

Chapter 5 : Uncertain Data Fusion

John Klein

Lille1 Université - CRIStAL UMR CNRS 9189



Chapter organization

- 1 Probabilistic data fusion
- 2 Probabilistic set-valued data fusion
- 3 Evidential data fusion
- 4 Appendix

- In this chapter, each datum delivered by a **source** is (possibly) tainted with **uncertainty**.
- The chapter is organized into sections w.r.t. the chosen **uncertainty** representation framework.

Uncertainty theories in the literature :

- The **probability theory** is by far the most frequently used and renowned one.
- Second most renowned is the **possibility theory** [6] which relies on Zadeh's fuzzy sets theory [5].
- Possibilistic approaches gained popularity in the second half of the 20th century because they provide a simple and flexible framework to describe uncertain situations that probabilities fail to fully grasp.
- **Ignorance**, and more generally **imprecise proposals**, are encoded awkwardly by probabilities.

Uncertainty theories in the literature :

- The **probability theory** is by far the most frequently used and renowned one.
- Second most renowned is the **possibility theory** [6] which relies on Zadeh's fuzzy sets theory [5].
- Possibilistic approaches gained popularity in the second half of the 20th century because they provide a simple and flexible framework to describe uncertain situations that probabilities fail to fully grasp.
- **Ignorance**, and more generally **imprecise proposals**, are encoded awkwardly by probabilities.

Uncertainty theories in the literature :

- The **probability theory** is by far the most frequently used and renowned one.
- Second most renowned is the **possibility theory** [6] which relies on Zadeh's fuzzy sets theory [5].
- Possibilistic approaches gained popularity in the second half of the 20th century because they provide a simple and flexible framework to describe uncertain situations that probabilities fail to fully grasp.
- **Ignorance**, and more generally **imprecise proposals**, are encoded awkwardly by probabilities.

Uncertainty theories in the literature :

- The **probability theory** is by far the most frequently used and renowned one.
- Second most renowned is the **possibility theory** [6] which relies on Zadeh's fuzzy sets theory [5].
- Possibilistic approaches gained popularity in the second half of the 20th century because they provide a simple and flexible framework to describe uncertain situations that probabilities fail to fully grasp.
- **Ignorance**, and more generally **imprecise proposals**, are encoded awkwardly by probabilities.

Uncertainty theories :

- Imprecision and uncertainty can be jointly handled by a more general framework introduced in the 90's by Walley [4] : imprecise probabilities.
- Imprecise probabilities are actually interval of probabilities. For instance, the statement $\mathbb{P}(A) \leq 0.9$ is an imprecise probability.
- Imprecise probabilities are also in correspondence with sets of probability measures. In our example, this is the set of all probability measures such that the measure of A is less than 0.9.

Uncertainty theories :

- Imprecision and uncertainty can be jointly handled by a more general framework introduced in the 90's by Walley [4] : imprecise probabilities.
- Imprecise probabilities are actually interval of probabilities. For instance, the statement $\mathbb{P}(A) \leq 0.9$ is an imprecise probability.
- Imprecise probabilities are also in correspondence with sets of probability measures. In our example, this is the set of all probability measures such that the measure of A is less than 0.9.

Uncertainty theories :

- Imprecision and uncertainty can be jointly handled by a more general framework introduced in the 90's by Walley [4] : imprecise probabilities.
- Imprecise probabilities are actually interval of probabilities. For instance, the statement $\mathbb{P}(A) \leq 0.9$ is an imprecise probability.
- Imprecise probabilities are also in correspondence with sets of probability measures. In our example, this is the set of all probability measures such that the measure of A is less than 0.9.

Uncertainty theories :

- We will also investigate a special case of **imprecise probabilities** known as **belief functions**.
- The **belief function** theory encompasses the **probability** theory, the **possibility** theory and Cantor's **set theory**.
- This latter framework is also known as **Dempster-Shafer theory of evidence** [1, 2].
- In the worst case, we only have **probabilities of probabilities** which are captured by **random measures**.

Uncertainty theories :

- We will also investigate a special case of imprecise probabilities known as belief functions.
- The belief function theory encompasses the probability theory, the possibility theory and Cantor's set theory.
- This latter framework is also known as Dempster-Shafer theory of evidence [1, 2].
- In the worst case, we only have probabilities of probabilities which are captured by random measures.

Uncertainty theories :

- We will also investigate a special case of **imprecise probabilities** known as **belief functions**.
- The **belief function** theory encompasses the **probability** theory, the **possibility** theory and Cantor's **set theory**.
- This latter framework is also known as **Dempster-Shafer theory of evidence** [1, 2].
- In the worst case, we only have **probabilities of probabilities** which are captured by **random measures**.

Uncertainty theories :

- We will also investigate a special case of imprecise probabilities known as belief functions.
- The belief function theory encompasses the probability theory, the possibility theory and Cantor's set theory.
- This latter framework is also known as Dempster-Shafer theory of evidence [1, 2].
- In the worst case, we only have probabilities of probabilities which are captured by random measures.

Uncertainty theories :

- The hierarchy of uncertainty theories is given in the following figure :

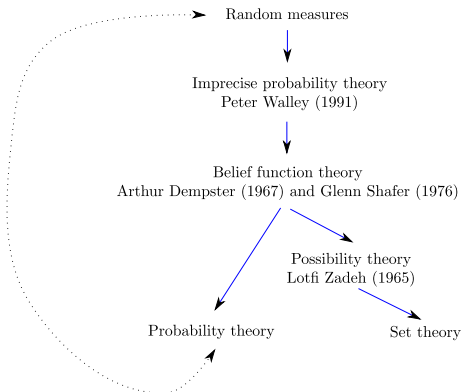


FIGURE : Uncertainty theories from most general (top) to most specific (bottom).

Uncertainty theories : notations

- Let $\mathcal{P}_{\mathcal{X}}$ denote the set of **probability** measures expressing the odds that a candidate value in \mathcal{X} is x .
- Let $\mathcal{B}_{\mathcal{X}}$ denote the set of **belief functions** expressing the odds that a candidate value in \mathcal{X} is x .
- Finally, let $2^{\mathcal{P}_{\mathcal{X}}}$ denote the **power set** of $\mathcal{P}_{\mathcal{X}}$.
- The hierarchy displayed in the previous figure implies that :

$$\mathcal{P}_{\mathcal{X}} \subsetneq \mathcal{B}_{\mathcal{X}} \subsetneq 2^{\mathcal{P}_{\mathcal{X}}}. \quad (1)$$

Generalized definitions of **conjunctive** and **disjunctive** operators :

- Time for a broader definition of the conjunctive and disjunctive nature of fusion operators.

Definition

Let us consider a (general) data fusion problem with \hat{x} is a given fusion operator. \hat{x} is said to be **conjunctive** if for any source with advocacy stating that $x = x'$ is (almost surely) impossible, then the aggregate also states so.

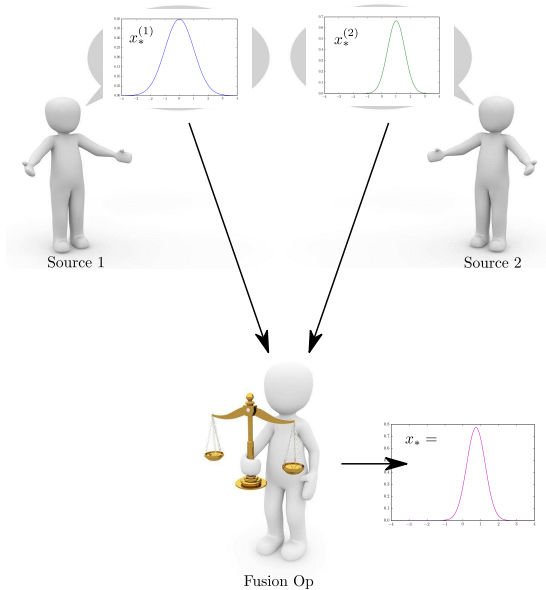
Definition

Let us consider a (general) data fusion problem with \hat{x} is a given fusion operator. \hat{x} is said to be **disjunctive** if for any source with advocacy stating that $x = x'$ is not (almost surely) impossible, then the aggregate also states so.

Chapter organization

- 1 Probabilistic data fusion
- 2 Probabilistic set-valued data fusion
- 3 Evidential data fusion
- 4 Appendix

Probabilistic data fusion setting :



Probabilistic data fusion :

- Let us give a formal definition of probabilistic data fusion problems :

Definition

Probabilistic data fusion is a subclass of data fusion where advocacies live in $\mathbb{X} = \mathcal{P}_{\mathcal{X}}$.

- In such problems, informations sources deliver a discrete distributions or a density¹.

1. existence of densities, or Radon Nikodym derivatives w.r.t. Lebesgue is always assumed in this course.

Probabilistic data fusion : paradox ?

- Suppose, one has only **two sources** to aggregate.
- The 1st one delivers a advocacy $x_*^{(1)} = P_x^{(1)}$ and the 2nd one delivers $x_*^{(2)} = P_x^{(2)}$.
- **Impossible** unless $P_x^{(1)} = P_x^{(2)} = P_x$! Indeed, there is only one probability for each possible value of x .
- There cannot be two laws for one r.v. otherwise at least one of the source is just **wrong**.

Probabilistic data fusion : paradox ?

- Suppose, one has only **two sources** to aggregate.
- The 1st one delivers a advocacy $x_*^{(1)} = P_x^{(1)}$ and the 2nd one delivers $x_*^{(2)} = P_x^{(2)}$.
- **Impossible** unless $P_x^{(1)} = P_x^{(2)} = P_x$! Indeed, there is only one probability for each possible value of x .
- There cannot be two laws for one r.v. otherwise at least one of the source is just **wrong**.

Probabilistic data fusion : paradox ?

- Suppose, one has only **two sources** to aggregate.
- The 1st one delivers a advocacy $x_*^{(1)} = P_x^{(1)}$ and the 2nd one delivers $x_*^{(2)} = P_x^{(2)}$.
- **Impossible** unless $P_x^{(1)} = P_x^{(2)} = P_x$! Indeed, there is only one probability for each possible value of x .
- There cannot be two laws for one r.v. otherwise at least one of the source is just **wrong**.

