Belief functions for the working scientist

A UAI 2015 Tutorial

F. Cuzzolin[1] and T. Denoeux[2]

[1]Department of Computing and Communication Technologies
Oxford Brookes University, UK

[2]Université de Technologie de Compiegne, France
HEUDIASYC (UMR CNRS 7253)

UAI 2015

Outline

of this tutorial

1 Rationale
2 Basic notions
3 Semantics
4 Inference
5 Conditioning
6 Computation
7 Propagation
8 Decisions
9 BFs on reals
10 Sister theories
11 Advances
12 Toolbox
13 Applications
14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
Future trends

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
Future trends

Some history

- a mathematical framework for representing and reasoning with uncertain information
- also known as **Dempster-Shafer (DS) theory** or **Evidence theory**
- originates from the work of Dempster (1968) in the context of **statistical inference**
- formalized by Shafer (1976) as a **theory of evidence**
- popularized and developed by Smets in the 1980's and 1990's under the name **Transferable Belief Model**.
- starting from the 1990's, **growing number of applications** in AI, information fusion, classification, reliability and risk analysis, etc.

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
Future trends

Rationale

1. a modeling language for representing **elementary items of evidence and combining them**, in order to form a representation of our beliefs about certain aspects of the world
2. the theory of belief function subsumes both the **set-based** and **probabilistic** approaches to uncertainty:
   - a belief function may be viewed both as a **generalized set** and as a **non additive measure**
   - basic mechanisms for reasoning with belief functions extend both probabilistic operations (such as marginalization and conditioning) and set-theoretic operations (such as intersection and union)
3. DS reasoning produces the same results as probabilistic reasoning or interval analysis when provided with the same information
4. however, its **greater expressive power** allows us to handle more general forms of information

Outline

1. Rationale
2. **Basic notions**
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Mass functions

or "Basic Probability Assignments"

- let ω be an unknown quantity with possible values in a **finite** domain Ω, called the **frame of discernment**
- a piece of evidence about ω may be represented by a **mass function** m on Ω, defined as a function $2^\Omega \to [0, 1]$, such that:

$$m(\emptyset) = 0 \quad \sum_{A \subseteq \Omega} m(A) = 1$$

- $\mathcal{P}(\Omega) = 2^\Omega$ is the set of all subsets of Ω
- any subset A of Ω such that $m(A) > 0$ is called a **focal element** (FE) or "focal set" of m
- special cases:
  - a **logical** (or "categorical") mass function has one focal set ($\sim$ set)
  - a **Bayesian** mass function has only focal sets of cardinality one ($\sim$ probability distribution)
- **complete ignorance** is represented by the vacuous mass function defined by $m_\Omega(\omega) = 1$

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Belief functions
Dempster's rule
Families of frames
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

Mass function

Example

- a mass function encodes evidence directly supporting **propositions** [Shafer, 1976] - let us see an example
- a murder has been committed. There are three suspects: $\Omega = \{Peter, John, Mary\}$
- available evidence: a witness saw the murderer going away, but he is short-sighted and he only saw that it was a man. We know that the witness is drunk 20 % of the time
- if the witness was not drunk, we know that $\omega \in \{Peter, John\}$
- otherwise, we only know that $\omega \in \Omega$. The first case holds with probability 0.8
- the corresponding mass function is:

$$m(\{Peter, John\}) = 0.8, \quad m(\Omega) = 0.2$$

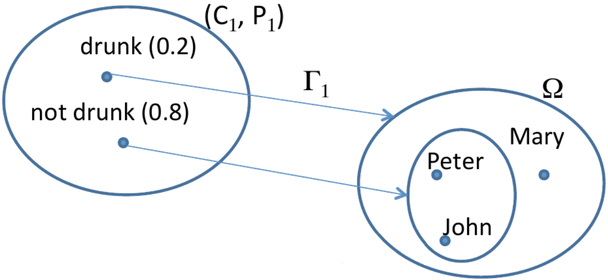
Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Belief functions
Dempster's rule
Families of frames
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

Random set interpretation

of belief functions



- a mass assignment on Ω is induced by a probability distribution P on a different domain $\mathcal{C}$, via a multi-valued mapping $\Gamma : \mathcal{C} \to 2^{\Omega}$
- Γ maps *elements* $c \in \mathcal{C}$ ("codes") to *subsets* of Ω
- this is a **random set**, i.e, a set-valued random variable
- in the example, the source of the evidence is the probability P_1 that the witness is drunk (or not)
- Γ_1 maps $\{not\ drunk\} \in \mathcal{C}_1$ to $\{Peter, John\} \subset \Omega$

Belief (BFs) and plausibility functions

induced by a mass function

- for any $A \subseteq \Omega$, we can define:
  - the total **degree of support (belief)** in A as the probability that the evidence implies A:

$$Bel(A) = P(\{c \in \mathcal{C} | \Gamma(c) \subseteq A\}) = \sum_{B \subseteq A} m(B)$$

  - the **plausibility** of A as the probability that the evidence does not contradict A:

$$Pl(A) = P(\{c \in \mathcal{C} | \Gamma(c) \cap A \neq \emptyset\}) = 1 - Bel(\bar{A})$$

- the uncertainty on the truth value of the proposition "$\omega \in A$" is represented by two numbers: $Bel(A)$ and $Pl(A)$, with

$$Bel(A) \leq Pl(A)$$

Special cases of belief functions

they generalise both probabilities and possibilities

- if all focal sets of m are singletons, then m is said to be **Bayesian**
- it is equivalent to a **probability distribution**, and $Bel = Pl$ is a probability measure
- if the focal sets of m are nested, then m is said to be **consonant**
- in that case Pl is a **possibility measure**, i.e.,

$$Pl(A \cup B) = \max(Pl(A), Pl(B)), \quad \forall A, B \subseteq \Omega,$$

  and Bel is the dual **necessity measure**
- the **contour function** $pl(\omega) = Pl(\{\omega\})$ corresponds to the possibility distribution (membership function)

