $P(x) > 0$ an $P'(x) = 0$, then $KL(P, P') = +\infty$.
- No special problem with $P(x) = 0$ and $P'(x) > 0$.

Kullback Leibler Divergence

- In probabilitic reasoning is more common to use Kullback Leibler Divergence as a measure of error.
- The definition is:

$$KL(P, P') = \sum_{x \in \Omega} P(x) \log\left(\frac{P(x)}{P'(x)}\right)$$

- $KL(P, P') = 0$ if and only if $P = P'$
- It is not symmetrical:

$$P \longleftarrow P'$$

Reference Estimation

- If $P(x) > 0$ an $P'(x) = 0$, then $KL(P, P') = +\infty$.
- No special problem with $P(x) = 0$ and $P'(x) > 0$.

Kullback Leibler Divergence

- In probabilitic reasoning is more common to use Kullback Leibler Divergence as a measure of error.
- The definition is:

$$KL(P,P') = \sum_{x\in\Omega} P(x)\log\left(\frac{P(x)}{P'(x)}\right)$$

- $KL(P,P') = 0$ if and only if $P = P'$
- It is not symmetrical:

$$P \longleftarrow P'$$

Reference Estimation

- If $P(x) > 0$ an $P'(x) = 0$, then $KL(P,P') = +\infty$.
- No special problem with $P(x) = 0$ and $P'(x) > 0$.

Kullback Leibler Divergence

- In probabilitic reasoning is more common to use Kullback Leibler Divergence as a measure of error.
- The definition is:

$$KL(P, P') = \sum_{x \in \Omega} P(x) \log\left(\frac{P(x)}{P'(x)}\right)$$

- $KL(P, P') = 0$ if and only if $P = P'$
- It is not symmetrical:

$$P \longleftarrow P'$$

Reference Estimation

- If $P(x) > 0$ an $P'(x) = 0$, then $KL(P, P') = +\infty$.
- No special problem with $P(x) = 0$ and $P'(x) > 0$.

Kullback Leibler Divergence

- In probabilitic reasoning is more common to use Kullback Leibler Divergence as a measure of error.
- The definition is:

$$KL(P, P') = \sum_{x \in \Omega} P(x) \log \left(\frac{P(x)}{P'(x)} \right)$$

- $KL(P, P') = 0$ if and only if $P = P'$
- It is not symmetrical:

$$P \longleftarrow P'$$

Reference Estimation

- If $P(x) > 0$ an $P'(x) = 0$, then $KL(P, P') = +\infty$.
- No special problem with $P(x) = 0$ and $P'(x) > 0$.

KL Divergence for Belief Functions

Straightforward Generalization

$$KL_X(m, m') = \sum_{A \subseteq \Omega} m(A) \log\left(\frac{m(A)}{m'(A)}\right)$$

- It can be very sensitive to small variations:

Example

$$\begin{array}{lll} m(\{x_1\}) = 0.6 & m(\{x_2\}) = 0.2 & m(\{x_1, x_2\}) = 0.2 \\ m'(\{x_1\}) = 0.6 & m'(\{x_2\}) = 0.2 & m'(\{x_1, x_2, x_3\}) = 0.2 \end{array}$$

$KL_X(m, m') = +\infty$.

Other Defintions

Shenoy, 2024

$$KL_S(m, m') = \sum_{A \subseteq \Omega} (-1)^{|A|+1} q(A) \log \left(\frac{q(A)}{q'(A)} \right)$$

If $q'(A) = 0$ and $q(A) > 0$, then it is $+\infty$.

Problem: $KL_S(m, m) = 0$, but we can have other masses for which $KL_S(m, m') < 0$.

A New Definition $KL(m, m')$

- First consider the minimal mass satisfying:

$$m^* = \arg \min_{m'' \sqsubseteq_s m'} KL_X(m, m'')$$

where $KL_X(m, m'') = \sum_{A \subseteq \Omega} m(A) \log \left(\frac{m(A)}{m''(A)} \right)$
There is an unique m^*.

- Then, our definition will have two components

$$KL_1(m, m') = (KL_X(m, m^*) \quad , \quad (NS(m') - NS(m^*)))$$

where NS is the non-specificity: $NS(m') = \sum_{A \subseteq \Omega} m'(A) \log(|A|)$

A New Definition $KL(m, m')$

- First consider the minimal mass satisfying:

$$m^* = \arg \min_{m'' \sqsubseteq_s m'} KL_X(m, m'')$$

where $KL_X(m, m'') = \sum_{A \subseteq \Omega} m(A) \log \left(\frac{m(A)}{m''(A)} \right)$
There is an unique m^*.

- Then, our definition will have two components

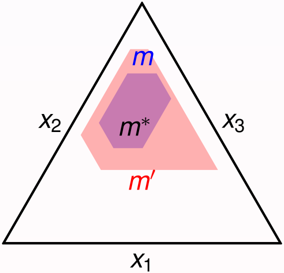
$$KL_1(m, m') = \left(KL_X(m, m^*) \quad , \quad (NS(m') - NS(m^*)) \right)$$

where NS is the non-specificity: $NS(m') = \sum_{A \subseteq \Omega} m'(A) \log(|A|)$

Properties

$$KL_1(m, m') = (\boxed{KL_X(m, m^*)} \quad , \quad (NS(m') - NS(m^*))$$

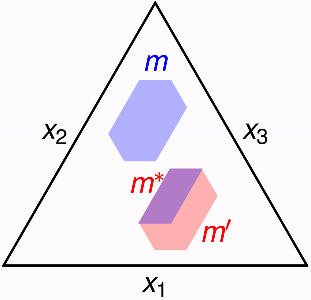
- It has two parts:
  - $KL_X(m, m^*)$
    This part penalizes the fact that m' is not less informative than m. If $m \sqsubseteq_s m'$, this value is 0.
    If this does not happen, then it is positive.



Properties

$$KL_1(m, m') = \left(\boxed{KL_X(m, m^*)} \quad , \quad (NS(m') - NS(m^*)) \right)$$

- $KL_X(m, m^*) = \sum_{A \subseteq \Omega} m(A) \log \left(\frac{m(A)}{m^*(A)} \right)$



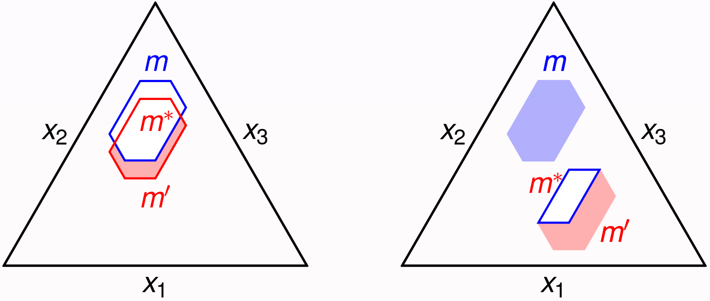
It satisfies that if $m(A) > 0$ and $q'(A) = 0$, then $KL_X(m, m^*) = +\infty$.

The Second Part

$$KL_1(m, m') = (KL_X(m, m^*) \quad , \quad \boxed{(NS(m') - NS(m^*))}$$

- $NS(m') - NS(m^*) = \sum_{A \subseteq \Omega} (m'(A) - m^*(A)) \log(|A|)$

It accounts for the extra in non-specificity of m' with respect to m^*.



It is bounded by $\log(n)$, being n the cardinal of Ω.

Examples

Example

$$
\begin{array}{ccc}
m(\{x_1\}) = 0.6 & m(\{x_2\}) = 0.2 & m(\{x_1, x_2\}) = 0.2 \\
m'(\{x_1\}) = 0.6 & m'(\{x_2\}) = 0.2 & m'(\{x_1, x_2, x_3\}) = 0.2 \\
m''(\{x_1\}) = 0.6 & m''(\{x_2\}) = 0.2 & m''(\{x_3\}) = 0.2
\end{array}
$$

- $KL_1(m, m') = (0, 0.2 \log(3/2))$
- $KL_1(m, m'') = (\infty, 0)$

Defining a Total Order

- We have defined a divergence with two components: $KL_1(m, m') = (x_1, x_2)$
- To define a total order, we have two options:
  - To consider a convex combination of the two components, for example:

$$(x_1, x_2) \leq (y_1, y_2) \Leftrightarrow 0.5x_1 + 0.5x_2 \leq 0.5y_1 + 0.5y_2$$

  - To consider the lexicographical order

$$(x_1, x_2) \leq (y_1, y_2) \Leftrightarrow x_1 < y_1 \text{ or } (x_1 = y_1, x_2 \leq y_2)$$

- We will consider the lexicographical order, then minimizing KL_1 is equivalent:
  - First, minimize the first component, computing the set of masses $\mathcal{B}_1$ minimizing the first component.
  - Then, compute the approximation m' minimizing the second component among those in $\mathcal{B}_1$.

Properties

$$KL_1(m, m') = (KL_X(m, m^*) \quad , \quad (NS(m') - NS(m^*))$$

Identifying $(x, 0) \equiv x$

- $KL_1(m, m') = 0 \Leftrightarrow m = m'$.
- If $m \sqsubseteq_s m'$, then

$$KL_1(m, m') = (0 \quad , \quad NS(m') - NS(m))$$

  Minimizing the divergence is equivalent to computing the most specific information (minimizing the non-specificity).
- If m and m' are Bayesian, then

$$KL_1(m, m') = \sum_{x \in \Omega} m(\{x\}) \log \left(\frac{m(\{x\})}{m'(\{x\})} \right)$$

  It is a true generalization of probabilistic KL divergence.

Basic Question

$$KL_1(m, m')$$

Original | Approximation
Reference | Estimation

¿Does it make sense the other way round?

$$KL_1(m', m)$$

Approximation | Original
Reference | Estimation

Probabilistic variational inference tries to approximate a density optimizing the second option.

Basic Question

$$KL_1(m, m')$$

Original — Reference

Approximation — Estimation

¿Does it make sense the other way round?

$$KL_1(m', m)$$

Approximation — Reference

Original — Estimation

Probabilistic variational inference tries to approximate a density optimizing the second option.

Basic Question

$$KL_1(m, m')$$

Original — Reference

Approximation — Estimation

¿Does it make sense the other way round?

$$KL_1(m', m)$$

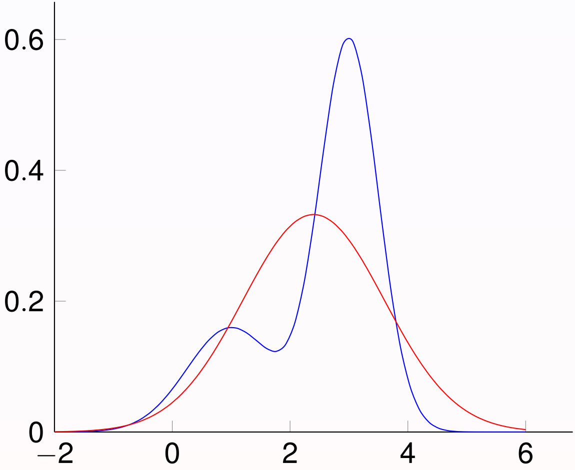
Approximation — Reference

Original — Estimation

Probabilistic variational inference tries to approximate a density optimizing the second option.