Probabilistic data fusion : paradox ?

- Suppose, one has only **two sources** to aggregate.
- The 1st one delivers a advocacy $x_*^{(1)} = P_x^{(1)}$ and the 2nd one delivers $x_*^{(2)} = P_x^{(2)}$.
- **Impossible** unless $P_x^{(1)} = P_x^{(2)} = P_x$! Indeed, there is only one probability for each possible value of x .
- There cannot be two laws for one r.v. otherwise at least one of the source is just **wrong**.

Probabilistic data fusion : paradox ?

- Suppose, one has only two sources to aggregate.
- The 1st one delivers a advocacy $x_*^{(1)} = P_x^{(1)}$ and the 2nd one delivers $x_*^{(2)} = P_x^{(2)}$.
- Impossible unless $P_x^{(1)} = P_x^{(2)} = P_x$! Indeed, there is only one probability for each possible value of x .
- There cannot be two laws for one r.v. otherwise at least one of the source is just wrong.

→ $P_x^{(1)}$ is just an estimate of P_x given the data seen by source i and the source model.

Model-free probabilistic data fusion :

- Buidling upon the preceding reasoning, let S_i denote a **latent r.v.** standing for hypotheses regarding source i .
- Since each source actually delivers a conditional distribution $P_{x|S_i}$, one actually aims at determining the aggregate $x_* = P_{x|S_1, \dots, S_{N_s}}$.
- Without loss of generality, suppose x is discrete. For any $x' \in \mathcal{X}$ and any $(t_1, \dots, t_{N_s}) \in \times_{i=1}^{N_s} \mathcal{T}_i$ such that $P_{S_1, \dots, S_{N_s}}(t_1, \dots, t_{N_s}) > 0$, one can write :

$$\begin{aligned}
 &P_{x|S_1=t_1, \dots, S_{N_s}=t_{N_s}}(x') \\
 &= \frac{P_{x, S_1, \dots, S_{N_s}}(x', t_1, \dots, t_{N_s})}{P_{S_1, \dots, S_{N_s}}(t_1, \dots, t_{N_s})},
 \end{aligned}$$

Model-free probabilistic data fusion :

- Building upon the preceding reasoning, let S_i denote a **latent r.v.** standing for hypotheses regarding source i .
- Since each source actually delivers a conditional distribution $P_{x|S_i}$, one actually aims at determining the aggregate $x_* = P_{x|S_1, \dots, S_{N_s}}$.
- Without loss of generality, suppose x is discrete. For any $x' \in \mathcal{X}$ and any $(t_1, \dots, t_{N_s}) \in \times_{i=1}^{N_s} \mathcal{T}_i$ such that $P_{S_1, \dots, S_{N_s}}(t_1, \dots, t_{N_s}) > 0$, one can write :

$$\begin{aligned}
 &P_{x|S_1=t_1, \dots, S_{N_s}=t_{N_s}}(x') \\
 &= \frac{P_{x, S_1, \dots, S_{N_s}}(x', t_1, \dots, t_{N_s})}{P_{S_1, \dots, S_{N_s}}(t_1, \dots, t_{N_s})},
 \end{aligned}$$

Model-free probabilistic data fusion :

- Building upon the preceding reasoning, let S_i denote a **latent r.v.** standing for hypotheses regarding source i .
- Since each source actually delivers a conditional distribution $P_{x|S_i}$, one actually aims at determining the aggregate $x_* = P_{x|S_1, \dots, S_{N_s}}$.
- Without loss of generality, suppose x is discrete. For any $x' \in \mathcal{X}$ and any $(t_1, \dots, t_{N_s}) \in \times_{i=1}^{N_s} \mathcal{T}_i$ such that $P_{S_1, \dots, S_{N_s}}(t_1, \dots, t_{N_s}) > 0$, one can write :

$$\begin{aligned}
 &P_{x|S_1=t_1, \dots, S_{N_s}=t_{N_s}}(x') \\
 &= \frac{P_{x, S_1, \dots, S_{N_s}}(x', t_1, \dots, t_{N_s})}{P_{S_1, \dots, S_{N_s}}(t_1, \dots, t_{N_s})},
 \end{aligned}$$

Model-free probabilistic data fusion : $P_{x|S_1=t_1,\dots,S_{N_s}=t_{N_s}}(x')$

$$\begin{aligned}
 &= \frac{P_{x,S_1,\dots,S_{N_s}}(x', t_1, \dots, t_{N_s})}{\prod_{i=1}^{N_s} P_{S_i}(t_i)}, \text{ (source indep.)} \\
 &= \frac{P_{S_1,\dots,S_{N_s}|x=x'}(t_1, \dots, t_{N_s}) P_x(x')}{\prod_{i=1}^{N_s} P_{S_i}(t_i)}, \text{ (Bayes)} \\
 &= \frac{P_{S_2,\dots,S_{N_s}|x=x',S_1=t_1}(t_2, \dots, t_{N_s}) P_{S_1|x=x'}(t_1) P_x(x')}{\prod_{i=1}^{N_s} P_{S_i}(t_i)}, \text{ (Bayes repeated)} \\
 &= \frac{P_{S_{N_s}|x=x',S_1=t_1,\dots,S_{N_s-1}=t_{N_s-1}}(t_{N_s}) \times \dots \times P_{S_2|x=x',S_1=t_1}(t_2) P_{S_1|x=x'}(t_1) P_x(x')}{\prod_{i=1}^{N_s} P_{S_i}(t_i)}, \\
 &= P_x(x') \prod_{i=1}^{N_s} \frac{P_{S_i|x=x'}(t_i)}{P_{S_i}(t_i)}, \text{ (source indep.)}
 \end{aligned}
 \tag{2}$$

Model-free probabilistic data fusion : $P_{x|S_1=t_1,\dots,S_{N_s}=t_{N_s}}(x')$

$$\begin{aligned}
 &= P_x(x') \prod_{i=1}^{N_s} \frac{\frac{P_{x|S_i=t_i}(x') P_{S_i}(t_i)}{P_x(x')}}{P_{S_i}(t_i)}, \text{ (Bayes)} \\
 &= \frac{\prod_{i=1}^{N_s} P_{x|S_i=t_i}(x')}{P_x(x')^{N_s-1}}.
 \end{aligned}$$

Model-free probabilistic data fusion :

Definition

Each advocacy is a conditional probability law : $x_*^{(i)} = P_{x|S_i}$. The **Bayes operator** \hat{x}_{bay} is defined as follows :

$$\hat{x}_{bay} \left(P_{x|S_1}, \dots, P_{x|S_{N_s}} \right) \rightarrow \frac{1}{a} \prod_{i=1}^{N_s} P_{x|S_i}, \quad (3)$$

with a a normalization constant so that x_* is a probability distribution.

- This kind of fusion operator is called **Bayes operator** because of the extensive use of Bayes formula in the derivation of this operator but this is not a proper Bayesian approach.
- The r.v. S_i could also contain **meta-data** on the source reliability.

Model-free probabilistic data fusion :

Definition

Each advocacy is a conditional probability law : $x_*^{(i)} = P_{x|S_i}$. The **Bayes operator** \hat{x}_{bay} is defined as follows :

$$\hat{x}_{bay} \left(P_{x|S_1}, \dots, P_{x|S_{N_s}} \right) \rightarrow \frac{1}{a} \prod_{i=1}^{N_s} P_{x|S_i}, \quad (3)$$

with a a normalization constant so that x_* is a probability distribution.

- This kind of fusion operator is called **Bayes operator** because of the extensive use of Bayes formula in the derivation of this operator but this is not a proper Bayesian approach.
- The r.v. S_i could also contain **meta-data** on the source reliability.

Model-free probabilistic data fusion :

Definition

Each advocacy is a conditional probability law : $x_*^{(i)} = P_{x|S_i}$. The **Bayes operator** \hat{x}_{bay} is defined as follows :

$$\hat{x}_{bay} \left(P_{x|S_1}, \dots, P_{x|S_{N_s}} \right) \rightarrow \frac{1}{a} \prod_{i=1}^{N_s} P_{x|S_i}, \quad (3)$$

with a a normalization constant so that x_* is a probability distribution.

- This kind of fusion operator is called **Bayes operator** because of the extensive use of Bayes formula in the derivation of this operator but this is not a proper Bayesian approach.
- The r.v. S_i could also contain **meta-data** on the source reliability.

Bayes op :

Property

The bayes operator \hat{x}_{bay} is a **conjunctive** fusion operator.

Proof

The proof is trivial. If one source is sure that x' is not a possible value for x , then this means that $x_*^{(i)}(x') = P_{x|S_i}(x') = 0$. According to equation (3), this implies that $x_*(x') = P_{x|S_1, \dots, S_{N_s}}(x') = 0$, which of course means that x' is not a possible value for x .

Bayes op : limitations

- The bayes operator can be criticized because equation (3) consists in multiplying rather small values. When the number of source N_s grows, machine precision may be reached.
- This is usually circumvented by using programming tricks like log-probabilities.
- Due to its conjunctive nature, the way advocacies are obtained by the sources must be handled with great care. Indeed, if a source states that $x = x'$ is impossible while this is untrue, then x' is permanently eliminated from the set of solutions.

Bayes op : limitations

- The bayes operator can be criticized because equation (3) consists in multiplying rather small values. When the number of source N_s grows, machine precision may be reached.
- This is usually circumvented by using programming tricks like log-probabilities.
- Due to its conjunctive nature, the way advocacies are obtained by the sources must be handled with great care. Indeed, if a source states that $x = x'$ is impossible while this is untrue, then x' is permanently eliminated from the set of solutions.

Bayes op : limitations

- The bayes operator can be criticized because equation (3) consists in multiplying rather small values. When the number of source N_s grows, machine precision may be reached.
- This is usually circumvented by using programming tricks like log-probabilities.
- Due to its conjunctive nature, the way advocacies are obtained by the sources must be handled with great care. Indeed, if a source states that $x = x'$ is impossible while this is untrue, then x' is permanently eliminated from the set of solutions.