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Belief functions
Dempster's rule
Families of frames
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

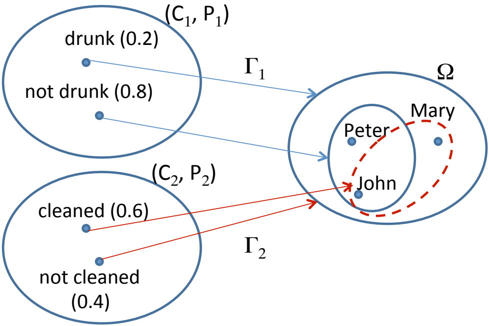
14 Future trends

Combination of evidence

Murder example continued

- when we have separate bodies of evidence, each represented by a belief function, **can we combine them** in order to estimate the state of the world, or make a decision?
- the first item of evidence gave us: $m_1(\{Peter, John\}) = 0.8$, $m_1(\Omega) = 0.2$
- new piece of evidence: a blond hair has been found
- also, there is a probability 0.6 that the room has been cleaned before the crime
- this second body of evidence is encoded by the mass assignment $m_2(\{John, Mary\}) = 0.6$, $m_2(\Omega) = 0.4$
- how to combine these two pieces of evidence?
- again, an answer can be given within the "random set" interpretation of belief functions

Combination of evidence



- if codes $c_1 \in \mathcal{C}_1$ and $c_2 \in \mathcal{C}_2$ were selected, $\omega \in \Gamma_1(c_1) \cap \Gamma_2(c_2)$
- if the codes are selected **independently**, then the probability that the pair (c_1, c_2) is selected is $P_1(\{c_1\})P_2(\{c_2\})$
- if $\Gamma_1(c_1) \cap \Gamma_2(c_2) = \emptyset$, (c_1, c_2) cannot be selected, hence:
- the joint probability distribution on $\mathcal{C}_1 \times \mathcal{C}_2$ must be conditioned to eliminate such pairs

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale

Basic notions
Belief functions
Dempster's rule
Families of frames

Semantics

Inference

Conditioning

Computation

Propagation

Decisions

BFs on reals

Sister theories

Advances

Toolbox

Applications

Dempster's rule

Definition

- under these assumptions we get Dempster's rule of combination
- let m_1 and m_2 be two mass functions on the same frame Ω, induced by two **independent pieces of evidence**
- their combination using **Dempster's rule** is defined as:

$$(m_1 \oplus m_2)(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall \emptyset \neq A \subseteq \Omega,$$

  where

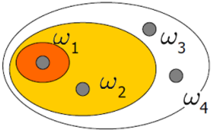
$$\kappa = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$$

  is the **degree of conflict** between m_1 and m_2
- their Dempster's sum $m_1 \oplus m_2$ exists iff $\kappa < 1$
- can be easily extended to any number of BFs

Rationale
Basic notions
Belief functions
Dempster's rule
Families of frames
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

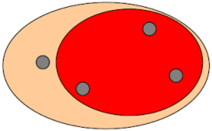
Dempster's rule

A simple numerical example



- Bel_1:

  $m(\{\omega_1\})=0.7$, $m(\{\omega_1,\omega_2\})=0.3$



- Bel_2:

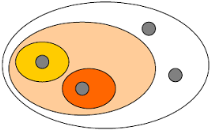
  $m(\Omega)=0.1$, $m(\{\omega_2,\omega_3,\omega_4\})=0.9$

- $Bel_1 \oplus Bel_2$:

  $m(\{\omega_1\}) = 0.7*0.1/0.37 = 0.19$

  $m(\{\omega_2\}) = 0.3*0.9/0.37 = 0.73$

  $m(\{\omega_1,\omega_2\}) = 0.3*0.1/0.37 = 0.08$



Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale

Basic notions

Belief functions

Dempster's rule

Families of frames

Semantics

Inference

Conditioning

Computation

Propagation

Decisions

BFs on reals

Sister theories

Advances

Toolbox

Applications

Dempster's rule

Properties

- Dempster's rule has some interesting properties:
- commutativity, associativity, existence of a neutral element: the **vacuous** BF m_Ω
- it generalises set-theoretical **intersection**: if m_A and m_B are logical mass functions and $A \cap B \neq \emptyset$, then

$$m_A \oplus m_B = m_{A \cap B}$$

- it generalises **probabilistic conditioning** via Bayes' rule: if m is a Bayesian mass function and m_A is a logical mass function, then $m \oplus m_A$ is a Bayesian mass function that corresponding to Bayes' conditioning of m by A

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Belief functions
Dempster's rule
Families of frames
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

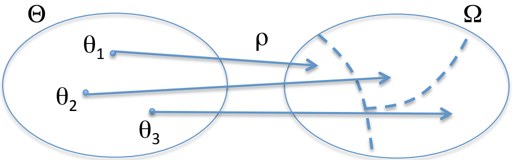
Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Families of frames

Refinements and coarsenings

- the theory allows us to **handle evidence impacting on different but related domains**
- assume we are interested in the nature of an object in a road scene. We could describe it, e.g., in the frame $\Theta = \{\text{vehicle}, \text{pedestrian}\}$, or in the finer frame $\Omega = \{\text{car}, \text{bicycle}, \text{motorcycle}, \text{pedestrian}\}$
- other example: different image features in pose estimation
- a frame Ω is a **refinement** of a frame Θ (or, equivalently, Θ is a coarsening of Ω) if elements of Ω can be obtained by splitting some or all of the elements of Θ



Rationale
Basic notions
Belief functions
Dempster's rule
Families of frames
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