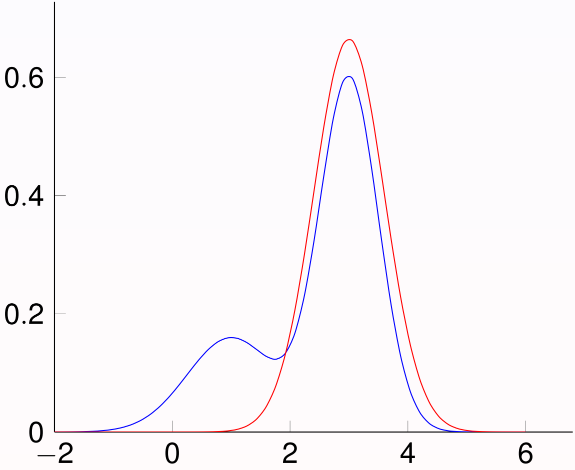
Two Options

$KL(P, P')$: Generalized Outer Approximation



Two Options

$KL(P', P)$: Gneralized Inner Approximation



With Mass Assignments

- Minimize $KL_1(m, m')$ where $m' \in \mathcal{B}$.
  - If $\mathcal{B}$ contains less informative masses; $KL_1(m, m') = (0, NS(m') - NS(m))$: minimum non-specificity among those that are less informative. Outer Approximations
  - In other cases, it will strongly penalize a degree of m' not being less informative than m. Generalized Outer Approximations
  - Appropriate when m is the 'true' belief if we want to transform it, for example, to make it simpler.
- Minimize $KL_1(m', m)$ where $m' \in \mathcal{B}$.
  - If $\mathcal{B}$ contains more informative masses; $KL_1(m', m) = (0, NS(m) - NS(m'))$ maximum non-specificity among those that are more informative. Inner Approximations
  - In other cases, it will penalize m' not being more informative than m. Generalized Inner Approximations
  - Appropriate when there is an underlying belief function m', that we only partially know through m (combination as approximation) or we want to compute a more specific representation (pignistic probability),

Complementing Kulback Leibler Definition

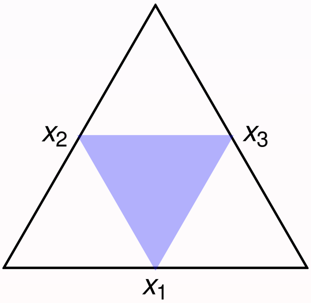
Very often, KL_1 does not provide an unique solution.
Imagine that we are looking for an outer approximation of the lower probability:

$$\underline{P}(\{x_i\}) = 0; \quad \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Restrictions: $Bel(A) \leq \underline{P}(A)$

Masses

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0	0	0	0.5	0	0.5
0	0	0	0	0.5	0.5
0	0	0	0.5	0.5	0



First component = 0, minimum non-specificity

There are 3 extreme solutions with minimum non-specificity

Complementing Kulback Leibler Definition

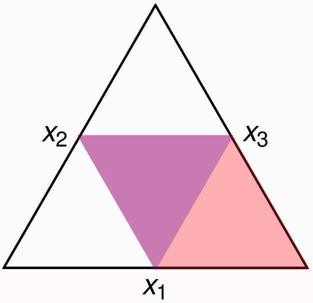
Very often, KL_1 does not provide an unique solution.
Imagine that we are looking for an outer approximation of the lower probability:

$$\underline{P}(\{x_i\}) = 0; \quad \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Restrictions: $Bel(A) \leq \underline{P}(A)$

Masses

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0	0	0	0.5	0	0.5
0	0	0	0	0.5	0.5
0	0	0	0.5	0.5	0



First component = 0, minimum non-specificity

There are 3 extreme solutions with minimum non-specificity

Complementing Kulback Leibler Definition

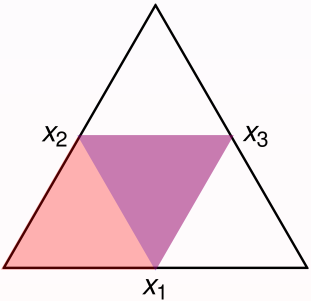
Very often, KL_1 does not provide an unique solution.
Imagine that we are looking for an outer approximation of the lower probability:

$$\underline{P}(\{x_i\}) = 0; \quad \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Restrictions: $Bel(A) \leq \underline{P}(A)$

Masses

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0	0	0	0.5	0	0.5
0	0	0	0	0.5	0.5
0	0	0	0.5	0.5	0



First component = 0, minimum non-specificity

There are 3 extreme solutions with minimum non-specificity

Complementing Kulback Leibler Definition

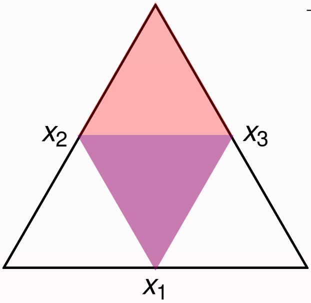
Very often, KL_1 does not provide an unique solution.
Imagine that we are looking for an outer approximation of the lower probability:

$$\underline{P}(\{x_i\}) = 0; \quad \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Restrictions: $Bel(A) \leq \underline{P}(A)$

Masses

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0	0	0	0.5	0	0.5
0	0	0	0	0.5	0.5
0	0	0	0.5	0.5	0



First component = 0, minimum non-specificity

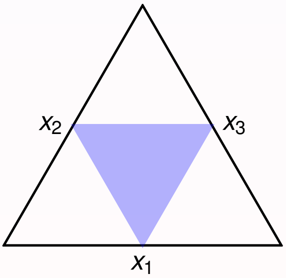
There are 3 extreme solutions with minimum non-specificity

Symmetry Principle

Moral, De Campos, 1993, Partially Specified Belief Functions

Intuitively, this principle says that if there are several possible solutions we should look for an intermediate solution among the extreme ones.

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0	0	0	0.5	0	0.5
0	0	0	0	0.5	0.5
0	0	0	0.5	0.5	0

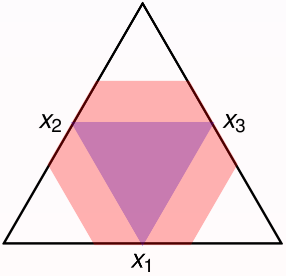


Symmetry Principle

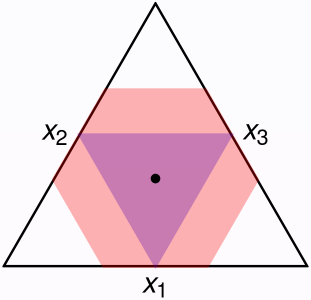
Moral, De Campos, 1993, Partially Specified Belief Functions

Intuitively, this principle says that if there are several possible solutions we should look for an intermediate solution among the extreme ones.

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0	0	0	0.5	0	0.5
0	0	0	0	0.5	0.5
0	0	0	0.5	0.5	0
0	0	0	1/3	1/3	1/3



Symmetry Principle



New Definition

$$KL_2(m, m') = (KL_X(m, m^*), (NS(m') - NS(m^*)), \boxed{\sum_{x \in \Omega} BetP(x) \log\left(\frac{BetP(x)}{BetP'(x)}\right)}$$

Perhaps, $d_J^2(m, m')$ or $d_{Bel}^2(m, m')$ is more appropriate.

Corrected divergence: KL_2

We can add a third factor to KL_1 divergence.

$$KL_2(m, m') = (KL_X(m, m^*), (NS(m') - NS(m^*)), \boxed{\sum_{x \in \Omega} BetP(x) \log\left(\frac{BetP(x)}{BetP'(x)}\right)}$$

We can consider any convex combination of the 3 factors to define a total order, but here we will use the lexicographical order.
Then, Minimizing KL_2 is equivalent to:

- First compute the set of masses M minimizing $KL_1(m, m')$
- If there is more than one, then compute the mass minimizing $\sum_{x \in \Omega} BetP(x) \log\left(\frac{BetP(x)}{BetP'(x)}\right)$ for the masses $m' \in M$.

Computation $KL_1(m, m')$

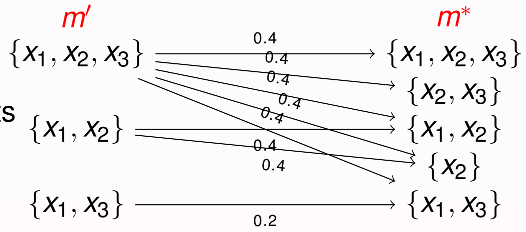
$$m^* = \arg \min_{m'' \sqsubseteq_s m'} KL_X(m, m'')$$

Algorithm with inspiration in Ford-Fulquenson for max-flow problem.

$$m(\{x_1, x_2\}) = 0.4, \;\; m(\{x_2, x_3\}) = 0.4, m(\{x_2\}) = 0.2$$

$$m'(\{x_1, x_2, x_3\}) = 0.4, \;\; m'(\{x_1, x_2\}) = 0.4, m'(\{x_1, x_3\}) = 0.2$$

Focal sets of m'



Focal sets of m' and m

Example

It is an incremental greedy algorithm, which converges to the optimal solution, and each step is based on solving a path problem in a graph.

$$m(\{x_1, x_2\}) = 0.4, \ \ m(\{x_2, x_3\}) = 0.4, \ \ m(\{x_2\}) = 0.2$$

$$m'(\{x_1, x_2, x_3\}) = 0.4, \ \ m'(\{x_1, x_2\}) = 0.4, m'(\{x_1, x_3\}) = 0.2$$

$$m^*(\{x_1,x_2\}){=}0.32, \ m^*(\{x_2,x_3\}){=}0.32, \ m^*(\{x_2\}){=}0.16, \ m^*(\{x_1,x_3\}){=}0.2$$

Approximating by a probability

- Assume that m is an arbitrary belief function and the approximation space, $\mathcal{B}$, is the family of all probability measures on Ω.
- It only makes sense to minimize $KL(P', m)$ (inner approximations).
- If KL_1 is used, any probability belonging to the credal set associated with m is such that $KL_1(P', m) = (0, NS(m))$. So, there is not an unique solution.
- If $KL_2(P', m)$ is used a term $KL(P', BetP_m)$ is added to KL_1 and then the solution is unique and equal to the betting probability.
- In summary: the best approximation of a belief function by a probability using K_2 is the betting probability.

Extending it to Interval Probabilities

Example

Imagine that we are looking for an outer approximation of the probability intervals on $\{x_1, x_2, x_3\}$:

$$\underline{P}(\{x_i\}) = 0; \quad \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

$$KL_1(m, m') = (KL_X(m, m^*), (NS(m') - NS(m^*)))$$

But now,

$$KL_1(\underline{P}, m') = (KL_X(\underline{P}, m^*), (NS(m') - NS(m^*)))$$

Divergence in Interval Probabilities

$$KL_1(\underline{P}, \underline{P}') = (KL_X(\underline{P}, \underline{P}^*), (NS(\underline{P}') - NS(\underline{P}^*)))$$

The Trick

To minimize $KL_1(\underline{P}, m') = (KL_X(\underline{P}, m^*), (NS(m') - NS(m^*)))$, I assume that m' is less informative and then,
$KL_1(\underline{P}, m') = (0, (NS(m') - NS(\underline{P})))$

- We can compute masses from lower probabilities, but if they are not belief functions, some masses are negative.
- Non-Specificity can be computed with these negative masses (Abellán, Moral, 2000).
- I do not know a direct generalization of $\underline{P}^*$, $KL_X(\underline{P}, \underline{P}^*)$ (negative masses).

Divergence in Interval Probabilities

$$KL_1(\underline{P}, \underline{P'}) = (KL_X(\underline{P}, \underline{P^*}), (NS(\underline{P'}) - NS(\underline{P^*})))$$

The Trick

To minimize $KL_1(\underline{P}, m') = (KL_X(\underline{P}, m^*), (NS(m') - NS(m^*)))$, I assume that m' is less informative and then,
$KL_1(\underline{P}, m') = (0, (NS(m') - NS(\underline{P})))$

- We can compute masses from lower probabilities, but if they are not belief functions, some masses are negative.
- Non-Specificity can be computed with these negative masses (Abellán, Moral, 2000).
- I do not know a direct generalization of $\underline{P^*}$, $KL_X(\underline{P}, \underline{P^*})$ (negative masses).