Parametric probabilistic data fusion :

- Instead of using Bayes and use independence assumption, one can also use a **parametric model** for the distribution $x_* = P_{x|S_1, \dots, S_{N_s}; \pi}$.
- A classical choice are **mixture models** :

Definition

Each advocacy is a conditional probability law : $x_*^{(i)} = P_{x|S_i}$. The mixture operator \hat{x}_{mix} is defined as follows :

$$\hat{x}_{mix} \left(P_{x|S_1}, \dots, P_{x|S_{N_s}} \right) \rightarrow \sum_{i=1}^{N_s} \pi^{(i)} P_{x|S_i}, \quad (4)$$

with $\pi^{(i)}$ positive coefficients summing to one : $\sum_{i=1}^{N_s} \pi^{(i)} = 1$.

Parametric probabilistic data fusion :

- Instead of using Bayes and use independence assumption, one can also use a **parametric model** for the distribution $x_* = P_{x|S_1, \dots, S_{N_s}; \pi}$.
- A classical choice are **mixture models** :

Definition

Each advocacy is a conditional probability law : $x_*^{(i)} = P_{x|S_i}$. The **mixture operator** \hat{x}_{mix} is defined as follows :

$$\hat{x}_{mix} \left(P_{x|S_1}, \dots, P_{x|S_{N_s}} \right) \rightarrow \sum_{i=1}^{N_s} \pi^{(i)} P_{x|S_i}, \quad (4)$$

with $\pi^{(i)}$ positive coefficients summing to one : $\sum_{i=1}^{N_s} \pi^{(i)} = 1$.

Parametric probabilistic data fusion :

- Instead of using Bayes and use independence assumption, one can also use a **parametric model** for the distribution $x_* = P_{x|S_1, \dots, S_{N_s}; \pi}$.
- A classical choice are **mixture models** :

Definition

Each advocacy is a conditional probability law : $x_*^{(i)} = P_{x|S_i}$. The **mixture operator** \hat{x}_{mix} is defined as follows :

$$\hat{x}_{mix} \left(P_{x|S_1}, \dots, P_{x|S_{N_s}} \right) \rightarrow \sum_{i=1}^{N_s} \pi^{(i)} P_{x|S_i}, \quad (4)$$

with $\pi^{(i)}$ positive coefficients summing to one : $\sum_{i=1}^{N_s} \pi^{(i)} = 1$.

Mixture op : limitations

- We need to **fit** the mixture operator!
- To fit, we **need data**.. not always available.
- If **each source** distribution is also a **parametric model** needing to be fit (ex : mixture of Gaussians), then I can use **EM** to fitting.
- If the fit is **bad** anyway then I may move to another parametric model (not always easy to choose).

Mixture op : limitations

- We need to **fit** the mixture operator!
- To fit, we **need data**.. not always available.
- If **each source** distribution is also a **parametric model** needing to be fit (ex : mixture of Gaussians), then I can use **EM** to fitting.
- If the fit is **bad** anyway then I may move to another parametric model (not always easy to choose).

Mixture op : limitations

- We need to **fit** the mixture operator!
- To fit, we **need data**.. not always available.
- If **each source** distribution is also a **parametric model** needing to be fit (ex : mixture of Gaussians), then I can use **EM** to fitting.
- If the fit is **bad** anyway then I may move to another parametric model (not always easy to choose).

Mixture op : limitations

- We need to **fit** the mixture operator!
- To fit, we **need data**.. not always available.
- If **each source** distribution is also a **parametric model** needing to be fit (ex : mixture of Gaussians), then I can use **EM** to fitting.
- If the fit is **bad** anyway then I may move to another parametric model (not always easy to choose).

Chapter organization

- 1 Probabilistic data fusion
- 2 Probabilistic set-valued data fusion**
- 3 Evidential data fusion
- 4 Appendix

The need for **sets of probabilities** : $\mathcal{X} = \{a, b, c\}$

- **Probabilities** are **not fully expressive** when it comes to represent a certain but **imprecise** piece of information like

$$x \in A = \{a, b\}.$$

- This **event** belongs to the σ -algebra $\sigma_{\mathcal{X}}$ on which probability measures are defined.
- We must have $P_X(A) = 1$ and $P_X(A^c) = 0..$
- But what can I say about $P_X(a)$? One can chose **any** $p \in [0; 1]$ and assign probabilities as $P_X(a) = p$ and $P_X(b) = 1 - p$.

The need for **sets of probabilities** : $\mathcal{X} = \{a, b, c\}$

- **Probabilities** are **not fully expressive** when it comes to represent a certain but **imprecise** piece of information like

$$x \in A = \{a, b\}.$$

- This **event** belongs to the σ -algebra $\sigma_{\mathcal{X}}$ on which probability measures are defined.
- We must have $P_X(A) = 1$ and $P_X(A^c) = 0..$
- But what can I say about $P_X(a)$? One can chose **any** $p \in [0; 1]$ and assign probabilities as $P_X(a) = p$ and $P_X(b) = 1 - p$.

The need for **sets of probabilities** : $\mathcal{X} = \{a, b, c\}$

- **Probabilities** are **not fully expressive** when it comes to represent a certain but **imprecise** piece of information like

$$x \in A = \{a, b\}.$$

- This **event** belongs to the σ -algebra $\sigma_{\mathcal{X}}$ on which probability measures are defined.
- We must have $P_x(A) = 1$ and $P_x(A^c) = 0..$
- But what can I say about $P_x(a)$? One can chose **any** $p \in [0; 1]$ and assign probabilities as $P_x(a) = p$ and $P_x(b) = 1 - p$.

The need for **sets of probabilities** : $\mathcal{X} = \{a, b, c\}$

- **Probabilities** are **not fully expressive** when it comes to represent a certain but **imprecise** piece of information like

$$x \in A = \{a, b\}.$$

- This **event** belongs to the σ -algebra $\sigma_{\mathcal{X}}$ on which probability measures are defined.
- We must have $P_x(A) = 1$ and $P_x(A^c) = 0$.
- But what can I say about $P_x(a)$? One can choose **any** $p \in [0; 1]$ and assign probabilities as $P_x(a) = p$ and $P_x(b) = 1 - p$.

The need for **sets of probabilities** : $\mathcal{X} = \{a, b, c\}$

- **Probabilities** are **not fully expressive** when it comes to represent a certain but **imprecise** piece of information like

$$x \in A = \{a, b\}.$$

- This **event** belongs to the σ -algebra $\sigma_{\mathcal{X}}$ on which probability measures are defined.
- We must have $P_x(A) = 1$ and $P_x(A^c) = 0$.
- But what can I say about $P_x(a)$? One can chose **any** $p \in [0; 1]$ and assign probabilities as $P_x(a) = p$ and $P_x(b) = 1 - p$.

→ There are **infinitely many** candidate distributions, i.e. a **set** of distributions.

The need for **sets of probabilities** : $\mathcal{X} = \{a, b, c\}$

- **1st idea** : circumvent this issue by **conditioning**, *i.e.* replacing the set of possible values \mathcal{X} with A (and update probabilities).
- Conditioning succeeds to deal with this situation but there is still **no actual mean to represent** $\{x \in A\}$ with a probability measure on $\sigma_{\mathcal{X}}$.
- Conditioning **does not apply** to advocacies like
 « $x \in A$ with prob. .8 and $x = c$ with prob. .2 ».
- **2nd idea** : use **another σ -algebra** : $\sigma' = \{\emptyset, A, A^c, \mathcal{X}\} \subset \sigma_{\mathcal{X}}$. The information can then be adequately represented but how can we combine measures defined on different σ -algebras?
- Let's give it a **try**!

The need for **sets of probabilities** : $\mathcal{X} = \{a, b, c\}$

- **1st idea** : circumvent this issue by **conditioning**, *i.e.* replacing the set of possible values \mathcal{X} with A (and update probabilities).
- Conditioning succeeds to deal with this situation but there is still **no actual mean to represent** $\{x \in A\}$ with a probability measure on $\sigma_{\mathcal{X}}$.
- Conditioning **does not apply** to advocacies like
 « $x \in A$ with prob. .8 and $x = c$ with prob. .2 ».
- **2nd idea** : use **another σ -algebra** : $\sigma' = \{\emptyset, A, A^c, \mathcal{X}\} \subset \sigma_{\mathcal{X}}$. The information can then be adequately represented but how can we combine measures defined on different σ -algebras?
- Let's give it a **try**!

The need for **sets of probabilities** : $\mathcal{X} = \{a, b, c\}$

- **1st idea** : circumvent this issue by **conditioning**, *i.e.* replacing the set of possible values \mathcal{X} with A (and update probabilities).
- Conditioning succeeds to deal with this situation but there is still **no actual mean to represent** $\{x \in A\}$ with a probability measure on $\sigma_{\mathcal{X}}$.
- Conditioning **does not apply** to advocacies like
 « $x \in A$ with prob. .8 and $x = c$ with prob. .2 ».
- **2nd idea** : use **another σ -algebra** : $\sigma' = \{\emptyset, A, A^c, \mathcal{X}\} \subset \sigma_{\mathcal{X}}$. The information can then be adequately represented but how can we combine measures defined on different σ -algebras?
- Let's give it a **try**!

The need for **sets of probabilities** : $\mathcal{X} = \{a, b, c\}$

- **1st idea** : circumvent this issue by **conditioning**, *i.e.* replacing the set of possible values \mathcal{X} with A (and update probabilities).
- Conditioning succeeds to deal with this situation but there is still **no actual mean to represent** $\{x \in A\}$ with a probability measure on $\sigma_{\mathcal{X}}$.
- Conditioning **does not apply** to advocacies like
 « $x \in A$ with prob. .8 and $x = c$ with prob. .2 ».
- **2nd idea** : use **another σ -algebra** : $\sigma' = \{\emptyset, A, A^c, \mathcal{X}\} \subset \sigma_{\mathcal{X}}$. The information can then be adequately represented but how can we combine measures defined on different σ -algebras?

- Let's give it a **try**!