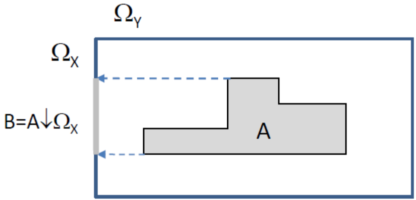
Compatible frames

Families of frames

- when Ω is a refinement for a collection $\Theta_1, ..., \Theta_N$ of other frames it is called their **common refinement**
- two frames are said to be **compatible** if they do have a common refinement
- compatible frames can be associated with **different variables/attributes/features**:
  - let $\Omega_X = \{\text{red}, \text{blue}, \text{green}\}$ and $\Omega_Y = \{\text{small}, \text{medium}, \text{large}\}$ be the domains of attributes X and Y describing, respectively, the color and the size of an object
  - in such a case the common refinement $\Omega_X \times \Omega_Y$ is Ω_X and Ω_Y
- or, they can be descriptions of the **same variable at different levels of granularity** (as in the road scene example)
- **evidence can be moved from one frame to another** within a family of compatible frames

Marginalization

- let Ω_X and Ω_Y be two compatible frames
- let m^{XY} be a mass function on $\Omega_X \times \Omega_Y$
- it can be expressed in the coarser frame Ω_X by transferring each mass $m^{XY}(A)$ to the **projection** of A on Ω_X:



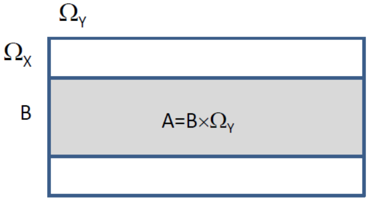
- we obtain a **marginal** mass function on Ω_X:

$$m^{XY\downarrow X}(B) = \sum_{\{A\subseteq\Omega_{XY}, A\downarrow\Omega_X = B\}} m^{XY}(A) \quad \forall B \subseteq \Omega_X$$

- (again, it generalizes both set projection and probabilistic marginalization)

Vacuous extension

- the "inverse" of marginalization
- a mass function m^X on Ω_X can be expressed in $\Omega_X \times \Omega_Y$ by transferring each mass $m_X(B)$ to the **cylindrical extension** of B:



- this operation is called the **vacuous extension** of m_X in $\Omega_X \times \Omega_Y$:

$$m^{X\uparrow XY}(A) = \begin{cases} m^X(B) & \text{if } A = B \times \Omega_Y \\ 0 & \text{otherwise} \end{cases}$$

- a strong feature of belief theory: the vacuous belief function (our representation of ignorance) **is left unchanged when moving from one compatible frame to another**

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Lower probabilities
Credal sets
Set functions
Generalised probabilities
Random sets
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale

Basic notions

Semantics

Lower probabilities

Credal sets

Set functions

Generalised probabilities

Random sets

Inference

Conditioning

Computation

Propagation

Decisions

BFs on reals

Sister theories

Advances

The multiple semantics

of belief functions

- being complex objects, belief functions have a number of (sometimes conflicting) semantics and mathematical interpretations
- original one [Dempster 1967]: lower probabilities induced by a multivalued mapping
  - the mathematical representation: **random set** framework
- Shafer's (1976): representations of pieces of evidence in favour of propositions within someone's subjective state of belief
  - represented as **set functions** on a finite domain Ω
- as **convex sets of probability measures**, in a robust Bayesian interpretation
  - mathematically, a credal set whose lower and upper envelopes are belief and plausibility functions
- other equivalent mathematical formulations:
  - as non-additive (generalised) probabilities
  - as monotone capacities
  - as inner measures (linked to the rough set idea)

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Belief functions for the working scientist

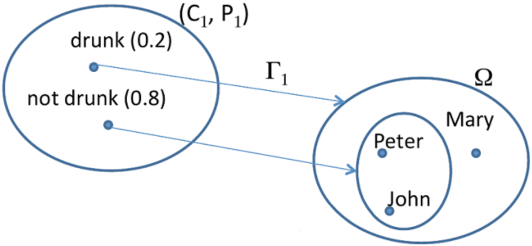
F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Lower probabilities
Credal sets
Set functions
Generalised probabilities
Random sets
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances

Upper and lower probabilities

induced by multi-valued mappings

- Dempster has shown that mapping a probability distribution via a multi-valued map yields an object more general than a probability distribution: a belief function



- belief and plausibility values are interpreted as **lower and upper bounds** to the values of an unknown, underlying probability measure: $Bel(A) \leq P(A) \leq Pl(A)$ for all $A \subseteq \Omega$

Outline

1 Rationale
2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions as credal sets

or convex sets of probabilities

- each focal element A of mass $m(A)$ as the indication of the existence of a mass $m(A)$ “floating" inside A
- **constraint on the probability measure** on Ω: a distribution is “consistent" with *Bel* if it is obtained by redistributing the mass of each focal element to its singletons
- **set of probabilities consistent with** b:

$$\mathcal{P}[Bel] \doteq \Big\{ P \in \mathcal{P} : P(A) \geq Bel(A) \; \forall A \subseteq \Omega \Big\}$$

- it is a polytope in the probability simplex, with vertices induced by permutations of the elements of Ω
- Shafer disavowed any probability-bound interpretation
- also criticized by Walley as incompatible with Dempster’s rule of combination

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Lower probabilities
Credal sets
Set functions
Generalised probabilities
Random sets
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances

Belief functions as set functions

- Shafer's definition in terms of set functions [Aigner]
- a **belief function** $Bel : 2^\Omega \to [0, 1]$ is such that,

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

where $m : 2^\Omega \to [0, 1]$ is a basic probability assignment s.t.

$$m(\emptyset) = 0, \sum_{A \subseteq \Omega} m(A) = 1, m(A) \geq 0 \ \forall A \subseteq \Omega$$

- operating with belief functions reduces to manipulating their focal elements
- in Shafer's framework, the mass assignment is derived by "impact of evidence" associated with propositions via an exponential-like relation

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Lower probabilities
Credal sets
Set functions
Generalised probabilities
Random sets
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Lower probabilities
Credal sets
Set functions
Generalised probabilities
Random sets
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances

As non-additive probabilities

or "generalised" probabilities

Probability measure

A function $P : \mathbf{F} \rightarrow [0,1]$ over a σ-field $\mathbf{F} \subset 2^\Omega$ such that

- $P(\emptyset) = 0$, $P(\Omega) = 1$;
- if $A \cap B = \emptyset$, $A, B \in \mathbf{F}$ then $P(A \cup B) = P(A) + P(B)$ (*additivity*).