Divergence in Interval Probabilities

$$KL_1(\underline{P}, \underline{P}') = (KL_X(\underline{P}, \underline{P}^*), (NS(\underline{P}') - NS(\underline{P}^*)))$$

The Trick

To minimize $KL_1(\underline{P}, m') = (KL_X(\underline{P}, m^*), (NS(m') - NS(m^*)))$, I assume that m' is less informative and then,
$KL_1(\underline{P}, m') = (0, (NS(m') - NS(\underline{P})))$

- We can compute masses from lower probabilities, but if they are not belief functions, some masses are negative.
- Non-Specificity can be computed with these negative masses (Abellán, Moral, 2000).
- I do not know a direct generalization of $\underline{P}^*$, $KL_X(\underline{P}, \underline{P}^*)$ (negative masses).

Pignistic Probability for Interval Probabilities

Question

Is there anything similar to the pignistic probability for interval probabilities? Yes, but it is more difficult to compute: average of the probabilities in the associated credal set with respect to the uniform distribution Smets (1994)

Computation with Extreme Points

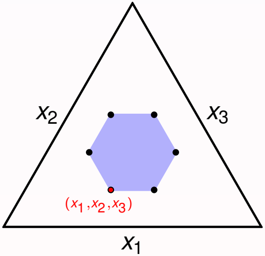
The pignistic probability can be computed in the following way:

- For each permutation σ of $\{1, \ldots, n\}$ compute the probability:

$$P_\sigma(x_{\sigma(1)}) = Bel(\{x_{\sigma(1)}\}),$$
$$P_\sigma(x_{\sigma(i)}) = Bel(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}) - Bel(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}), i = 2, \ldots, n$$

- *BetP* is the average of these probabilities:

$$BetP = \frac{1}{n!} \sum_{\sigma \in Per} P_\sigma$$



Permutation Probabilities for Imprecise Probabilities

- Walner (2007) proved that interval probabilities have at most $n!$ extreme points.
- Are they associated with permutations?
- I did not know it, but they are.

Putting a limit to the number of focal elements

Frédéric Pichon Lecture

Problem: Limit in the Number of Focal Sets

- After combining belief functions the number of focal elements increases in a combinatorial way.
- We can consider after each combination, to approximate the resulting mass assignment m with r focal sets, by another simpler mass m' with a limited number of elements K. The approximation $\mathcal{B}$ is the set of belief functions with K or less focal sets.
- In this problem, we have m as reference and m' as approximation. If $KL_1(m, m')$ is minimized, then we should look for the mass assignment strongly less informative than m ($m \sqsubseteq_s m'$) with minimum non-specificity in the space

$$\mathcal{B} = \{m' \ : \ |\mathcal{F}_{m'}| \leq K\}$$

where $|\mathcal{F}_{m'}|$ is the number of focal elements of m'.

Basic Approaches

- The global partition of the focal sets approach (Harmanec, 1999; Denœux, 2001)
- The hierarchical greedy approach (Harmanec, 1999; Moral, Salmerón, 1999)
- Frame Coarsening (Denœux, Ben Yaghlane, 2001)

Example

$K = 2$

$m(\{x_1\})=0.1$ $m(\{x_2\})=0.1$ $m(\{x_1,x_2\})=0.4$ $m(\{x_3\})=0.3$ $m(\{x_3,x_4\})=0.1$

Example

$K = 2$

$m(\{x_1\})=0.1$ $m(\{x_2\})=0.1$ $m(\{x_1,x_2\})=0.4$ $m(\{x_3\})=0.3$ $m(\{x_3,x_4\})=0.1$

$m'(\{x_1,x_2\})=0.6$ $m'(\{x_3,x_4\})=0.4$

Increment of non-specificity $5\log(2)$

Example

$K = 2$

$m(\{x_1\})$=0.1 $m(\{x_2\})$=0.1 $m(\{x_1, x_2\})$=0.4 $m(\{x_3\})$=0.3 $m(\{x_3, x_4\})$=0.1

Problem

The number of partitions is too high.

Example

$K = 2$

$m(\{x_1\})=0.1$ $m(\{x_2\})=0.1$ $m(\{x_1,x_2\})=0.4$ $m(\{x_3\})=0.3$ $m(\{x_3,x_4\})=0.1$

Problem

The number of partitions is too high.

NP-complete

¿Is there a partition of size K, such that the increasing in non-specificity is less or equal than R?
NP-complete problem: reduction from edge partitioning in graphs.

Incremental Greedy Approach, Harmanec (1999), Moral, Salmerón (1999)

- If we must go from 6 to 3 focal sets, instead of doing it in one step, we do it in a greedy way in 3 steps, by reducing one focal set in each step.

$m(A_1)=a_1 \quad m(A_2)=a_2 \quad m(A_3)=a_3 \quad m(A_4)=a_4 \quad m(A_5)=a_5 \quad m(A_6)=a_6$

Incremental Greedy Approach, Harmanec (1999), Moral, Salmerón (1999)

- If we must go from 6 to 3 focal sets, instead of doing it in one step, we do it in a greedy way in 3 steps, by reducing one focal set in each step.

$$
\begin{array}{cccccc}
m(A_1)=a_1 & m(A_2)=a_2 & m(A_3)=a_3 & m(A_4)=a_4 & m(A_5)=a_5 & m(A_6)=a_6
\end{array}
$$

$$m'(A_2 \cup A_3)=a_2+a_3$$

Incremental Greedy Approach, Harmanec (1999), Moral, Salmerón (1999)

- If we must go from 6 to 3 focal sets, instead of doing it in one step, we do it in a greedy way in 3 steps, by reducing one focal set in each step.

$$m(A_1)=a_1 \quad m(A_2)=a_2 \quad m(A_3)=a_3 \quad m(A_4)=a_4 \quad m(A_5)=a_5 \quad m(A_6)=a_6$$

$$m'(A_2 \cup A_3)=a_2+a_3 \qquad m'(A_4 \cup A_5)=a_4+a_5$$

Incremental Greedy Approach, Harmanec (1999), Moral, Salmerón (1999)

- If we must go from 6 to 3 focal sets, instead of doing it in one step, we do it in a greedy way in 3 steps, by reducing one focal set in each step.

$m(A_1)=a_1 \quad m(A_2)=a_2 \quad m(A_3)=a_3 \quad m(A_4)=a_4 \quad m(A_5)=a_5 \quad m'(A_6)=a_6$

$m(A_2 \cup A_3)=a_2+a_3 \quad m'(A_4 \cup A_5)=a_4+a_5$

$m'(A_1 \cup A_2 \cup A_3)=a_1+a_2+a_3$

KL_1 or KL_2?

- In practice KL_1 has been used: minimizing non-specificity.
- Example:

$$m(\{x_1,x_2\})=0.25,\ m(\{x_1,x_3\})=0.25,\ m(\{x_4,x_5\})=0.3,\ m(\{x_4,x_6\})=0.2$$

  and we want to reduce one focal set.
- Using KL_1 there are two options with minimum non-specificity:
  - Join $\{x_1, x_2\}$ and $\{x_1, x_3\}$, producing $\{x_1, x_2, x_3\}$ with mass 0.5
  - Join $\{x_4, x_5\}$ and $\{x_4, x_6\}$, producing $\{x_4, x_5, x_6\}$ with mass 0.5
- If KL_2 is used, only the first option is possible, as it joins two focal sets with the same mass.
- The computation of KL_2 can take some time, but it can be done by taking only into account the elements in the joined focal sets.

Outer Approximations of Interval Probabilities

Outer Approximations of Interval Probabilities

Montes, Miranda, Vicig, 2019

- We have coherent lower probabilities $\underline{P}$, and we want to approximate by a belief function (outer approximation): $Bel(E) \leq \underline{P}(E)$.
- The problem is solved by minimizing:

$$\begin{array}{ll} \min & \sum_{E\subseteq\Omega} \left(\underline{P}(E) - \overbrace{\left(\sum_{A\subseteq E} m(E)\right)}^{Bel(E)} \right) \\ s.t. & \sum_{A\subseteq E} m(E) \leq \underline{P}(E), \ \ E \subseteq \Omega \\ & m(E) \geq 0, \ \ E \subseteq \Omega \\ & \sum_{E\subseteq\Omega} m(E) = 1 \end{array}$$

Comments

- It is a linear programming problem
- The solution is not always unique
- Minimize this objective is equivalent to maximize

$$\sum_{E\subseteq\Omega} \frac{m(E)}{2^{|E|}}$$

  This can be considered as a specificity measure, penalizing a lot larger focal sets.
- This measure goes from 1 (for probabilities) to $1/2^{|\Omega|}$ for the vacuous belief.
- It is similar to minimize non-specificity: maximizing specificity – minimizing non-specificity.

Non-Dominated: maximality

- Minimizing the non-specificity or maximazing specificity we always obtain maximal approximations, but not all of them.
- The set of maximal approximations is not convex.
- The set of solutions to a linear optimization problem is always convex.

Example. Montes, I., E. Miranda, and P. Vicig, 2019

A	$\underline{P}$	Bel_0	Bel_1	Bel_2
$\{x_1\}$	0.1	0.1	0.1	0.1
$\{x_2\}$	0	0	0	0
$\{x_3\}$	0	0	0	0
$\{x_4\}$	0.3	0.3	0.3	0.3
$\{x_1, x_2\}$	0.1	0.1	0.1	0.1
$\{x_1, x_3\}$	0.3	0.2	0.3	0.1
$\{x_1, x_4\}$	0.6	0.6	0.5	0.6
$\{x_2, x_3\}$	0.3	0.3	0.2	0.2
$\{x_2, x_4\}$	0.4	0.3	0.4	0.3
$\{x_3, x_4\}$	0.4	0.3	0.3	0.4
$\{x_1, x_2, x_3\}$	0.5	0.5	0.5	0.4
$\{x_1, x_2, x_4\}$	0.6	0.6	0.6	0.6
$\{x_1, x_3, x_4\}$	0.7	0.7	0.7	0.7
$\{x_2, x_3, x_4\}$	0.6	0.6	0.6	0.6
Ω	1	1	1	1

Bel_0, Bel_1 are optimal (maximizing specificity) and its convex combinations.

Bel_2 is not optimal, but it is minimal.

Example. Montes, I., E. Miranda, and P. Vicig, 2019

A	$\underline{P}$	m_0	m_1	m_2
$\{x_1\}$	0.1	0.1	0.1	0.1
$\{x_2\}$	0	0	0	0
$\{x_3\}$	0	0	0	0
$\{x_4\}$	0.3	0.3	0.3	0.3
$\{x_1, x_2\}$	0.1	0	0	0
$\{x_1, x_3\}$	0.3	0.1	0.2	0
$\{x_1, x_4\}$	0.6	0.2	0.1	0.2
$\{x_2, x_3\}$	0.3	0.3	0.2	0.2
$\{x_2, x_4\}$	0.4	0	0.1	0
$\{x_3, x_4\}$	0.4	0	0	0.1
$\{x_1, x_2, x_3\}$	0.5	0	0	0.1
$\{x_1, x_2, x_4\}$	0.6	0	0	0
$\{x_1, x_3, x_4\}$	0.7	0	0	0
$\{x_2, x_3, x_4\}$	0.6	0	0	0
Ω	0	0	0	0

Masses of the same example. The two solutions, m_0 and m_1 also minimize non-specificity.