The need for **sets of probabilities** : $\mathcal{X} = \{a, b, c\}$

- **1st idea** : circumvent this issue by **conditioning**, *i.e.* replacing the set of possible values \mathcal{X} with A (and update probabilities).
- Conditioning succeeds to deal with this situation but there is still **no actual mean to represent** $\{x \in A\}$ with a probability measure on $\sigma_{\mathcal{X}}$.
- Conditioning **does not apply** to advocacies like
 « $x \in A$ with prob. .8 and $x = c$ with prob. .2 ».
- **2nd idea** : use **another σ -algebra** : $\sigma' = \{\emptyset, A, A^c, \mathcal{X}\} \subset \sigma_{\mathcal{X}}$. The information can then be adequately represented but how can we combine measures defined on different σ -algebras?
- Let's give it a **try**!

Example : Dice throw

Suppose $\mathcal{X} = \{1; 2; 3; 4; 5; 6\}$ is the set of outcomes of a **dice throw**. Three sources provide a (conditional) probability distribution over a probability space $(\Omega, \sigma_\Omega, \mathbb{P})$:

- S_1 provides $P_{x|S_1}$ whose codomain the following measurable space (\mathcal{X}, σ_1) ,
- S_2 provides $P_{x|S_2}$ whose codomain the following measurable space (\mathcal{X}, σ_2) ,
- S_3 provides $P_{x|S_3}$ whose codomain the following measurable space (\mathcal{X}, σ_3) ,

Suppose also that $\sigma_1 = \sigma_2 = \sigma_{\mathcal{X}} = 2^{\mathcal{X}}$ while $\sigma_3 = \{\emptyset, \{2; 4; 6\}, \{1; 3; 5\}, \mathcal{X}\}$. In other words, S_1 and S_2 can express the odds of any outcome and **S_3 can only discriminate odd numbers from even ones.**

Example : Dice throw

In addition, we have

element $i \in \mathcal{X}$	1	2	3	4	5	6
$P_{x S_1}(i)$	0	0.3	0.4	0.2	0.1	0
$P_{x S_2}(i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

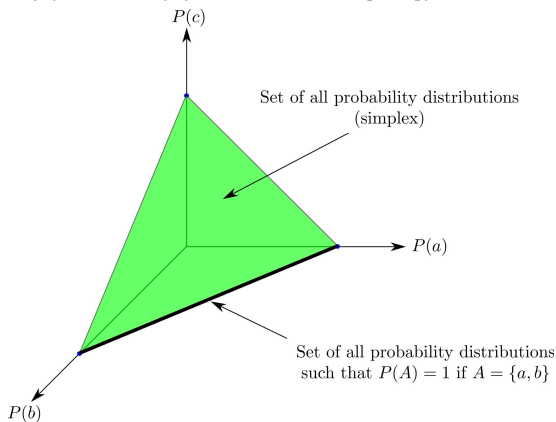
set A	$\{2; 4; 6\}$	$\{1; 3; 5\}$
$P_{x S_3}(A)$	0.8	0.2

Example : 1st solution - Break the problem into pieces!

Example : 2nd solution - Choose a distribution on \mathcal{X} that best represents the 3rd source.

The need for **sets of probabilities** :

- We came to the conclusion that such a source supporting $x \in A$ is adequately represented by the **set of probability measures** such that A has probability 1.
- In our example, this set is $\{P \in \mathcal{P}_{\mathcal{X}} \mid P(a) = p, P(b) = 1 - p, p \in [0; 1]\}$.



The need for **sets of probabilities** :

- We will therefore now consider that advocacies are sets of probability measures, *i.e.* $\mathbb{X} = 2^{\mathcal{P}_{\mathcal{X}}}$, hence the following definition :

Definition

Probabilistic set-valued data fusion is a subclass of data fusion where **advocacies** live in $\mathbb{X} = 2^{\mathcal{P}_{\mathcal{X}}}$ ^a.

a. $2^{\mathcal{P}_{\mathcal{X}}}$ is the power set of probability measures defined on $(\mathcal{X}, \sigma_{\mathcal{X}})$.

Probabilistic set-valued data fusion :

- We focus on the countable finite case $|\mathcal{X}| < \infty$ and $\sigma_{\mathcal{X}} = 2^{\mathcal{X}}$.
- In general, we will denote by \mathcal{P}_i a set of probability measures :
 $\mathcal{P}_i \in 2^{\mathcal{P}_{\mathcal{X}}}$.
- We will call such sets **p.m.-sets** for short.

Probabilistic set-valued data fusion :

- We focus on the countable finite case $|\mathcal{X}| < \infty$ and $\sigma_{\mathcal{X}} = 2^{\mathcal{X}}$.
- In general, we will denote by \mathcal{P}_i a set of probability measures :
 $\mathcal{P}_i \in 2^{\mathcal{P}_{\mathcal{X}}}$.
- We will call such sets **p.m.-sets** for short.

Probabilistic set-valued data fusion :

- We focus on the countable finite case $|\mathcal{X}| < \infty$ and $\sigma_{\mathcal{X}} = 2^{\mathcal{X}}$.
- In general, we will denote by \mathcal{P}_i a set of probability measures :
 $\mathcal{P}_i \in 2^{\mathcal{P}_{\mathcal{X}}}$.
- We will call such sets **p.m.-sets** for short.

Probabilistic set-valued data fusion : 2 alternative representations

- In the **dice throwing example**, a source delivered the following advocacy :

« $x \in A$ with prob. .8 and $x = c$ with prob. .2 ».

- The corresponding p.m.-set is

$$\mathcal{P}_i = \{P \in \mathcal{P}_{\mathcal{X}} \mid P(a) = p, P(b) = .8 - p, p \in [0; .8]\}.$$

- Observe that the same information can be encoded by stating that

$$0 \leq P(a) \leq .8, \quad 0 \leq P(b) \leq .8 \quad \text{and} \quad P(c) = .2.$$

- In this case, the p.m.-set can be equivalently represented by **probability bounds**.

Probabilistic set-valued data fusion : 2 alternative representations

- In the **dice throwing example**, a source delivered the following advocacy :

« $x \in A$ with prob. .8 and $x = c$ with prob. .2 ».

- The corresponding p.m.-set is

$$\mathcal{P}_i = \{P \in \mathcal{P}_{\mathcal{X}} \mid P(a) = p, P(b) = .8 - p, p \in [0; .8]\}.$$

- Observe that the same information can be encoded by stating that

$$0 \leq P(a) \leq .8, \quad 0 \leq P(b) \leq .8 \quad \text{and} \quad P(c) = .2.$$

- In this case, the p.m.-set can be equivalently represented by **probability bounds**.

Probabilistic set-valued data fusion : 2 alternative representations

- In the **dice throwing example**, a source delivered the following advocacy :

« $x \in A$ with prob. .8 and $x = c$ with prob. .2 ».

- The corresponding p.m.-set is

$$\mathcal{P}_i = \{P \in \mathcal{P}_{\mathcal{X}} \mid P(a) = p, P(b) = .8 - p, p \in [0; .8]\}.$$

- Observe that the same information can be encoded by stating that

$$0 \leq P(a) \leq .8, \quad 0 \leq P(b) \leq .8 \quad \text{and} \quad P(c) = .2.$$

- In this case, the p.m.-set can be equivalently represented by **probability bounds**.

Probabilistic set-valued data fusion : 2 alternative representations

- In the **dice throwing example**, a source delivered the following advocacy :

« $x \in A$ with prob. .8 and $x = c$ with prob. .2 ».

- The corresponding p.m.-set is

$$\mathcal{P}_i = \{P \in \mathcal{P}_{\mathcal{X}} \mid P(a) = p, P(b) = .8 - p, p \in [0; .8]\}.$$

- Observe that the same information can be encoded by stating that

$$0 \leq P(a) \leq .8, \quad 0 \leq P(b) \leq .8 \quad \text{and} \quad P(c) = .2.$$

- In this case, the p.m.-set can be equivalently represented by **probability bounds**.

Convenient objects to encode **probability bounds** : Capacities

Definition

Let ν denote a set-function from $2^{\mathcal{X}}$ to \mathbb{R} . ν is said to be a **capacity** if it has the following properties :

- $\nu(\emptyset) = 0$,
- $\nu(\mathcal{X}) = 1$,
- $A \subseteq B \Rightarrow \nu(A) \leq \nu(B)$, for any A, B in $2^{\mathcal{X}}$ (monotony).

Capacities vs Probability measures :

Probability measure :

- $\mu(\emptyset) = 0$,
- $\mu(\mathcal{X}) = 1$,
- $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$,
- $A \cap B = \emptyset \Rightarrow \mu(A \cup B) = \mu(A) + \mu(B)$.

Capacity :

- $\nu(\emptyset) = 0$,
- $\nu(\mathcal{X}) = 1$,
- $A \subseteq B \Rightarrow \nu(A) \leq \nu(B)$.

- Probability measures = additive capacities.
- Capacities = non-additive measures, (bit awkward because the additivity property belongs to the definition of measures).

Capacities : properties

Definition

A probability measure $P : (\mathcal{X}, \sigma_{\mathcal{X}} = 2^{\mathcal{X}}) \rightarrow [0; 1]$ is said to **dominate** a capacity $\nu : 2^{\mathcal{X}} \rightarrow [0; 1]$ if for all $A \in 2^{\mathcal{X}}$, $\nu(A) \leq P(A)$.

Definition

The **core** \mathcal{P}_{ν} of a capacity ν is the set of probability measures dominating ν .

Property

All cores are **convex closed** subsets of $\mathcal{P}_{\mathcal{X}}$.

Capacities : properties

Definition

A capacity $\nu : 2^{\mathcal{X}} \rightarrow [0; 1]$ is said to be **n-monotonic**, with $n \in \mathbb{N}^*$, if and only if for any family of n events $\mathcal{A} = (A_i)_{i=1}^n$, one has :

$$\sum_{I \subseteq \mathcal{A}} (-1)^{|I|} \nu \left(\bigcap_{A \in I} A \right) \leq \nu \left(\bigcup_{1 \leq i \leq n} A_i \right). \quad (5)$$

- The interest of the n -monotonicity is not intuitive at all.
- This property will help us to identify classes of capacities with interesting properties.
- Note that 1-monotonicity is just monotonicity as described in the definition of capacities 7.