- if we relax the third constraint to allow the function to meet additivity **only as a lower bound** we obtain a:

Belief function

A function $Bel : 2^\Omega \rightarrow [0,1]$ from the power set 2^Ω to $[0,1]$ such that:

- $Bel(\emptyset) = 0$, $Bel(\Omega) = 1$;
- for every n and for every collection $A_1, ..., A_n \in 2^\Omega$ we have that:

$$Bel(A_1 \cup ... \cup A_n) \geq \sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + \cdots + (-1)^{n+1} Bel(A_1 \cap ... \cap A_n)$$

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Lower probabilities
Credal sets
Set functions
Generalised probabilities
Random sets
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances

Belief functions as completely monotone capacities

- a function $Bel : 2^\Omega \rightarrow [0,1]$ is a **completely monotone capacity**, i.e., it verifies $Bel(\emptyset) = 0$, $Bel(\Omega) = 1$ and

$$Bel\left(\bigcup_{i=1}^{k} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} A_i\right)$$

for any $k \geq 2$ and for any family $A_1, \ldots, A_k$ in 2^Ω

- conversely, to any completely monotone capacity *Bel* corresponds a **unique mass function** *m* such that:

$$m(A) = \sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} Bel(B), \quad \forall A \subseteq \Omega$$

- in combinatorics this is called **Moebius transform**
- *m*, *Bel* and *Pl* are thus **equivalent representations** of the same piece of evidence

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Lower probabilities
Credal sets
Set functions
Generalised probabilities
Random sets
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Lower probabilities
Credal sets
Set functions
Generalised probabilities
Random sets
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances

Belief functions as random sets

Rationale

- given a multi-valued mapping Γ, a straightforward step is to consider the probability value $P(c)$ as attached to the subset $\Gamma(c) \subseteq \Omega$
- what we obtain is a **random set** in Ω, i.e., a probability measure on a collection of subsets
- roughly speaking, a random set is a set-valued random variable
- the degree of belief $Bel(A)$ of an event A becomes the **cumulative distribution function (CDF) of the open interval of sets** $\{B \subseteq A\}$
- this approach has been emphasized in particular by [Nguyen,1978] and [Hestir,1991] and [Shafer,1987]
- example: **a dice where one or more of faces are covered** so that we do not know what's beneath is a random variable which "spits" subsets of possible outcomes: a random set

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Lower probabilities
Credal sets
Set functions
Generalised probabilities
Random sets
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances

Belief functions as random sets

Mathematics

- the **lower inverse** of Γ is defined as:

$$\Gamma_*(A) \doteq \left\{ c \in \mathcal{C} : \Gamma(c) \subset A, \Gamma(c) \neq \emptyset \right\}$$

  while its **upper inverse** is

$$\Gamma^*(A) \doteq \left\{ c \in \mathcal{C} : \Gamma(c) \cap A \neq \emptyset \right\}$$

- given two σ-fields $\mathcal{A}, \mathcal{B}$ on $\mathcal{C}, \Omega$ respectively, Γ is said **strongly measurable** iff $\forall B \in \mathcal{B}$, $\Gamma^*(B) \in \mathcal{A}$
- the **lower probability measure** on $\mathcal{B}$ is defined as $P_*(B) \doteq P(\Gamma_*(B))$ for all $B \in \mathcal{B}$ - this is nothing but a belief function!
- Nguyen proved that, if Γ is strongly measurable, the CDF $\hat{P}$ of the random set coincides with the lower probability measure:

$$\hat{P}[I(B)] = P_*(B) \quad \forall B \in \mathcal{B}, \qquad I(B) \doteq \{C \in \mathcal{B}, C \subseteq B\}$$

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
   - Imprecise probability
   - Fuzzy sets and possibility
   - p-Boxes
11. Advances
   - Matrix representation
   - Geometry
   - Distances
   - Combinatorics
   - Uncertainty measures
12. Toolbox
   - Classification
   - Ranking aggregation
13. Applications
   - A brief survey
   - Climate change
   - Pose estimation
14. Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Dempster's approach
Likelihood-based
From preferences
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

The inference problem

Building belief functions from the available data

- first step in any estimation/decision problem: **constructing a belief function from the available evidence**
- belief functions can be constructed from both statistical data (quantitative inference) and experts' preferences (qualitative inference)
- inference from **statistical data**: we will see two
  - Dempster's approach
  - likelihood-based approach
- inference from **qualitative data**
  - Wong and Lingras's perceptron idea
  - Qualitative Discrimination Process (QDP)
  - Ben Yaghlane's constrained optimisation

Inferring belief functions from statistical data

- consider a **statistical model**

$$\left\{ f(x;\theta), x \in \mathbb{X}, \theta \in \Theta \right\},$$

  where $\mathbb{X}$ is the sample space and Θ the parameter space
- having observed x, how to **quantify the uncertainty about the parameter** θ, without specifying a prior probability distribution?
- two main approaches using belief functions:
  1. **Dempster's approach** based on an auxiliary variable with a pivotal probability distribution [Dempster, 1967]
  2. **Likelihood-based approach** [Shafer, 1976, Wasserman 1990]

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Dempster's approach
Likelihood-based
From preferences
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Dempster's approach
Likelihood-based
From preferences
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