Heuristic: Iterative Rescaling Method (IRM)

Hall, Lawry; 2004

- For each $A \subseteq \Omega$ in increasing size order:
  - Compute $m'(A) = \underline{P}(A) - \sum_{B \subset A} m'(B)$
  - If $m'(A) < 0$ then
    - $\alpha = -m'(A)$
    - Rescale(m', A, α)

Rescale(m, A, α) is a function that decreases some of the masses $m'(B)$ for $B \subset A$ so that $m'(A)$ becomes 0 (larger sets better).

Example

A	P	m_0	m_1	m'
$\{x_1\}$	0.1	0.1	0.1	
$\{x_2\}$	0	0	0	
$\{x_3\}$	0	0	0	
$\{x_4\}$	0.3	0.3	0.3	
$\{x_1, x_2\}$	0.1	0	0	
$\{x_1, x_3\}$	0.3	0.1	0.2	
$\{x_1, x_4\}$	0.6	0.2	0.1	
$\{x_2, x_3\}$	0.3	0.3	0.2	
$\{x_2, x_4\}$	0.4	0	0.1	
$\{x_3, x_4\}$	0.4	0	0	
$\{x_1, x_2, x_3\}$	0.5	0	0	
$\{x_1, x_2, x_4\}$	0.6	0	0	
$\{x_1, x_3, x_4\}$	0.7	0	0	...
$\{x_2, x_3, x_4\}$	0.6	0	0	...
Ω	0	0	0	...

Example

A	P	m_0	m_1	m'
$\{x_1\}$	0.1	0.1	0.1	0.1
$\{x_2\}$	0	0	0	0
$\{x_3\}$	0	0	0	0
$\{x_4\}$	0.3	0.3	0.3	0.3
$\{x_1, x_2\}$	0.1	0	0	
$\{x_1, x_3\}$	0.3	0.1	0.2	
$\{x_1, x_4\}$	0.6	0.2	0.1	
$\{x_2, x_3\}$	0.3	0.3	0.2	
$\{x_2, x_4\}$	0.4	0	0.1	
$\{x_3, x_4\}$	0.4	0	0	
$\{x_1, x_2, x_3\}$	0.5	0	0	
$\{x_1, x_2, x_4\}$	0.6	0	0	
$\{x_1, x_3, x_4\}$	0.7	0	0	...
$\{x_2, x_3, x_4\}$	0.6	0	0	...
Ω	0	0	0	...

Example

A	P	m_0	m_1	m'
$\{x_1\}$	0.1	0.1	0.1	0.1
$\{x_2\}$	0	0	0	0
$\{x_3\}$	0	0	0	0
$\{x_4\}$	0.3	0.3	0.3	0.3
$\{x_1, x_2\}$	0.1	0	0	0.1
$\{x_1, x_3\}$	0.3	0.1	0.2	0.2
$\{x_1, x_4\}$	0.6	0.2	0.1	0.2
$\{x_2, x_3\}$	0.3	0.3	0.2	0.3
$\{x_2, x_4\}$	0.4	0	0.1	0.1
$\{x_3, x_4\}$	0.4	0	0	0.1
$\{x_1, x_2, x_3\}$	0.5	0	0	
$\{x_1, x_2, x_4\}$	0.6	0	0	
$\{x_1, x_3, x_4\}$	0.7	0	0	...
$\{x_2, x_3, x_4\}$	0.6	0	0	...
Ω	0	0	0	...

Example

A	P	m_0	m_1	m'
$\{x_1\}$	0.1	0.1	0.1	0.1
$\{x_2\}$	0	0	0	0
$\{x_3\}$	0	0	0	0
$\{x_4\}$	0.3	0.3	0.3	0.3
$\{x_1, x_2\}$	0.1	0	0	0.1
$\{x_1, x_3\}$	0.3	0.1	0.2	0.2
$\{x_1, x_4\}$	0.6	0.2	0.1	0.2
$\{x_2, x_3\}$	0.3	0.3	0.2	0.3
$\{x_2, x_4\}$	0.4	0	0.1	0.1
$\{x_3, x_4\}$	0.4	0	0	0.1
$\{x_1, x_2, x_3\}$	0.5	0	0	-0.2
$\{x_1, x_2, x_4\}$	0.6	0	0	
$\{x_1, x_3, x_4\}$	0.7	0	0	...
$\{x_2, x_3, x_4\}$	0.6	0	0	...
Ω	0	0	0	...

Example

A	P	m_0	m_1	m'
$\{x_1\}$	0.1	0.1	0.1	0.1
$\{x_2\}$	0	0	0	0
$\{x_3\}$	0	0	0	0
$\{x_4\}$	0.3	0.3	0.3	0.3
$\{x_1, x_2\}$	0.1	0	0	$0.2/3$
$\{x_1, x_3\}$	0.3	0.1	0.2	$0.4/3$
$\{x_1, x_4\}$	0.6	0.2	0.1	0.2
$\{x_2, x_3\}$	0.3	0.3	0.2	0.2
$\{x_2, x_4\}$	0.4	0	0.1	0.1
$\{x_3, x_4\}$	0.4	0	0	0.1
$\{x_1, x_2, x_3\}$	0.5	0	0	0
$\{x_1, x_2, x_4\}$	0.6	0	0	
$\{x_1, x_3, x_4\}$	0.7	0	0	...
$\{x_2, x_3, x_4\}$	0.6	0	0	...
Ω	0	0	0	...

Example

A	P	m_0	m_1	m'
$\{x_1\}$	0.1	0.1	0.1	0.1
$\{x_2\}$	0	0	0	0
$\{x_3\}$	0	0	0	0
$\{x_4\}$	0.3	0.3	0.3	0.3
$\{x_1, x_2\}$	0.1	0	0	$0.2/3$
$\{x_1, x_3\}$	0.3	0.1	0.2	$0.4/3$
$\{x_1, x_4\}$	0.6	0.2	0.1	0.2
$\{x_2, x_3\}$	0.3	0.3	0.2	0.2
$\{x_2, x_4\}$	0.4	0	0.1	0.1
$\{x_3, x_4\}$	0.4	0	0	0.1
$\{x_1, x_2, x_3\}$	0.5	0	0	0
$\{x_1, x_2, x_4\}$	0.6	0	0	$-0.5/3$
$\{x_1, x_3, x_4\}$	0.7	0	0	...
$\{x_2, x_3, x_4\}$	0.6	0	0	...
Ω	0	0	0	...

Example

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1,2,3;\ \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Mass assignment:

$$m'(\{x_1\}) = m'(\{x_2\}) == m'(\{x_3\}) = 0.0$$
$$m'(\{x_1, x_2\}) = m'(\{x_1, x_3\}) = m'(\{x_2, x_3\}) = 0.5$$
$$m'(\{x_1, x_2, x_3\}) = -0.5$$

Rescaling:

$$m'(\{x_1\} = m'(\{x_2\}) == m'(\{x_3\}) = 0.0$$
$$m'(\{x_1, x_2\}) = m'(\{x_1, x_3\}) = m'(\{x_2, x_3\}) = 1/3$$
$$m'(\{x_1, x_2, x_3\}) = 0.0$$

Example

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1,2,3; \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Mass assignment:

$$m'(\{x_1\}) = m'(\{x_2\}) == m'(\{x_3\}) = 0.0$$
$$m'(\{x_1, x_2\}) = m'(\{x_1, x_3\}) = m'(\{x_2, x_3\}) = 0.5$$
$$m'(\{x_1, x_2, x_3\}) = -0.5$$

Rescaling:

$$m'(\{x_1\} = m'(\{x_2\}) == m'(\{x_3\}) = 0.0$$
$$m'(\{x_1, x_2\}) = m'(\{x_1, x_3\}) = m'(\{x_2, x_3\}) = 1/3$$
$$m'(\{x_1, x_2, x_3\}) = 0.0$$

Example

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Mass assignment:

$$m'(\{x_1\}) = m'(\{x_2\}) == m'(\{x_3\}) = 0.0$$
$$m'(\{x_1, x_2\}) = m'(\{x_1, x_3\}) = m'(\{x_2, x_3\}) = 0.5$$
$$m'(\{x_1, x_2, x_3\}) = -0.5$$

Rescaling:

$$m'(\{x_1\} = m'(\{x_2\}) == m'(\{x_3\}) = 0.0$$
$$m'(\{x_1, x_2\}) = m'(\{x_1, x_3\}) = m'(\{x_2, x_3\}) = 1/3$$
$$m'(\{x_1, x_2, x_3\}) = 0.0$$

Problem, Quaeqhebeur, 2011

It is sensible to the order in which sets are considered.

$$\underline{P}(\{x_i\}) = 0.0, \ \ i = 1,2,3;4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0, \ \ i = 1,2,3;4$$

$$m'(\{x_1,x_2\})=,m'(\{x_1,x_3\})=,m'(\{x_1,x_4\})=,$$

$$m'(\{x_2,x_3\})=,m'(\{x_2,x_4\})=1/3,m'(\{x_3,x_4\})=1/3$$

,

Problem, Quaeqhebeur, 2011

It is sensible to the order in which sets are considered.

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; 4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; 4$$

$$m'(\{x_1,x_2\})=1/3, m'(\{x_1,x_3\})=1/3, m'(\{x_1,x_4\})=1/3,$$

$$m'(\{x_2,x_3\})=1/3, m'(\{x_2,x_4\})=1/3, m'(\{x_3,x_4\})=1/3$$

,

Problem, Quaeqhebeur, 2011

It is sensible to the order in which sets are considered.

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; 4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; 4$$

$$m'(\{x_1,x_2\})=1/3, m'(\{x_1,x_3\})=1/3, m'(\{x_1,x_4\})=1/3,$$

$$m'(\{x_2,x_3\})=1/3, m'(\{x_2,x_4\})=1/3, m'(\{x_3,x_4\})=1/3$$

$$m'(\{x_1,x_2,x_3\})=-1/3,$$

Problem, Quaeqhebeur, 2011

It is sensible to the order in which sets are considered.

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; 4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; 4$$

$$m'(\{x_1,x_2\})=2/9, m'(\{x_1,x_3\})=2/9, m'(\{x_1,x_4\})=1/3,$$

$$m'(\{x_2,x_3\})=2/9, m'(\{x_2,x_4\})=1/3, m'(\{x_3,x_4\})=1/3$$

$$m'(\{x_1,x_2,x_3\})=0, m'(\{x_1,x_2,x_4\})=-2/9$$

Problem, Quaeqhebeur, 2011

It is sensible to the order in which sets are considered.

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; 4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; 4$$

$$m'(\{x_1,x_2\})=1/6, m'(\{x_1,x_3\})=2/9, m'(\{x_1,x_4\})=2/9,$$

$$m'(\{x_2,x_3\})=2/9, m'(\{x_2,x_4\})=2/9, m'(\{x_3,x_4\})=1/3$$

$$m'(\{x_1,x_2,x_3\})=0, m'(\{x_1,x_2,x_4\})=0, m'(\{x_1,x_3,x_4\})=-2/9$$

Problem, Quaeqhebeur, 2011

It is sensible to the order in which sets are considered.