Capacities : Mass functions

Definition

Given a capacity $\nu : 2^{\mathcal{X}} \rightarrow [0; 1]$, its **Möbius transform** is the set-function m from the power set $2^{\mathcal{X}}$ to the real line \mathbb{R} defined as follows :

$$m : 2^{\mathcal{X}} \rightarrow \mathbb{R}, \quad (6)$$

$$A \rightarrow \sum_{B \subseteq A} (-1)^{|A \setminus B|} \nu(B). \quad (7)$$

The function m is called **mass function**.

- Due to the boundary conditions in the definition of capacities, any **mass function** m is such that :
 - $\sum_{A \subseteq \mathcal{X}} m(A) = 1,$
 - $m(\emptyset) = 0.$

Capacities : Mass functions

- The Möbius transform can be **reversed**. For any capacity ν and its mass function m , we have :

$$\nu(A) = \sum_{B \subseteq A} m(B), \forall A \subseteq \mathcal{X} \text{ (inverse Möbius)}. \quad (8)$$

- This equation looks very much alike a Cramer system of $2^{|\mathcal{X}|}$ equations which intuitively guarantee that the **bijective correspondence** between capacities and mass functions.

Lemma

Let $\nu : 2^{\mathcal{X}} \rightarrow [0; 1]$ denote an ∞ -monotonic capacity. Then the codomain of its mass function m is $[0; 1] : \forall A \subseteq \mathcal{X}$,

$$0 \leq m(A) \leq 1.$$

Upper and lower probabilities :

- Next : How do we relate **capacities** with **probability bounds**?

Definition

Let \mathcal{P}_i denote a p.m.-set. The **lower envelope** $\underline{\nu}_{\mathcal{P}_i}$ of \mathcal{P}_i is a mapping defined as follows :

$$\begin{aligned} \underline{\nu}_{\mathcal{P}_i} : 2^{\mathcal{X}} &\rightarrow [0; 1], \\ A &\rightarrow \min_{\mu \in \mathcal{P}_i} \{\mu(A)\}. \end{aligned} \quad (9)$$

Upper and lower probabilities :

Definition

Let \mathcal{P}_i denote a p.m.-set. The **upper envelope** $\overline{\nu}_{\mathcal{P}_i}$ of \mathcal{P}_i is a mapping defined as follows :

$$\begin{aligned} \overline{\nu}_{\mathcal{P}_i} : 2^{\mathcal{X}} &\rightarrow [0; 1], \\ A &\rightarrow \max_{\mu \in \mathcal{P}_i} \{\mu(A)\}. \end{aligned} \quad (10)$$

- This means that for any $A \in 2^{\mathcal{X}}$ and any $\mu \in \mathcal{P}_i$, we have :

$$\underline{\nu}_{\mathcal{P}_i}(A) \leq \mu(A) \leq \overline{\nu}_{\mathcal{P}_i}(A). \quad (11)$$

Upper and lower probabilities :

- In this context, lower envelopes are often called **lower probabilities** and upper envelopes are called **upper probabilities**.

Theorem

Let \mathcal{P}_i denote a p.m.-set with lower envelope $\underline{\nu}_{\mathcal{P}_i}$. Then the upper envelope $\overline{\nu}_{\mathcal{P}_i}$ of \mathcal{P}_i is the **conjugate** of its lower envelope :

$$\overline{\nu}_{\mathcal{P}_i} = (\underline{\nu}_{\mathcal{P}_i})^c. \quad (12)$$

Upper and lower probabilities : Example with $\mathcal{X} = \{a, b, c\}$

Source i delivers the p.m.-set $\mathcal{P}_i = \{\mu_1, \mu_2\}$.

set $A \in 2^{\mathcal{X}}$	$\mu_1(A)$	$\mu_2(A)$	$\underline{\nu}_{\mathcal{P}_i}(A)$	$\overline{\nu}_{\mathcal{P}_i}(A)$	$\overline{\nu}_{\mathcal{P}_i}(A^c)$	$(\overline{\nu}_{\mathcal{P}_i})^c(A) = 1 - \overline{\nu}_{\mathcal{P}_i}(A^c)$
\emptyset	0	0	0	0	1	0
$\{a\}$	1/3	1/2	1/3	1/2	2/3	1/3
$\{b\}$	1/3	1/2	1/3	1/2	2/3	1/3
$\{a, b\}$	2/3	1	2/3	1	1/3	2/3
$\{c\}$	1/3	0	0	1/3	1	0
$\{a, c\}$	2/3	1/2	1/2	2/3	1/2	1/2
$\{b, c\}$	2/3	1/2	1/2	2/3	1/2	1/2
$\{a, b, c\} = \mathcal{X}$	1	1	1	1	0	1

The 4th = lower envelope of \mathcal{P}_i .

The 5th = upper envelope of \mathcal{P}_i .

Last column = conjugate of the upper envelope = lower envelope.

Upper and lower probabilities :

- The preceding theorem is very interesting in the sense that it is **unnecessary** to study both the lower and upper envelopes since they are in bijective correspondence.
- In general, a p.m.-set \mathcal{P}_i is **included** in the core $\mathcal{P}_{\underline{\nu}_{\mathcal{P}_i}}$ induced by its lower envelope :

$$\mathcal{P}_i \subseteq \mathcal{P}_{\underline{\nu}_{\mathcal{P}_i}}. \quad (13)$$

- However, they are not always equal.

Upper and lower probabilities :

- The preceding theorem is very interesting in the sense that it is **unnecessary** to study both the lower and upper envelopes since they are in bijective correspondence.
- In general, a p.m.-set \mathcal{P}_i is **included** in the core $\mathcal{P}_{\underline{\nu}_{\mathcal{P}_i}}$ induced by its lower envelope :

$$\mathcal{P}_i \subseteq \mathcal{P}_{\underline{\nu}_{\mathcal{P}_i}}. \quad (13)$$

- However, they are not always equal.

Upper and lower probabilities :

- The preceding theorem is very interesting in the sense that it is **unnecessary** to study both the lower and upper envelopes since they are in bijective correspondence.
- In general, a p.m.-set \mathcal{P}_i is **included** in the core $\mathcal{P}_{\underline{\nu}_{\mathcal{P}_i}}$ induced by its lower envelope :

$$\mathcal{P}_i \subseteq \mathcal{P}_{\underline{\nu}_{\mathcal{P}_i}}. \quad (13)$$

- However, they are not always equal.

Upper and lower probabilities : Example with $\mathcal{X} = \{a, b, c\}$

Source i delivers the p.m.-set $\mathcal{P}_i = \{\mu_1, \mu_2\}$.

set $A \in 2^{\mathcal{X}}$	$\mu_1(A)$	$\mu_2(A)$	$\underline{\nu}_{\mathcal{P}_i}(A)$	$\mu_3(A)$
\emptyset	0	0	0	0
$\{a\}$	1/3	1/2	1/3	5/12
$\{b\}$	1/3	1/2	1/3	5/12
$\{a, b\}$	2/3	1	2/3	5/6
$\{c\}$	1/3	0	0	1/6
$\{a, c\}$	2/3	1/2	1/2	7/12
$\{b, c\}$	2/3	1/2	1/2	7/12
$\{a, b, c\} = \mathcal{X}$	1	1	1	1

4th column = lower envelope of \mathcal{P}_i .

Last column = values of the measure $\mu_3 = \frac{\mu_1 + \mu_2}{2}$.

We have $\mu_3 \notin \mathcal{P}_i$ while $\mu_3 \in \underline{\mathcal{P}}_{\mathcal{P}_i}$ because μ_3 dominates $\underline{\nu}_{\mathcal{P}_i}$.

Fusion op for p.m.-sets :

- Since the source advocacies are p.m.-sets, they are also just **sets** and consequently, all **fusion operators** introduced in **chapter 3** apply.

Definition

Each advocacy is a p.m.-set : $x_*^{(i)} = \mathcal{P}_i$. The **conjunctive operator** \hat{x}_\cap is defined as follows :

$$\hat{x}_\cap (\mathcal{P}_1, \dots, \mathcal{P}_{N_s}) \rightarrow \bigcap_{i=1}^{N_s} \mathcal{P}_i. \quad (14)$$

Conjunctive op for p.m.-sets :

Property

The conjunctive operator \hat{x}_{\cap} is conjunctive.

Property

If \mathcal{P}_1 and \mathcal{P}_2 are two **cores** with respective lower envelopes ν_1 and ν_2 , then the p.m.-set $\mathcal{P} = \hat{x}_{\cap}(\mathcal{P}_1, \mathcal{P}_2)$ is also a core whose lower probabilities are

$$\nu = \text{max} \{ \nu_1; \nu_2 \} \quad (\text{entrywise max}).$$

The upper probabilities are obtained using an entrywise min.

Chapter organization

- 1 Probabilistic data fusion
- 2 Probabilistic set-valued data fusion
- 3 Evidential data fusion**
- 4 Appendix

Theory of belief functions :

- Belief functions are ∞ -monotonic lower probabilities of a p.m.-set.
- However, they were introduced by plugging a probability measure into a multi-valued mapping.

Definition

Evidential data fusion is a subclass of data fusion where advocacies live in the set of belief functions $\mathbb{X} = \mathcal{B}_{\mathcal{X}}$.

- All p.m.-sets encoded by belief functions (seen as a lower envelope) is a core.
- If a source is reliable, this core should contain the (true) probability distribution of x .

Belief functions : basics

- Recalling lemma 1, **mass functions** obtained via Möbius transform from a belief function are **positive**.
- Mass functions** are then reminiscent of probability distribution except that masses are distributed on $2^{\mathcal{X}}$ instead of \mathcal{X} .
- In the theory of **belief functions**, a **focal element** of a mass function m_i is a set $A \subseteq \mathcal{X}$ such that $m_i(A) > 0$ meaning that the i^{th} collected **evidence** supports the event $\{\theta \in A\}$.
- This evidence viewpoint of **belief functions** accounts for the fact that the theory of belief functions is also frequently called the **evidence theory**. Approaches developed within this framework are thus called **evidential**.