Sampling model

Dempster's approach to statistical inference

- suppose that the sampling model $X \sim f(x;\theta)$ can be represented by an "a-equation" of the form

$$X = a(\theta, U)$$

  where $U \in \mathbb{U}$ is an **(unobserved) auxiliary variable** with known probability distribution μ independent of θ
- this representation is quite natural in the context of sampling and **data generation**
- for instance, to generate a continuous random variable X with cumulative distribution function (CDF) F_θ, one might draw U from $\mathcal{U}([0,1])$ and set

$$X = F_\theta^{-1}(U)$$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Dempster's approach
Likelihood-based
From preferences
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

From *a*-equations to belief functions

- the equation $X = a(\theta, U)$ defines a multi-valued mapping (a "compatibility relation")

$$\Gamma : U \to \Gamma(U) = \Big\{ (X, \theta) \in \mathbb{X} \times \Theta \Big| X = a(\theta, U) \Big\}$$

- under the usual measurability conditions, the probability space $(\mathbb{U}, \mathcal{B}(\mathbb{U}), \mu)$ and the multi-valued mapping Γ induce a belief function $Bel_{\Theta\times\mathbb{X}}$ on $\mathbb{X} \times \Theta$
- conditioning (by Dempster's rule) $Bel_{\Theta\times\mathbb{X}}$ on θ yields the desired sampling distribution $f(\cdot;\theta)$ on $\mathbb{X}$
- conditioning it on $X = x$ gives a belief function $Bel_\Theta(\cdot; x)$ on Θ

Example: Bernoulli sample

Dempster's approach to inference

- let $X = (X_1, \ldots, X_n)$ consist of **independent Bernoulli observations** and $\theta \in \Theta = [0, 1]$ is the probability of success
- sampling model:

$$X_i = \begin{cases} 1 & \text{if } U_i \leq \theta \\ 0 & \text{otherwise,} \end{cases}$$

  where $U = (U_1, \ldots, U_n)$ has pivotal measure $\mu = \mathcal{U}([0, 1]^n)$
- having observed the number of successes $y = \sum_{i=1}^n x_i$, the belief function $Bel_\Theta(\cdot; x)$ is induced by a **random closed interval**

$$[U_{(y)}, U_{(y+1)}],$$

  where $U_{(i)}$ denotes the i-th order statistics from $U_1, \ldots, U_n$
- quantities like $Bel_\Theta([a, b]; x)$ or $Pl_\Theta([a, b]; x)$ are readily calculated

Discussion

Dempster's method

- Dempster's model has several nice features:
  - it allows us to quantify the uncertainty on Θ after observing the data, without having to specify a prior distribution on Θ
  - when a Bayesian prior P_0 is available, **combining it with** $Bel_\Theta(\cdot; x)$ **using Dempster's rule yields the Bayesian posterior**:

$$Bel_\Theta(\cdot; x) \oplus P_0 = P(\cdot|x)$$

- it also has some drawbacks:
  - it often leads to **cumbersome or even intractable calculations** except for very simple models, which imposes the use of Monte-Carlo simulations (see Computation later)
  - more fundamentally, **the analysis depends on the a-equation** $X = a(\theta, U)$ **and the auxiliary variable** U, which are not unique for a given statistical model $\{f(\cdot; \theta), \theta \in \Theta\}$
  - As U is not observed, how can we argue for an a-equation or another?

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Dempster's approach
Likelihood-based
From preferences
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Likelihood-based belief function

Requirements

1. **Likelihood principle**: $Bel_\Theta(\cdot; x)$ should be based only on the likelihood function $L(\theta; x) = f(x; \theta)$
2. **Compatibility with Bayesian inference**: when a Bayesian prior P_0 is available, combining it with $Bel_\Theta(\cdot, x)$ using Dempster's rule should yield the Bayesian posterior:
$$Bel_\Theta(\cdot; x) \oplus P_0 = P(\cdot|x)$$
3. **Principle of minimal commitment**: among all the belief functions satisfying the previous two requirements, $Bel_\Theta(\cdot; x)$ should be the least committed (least informative)

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Dempster's approach
Likelihood-based
From preferences
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

Likelihood-based belief function

Solution

- $Bel_\Theta(\cdot; x)$ is the **consonant belief function** with contour function (plausibility of singletons) equal to the **normalized likelihood**:

$$pl(\theta; x) = \frac{L(\theta; x)}{\sup_{\theta' \in \Theta} L(\theta'; x)}$$

- the corresponding plausibility function is:

$$Pl_\Theta(A; x) = \sup_{\theta \in A} pl(\theta; x) = \frac{\sup_{\theta \in A} L(\theta; x)}{\sup_{\theta \in \Theta} L(\theta; x)}, \quad \forall A \subseteq \Theta$$

- the corresponding random set is: $(\Omega, \mathcal{B}(\Omega), \mu, \Gamma_x)$ with $\Omega = [0, 1]$, $\mu = \mathcal{U}([0, 1])$ and

$$\Gamma_x(\omega) = \left\{ \theta \in \Theta \middle| pl(\theta; x) \geq \omega \right\}$$

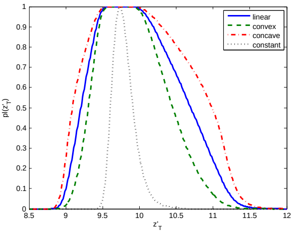
Rationale
Basic notions
Semantics
Inference
Dempster's approach
Likelihood-based
From preferences
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

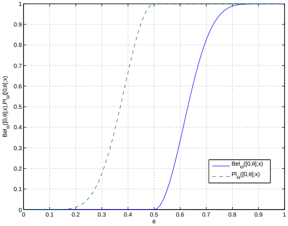
Example: Bernoulli sample

- let $X = (X_1, \ldots, X_n)$ consist of independent Bernoulli observations and $\theta \in \Theta = [0, 1]$ is the probability of success
- we get

$$pl(\theta; x) = \frac{\theta^y (1-\theta)^{n-y}}{\hat{\theta}^y (1-\hat{\theta})^{n-y}},$$

  where $y = \sum_{i=1}^n x_i$ and $\hat{\theta}$ is the MLE
- example for $n = 20$ and $y = 10$:





Discussion

Likelihood method

- the likelihood-based method is much simpler to implement than Dempster's method, even for complex models.
- by construction, it **boils down to Bayesian inference when a Bayesian prior is available**
- it is compatible with usual likelihood-based inference:
  - assume that $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$ and θ_2 is a **nuisance parameter**. The marginal contour function on Θ_1

    $$pl(\theta_1; x) = \sup_{\theta_2 \in \Theta_2} pl(\theta_1, \theta_2; x) = \frac{\sup_{\theta_2 \in \Theta_2} L(\theta_1, \theta_2; x)}{\sup_{(\theta_1, \theta_2) \in \Theta} L(\theta_1, \theta_2; x)}$$

    is the relative **profile likelihood** function
  - the plausibility of a composite hypothesis $H_0 \subset \Theta$

    $$Pl(H_0; x) = \frac{\sup_{\theta \in H_0} L(\theta; x)}{\sup_{\theta \in \Theta} L(\theta; x)}$$

    is the usual **likelihood ratio statistics** $\Lambda(x)$

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Wong and Lingras

Building belief functions from preferences

- Wong and Lingras [16] proposed a method for generating BFs from a body of qualitative preference relations between propositions
- two binary relations: preference $\cdot >$ and indifference $\sim$
- goal: to build a belief function Bel such that $A \cdot > B$ iff $Bel(A) > Bel(B)$ and $A \sim B$ iff $Bel(A) = Bel(B)$
- exists if $\cdot >$ is a weak order and $\sim$ an equivalence relation
- **Algorithm**
  1. consider all propositions that appear in the preference relations as potential focal elements (FEs)
  2. elimination: if $A \sim B$ for some $B \subset A$ then A is not a FE
  3. a perceptron algorithm is used to generate the mass m by solving the system of remaining equalities and disequalities
- however: it selects arbitrarily one solution over many
- does not address possible inconsistency in the given preferences

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Dempster's approach
Likelihood-based
From preferences
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications

Ben Yaghlane's constrained optimisation approach

Building belief functions from preferences

- uses preferences and indifferences as in Wong and Lingras, with same axioms..
- .. but converts them into a **constrained optimisation problem**
- objective function: maximise the entropy/uncertainty of the BF to generate (least informative result)
- constraints derived from input preferences/indifferences, i.e.

$$A \cdot > B \leftrightarrow Bel(A) - Bel(B) \geq \epsilon, \quad A \sim B \leftrightarrow |Bel(A) - Bel(B)| \leq \epsilon$$

- ϵ is a constant specified by the expert
- various uncertainty measures can be plugged in (see Advances)

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Conditional belief functions

A variety of proposals

- many different approaches to conditioning belief functions have been proposed
- a non-exhaustive list:
  - original Dempster's conditioning
  - **lower and upper envelopes** of conditional probabilities [Fagin and Halpern]
  - **geometric conditioning** [Suppes]
  - **unnormalized** conditional belief functions [Smets]
  - generalised Jeffrey's rules [Smets]
  - **sets of equivalent events** under multi-valued mappings [Spies]
  - conditioning by distance minimisation [Cuzzolin]
- implications of the notion of conditional belief function:
  - **generalised Bayes theorem** [Smets]
  - the generalisation of the **total probability** theorem

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

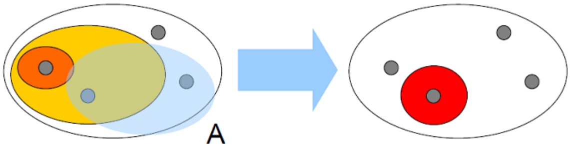
F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Dempster's conditioning

Conditioning

- Dempster's rule of combination is associated with a conditioning operator
- suppose we have an "a-priori" BF *Bel*
- given a new event *A*, the "logical" belief function such that $m(A) = 1$ can be defined ...
- ... and combined with *Bel* using Dempster's rule
- the resulting BF is the conditional belief function given *A*, *a la Dempster*



Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Lower conditional envelopes

Conditioning

- Fagin and Halpern proposed an approach based on interpretation of a belief function as the lower envelope of the family of probabilities consistent with it (robust Bayesian)

$$Bel(A) = \inf_{P \in \mathcal{P}[Bel]} P(A)$$

- they define the conditional belief as the **lower envelope** (that is, the infimum) **of the family of conditional probability functions** $P(A|B)$, where P is consistent with Bel:

$$Bel(A|B) \doteq \inf_{P \in \mathcal{P}[Bel]} P(A|B), \quad Pl(A|B) \doteq \sup_{P \in \mathcal{P}[Bel]} P(A|B)$$

- trivially generalises conditional probability
- have been considered by other authors too, e.g. Dempster 1967 and Walley 1981

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale

Basic notions

Semantics

Inference

Conditioning

Dempster's conditioning

Lower conditional envelopes

Unnormalised vs geometric conditioning

Conditional events as equivalence classes

Generalised Bayes theorem

The total belief theorem

Computation

Propagation

Decisions

Lower conditional envelopes

versus Dempster's

- the authors provide a closed-form expression for it:

$$Bel(A|B) = \frac{Bel(A \cap B)}{Bel(A \cap B + Pl(\bar{A} \cap B)}, \quad Pl(A|B) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(\bar{A} \cap B)}$$

- lower/upper envelopes of arbitrary sets of probabilities are not in general belief functions, but **these actually are**, as Fagin and Halpern have proven
- they are quite different from Dempster's conditioning:

$$Bel_{\oplus}(A|B) = \frac{Bel(A \cup \bar{B})}{1 - Bel(\bar{B})}, \quad Pl_{\oplus}(A|B) = \frac{Pl(A \cap B)}{Pl(B)}$$

- in fact, they provide a **more conservative estimate**:

$$Bel(A|B) \leq Bel_{\oplus}(A|B) \leq Pl_{\oplus}(A|B) \leq Pl(A|B)$$

- Fagin and Halpern argue that Dempster's conditioning behaves unreasonably on their "three prisoners" example

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Revision versus focussing

in belief as opposed to probability theory

Focussing

No new information is introduced, we merely focus on a specific subset of the original set.