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; 4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0,\ \ i = 1, 2, 3; 4$$

$$m'(\{x_1,x_2\})=1/6, m'(\{x_1,x_3\})=7/36, m'(\{x_1,x_4\})=7/36,$$

$$m'(\{x_2,x_3\})=2/9, m'(\{x_2,x_4\})=2/9, m'(\{x_3,x_4\})=7/24$$

$$m'(\{x_1,x_2,x_3\})=0, m'(\{x_1,x_2,x_4\})=0, m'(\{x_1,x_3,x_4\})=0, m'(\{x_2,x_3,x_4\})=-1/8$$

Example

$$\underline{P}(\{x_i\}) = 0.0, \; i = 1, 2, 3; 4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0, \; i = 1, 2, 3; 4$$

$$m'(\{x_1,x_2\})=0.167, m'(\{x_1,x_3\})=0.194, m'(\{x_1,x_4\})=0.194,$$

$$m'(\{x_2,x_3\})=0.184, m'(\{x_3,x_4\})=0.184, m'(\{x_3,x_4\})=0.244$$

$$m'(\{x_1,x_2,x_3\})=0, m'(\{x_1,x_2,x_4\})=0, m'(\{x_1,x_3,x_4\})=0, m'(\{x_2,x_3,x_4\})=0$$

$$m'(\{x_1, x_2, x_3, x_4\} = -0167$$

Example

$$\underline{P}(\{x_i\}) = 0.0, \ \ i = 1, 2, 3; 4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0, \ \ i = 1, 2, 3; 4$$

$$m'(\{x_1,x_2\})=0.143, m'(\{x_1,x_3\})=0.166, m'(\{x_1,x_4\})=0.166,$$

$$m'(\{x_2,x_3\})=0.158, m'(\{x_3,x_4\})=0.158, m'(\{x_3,x_4\})=0.209$$

$$m'(\{x_1,x_2,x_3\})=0, m'(\{x_1,x_2,x_4\})=0, m'(\{x_1,x_3,x_4\})=0, m'(\{x_2,x_3,x_4\})=0$$

$$m'(\{x_1, x_2, x_3, x_4\} = 0$$

Quaeghebeur, 2011, Solution

All the negative masses of the sets with the same size are decreased to 0, simultaneously.

$$\underline{P}(\{x_i\}) = 0.0, \;\; i = 1, 2, 3; 4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0, \;\; i = 1, 2, 3; 4$$

$$m'(\{x_1,x_2\})=1/3, m'(\{x_1,x_3\})=1/3, m'(\{x_1,x_4\})=1/3,$$

$$m'(\{x_2,x_3\})=1/3, m'(\{x_3,x_4\})=1/3, m'(\{x_3,x_4\})=1/3$$

$$m'(\{x_1,x_2,x_3\})=-1/3, m'(\{x_1,x_2,x_4\})=-1/3, m'(\{x_1,x_3,x_4\})=-1/3, m'(\{x_2,x_3,x_4\})=-1/3$$

Quaeghebeur, 2011, Solution

All the negative masses of the sets with the same size are decreased to 0, simultaneously.

$$\underline{P}(\{x_i\}) = 0.0, \; i = 1,2,3;4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0, \; i = 1,2,3;4$$

$$m'(\{x_1,x_2\})=2/9, m'(\{x_1,x_3\})=2/9, m'(\{x_1,x_4\})=2/9,$$

$$m'(\{x_2,x_3\})=2/9, m'(\{x_3,x_4\})=2/9, m'(\{x_3,x_4\})=2/9$$

$$m'(\{x_1,x_2,x_3\})=0, m'(\{x_1,x_2,x_4\})=0, m'(\{x_1,x_3,x_4\})=0, m'(\{x_2,x_3,x_4\})=0$$

$$m'(\{x_1,x_2,x_3,x_4\} = -1/3$$

Quaeghebeur, 2011, Solution

All the negative masses of the sets with the same size are decreased to 0, simultaneously.

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1,2,3;4$$

$$\underline{P}(\{x_1,x_2\})=\underline{P}(\{x_1,x_3\})=\underline{P}(\{x_1,x_4\})=\underline{P}(\{x_2,x_3\})=\underline{P}(\{x_2,x_4\})=\underline{P}(\{x_3,x_4\})=1/3$$

$$\underline{P}(\{x_1,x_2,x_3\})=\underline{P}(\{x_1,x_2,x_4\})=\underline{P}(\{x_1,x_3,x_4\})=\underline{P}(\{x_2,x_3,x_4\})=2/3$$

$$m'(\{x_i\}) = 0.0,\ \ i = 1,2,3;4$$

$$m'(\{x_1,x_2\})=1/6, m'(\{x_1,x_3\})=1/6, m'(\{x_1,x_4\})=1/6,$$

$$m'(\{x_2,x_3\})=1/6, m'(\{x_3,x_4\})=1/6, m'(\{x_3,x_4\})=1/6$$

$$m'(\{x_1,x_2,x_3\})=0, m'(\{x_1,x_2,x_4\})=0, m'(\{x_1,x_3,x_4\})=0, m'(\{x_2,x_3,x_4\})=0$$

$$m'(\{x_1,x_2,x_3,x_4\} = 0$$

One Solution? The case of KL_1

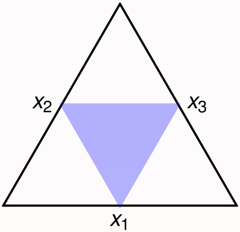
Consider the optimization problem (with non-specificity)

$$\begin{array}{ll} \min & NS(m) \\ s.t. & \sum_{A\subseteq E} m(A) \leq \underline{P}(E),\ \ E\subseteq\Omega \\ & m(E)\geq 0,\ \ E\subseteq\Omega \\ & \sum_{E\subseteq\Omega} m(E) = 1 \end{array}$$

The solution is not unique. In the example:

$$\underline{P}(\{x_i\}) = 0.0,\ \ i=1,2,3;\ \underline{P}(\{x_i,x_j\}) = 0.5,\ \ i\neq j$$

We obtain different solutions:



One Solution? The case of KL_1

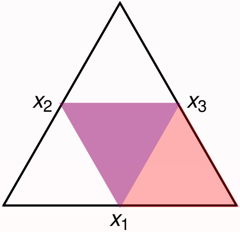
Consider the optimization problem (with non-specificity)

$$\begin{array}{ll} \min & NS(m) \\ s.t. & \sum_{A\subseteq E} m(A) \leq \underline{P}(E),\ \ E \subseteq \Omega \\ & m(E) \geq 0,\ \ E \subseteq \Omega \\ & \sum_{E\subseteq \Omega} m(E) = 1 \end{array}$$

The solution is not unique. In the example:

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1,2,3; \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

We obtain different solutions:



One Solution? The case of KL_1

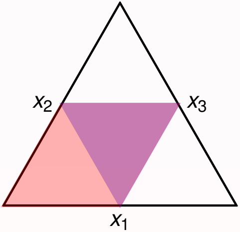
Consider the optimization problem (with non-specificity)

$$\begin{array}{ll} \min & NS(m) \\ s.t. & \sum_{A\subseteq E} m(A) \leq \underline{P}(E),\ \ E\subseteq \Omega \\ & m(E) \geq 0,\ \ E\subseteq\Omega \\ & \sum_{E\subseteq\Omega} m(E) = 1 \end{array}$$

The solution is not unique. In the example:

$$\underline{P}(\{x_i\}) = 0.0,\ \ i = 1,2,3;\ \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

We obtain different solutions:



One Solution? The case of KL_1

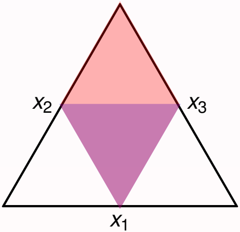
Consider the optimization problem (with non-specificity)

$$\begin{array}{ll} \min & NS(m) \\ s.t. & \sum_{A \subseteq E} m(A) \leq \underline{P}(E), \;\; E \subseteq \Omega \\ & m(E) \geq 0, \;\; E \subseteq \Omega \\ & \sum_{E \subseteq \Omega} m(E) = 1 \end{array}$$

The solution is not unique. In the example:

$$\underline{P}(\{x_i\}) = 0.0, \;\; i = 1, 2, 3; \; \underline{P}(\{x_i, x_j\}) = 0.5, \; i \neq j$$

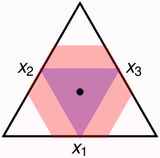
We obtain different solutions:



Unique Solution

To obtain an unique solution we need to change the objective function. There are several possibilities:

- To consider a quadratic function like $\sum_{E\subseteq\Omega}\left(\underline{P}(E) - \sum_{A\subseteq E} m(A)\right)^2$
- But also $d_J^2(\underline{P}, m) = \sqrt{(m_{\underline{P}} - m)Jac(m_{\underline{P}} - m)}$. A Lower probability has always a Moebius inverse (some masses can be negative).
- Or to add the Kullback-Leibler divergence between betting probabilities to the non-specificity minimizing KL_2. I do not have a proof that the solution is unique.
- Iterative rescaling produces an unique solution that follows the 'symmetry principle' (it is a compromise solution).



Example: generalized outer approximations

- Imagine that $\mathcal{C}$ is the space of belief functions.
- And that $\mathcal{B}$ is the space of belief functions with $m'(\Omega) \leq 0.1$
- Given a $m \in \mathcal{C}$, we compute the approximation m' minimizing $KL_1(m, m')$.
- if $m(\Omega) > 0.1$, we do not have a mass in $\mathcal{B}$, such that $m \sqsubseteq_s m'$, but there is always one minimizing $KL_1(m, m')$:
  - If $m(\Omega) \leq 0.1$, then $m' = m$
  - If $1 > m(\Omega) > 0.1$, then

$$m'(\Omega) = 0.1$$

$$m'(A) = \frac{0.9m(A)}{1-m(\Omega)} \quad \text{if } A \neq \Omega$$

- If $m(\Omega) > 0.1$, then the first component of KL_1 is positive (the approximation is not less informative) than the original one.

Inner Approximating Coherent Interval Probabilities

Inner Approximating Coherent Interval Probabilities

Miranda, Montes, Presa, 2023

Given a lower probability $\underline{P}$ they look for a more informative m' in the space of belief functions,

$$d(m', \underline{P}) = \sum_{A \subseteq \Omega} |Bel'(A) - \underline{P}(A)| = \sum_{A \subseteq \Omega} \left| \left(\sum_{E \subseteq A} m'(E) \right) - \underline{P}(A) \right|$$

The restrictions are:

$$m'(E) \geq 0$$
$$\sum_{E \subseteq A} m'(E) \geq \underline{P}(A)$$

The Objective

- The objective is equivalent to minimize:

$$\sum_{E \subseteq \Omega} \frac{m'(E)}{2^{|E|}}$$

which is a measure of specificity.