Belief functions : basics

- Recalling lemma 1, **mass functions** obtained via Möbius transform from a belief function are **positive**.
- Mass functions** are then reminiscent of probability distribution except that masses are distributed on $2^{\mathcal{X}}$ instead of \mathcal{X} .
- In the theory of **belief functions**, a **focal element** of a mass function m_i is a set $A \subseteq \mathcal{X}$ such that $m_i(A) > 0$ meaning that the i^{th} collected **evidence** supports the event $\{\theta \in A\}$.
- This evidence viewpoint of **belief functions** accounts for the fact that the theory of belief functions is also frequently called the **evidence theory**. Approaches developed within this framework are thus called **evidential**.

Belief functions : basics

- Recalling lemma 1, **mass functions** obtained via Möbius transform from a belief function are **positive**.
- Mass functions** are then reminiscent of probability distribution except that masses are distributed on $2^{\mathcal{X}}$ instead of \mathcal{X} .
- In the theory of **belief functions**, a **focal element** of a mass function m_i is a set $A \subseteq \mathcal{X}$ such that $m_i(A) > 0$ meaning that the i^{th} collected **evidence** supports the event $\{\theta \in A\}$.
- This evidence viewpoint of **belief functions** accounts for the fact that the theory of belief functions is also frequently called the **evidence theory**. Approaches developed within this framework are thus called **evidential**.

Belief functions : basics

- Recalling lemma 1, **mass functions** obtained via Möbius transform from a belief function are **positive**.
- Mass functions** are then reminiscent of probability distribution except that masses are distributed on $2^{\mathcal{X}}$ instead of \mathcal{X} .
- In the theory of **belief functions**, a **focal element** of a mass function m_i is a set $A \subseteq \mathcal{X}$ such that $m_i(A) > 0$ meaning that the i^{th} collected **evidence** supports the event $\{\theta \in A\}$.
- This evidence viewpoint of **belief functions** accounts for the fact that the theory of belief functions is also frequently called the **evidence theory**. Approaches developed within this framework are thus called **evidential**.

Evidential data fusion :

Definition

Each advocacy is a mass function : $x_*^{(i)} = m_i$.

The **Dempster's rule operator** \hat{x}_\otimes is defined as follows :

$$\hat{x}_\otimes : (m_1, \dots, m_{N_s}) \rightarrow m_*, \quad (15)$$

with m_* a mass function such that for any $A \in 2^{\mathcal{X}}$, one has :

$$m_*(A) = \begin{cases} \frac{1}{1-\kappa} \sum_{A_1, \dots, A_{N_s} \in 2^{\mathcal{X}}} \prod_{i=1}^{N_s} m_i(A_i) & \text{if } \kappa < 1 \\ \text{s.t. } \bigcap_{i=1}^{N_s} A_i = A & \\ \emptyset & \text{otherwise} \end{cases} . \quad (16)$$

Evidential data fusion :

Definition

The parameter κ is called **Dempster's degree of conflict**. It is defined as follows :

$$\kappa = \sum_{\substack{A_1, \dots, A_{N_s} \in 2^{\mathcal{X}} \\ \text{s.t. } \bigcap_{i=1}^{N_s} A_i = \emptyset}} \prod_{i=1}^{N_s} m_i(A_i). \quad (17)$$

- This parameter is the total mass assigned to **incompatible** pieces of evidence.

Dempster's rule : properties

Property

- Dempster's rule is **commutative** : $\hat{x}_{\otimes}(m_1, m_2) = \hat{x}_{\otimes}(m_2, m_1)$, for any mass functions m_1 and m_2 .
- Dempster's rule is **associative** :
 $\hat{x}_{\otimes}(m_1, \hat{x}_{\otimes}(m_2, m_3)) = \hat{x}_{\otimes}(\hat{x}_{\otimes}(m_1, m_2), m_3)$, for any mass functions m_1 , m_2 and m_3 .
- Unique **neutral element** : If $m(\mathcal{X}) = 1$ then $\hat{x}_{\otimes}(m_1, m) = m_1$, for any mass function m_1 .
- Note that, however, Dempster rule is in general **not idempotent** :
 $\hat{x}_{\otimes}(m, m) \neq m$. This latter point is sometimes criticized because when two sources are saying the exact same thing, it may be desirable that the aggregate is equal to their proposal.

Dempster's rule : properties

- Since Dempster's rule is associative and commutative, it is **unnecessary** to compute it in batch mode.
- It is far less time consuming to do **pairwise combinations**. In addition, when $N_s = 2$, Dempster's rule is a bit more easy to grasp : let $m_{1 \otimes 2} = \hat{x}_{\otimes} (m_1, m_2)$. For any $A \in 2^{\mathcal{X}}$, we have :

$$m_{1 \otimes 2} (A) = \begin{cases} \frac{1}{1-\kappa} \sum_{\substack{A_1, A_2 \in 2^{\mathcal{X}} \\ \text{s.t. } A_1 \cap A_2 = A}} m_1 (A_1) m_2 (A_2) & \text{if } \kappa < 1 \\ \emptyset & \text{otherwise} \end{cases} \quad (18)$$

$$\text{with } \kappa = \sum_{\substack{A_1, A_2 \in 2^{\mathcal{X}} \\ \text{s.t. } A_1 \cap A_2 = \emptyset}} m_1 (A_1) m_2 (A_2).$$

Dempster's rule : properties

Property

Curse of conflict :

Let $(m_i)_{i=1}^N$ be a sequence of mass functions. Let κ_n denote Dempster's degree of conflict computed from the n 1st members of the sequence.

The sequence κ_n is **non-decreasing**.

Dempster's rule : justifications

- Suppose $m_2(B) = 1$, the result of the conjunctive combination of m_1 and m_2 is denoted by $m_{1|B}$. For any $A \in 2^{\mathcal{X}}$, one has :

$$m_{1|B}(A) = \frac{\kappa}{1 - \kappa} \sum_{\substack{C \in 2^{\mathcal{X}} \\ \text{s.t. } B \cap C = A}} m_1(C) \text{ (ev. conditioning)}. \quad (19)$$

- 1st justification** : Evidential conditioning is **compliant** with Bayesian conditioning in the sense that if m_1 is a Bayesian mass function with $\mathcal{P}_{bel_1} = \{P_1\}$, then the only element in the p.m.-set induced by $bel_{1|B}$ is the conditional probability measure $P_{1|x \in B}$.
- 2nd justification** : Dempster's rule just relies on the (usual) statistical independence assumption among probability measures of the sources.

Dempster's rule : justifications

- Suppose $m_2(B) = 1$, the result of the conjunctive combination of m_1 and m_2 is denoted by $m_{1|B}$. For any $A \in 2^{\mathcal{X}}$, one has :

$$m_{1|B}(A) = \frac{\kappa}{1 - \kappa} \sum_{\substack{C \in 2^{\mathcal{X}} \\ \text{s.t. } B \cap C = A}} m_1(C) \text{ (ev. conditioning)}. \quad (19)$$

- 1st justification** : Evidential conditioning is **compliant** with Bayesian conditioning in the sense that if m_1 is a Bayesian mass function with $\mathcal{P}_{bel_1} = \{P_1\}$, then the only element in the p.m.-set induced by $bel_{1|B}$ is the conditional probability measure $P_{1|x \in B}$.
- 2nd justification** : Dempster's rule just relies on the (usual) statistical independence assumption among probability measures of the sources.

Dempster's rule : justifications

- Suppose $m_2(B) = 1$, the result of the conjunctive combination of m_1 and m_2 is denoted by $m_{1|B}$. For any $A \in 2^{\mathcal{X}}$, one has :

$$m_{1|B}(A) = \frac{\kappa}{1 - \kappa} \sum_{\substack{C \in 2^{\mathcal{X}} \\ \text{s.t. } B \cap C = A}} m_1(C) \text{ (ev. conditioning)}. \quad (19)$$

- 1st justification** : Evidential conditioning is **compliant** with Bayesian conditioning in the sense that if m_1 is a Bayesian mass function with $\mathcal{P}_{bel_1} = \{P_1\}$, then the only element in the p.m.-set induced by $bel_{1|B}$ is the conditional probability measure $P_{1|x \in B}$.
- 2nd justification** : Dempster's rule just relies on the (usual) statistical independence assumption among probability measures of the sources.

Decision making with belief functions :

- If the aggregate x_* is a mass function, usual decision making do not apply.
- Solutions :
 - In practice, the restriction of upper probabilities on singletons² work well.
 - I can look for the probability distribution that best represents the p.m.-set (center of gravity - pignistic transform).
 - I can compute lower and upper expectations and end up with an interval-valued solution.

2. A singleton is a set with unit cardinality.

Conjunctive op and Dempster rule in action :

Example : the trial

3 suspects for a murder case : $\mathcal{X} = \{peter; paul; mary\}$ but only 1 is guilty.
10 witnesses :

- 1st source : 8 are supporting the fact that the culprit is a man. 1 supports the opposite. 1 is undecided.
- 2nd source : 5 are supporting the fact that the culprit is dark haired. 1 is supporting the fact that the culprit is red haired. 4 are undecided. Peter and Mary are dark haired while Paul is red haired.

What can be inferred about the murderer identity ?

Chapter organization

- 1 Probabilistic data fusion
- 2 Probabilistic set-valued data fusion
- 3 Evidential data fusion
- 4 Appendix**

Capacities : properties

- Capacities can have many properties, some of which are given in the sequel.

Definition

A capacity ν is called **super-additive** if for all sets A and B in $2^{\mathcal{X}}$, one has :

$$\nu(A) + \nu(B) \leq \nu(A \cup B). \quad (20)$$

A capacity ν is called **sub-additive** if for all sets A and B in $2^{\mathcal{X}}$, one has :

$$\nu(A) + \nu(B) \geq \nu(A \cup B). \quad (21)$$

A capacity that is both super and sub-additive is a probability measure.