Belief revision

A state of belief is modified to take into account a new piece of information.

- in probability theory, **both are expressed by Bayes' rule**, but they are conceptually different operations
- in belief theory, these principles **lead to different conditioning rules**
- the application of revision and focussing to belief theory has been explored by Smets in his Transferable Belief Model (TBM)
- here we are not assuming any random set generating *Bel*, nor any underlying convex sets of probabilities

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Suppes' geometric conditioning

Conditioning

- the **geometric conditioning** proposed by Suppes and Zanotti

$$Bel_G(A|B) = \frac{Bel(A \cap B)}{Bel(B)},$$

  is indeed a consequence of the **focussing** idea
- (this was proved by Smets using the "probability of provability" interpretation of belief functions, yes, yet another one!)
- somewhat dual to Dempster's conditioning, as it **replaces probability with belief** in Bayes' rule
- remember that Dempster's rule dually **replaces probability with plausibility** in Bayes' rule

$$Pl_\oplus(A|B) = \frac{Pl(A \cap B)}{Pl(B)} \quad \leftrightarrow \quad Bel_G(A|B) = \frac{Bel(A \cap B)}{Bel(B)}$$

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Smets' unnormalised rule
of conditioning

- to rebuke Bayesian criticisms, in his TBM Smets rejects the existence of a probability measure on a parent space $\mathcal{C}$
- Smet's (Dempster's) **unnormalized conditional belief function**:

$$m_U(.|B) = \begin{cases} \sum_{X \subseteq B^c} m(A \cup X) & \textit{if } A \subseteq B \\ 0 & \textit{elsewhere} \end{cases}$$

- (in the TBM BFs which assign mass to $\emptyset$ can exist, under the "open world" assumption)
- in terms of plausibilities: $Pl_U(A|B) = Pl(A \cap B)$ - in the TBM the mass $m(A)$ is transferred by conditioning on B to $A \cap B$
- it is a consequence of **belief revision** principles [Gilboa, Perea]

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Spies' sets of equivalent events under multi-valued mappings

Conditioning

- intriguing approach to conditioning, in the random set interpretation with $(\mathcal{C}, \mathcal{F}, P)$ and $\Gamma : \mathcal{C} \to 2^\Omega$
- null sets for $P(.|A)$: $\mathcal{N}(P(.|A)) = \{B \in \mathcal{A} : P(B|A) = 0\}$
- let $\triangle$ be the symmetric difference $A \triangle B = (A \cap \bar{B}) \cup (\bar{A} \cap B)$
- two events have the same conditional probability if they both are the symmetric difference between a same event and some null set
- a **conditional event** $[B|A]$ with $A, B \subseteq \Omega$ is a set of events with the same conditional probability $P(B|A)$:

$$[B|A] = B \triangle \mathcal{N}(P_A)$$

- you can prove that $[B|A] = \{C : A \cap B \subseteq C \subseteq \bar{A} \cup B\}$

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Conditional belief functions

Spies' approach

- by applying to conditional events a multivalued mapping Spies gave a new definition of conditional belief function
- **conditional multivalued mapping** for $B \subseteq \Omega$: $\Gamma_B(c) = [\Gamma(c)|B]$, where $\Gamma : \mathcal{C} \rightarrow 2^\Omega$
- (if $A = \Gamma(c)$, Γ_B maps c to $[A|B]$)
- consequence: to all elements of each conditioning event (an equivalence class) must be assigned equal belief/plausibility
- a **conditional belief function** is then a "second-order" BF with values on *collections* of focal elements (the conditional events)

$$Bel([C|B]) = P(\{c : \Gamma_B(c) = [C|B]\}) = \frac{1}{K} \sum_{A \in [C|B]} m(A)$$

- it is *not* a BF on the sub-algebra $\{Y = C \cap B, C \subseteq \Omega\}$
- Spies' conditional belief functions are **closed under Dempster's rule** of combination

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Jeffrey's rule of conditioning

or Total Probability Theorem

- suppose P is defined on a σ-algebra $\mathbb{A}$
- there is a new prob measure P' on a sub-algebra $\mathbb{B}$ of $\mathbb{A}$, and the updated probability P'' has to:
  1. meet the prob values specified by P' for events in $\mathbb{B}$
  2. be such that $\forall\ B \in \mathbb{B}, X, Y \subset B, X, Y \in \mathbb{A}$

$$\frac{P''(X)}{P''(Y)} = \begin{cases} \frac{P(X)}{P(Y)} & \textit{if } P(Y) > 0 \\ 0 & \textit{if } P(Y) = 0 \end{cases}$$

- there is a unique solution:

$$P''(A) = \sum_{B \in \mathbb{B}} P(A|B)P'(B)$$

- meaning: the initial probability **stands corrected by the second one** on a number of events
- generalises conditioning (obtained when $P'(B) = 1$ for some B)

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Jeffrey's rule generalised

Jeffrey's rule for belief functions

1. Let $\Pi = \{B_1, ..., B_n\}$ a disjoint partition of Ω;
2. $m_1, ..., m_n$ the BPAs of BFs conditional on $B_1, ..., B_n$ respectively;
3. m_B an unconditional belief function on the coarsening associated with the partition Π

Then the belief function $Bel_{tot}(A) = \sum_{C \subseteq A} \left(m_B \oplus \oplus_i^n m_{B_i} \right)(C)$ is a marginal belief function on Ω, and if all BFs are probabilities is reduces to the result of Jeffrey's rule of total probability