- If the objective is $KL_1(m', \underline{P})$, then it is equivalent to minimize $NS(\underline{P}) - NS(m')$, i.e. to maximize the non-specificity $NS(m')$.
- In both cases, we have a linear programming problem with a non-unique solution.
- To obtain an unique solution, we could use a quadratic distance as $d^2(m', \underline{P}) = \sqrt{(\sum_{E \subseteq \Omega} Bel'(E) - \underline{P}(E))^2}$ or to add a complement to KL_1 using KL_2

Example

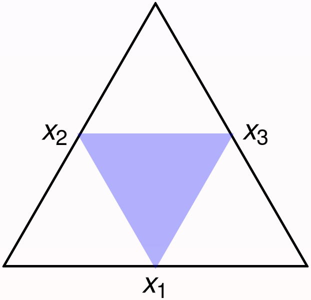
Very often, KL_1 does not provide an unique solution.
Imagine that we are looking for an outer approximation of the probability intervals:

$$\underline{P}(\{x_i\}) = 0; \quad \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Restrictions: $Bel'(A) \geq \underline{P}(A)$

Masses

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0.5	0	0	0.0	0	0.5
0	0.5	0	0.0	0.5	0.0
0	0	0.5	0.5	0.0	0



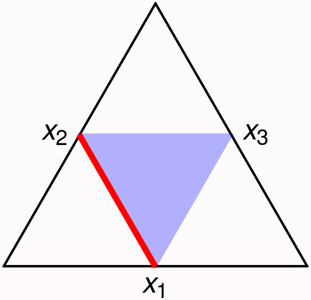
There are 3 extreme solutions with maximum non-specificity

Example

Very often, KL_1 does not provide an unique solution.
Imagine that we are looking for an outer approximation of the probability intervals:

$$\underline{P}(\{x_i\}) = 0; \quad \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Restrictions: $Bel'(A) \geq \underline{P}(A)$



Masses

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0.5	0	0	0.0	0	0.5
0	0.5	0	0.0	0.5	0.0
0	0	0.5	0.5	0.0	0

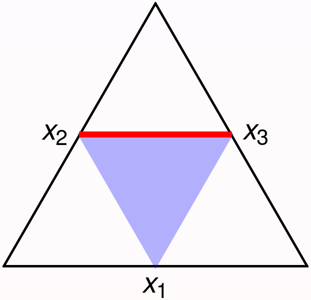
There are 3 extreme solutions with maximum non-specificity

Example

Very often, KL_1 does not provide an unique solution.
Imagine that we are looking for an outer approximation of the probability intervals:

$$\underline{P}(\{x_i\}) = 0; \quad \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Restrictions: $Bel'(A) \geq \underline{P}(A)$



Masses

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0.5	0	0	0.0	0	0.5
0	0.5	0	0.0	0.5	0.0
0	0	0.5	0.5	0.0	0

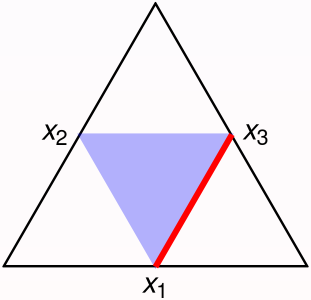
There are 3 extreme solutions with maximum non-specificity

Example

Very often, KL_1 does not provide an unique solution.
Imagine that we are looking for an outer approximation of the probability intervals:

$$\underline{P}(\{x_i\}) = 0; \quad \underline{P}(\{x_i, x_j\}) = 0.5,\ i \neq j$$

Restrictions: $Bel'(A) \geq \underline{P}(A)$



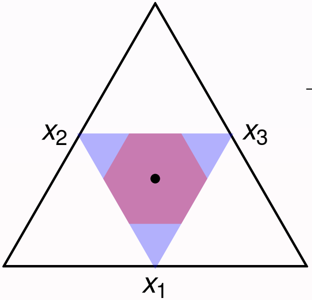
Masses

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0.5	0	0	0.0	0	0.5
0	0.5	0	0.0	0.5	0.0
0	0	0.5	0.5	0.0	0

There are 3 extreme solutions with maximum non-specificity

Unique solution

Using any of the procedures to obtain an unique solution, we obtain the average:



Masses

$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
0.5	0	0	0.0	0	0.5
0	0.5	0	0.0	0.5	0.0
0	0	0.5	0.5	0.0	0
0.5/3	0.5/3	0.5/3	0.5/3	0.5/3	0.5/3

Inner Approximations: Partially Specified Belief Functions

Inner Approximations: Partially Specified Belief Functions; Moral De Campos, 1993

- We have a series of restrictions $Bel_*(A_i) = a_i, i = 1, \ldots, k$ and we want to build a full belief function Bel.
- This problem is similar to probability intervals, when we have a set of lower probability values $P_*(A_i) = b_i, i = 1, \ldots, k$.
- The basic tool in probability intervals are the concepts of avoiding sure loss and natural extension:
  - Avoiding sure loss is equivalent to the existence of a full coherent lower probability $\underline{P}$ satisfying such that $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$, which is equivalent to the existence of a probability P satisfying $P(A_i) \geq b_i$
  - Natural extension is the least informative (lower probability ordering) of the coherent probabilities satisfying $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$
  - Natural extension $\underline{P}$ do not necessarily satisfies $\underline{P}(A_i) = b_i$, but $\underline{P}(A_i) \geq b_i$ (more informative or inner approximation).

Inner Approximations: Partially Specified Belief Functions; Moral De Campos, 1993

- We have a series of restrictions $Bel_*(A_i) = a_i, i = 1, \ldots, k$ and we want to build a full belief function Bel.
- This problem is similar to probability intervals, when we have a set of lower probability values $P_*(A_i) = b_i, i = 1, \ldots, k$.
- The basic tool in probability intervals are the concepts of avoiding sure loss and natural extension:
  - Avoiding sure loss is equivalent to the existence of a full coherent lower probability $\underline{P}$ satisfying such that $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$, which is equivalent to the existence of a probability P satisfying $P(A_i) \geq b_i$.
  - Natural extension is the least informative (lower probability ordering) of the coherent probabilities satisfying $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$.
  - Natural extension $\underline{P}$ do not necessarily satisfies $\underline{P}(A_i) = b_i$ but $\underline{P}(A_i) \geq b_i$ (more informative or inner approximation).

Inner Approximations: Partially Specified Belief Functions; Moral De Campos, 1993

- We have a series of restrictions $Bel_*(A_i) = a_i, i = 1, \ldots, k$ and we want to build a full belief function Bel.
- This problem is similar to probability intervals, when we have a set of lower probability values $P_*(A_i) = b_i, i = 1, \ldots, k$.
- The basic tool in probability intervals are the concepts of avoiding sure loss and natural extension:
  - Avoiding sure loss is equivalent to the existence of a full coherent lower probability $\underline{P}$ satisfying such that $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$, which is equivalent to the existence of a probability P satisfying $P(A_i) \geq b_i$.
  - Natural extension is the least informative (lower probability ordering) of the coherent probabilities satisfying $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$
  - Natural extension $\underline{P}$ do not necessarily satisfies $\underline{P}(A_i) = b_i$, but $\underline{P}(A_i) \geq b_i$ (more informative or inner approximation).

Inner Approximations: Partially Specified Belief Functions; Moral De Campos, 1993

- We have a series of restrictions $Bel_*(A_i) = a_i, i = 1, \ldots, k$ and we want to build a full belief function Bel.
- This problem is similar to probability intervals, when we have a set of lower probability values $P_*(A_i) = b_i, i = 1, \ldots, k$.
- The basic tool in probability intervals are the concepts of avoiding sure loss and natural extension:
  - Avoiding sure loss is equivalent to the existence of a full coherent lower probability $\underline{P}$ satisfying such that $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$, which is equivalent to the existence of a probability P satisfying $P(A_i) \geq b_i$.
  - Natural extension is the least informative (lower probability ordering) of the coherent probabilities satisfying $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$
  - Natural extension $\underline{P}$ do not necessarily satisfies $\underline{P}(A_i) = b_i$, but $\underline{P}(A_i) \geq b_i$ (more informative or inner approximation).

Inner Approximations: Partially Specified Belief Functions; Moral De Campos, 1993

- We have a series of restrictions $Bel_*(A_i) = a_i, i = 1, \ldots, k$ and we want to build a full belief function Bel.
- This problem is similar to probability intervals, when we have a set of lower probability values $P_*(A_i) = b_i, i = 1, \ldots, k$.
- The basic tool in probability intervals are the concepts of avoiding sure loss and natural extension:
  - Avoiding sure loss is equivalent to the existence of a full coherent lower probability $\underline{P}$ satisfying such that $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$, which is equivalent to the existence of a probability P satisfying $P(A_i) \geq b_i$.
  - Natural extension is the least informative (lower probability ordering) of the coherent probabilities satisfying $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$
  - Natural extension $\underline{P}$ do not necessarily satisfies $\underline{P}(A_i) = b_i$, but $\underline{P}(A_i) \geq b_i$ (more informative or inner approximation).

Inner Approximations: Partially Specified Belief Functions; Moral De Campos, 1993

- We have a series of restrictions $Bel_*(A_i) = a_i, i = 1, \ldots, k$ and we want to build a full belief function Bel.
- This problem is similar to probability intervals, when we have a set of lower probability values $P_*(A_i) = b_i, i = 1, \ldots, k$.
- The basic tool in probability intervals are the concepts of avoiding sure loss and natural extension:
  - Avoiding sure loss is equivalent to the existence of a full coherent lower probability $\underline{P}$ satisfying such that $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$, which is equivalent to the existence of a probability P satisfying $P(A_i) \geq b_i$.
  - Natural extension is the least informative (lower probability ordering) of the coherent probabilities satisfying $\underline{P}(A_i) \geq b_i, i = 1, \ldots, k$
  - Natural extension $\underline{P}$ do not necessarily satisfies $\underline{P}(A_i) = b_i$, but $\underline{P}(A_i) \geq b_i$ (more informative or inner approximation).

Avoiding Sure Loss and 'Natural' Extension for Belief Functions

Avoiding Sure Loss

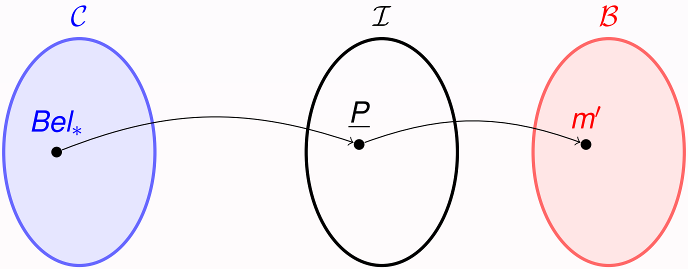
This concept can be defined in a similar way. If we have a partial assessment $Bel_*(A_i) = a_i, i = 1, \ldots, k$, we can say that it avoids sure loss if there is a full belief function Bel, satisfying $Bel(A_i) \geq a_i,\ i = 1, \ldots, k$, which is equivalent to the same concept as lower probability, there is a probability P satisfying $P(A_i) \geq a_i$.

'Natural' Extension

The problem with this concept is that in general there is not a 'least informative' belief function Bel satisfying $Bel(A_i) \geq a_i, i = 1, \ldots, k$. In general, there are several minimal elements.

Extension

A proposal to define an extension of a partially specified belief function: Assume that Bel_* is a partial specification and $\mathcal{C}$ is a partial specification of a belief function, then the proposal is, first make the natural extension building a coherent lower probability $\underline{P}$ in the space $\mathcal{I}$ and finally to build a belief function m' in the space $\mathcal{B}$ of belief functions



Objective Function

- We assume that m' is a true underlying belief function and that Bel_* is a partial knowledge about this belief function, so we have to minimize $KL(m', Bel_*)$, where KL is KL_1 or KL_2.
- In our proposal, we assume that this distance is equal to $KL(m', \underline{P})$, where $\underline{P}$ is the natural extension of Bel_*
- If the initial assessment satisfies avoiding sure loss, there is always a more informative belief (at least one probability) and then the minimum of the first component is 0, and we should look for inner approximations minimizing $KL_1(m', \underline{P}) = (0, NS(\underline{P}) - NS(m'))$, i.e. maximizing non-specificity (least commitment principle).
- If KL_2 is used, then if the solution maximizing non-specificity is not unique, then we have to select the one minimizing:

$$KL(BetP_{m'}, BetP_{\underline{P}})$$

Extension vs Natural Extension

Property of Natural Extension

If we have a partial specification $P_*(A_i) = a_i, i = 1, \dots, k$ and there is a coherent probability $\underline{P'}$ satisfying $\underline{P'}(A_i) = a_i, i = 1, \dots, k$. then the natural extension also satisfies $\underline{P}(A_i) = a_i$.

Extension

If we have a partial specification $Bel_*(A_i) = a_i, i = 1, \dots, k$ and there is a belief function Bel satisfying $Bel(A_i) = a_i, i = 1, \dots, k$, then it is possible that extension satisfies $Bel_{m'}(A_i) > a_i$.