Capacities : properties

- It is noteworthy that in particular :
 - if ν is super-additive, $\nu(A) + \nu(A^c) \leq 1$,
 - if ν is sub-additive, $\nu(A) + \nu(A^c) \geq 1$,

Definition

The **conjugate capacity** ν^c of a capacity ν is such that for all $A \in 2^{\mathcal{X}}$, $\nu^c(A) = 1 - \nu(A^c)$.

Capacities : properties

Property

Any n -monotonic capacity is also $(n - 1)$ -monotonic.

- The above property obviously implies that the set of ∞ -monotonic capacities is included in the set of n -monotonic capacities which in turn is included in the set of 1-monotonic capacities.

Lemma

For any 2-monotonic capacity ν , we have $\mathcal{P}_\nu \neq \emptyset$.

- Since the above lemma holds for 2-monotonic capacities, it is also true for n -monotonic capacities ($n \geq 2$).
- This property is quite handy in some circumstances.

Upper and lower probabilities :

- Enveloppes are characterizing p.m.-sets but we would like preferably that they **entirely encode** the same information as p.m.-sets.
- This is tantamount to have $\mathcal{P}_i = \mathcal{P}_{\underline{\nu}_{\mathcal{P}_i}}$ because $\underline{\nu}_{\mathcal{P}_i}$ uniquely defines $\mathcal{P}_{\underline{\nu}_{\mathcal{P}_i}}$.

Definition

A p.m.-set \mathcal{P}_i with lower envelope $\underline{\nu}_{\mathcal{P}_i}$ is said to be **coherent** if $\mathcal{P}_i = \mathcal{P}_{\underline{\nu}_{\mathcal{P}_i}}$.

Theorem

If the lower envelope $\underline{\nu}_{\mathcal{P}_i}$ of a p.m.-set \mathcal{P}_i is 2-monotonic, then this p.m.-set is coherent.

- From the above theorem, we know that **2-monotonic super-additive capacities** entirely characterize a **closed convex p.m.-set**.

Upper and lower probabilities :

Theorem

Any lower (resp. upper) envelope of p.m.-set is a super-additive (resp. sub-additive) capacity.

- In this context, super-additive capacities are often called **lower probabilities** and sub-additive capacities are called **upper probabilities**.

Theorem

Let \mathcal{P}_i denote a p.m.-set with lower envelope $\underline{\nu}_{\mathcal{P}_i}$. Then the upper envelope $\overline{\nu}_{\mathcal{P}_i}$ of \mathcal{P}_i is the conjugate of its lower envelope :

$$\overline{\nu}_{\mathcal{P}_i} = (\underline{\nu}_{\mathcal{P}_i})^c. \quad (22)$$

Belief functions : basics

- There are several noteworthy **sub-classes** of mass functions.
 - A mass function having only one focal element A is called a **categorical mass function** and it is denoted by m_A . The categorical mass function $m_{\mathcal{X}}$ is called the **vacuous mass function** because it carries no information.
 - A **simple mass function** m_A^w is the convex combination of $m_{\mathcal{X}}$ with a categorical mass function m_A with $A \neq \Omega$: $m_A^w = (1 - w) m_A + w m_{\mathcal{X}}$ with $w \in [0; 1]$.
 - A mass function whose focal elements have unit cardinality is a **bayesian mass function**. Such a mass function is formally equivalent to a probability distribution. The underlying p.m.-set has thus also unit cardinality. This also shows that the theory of belief functions encompasses the probability theory.
 - A **consonant mass function** is such that for any pair of focal elements (A, B) , one has either $A \subseteq B$ or $B \subseteq A$. The inclusion is thus a total order relation for focal elements of consonant mass functions.

Belief functions : basics

- Several alternatives for evidence representation are commonly used :
 - the **belief** bel_i which is the inverse Möbius transform of the mass function m_i .
 - the **commonality** function q_i which is the inverse co-Möbius transforms of the mass function m_i . We have that :

$$q_i(A) = \sum_{B \supseteq A} m_i(B), \forall A \in 2^{\mathcal{X}}. \quad (23)$$

- the **plausibility** function pl_i is the conjugate of bel_i : $pl_i = (bel_i)^c$. The plausibility function is consequently viewed as an upper probability. For any $A \in 2^{\mathcal{X}}$ and any $\mu \in \mathcal{P}_{bel_i}$, we have :

$$bel_i(A) \leq \mu(A) \leq pl_i(A). \quad (24)$$

In addition, we also have :

$$pl_i(A) = \sum_{\substack{B \subseteq \Omega \\ B \cap A \neq \emptyset}} m_i(B), \forall A \in 2^{\mathcal{X}}. \quad (25)$$

Belief functions : basics

- Several alternatives for evidence representation are commonly used :
 - the **implicability** function b_i is such that $\forall A \subseteq \mathcal{X}$,
 $b_i(A) = bel_i(A) + m_i(\emptyset)$.
- There is a **one-to-one correspondence** between a mass function m_i and any of these four functions.
- If the truthfulness of the evidence encoded in a mass function can be evaluated through a coefficient $\alpha \in [0, 1]$, then a so-called **discounting** operation on m can be performed. A discounted mass function is denoted by m^α and we have :

$$m^\alpha = (1 - \alpha)m + \alpha m_{\mathcal{X}}. \quad (26)$$

- α is called the **discounting rate**. Since $m_{\mathcal{X}}$ represents a state of ignorance, setting $\alpha = 1$ discards it from further processing. Note that the notation is coherent with that of simple mass functions. Indeed, a simple mass function m_A^w is the categorical mass function m_A discounting with rate w .

Belief functions : basics

- Another useful concept is the **complement** \overline{m}_i of a mass function m_i . The function \overline{m}_i is such that $\forall A \subseteq \mathcal{X}, \overline{m}_i(A) = m_i(A^c)$.
- Important remark : many authors relax the constraint $m_i(\emptyset) = 0$. A mass function such that $m_i(\emptyset) > 0$ is said to be **unnormalized**.
- In this case, the set of belief functions is not a subset of capacities anymore because the boundary condition $bel_i(\mathcal{X}) = 1$ is not no longer valid.
- Nonetheless, this is not much a problem because one can retrieve a regular mass function by deleting the mass allocated to the empty set and **re-normalize** the function so that it sums to one.
- The mass assigned to the empty set is often regarded as an indicator that the solution of the problem does not belong to \mathcal{X} , which implies that the problem is **ill-posed**.

Evidential data fusion :

- The **unnormalized** version of Dempster's rule is known as the **conjunctive rule** (same rule without normalizing constant $\frac{1}{1-\kappa}$).

Definition

Let us consider a data fusion problem as defined in 15. Each advocacy is a mass function : $x_*^{(i)} = m_i$. The **conjunctive rule operator** \hat{x}_{\odot} is defined as follows :

$$\hat{x}_{\odot} : (m_1, \dots, m_{N_s}) \rightarrow m_*, \quad (27)$$

with m_* a mass function such that for any $A \in 2^{\mathcal{X}}$, one has :

$$m_*(A) = \sum_{\substack{A_1, \dots, A_{N_s} \in 2^{\mathcal{X}} \\ \text{s.t. } \bigcap_{i=1}^{N_s} A_i = A}} \prod_{i=1}^{N_s} m_i(A_i). \quad (28)$$

Evidential data fusion :

- The conjunctive rule has the same properties as Dempster's rule, plus one more :

Property

m_\emptyset is the unique **absorbing element** of $\hat{\hat{\cap}}$: $\hat{\hat{\cap}} (m, m_\emptyset) = m_\emptyset$, for any mass function m .

In addition, the conjunctive rule has the **conflict curse** : for any mass functions m_1 and m_2 , their conjunctive combination $m_{1 \cap 2} = \hat{\hat{\cap}} (m_1, m_2)$ is such that :

$$\max \{m_1(\emptyset); m_2(\emptyset)\} \leq m_{1 \cap 2}(\emptyset). \quad (29)$$

- This means that **conflict** can only **grow** as N_s increases.

Evidential data fusion :

- A mass function $m_{1 \cap 2}$ obtained by **pairwise combination** of two mass functions m_1 and m_2 is such that for any $A \in 2^{\mathcal{X}}$, one has :

$$m_{1 \cap 2}(A) = \sum_{\substack{A_1, A_2 \in 2^{\mathcal{X}} \\ \text{s.t. } A_1 \cap A_2 = A}} m_1(A_1) m_2(A_2). \quad (30)$$

Property

Let q_1 and q_2 denote two commonality functions computed from two mass functions m_1 and m_2 using equation (23). Let $q_{1 \cap 2}$ denote the commonality function obtained using the same equation on the mass function $m_{1 \cap 2} = \hat{x}_{\odot}(m_1, m_2)$. For any $A \in 2^{\mathcal{X}}$, we have :

$$q_{1 \cap 2}(A) = q_1(A) q_2(A). \quad (31)$$

- This last property shows that conjunctive combination is very **easy** to perform in the commonality space.

Fusion with belief functions :

- Unsurprisingly, a **disjunctive rule** also exists in the theory of belief functions. It is often seen as the *dual* rule of the conjunctive one.

Definition

Let us consider a data fusion problem as defined in 15. Each advocacy is a mass function : $x_*^{(i)} = m_i$. The **disjunctive rule operator** \hat{x}_{\cup} is defined as follows :

$$\hat{x}_{\cup} : \left(x_*^{(1)}, \dots, x_*^{(N_s)} \right) \rightarrow x_*, \quad (32)$$

$$(m_1, \dots, m_{N_s}) \rightarrow m_*, \quad (33)$$

with m_* a mass function such that for any $A \in 2^{\mathcal{X}}$, one has :

$$m_*(A) = \sum_{\substack{A_1, \dots, A_{N_s} \in 2^{\mathcal{X}} \\ \text{s.t. } \bigcup_{i=1}^{N_s} A_i = A}} \prod_{i=1}^{N_s} m_i(A_i). \quad (34)$$

Fusion with belief functions :

- The disjunctive rule can process standard or unnormalized mass functions.