Extension vs Natural Extension

Property of Natural Extension

If we have a partial specification $P_*(A_i) = a_i, i = 1, \dots, k$ and there is a coherent probability $\underline{P'}$ satisfying $\underline{P'}(A_i) = a_i, i = 1, \dots, k$. then the natural extension also satisfies $\underline{P}(A_i) = a_i$.

Extension

If we have a partial specification $Bel_*(A_i) = a_i, i = 1, \dots, k$ and there is a belief function Bel satisfying $Bel(A_i) = a_i, i = 1, \dots, k$, then it is possible that extension satisfies $Bel_{m'}(A_i) > a_i$.

Example

$$Bel_*(\{x_1,x_2\}) = Bel_*(\{x_1,x_3,x_4\}) = Bel_*(\{x_2,x_3,x_4\}) = 0.1,$$
$$Bel_*(\{x_1,x_2,x_3,x_4\}) = 0.2, Bel_*(\{x_1,x_2,x_3,x_4,x_5\}) = 1$$

- The following mass assignment is such that all restrictions are satisfied with equality:

$$m'_1(\{x_1\}) = 0.1, m'_1(\{x_2,x_3,x_4\}) = 0.1, m'_1(\{x_1,x_2,x_3,x_4,x_5\}) = 0.8$$

- But the following one should be chosen first if we consider non-specificity

$$m'_2(\{x_1,x_2\}) = m'_2(\{x_1,x_3,x_4\}) = m'_2(\{x_2,x_3,x_4\}) = 0.1,$$
$$m'_2(\{x_1,x_2,x_3,x_4,x_5\}) = 0.7$$

- $NS(m'_2)=0.1\log(2)+0.2\log(3)+0.7\log(5)>0.1\log(3)+0.8\log(5)=NS(m'_1)$
- But $Bel'_2(\{x_1,x_2,x_3,x_4\}) = 0.3 > 0.2$

Restrictions

- If we have a partial specification of interval probabilities (or a belief function), $P_*(A_i) = a_i,\ i = 1, \ldots, k$, we have considered that we are going to approximate its natural extension $\underline{P}$ and the restrictions are:

$$\sum_{E \subseteq A} m'(E) \geq \underline{P}(A),\ A \subseteq \Omega$$

- However, it is enough to consider a smaller set of restrictions. In particular,

$$\sum_{E \subseteq A_i} m'(E) \geq a_i,\ i = 1, \ldots, k$$

- The proof is simple, taking into account that Bel' is always a coherent lower probability: if the simplified set of restrictions is satisfied, then it will be more informative than the natural extension (least informative lower probability satisfying these restrictions) and the original restrictions are satisfied.

The Focal Sets

- Following, Moral de Campos (1993), if we have a partial specification $Bel_*(A_i) = a_i$, $i = 1, \ldots, k$. if there is a belief function satisfying

$$\sum_{A \subseteq A_i} m'(A) \geq a_i, \; i = 1, \ldots, k$$

then there is another less informative one (strong ordering) with focal elements in the set

$$\mathcal{F} = \{A_{i_1} \cap \cdots \cap A_{i_l} \,|\, l = 1, \ldots, k, \; 1 \leq i_j \leq k\}$$

- So we can concentrate in these focal sets, reducing the number of variables.

The Focusing Principle

The Principle

If we are given $Bel_*(A_i) = a_i$, $i = 1, \ldots, k$, we should first try to concentrate in finding a belief function with $A_1, \ldots, A_k$ as focal elements: the most relevant sets.

If there is not a solution, we should add intersections of two sets: $A_i \cap A_j$; and continue while there is not solution, increasing the size of the intersections.

Additional Intuition

We start looking for a belief function using the largest possible focal sets and then we are adding smaller sets, if there is not solution with the current ones.

Iterative Heuristic Procedure

Basic question

Is there anything similar to the iterative rescaling procedure for outer approximations?

I do not know it, but I believe that it is possible to define some procedure, that follows the lines of the focusing principle and symmetry.

Procedure (assuming that values are corrected by natural extension)

In order of size of sets A, build m' as follows, if $A = A_i$

$$m'(A_i) = Bel_*(A_i) - \sum_{B \subset A_i} m'(B)$$

It is 0 if $A \neq A_i, i = 1, \ldots, k$

Iterative Heuristic Procedure

Basic question

Is there anything similar to the iterative rescaling procedure for outer approximations?

I do not know it, but I believe that it is possible to define some procedure, that follows the lines of the focusing principle and symmetry.

Procedure (assuming that values are corrected by natural extension)

In order of size of sets A, build m' as follows, if $A = A_i$

$$m'(A_i) = Bel_*(A_i) - \sum_{B \subset A_i} m'(B)$$

It is 0 if $A \neq A_i, i = 1, \ldots, k$

Basic Procedure

- The problem is when we are going to compute $m'(A)$ with $A = A_i$ and we have

$$Bel_*(A_i) < \sum_{B \subset A_i} m'(B)$$

then the mass is negative and we do not obtain a belief function.
- In outer approximations we rescaled the largest elements in $\{B \subset A_i \mid m'(B) > 0\}$, by reducing their masses.
- Here, as we are in inner approximations we should increase the masses.

Two ways

$$Bel_*(A_i) < \sum_{B \subset A_i} m'(B)$$

- Method 1: Make $m'(A_i) = 0$, in this way we are increasing $Bel'(A_i)$ to $\sum_{B \subset A_i} m'(B)$
- Method 2: Take $B_1, B_2 \in \{B \subset A_i \mid m(B) > 0\}$, make:

$$m'(B_1 \cap B_2) \leftarrow m'(B_1 \cap B_2) + \alpha$$
$$m'(B_1) \leftarrow m'(B_1) - \alpha, \quad m'(B_2) \leftarrow m'(B_2) - \alpha$$

In this way, beliefs are increased and $\sum_{B \subset A_i} m'(B)$ decreases by α

Method 1 is the preferred (original focal sets) and largest non-specificity (intuition), but it can not always applied: it is not possible if $\sum_{B \subset A_i} m'(B) > 1$, as then the belief is increased to a value larger than 1.0.

The 2 methods

$$Bel_*(\{x_1, x_2, x_3\}) = 0.4$$
$$m'(\{x_1, x_2\}) = 0.3, \;\; m'(\{x_2, x_3\}) = 0.3$$

Method 1

$$m'(\{x_1, x_2\}) = 0.3, \;\; m'(\{x_2, x_3\}) = 0.3, \quad m'(\{x_1, x_2, x_3\}) = 0$$

$$Bel'(\{x_1, x_2, x_3\}) = 0.6 > 0.4,$$

Method 2

$$m'(\{x_1, x_2, x_3\}) = 0,$$
$$m'(\{x_1, x_2\}) = 0.1, \;\; m'(\{x_2, x_3\}) = 0.1, \quad m'(\{x_2\}) = 0.2$$
$$Bel'(\{x_2\}) = 0.2 > 0.0$$

I believe that Method 1 should be preferred but not always possible.

Principles

- Focusing Principle: it tries to use focal elements as close as possible to the original, adding intersections only when necessary,
- Symmetry Principle: When method 2 is applied, all the transformations with the largest $B_i \cap B_j$ should be done simultaneously,

Example

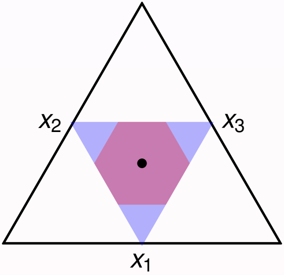
With $\Omega = \{x_1, x_2, x_3\}$ if we have
$Bel_*(\{x_1, x_2\}) = Bel_*(\{x_1, x_3\}) = Bel_*(\{x_2, x_3\}) = 0.5$
We first obtain, $m'(\{x_1, x_2\}) = m'(\{x_1, x_3\}) = m'(\{x_2, x_3\}) = 0.5$
When computing $m'(\{x_1, x_2, x_3\})$ we obtain a negative mass -0.5, and only Method 2 can be applied. All the intersections of the same size are $\{x_1\}, \{x_2\}, \{x_3\}$, we should increase the masses of these sets, and decrease the already assigned masses, obtaining;

$$m'(\{x_1, x_2\}) = m'(\{x_1, x_3\}) = m'(\{x_2, x_3\}) = 0.5/3$$
$$m'(\{x_1\}) = m'(\{x_2\}) = m'(\{x_3\}) = 0.5/3$$

Example

With $\Omega = \{x_1, x_2, x_3\}$ if we have
$Bel_*(\{x_1, x_2\}) = Bel_*(\{x_1, x_3\}) = Bel_*(\{x_2, x_3\}) = 0.5$
We first obtain, $m'(\{x_1, x_2\}) = m'(\{x_1, x_3\}) = m'(\{x_2, x_3\}) = 0.5$
And finally,

$$m'(\{x_1, x_2\}) = m'(\{x_1, x_3\}) = m'(\{x_2, x_3\}) = 0.5/3$$
$$m'(\{x_1\}) = m'(\{x_2\}) = m'(\{x_3\}) = 0.5/3$$



Combination as Approximation

Combination as Approximation

- If we have two belief functions m_1, m_2, then an idempotent combination rule can be defined as the problem of finding an underlying belief function m' such as m_1, m_2, are the best approximations of it, i.e. m' is computed by minimizing:

$$KL(m', m_1) + KL(m', m_2)$$

- Should m' be more informative than m_1 and m_2? desirable, but not always possible: for example if m_1 and m_2 are Bayesian, then it is not possible, except if they are equal.

Combination as Approximation

- If we have two belief functions m_1, m_2, then an idempotent combination rule can be defined as the problem of finding an underlying belief function m' such as m_1, m_2, are the best approximations of it, i.e. m' is computed by minimizing:

$$KL(m', m_1) + KL(m', m_2)$$

- Should m' be more informative than m_1 and m_2? desirable, but not always possible: for example if m_1 and m_2 are Bayesian, then it is not possible, except if they are equal.

Computation (not simple in general)

- After some transformations, we should select (minimizing first component) m' proportional to $\sqrt{m_1^*.m_2^*}$ (pointwise multiplication) where $m_1^* \sqsubseteq_s m_1$ and $m_2^* \sqsubseteq_s m_2$, maximizing

$$\sum_{A \subseteq \Omega} \sqrt{m_1^*(A).m_2^*(A)}$$

- If there is a belief m that is simultaneously $m \sqsubseteq_s m_1$, $m \sqsubseteq_s m_2$, the optimum is 1 and it is only obtained for the masses with this property.
- Then, between all the m' computed in this way, we should select one with maximum non-specificity.
- Finally, if the solution is not unique, we should select one for example, minimizing the sum divergences of the pignistic probabilities $BetP_1$ and $BetP_2$ to the pignistic probabilily $BetP'$.

Particular Cases: Probabilistic

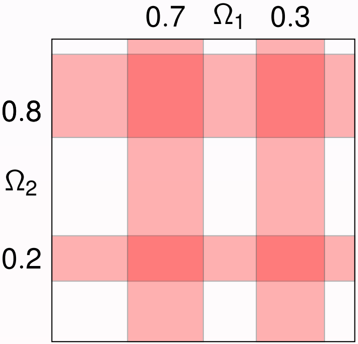
- When m_1 and m_2 are Bayesian, we obtain

$$m'(\{x\}) = \frac{\sqrt{m_1(\{x\})m_2(\{x\})}}{\sum_{y\in\Omega}\sqrt{m_1(\{y\})m_2(\{y\})}}$$

Multiplicative opinion pool rule (aggregating probabilities).