Property

We have the following properties for the disjunctive rule :

- The disjunctive rule is commutative : $\hat{x}_{\cup} (m_1, m_2) = \hat{x}_{\cup} (m_2, m_1)$, for any mass functions m_1 and m_2 .
- The disjunctive rule is associative :
 $\hat{x}_{\cup} (m_1, \hat{x}_{\cup} (m_2, m_3)) = \hat{x}_{\cup} (\hat{x}_{\cup} (m_1, m_2), m_3)$, for any mass functions m_1 , m_2 and m_3 .
- m_{\emptyset} is the unique neutral element of the disjunctive : $\hat{x}_{\cup} (m_1, m_{\emptyset}) = m_1$, for any mass function m_1 .

Fusion with belief functions :

Property

We have the following properties for the disjunctive rule :

- $m_{\mathcal{X}}$ is the unique absorbing element of the disjunctive :

$$\hat{x}_{\cup} (m_1, m_{\mathcal{X}}) = m_{\mathcal{X}}, \text{ for any mass function } m_1.$$

In addition, the disjunctive rule has the **ignorance curse** : for any mass functions m_1 and m_2 , their disjunctive combination $m_{1 \cup 2} = \hat{x}_{\cup} (m_1, m_2)$ is such that :

$$\max \{m_1(\mathcal{X}); m_2(\mathcal{X})\} \leq m_{1 \cup 2}(\mathcal{X}). \quad (35)$$

- This means that **ignorance** can only **grow** as N_s increases.

Fusion with belief functions :

Property

Let b_1 and b_2 denote two implicability functions computed from two mass functions m_1 and m_2 . Let $b_{1 \cap 2}$ denote the commonality function obtained from the mass function $m_{1 \cup 2} = \hat{x}_{\odot} (m_1, m_2)$. For any $A \in 2^{\mathcal{X}}$, we have :

$$b_{1 \cup 2} (A) = b_1 (A) b_2 (A). \quad (36)$$

- This last property shows that disjunctive combination is very **easy** to perform in the implicability space.

Fusion with belief functions :

- The underlying **duality** between the conjunctive and disjunctive rules has its roots in the following property :

Property

De Morgan laws property

Let \odot and \oslash denote the binary operations corresponding to the conjunctive and disjunctive rules respectively. For any mass function m_1 and m_2 , we have :

$$\overline{m_1 \odot m_2} = \overline{m_1} \oslash \overline{m_2}, \quad (37)$$

$$\overline{m_1 \oslash m_2} = \overline{m_1} \odot \overline{m_2}. \quad (38)$$

Fusion with belief functions :

Property

The conjunctive rule is **distributive** over the disjunctive rule. Let \odot and \oplus denote the binary operations corresponding to the conjunctive and disjunctive rules respectively. For any mass function m_1 , m_2 and m_3 , we have :

$$m_1 \odot (m_2 \oplus m_3) = (m_1 \odot m_2) \oplus (m_1 \odot m_3). \quad (39)$$

Property

The conjunctive operator \hat{x}_{\odot} and Dempster's operator \hat{x}_{\otimes} are **conjunctive** fusion operators. The disjunctive operator \hat{x}_{\oplus} is a **disjunctive** fusion operator.

Decision making with belief functions :

- Most decision making approaches are fed with a probability measure, so the lead idea in evidence theory is to compute a **probability measure** μ_* that is the best representative of the p.m.-set \mathcal{P}_{bel_*} .
- The notion of *best representative* is of course subject to debate and is usually justified with respect to a given **criterion**.
- Smets [3] introduced the **pignistic transform** which maps the set of belief functions $\mathcal{B}_{\mathcal{X}}$ to the set of probability measures $\mathcal{P}_{\mathcal{X}}$.

Decision making with belief functions :

Definition

Given a belief function $bel : 2^{\mathcal{X}} \rightarrow [0; 1]$ with associated mass function m_i , its **Pignistic transform** is the probability measure $betp$ from the power set $2^{\mathcal{X}}$ to $[0; 1]$ defined as follows :

$$betp : 2^{\mathcal{X}} \rightarrow [0; 1],$$

$$A \rightarrow \sum_{x_0 \in A} \sum_{B \subseteq \mathcal{X} \mid x_0 \in B} \frac{1}{|B|} \frac{m_i(B)}{1 - m_i(\emptyset)}. \quad (40)$$

The probability measure $betp$ is called the **pignistic measure** of bel and it is the **barycenter** of \mathcal{P}_{bel} .

- Since $betp$ is a probability measure on a finite space, it is sufficient to compute its probability **distribution** solely.

Decision making with belief functions :

- Since $\mathcal{P}_{\mathcal{X}} \subsetneq \mathcal{B}_{\mathcal{X}}$, the pignistic transform is **lossy**.
- There are many belief functions sharing the same pignistic measure.

Property

For any belief function bel with associated pignistic measure $betp$, it holds that $betp \in \mathcal{P}_{bel}$, i.e. for all $A \in 2^{\mathcal{X}}$,

$$bel(A) \leq betp(A) \leq pl(A). \quad (41)$$

Decision making with belief functions : example

Suppose the solution set $\mathcal{X} = \{a, b, c\}$. The advocacy of the i^{th} source is a belief function bel_i .

set $A \in 2^{\mathcal{X}}$	$bel_i(A)$	$pl_i(A)$	$m_i(A)$	$betp_i(A)$
\emptyset	0	0	0	0
$\{a\}$	$1/3$	$1/2$	$1/3$	$5/12$
$\{b\}$	$1/3$	$1/2$	$1/3$	$5/12$
$\{a, b\}$	$2/3$	1	0	$5/6$
$\{c\}$	0	$1/3$	0	$1/6$
$\{a, c\}$	$1/2$	$2/3$	$1/6$	$7/12$
$\{b, c\}$	$1/2$	$2/3$	$1/6$	$7/12$
$\{a, b, c\} = \mathcal{X}$	1	1	0	1

The 4th column contains the pignistic measure $betp_i$.

Under equal decision costs, the maximum *a posteriori* solution is either $x = a$ or $x = b$.

Decision making with belief functions :

- There are also decision processes that can be designed from **plausibility** functions.
- More generally, it is also possible to compute **lower and upper expectation** for p.m.-sets which can lead to robust decisions too.

Definition

Let \mathcal{P}_i denote a p.m.-set. The **lower expectation** $\underline{E}_{\mathcal{P}_i}$ of \mathcal{P}_i is defined as follows :

$$\underline{E}_{\mathcal{P}_i} [f] = \min_{\mu \in \mathcal{P}_i} \{E_{\mu} [f]\}, \quad (42)$$

with f a measurable mapping.

Decision making with belief functions :

Definition

Let \mathcal{P}_i denote a p.m.-set. The **upper expectation** $\overline{E}_{\mathcal{P}_i}$ of \mathcal{P}_i is defined as follows :

$$\overline{E}_{\mathcal{P}_i}[f] = \max_{\mu \in \mathcal{P}_i} \{E_{\mu}[f]\}, \quad (43)$$

with f a measurable mapping.

Decision making with belief functions :

- In particular, when the p.m.-set \mathcal{P}_i is governed by a belief function bel , or equivalently by its conjugate plausibility function pl , it holds that :

$$\overline{E}_{\mathcal{P}_i} [1_{\{x_k\}}] = pl(\{x_k\}), \quad (44)$$

for any $x_k \in \mathcal{X}$.

- This is all the more interesting as making decision with the upper expectation requires only to compute plausibility values on \mathcal{X} instead of $2^{\mathcal{X}}$.

Fusion with belief functions :

- There are of course many **other combination rules** in the literature (convex combination rules, idempotent rules, *etc.*).
- Note that the rules introduced in this section have nice **justifications** but these justifications are not expressed in terms of operations on the underlying **p.m.-sets** induced by the belief functions involved in the fusion. For example, the operator \hat{x}_{\odot} is not equivalent to the operator \hat{x}_{\cap} for p.m.-sets.

Belief functions and random sets

- It is possible to relate the belief function theory with random set theory.
- Let us first give a definition of a finite random set.

Definition

Let $(\Omega, \sigma_\Omega, \mu)$ denote a probability space. Let $2^{\mathcal{X}}$ denote the power set of a space \mathcal{X} . A **random set** Γ is a multi-valued mapping from Ω to $2^{\mathcal{X}}$ such that for any $B \in 2^{\mathcal{X}}$, one has

$$\{\omega : \Gamma(\omega) \cap B \neq \emptyset\} \in \sigma_\Omega. \quad (45)$$

The above property is called **strong measurability**.

Belief functions and random sets

- The preceding definition implies that random sets (r.s.) are a generalization of random variables.
- A random set is a random variable if $|\Gamma(\omega)| = 1$ for $\omega \in \Omega$.
- For any mass function m defined within the framework of the theory of belief functions, it is always possible to suppose the **existence** of a given probability space such that there exists a random set Γ mapping this probability space with $2^{\mathcal{X}}$.
- Conversely, for any random set Γ , the function $m = \mu \circ \Gamma^{-1}$ is a mass function as described in the theory of belief functions.
- Suppose a r.v. $\phi : \Omega \rightarrow \mathcal{X}$ is imprecisely known, *i.e.* we know that $\phi(\omega) \in B$ but we do not know which element of B is the image of ω through ϕ . Such information can be encoded using a random set Γ . (Dempster)



A.P. Dempster.

Upper and lower probabilities induced by a multiple valued mapping.
Annals of Mathematical Statistics, 38(2) :325–339, 1967.



G. Shafer.

A Mathematical Theory of Evidence.

Princeton University press, Princeton (NJ), USA, 1976.



P. Smets and R. Kennes.

The transferable belief model.

Artificial Intelligence, 66(2) :191–234, 1994.



Peter Walley.

Statistical Reasoning with Imprecise Probabilities.

Chapman and Hall, London, 1991.



L. A. Zadeh.

Fuzzy sets.

Journal of Information and Control, 8 :338–353, 1965.



L.A. Zadeh.

Fuzzy sets as a basis for a theory of possibility.

Fuzzy Sets and Systems, 1 :3–28, 1978.