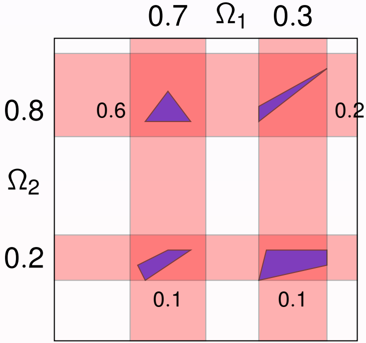
The Marginal Problem

- We have two frames Ω_1 and Ω_2 and m_1, m_2, defined on Ω_1 and Ω_2, respectively.
- We want to compute a belief m' in $\Omega_1 \times \Omega_2$.



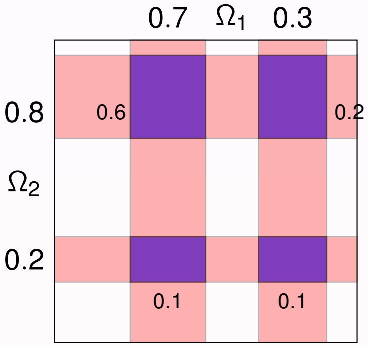
The Marginal Problem: Minimizing the first component

We must select a belief function such that its marginals are more informative than the original ones.



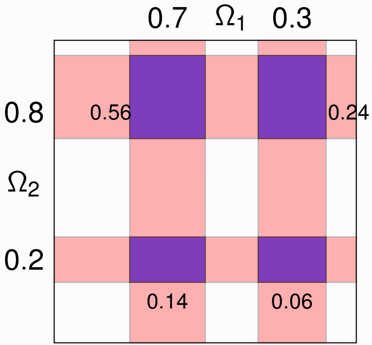
The Marginal Problem: Minimizing the second component

Maximum non-specificity: concentration in rectangles



The Marginal Problem: Minimizing the third component

Minimum divergence betting probabilities: Dempster's rule



Outer Approximations by Necessity-Possibility

Outer Approximation by a Necessity-Possibility

- We have that $\mathcal{C}$ is the space of belief functions.
- $\mathcal{B}$ is the space of possibility measures.
- Given a mass assignment m we look for a consonant mass assignment m' minimizing $KL(m, m')$.
- As there is always a consonant belief (the vacuous belief m_0) that is less informative than m, we should look for a m which is less informative and minimizing non-specificity.

Outer Approximation by a Possibility

- Optimal Mass Allocation Procedure (Dubois, Prade, 1990)
- We start with a permutation σ on $\{1, \dots, n\}$ of the elements of $\Omega = \{x_1, \dots, x_n\}$
- The following nested focal elements are considered:

$$
\begin{array}{l}
\{x_{\sigma(1)}\} \\
\{x_{\sigma(1)}, x_{\sigma(2)}\} \\
\{x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)}\} \\
\{x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)}, x_{\sigma(4)}\}
\end{array}
$$

- For each $i = 1, \dots, n$ and for each focal set A, if A contains $x_{\sigma(j)}, j > i$, and not $x_{\sigma(i)}$, then add $x_{\sigma(i)}$ to A.
- In this way, less informative (outer) approximations are obtained.

Basic Step

$$\Omega = \{x_1, x_2, x_3, x_4\}$$

- Order $\sigma = (3, 2, 4, 1)$
- If I have focal set $\{x_1, x_2\}$, I have to add x_3 to it, obtaining $\{x_1, x_2, x_3\}$
- If I have focal set: $\{x_1, x_2, x_3\}$, I have to add x_4 to it, obtaining $\{x_1, x_2, x_3, x_4\}$

Example, Montes, Miranda, Vicig

$$m(\{x_1\}) = 0.3, m(\{x_2\}) = 0.3, m(\{x_1, x_2\}) = 0.1, m(\{x_2, x_3\}) = 0.3$$

Consider the permutation $\sigma = (1, 2, 3)$

	m_σ
$\{x_1\}$	0.3
$\{x_1, x_2\}$	0.4
$\{x_1, x_2, x_3\}$	0.3

- $\pi_\sigma(x_1) = 1, \quad \pi_\sigma(x_2) = 0.7, \quad \pi_\sigma(x_3) = 0.3$
- It is a maximal outer approximation.

Example

$$m(\{x_1\}) = 0.3, m(\{x_2\}) = 0.3, m(\{x_1, x_2\}) = 0.1, m(\{x_2, x_3\}) = 0.3$$

Consider the permutation $\sigma' = (1, 3, 2)$

	$m_{\sigma'}$
$\{x_1\}$	0.3
$\{x_1, x_3\}$	0.0
$\{x_1, x_2, x_3\}$	0.7

- $\pi_{\sigma'}(x_1) = 1, \quad \pi_{\sigma'}(x_2) = 0.7, \quad \pi_{\sigma'}(x_3) = 0.7$
- It is not a maximal outer approximation.
- $\pi_{\sigma}(x_1) = 1, \quad \pi_{\sigma}(x_2) = 0.7, \quad \pi_{\sigma}(x_3) = 0.3$

Example

$$m(\{x_1\}) = 0.3, m(\{x_2\}) = 0.3, m(\{x_1, x_2\}) = 0.1, m(\{x_2, x_3\}) = 0.3$$

The Rule

If x_i, x_j is such that any focal set containing x_j also contains x_i but not the other way round, then i should precede j.

So the order $\sigma' = (1, 3, 2)$ should not be considered.

Example

Heuristic Procedure

- $m' = m$
- For $i = 1, \ldots, n$
  - Select $x_\sigma(i) \in A_i = \Omega \setminus \{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}$ minimizing:

$$\sum_{\substack{A \cap (A_i \setminus \{x_{\sigma(i)}\}) \neq \emptyset \\ x_{\sigma(i)} \notin A}} m'(A) \log\left(\frac{|A|}{|A|+1}\right)$$

  - Add $x_{\sigma(i)}$ to the focal elements A, such that $A \cap (A_i \setminus \{x_{\sigma(i)}\}) \neq \emptyset$

$$m(\{x_1\}) = 0.3, m(\{x_2\}) = 0.3, m(\{x_1, x_2\}) = 0.1, m(\{x_2, x_3\}) = 0.3$$

It selects the permutation: $\sigma'' = (2, 1, 3)$, with associated possibility:

$\pi_{\sigma''}(x_1) = 0.7 \quad \pi_{\sigma''}(x_2) = 1.0, \quad \pi_{\sigma''}(x_3) = 0.3$

Inner Aproximation: Approximating with Possibility Measures

Inner Approximation of Belief by Possibility

- Given m we propose to compute a consonant belief function m' minimizing: $KL(m', m)$
- There is not always a consonant belief function m' which is more informative than m, A necessary and sufficient condition is that there is $x \in \Omega$ with $Pl(\{x\}) = 1$.
- But this only means that the first compoment of KL is not always 0.
- The minimum of the first component is

$$\max_{x \in \Omega} \log(Pl(\{x\}))$$

and it is obtained by any m' which is strongly included in the belief function obtained by removing all the focal elements not containing x and normalizing afterwards.

Heuristic Procedure

To inner approximate m minimizing $KL(m', m)$, an order is computed in which the first element is x maximizing $Pl(\{x\})$.

- Select $x_j \in \Omega$ maximizing $Pl(\{x\})$, let $\sigma(1) = j$.
- Let m' equal to m removing all the focal elements not containing x_j. Normalize
- For $i = 2$ to n
  - Select $\sigma(i) = x_j \in A_i = \Omega \setminus \{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}$
  - For any focal set A with $A \cap A_i \neq \emptyset$, $x_j \notin A$, move its mass to $A_i = \{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}$.

Example

Already in the order $(1, 2)$ we are going to assign the third element: if $\sigma(3) = 3$, then we should remove x_4 from focal set $\{x_1, x_2, x_4\}$.

Hewuristic procedure

For $i = 2$ to n

- Select $\sigma(i)$ equal to x_j minimizing:

$$\sum_{\substack{A \cap A_i \neq \emptyset \\ x_j \notin A}} m'(A) \log(\frac{|A|}{i-1})$$

  where $A_i = \Omega \setminus \{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}$
- For any A with $A \cap A_i \neq \emptyset$, $x_j \notin A$, move its mass to $A_i = \{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}$.

Example

$$m(\{x_1\}) = 0.3, m(\{x_2\}) = 0.3, m(\{x_1, x_2\}) = 0.1, m(\{x_2, x_3\}) = 0.3$$

- $\sigma(1) = 2$, remove masses not containing x_2 and normalize

$$m'(\{x_2\}) = 3/7, m'(\{x_1, x_2\}) = 1/7, m'(\{x_2, x_3\}) = 3/7$$

- Now, if $\sigma(2) = 1$, the mass from $\{x_2, x_3\}$ goes to $\{x_2\}$
  If $\sigma(2) = 3$, the mass from $\{x_1, x_2\}$ goes to $\{x_2\}$
- The second option is better, Final result

$$m'(\{x_2\}) = 4/7, m'(\{x_2, x_3\}) = 3/7$$

- Associated possibility

$$\pi'(x_2) = 1, \pi'(x_3) = 3/7, \pi'(x_1) = 0$$

Other Approximation Problems and Techniques

- Inner and Outer approximations with special cases: p-boxes, distorsion models,...
  Example: Given a mass m on $\Omega_1 \times \Omega_2$, compute m_1 on Ω_1, m_2 on Ω_2 minimizing $KL(m_1 \oplus m_2, m)$: variational evidential inference.
- Monte Carlo methods for Dempster's combination: importance sampling, Markov Chain Monte Carlo.

References

- Dubois, D„ and H. Prade. "Consonant approximations of belief functions." International Journal of Approximate Reasoning 4.5-6 (1990): 419-449.
- D. Harmanec. Faithful approximations of belief functions. In K. B. Laskey and H. Prade, editors, Uncertainty in Artificial Intelligence 15 (UAI99), Stockholm, Sweden, 1999.
- Hall, J. W., and J, Lawry. "Generation, combination and extension of random set approximations to coherent lower and upper probabilities." Reliability Engineering & System Safety 85.1-3 (2004): 89-101.
- Miranda, E., I, Montes, A. Presa. "Inner approximations of coherent lower probabilities and their application to decision making problems." Annals of Operations Research (2023): 1-39.
- Montes, I., E. Miranda, and P. Vicig. "Outer approximating coherent lower probabilities with belief functions." International Journal of Approximate Reasoning 110 (2019): 1-30.

References

- Moral, S., and L.M. De Campos. "Partially specified belief functions." Uncertainty in Artificial Intelligence. Morgan Kaufmann, 1993.
- Quaeghebeur, E. "Completely Monotone Outer Approximations of Lower Probabilities on Finite Possibility Spaces." NL-MUA. 2011.
- Shenoy, P.P. (2024). Mutual Information and Kullback-Leibler Divergence in the Dempster-Shafer Theory. In: Bi, Y., Jousselme, AL., Denoeux, T. (eds) Belief Functions: Theory and Applications. BELIEF 2024. Lecture Notes in Computer Science(), vol 14909. Springer, Cham. https://doi.org/10.1007/978-3-031-67977-3_24
- Y. Song and Y. Deng, Divergence Measure of Belief Function and Its Application in Data Fusion, in IEEE Access, vol. 7, pp. 107465-107472, 2019.
- Soubaras, Hélène. "Towards an axiomatization for the generalization of the Kullback-Leibler divergence to belief functions." In Proceedings of the 7th conference of the European Society for Fuzzy Logic and Technology, pp. 1090-1097. Atlantis Press, 2011.

References

- A. Wallner, Extreme points of coherent probabilities in finite spaces, Int. J. Approx. Reason. 44 (3), 2007, 339–357.
- F. Xiao, Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy, Inf. Fusion, vol. 46, pp. 23–32, Mar. 2019.