- combining the a-priori with all the conditionals we get a marginal
- is this the only solution? we will discuss this later (total belief theorem)

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Conditioning approaches

A summary

- various approaches to conditioning have been proposed:
  - **Dempster's** (normalised) conditioning:
  $$Bel_{\oplus}(A|B) = \frac{Bel(A \cup \bar{B})}{1 - Bel(\bar{B})}, \quad Pl_{\oplus}(A|B) = \frac{Pl(A \cap B)}{Pl(B)}$$
  - lower and upper **conditional envelopes**:
  $$Bel(A|B) = \frac{Bel(A \cap B)}{Bel(A \cap B + Pl(\bar{A} \cap B)}, \quad Pl(A|B) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(\bar{A} \cap B)}$$
  - **geometric** conditioning:
  $$Bel_G(A|B) = \frac{Bel(A \cap B)}{Bel(B)},$$
  - Smets' **unnnormalised** rule: $Pl_U(A|B) = Pl(A \cap B)$
  - Spies' conditioning: $Bel[A|B] \propto \sum_{X: A \cap B \subseteq X \subseteq A \cup \bar{B}} m(X)$
- derived from **different revision principles**
- follow **different semantic interpretations** (TBM rather than random set, open versus closed world assumption, robust Bayesian)

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Bayes' theorem

generalised to belief functions

- have a conditional probability $P(x|\theta_i)$ over observations $x \in X$, and an a-priori probability P_0 over a set of hidden variables $\theta_i \in \Theta$
- (for instance, x is a symptom and θ_i a disease)
- after observing x, the probability distribution on Θ is updated to the posterior via Bayes's theorem:

$$P(\theta_i|x) = \frac{P(x|\theta_i)P_0(\theta_i)}{\sum_j P(x|\theta_j)P_0(\theta_j)} \quad \forall \theta_j \in \Theta$$

- the **GBT is a generalisation of Bayes' theorem for conditional BFs, when the a-priori BF on Θ is vacuous**
- Dempster's normalised/unnormalised conditioning is assumed
- (a further generalisation for non-vacuous priors is proposed in Smets' work) [Smets 1993]

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Cognitive independence

- consider a belief function over the product space $X \times Y$
- the two variables are **cognitively independent** if

$$pl_{X \times Y}(x \cap y) = pl_X(x) pl_Y(y) \quad \forall x \subseteq X, y \subseteq Y$$

- cognitive independence extends stochastic independence
- **conditional** cognitive independence reads as

$$pl_{X \times Y}(x \cap y|\theta_i) = pl_X(x|\theta_i) pl_Y(y|\theta_i) \quad \forall x, y, \theta_i$$

and implies that the ratio of plausibility/belief on X does not depend on Y:

$$\frac{pl_X(x_1|y)}{pl_X(x_2|y)} = \frac{pl_X(x_1)}{pl_X(x_2)}, \quad \frac{Bel_X(x_1|y)}{Bel_X(x_2|y)} = \frac{Bel_X(x_1)}{Bel_X(x_2)}$$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Likelihood principle

Edwards, 1929

- the likelihood of an hypothesis given the data amounts to the conditional probability of the data given the hypothesis:

$$l(\theta_i|x) = p(x|\theta_i)$$

and, for unions of singleton hypotheses:

$$l(\theta = \{\theta_1, ..., \theta_k\}|x) = \max\left\{l(\theta_i|x) : \theta_i \in \theta\right\}$$

- Shafer's somewhat similar proposal for statistical inference (see inference-likelihood method):

$$pl(\theta|x) = \max_{\theta_i \in \theta} pl(\theta_i|x)'$$

was rejected by Smets, for not satisfying the condition that, if two pieces of evidence are conditionally independent, $Bel_\Theta(.|x, y)$ is the conjunctive combination of $Bel_\Theta(.|x)$ and $Bel_\Theta(.|y)$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Generalised Likelihood Principle

Smets' Generalised Likelihood Principle (GLP)

1. $pl_\Theta(\theta|x) = pl_X(x|\theta)$
2. For all x, θ the plausibility of data $pl(x|\theta)$ given a compound hypothesis $\theta = \{\theta_1, ..., \theta_m\}$ is a function of only

$$\{pl(x|\theta_i), pl(\bar{x}|\theta_i) : \theta_i \in \theta\}$$

- the form of the function is not assumed (not necessarily the max)
- both $pl(x|\theta_i)$ and $pl(\bar{x}|\theta_i)$ are necessary because of the non-additivity of belief functions
- justified by the following requirements:
  - $pl(x|\theta)$ remains the same on the coarsening of X formed by just x and $\bar{x}$
  - plausibilities for $\theta_j \notin \theta$ are irrelevant for $pl(x|\theta)$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Generalised Bayesian Theorem

and the disjunctive rule of combination

- under conditional cognitive independence and the Generalised Likelihood Principle (2), $Bel_X(.|\theta), \theta \subset \Theta$ is generated from the $\{Bel_X(.|\theta_i), \theta_i \in \Theta\}$ by **disjunctive rule of combination**

$$Pl_X(x|\theta) = 1 - \prod_{\theta_i \in \theta} (1 - Pl_X(x|\theta_i)), \quad Bel_X(x|\theta) = \prod_{\theta_i \in \theta} Bel_X(x|\theta_i)$$

- then, condition (1) of the GLP $pl_\Theta(\theta|x) = pl_X(x|\theta)$ implies the **generalised Bayes theorem**:

$$\begin{aligned} Pl_\Theta(\theta|x) &= \frac{1}{K}\Big(1 - \prod_{\theta_i \in \theta}(1 - pl_X(x|\theta_i))\Big) \\ Bel_\Theta(\theta|x) &= \frac{1}{K}\Big(\prod_{\theta_i \in \bar{\theta}} Bel_X(\bar{x}|\theta_i) - \prod_{\theta_i \in \Theta} Bel_X(\bar{x}|\theta_i)\Big) \end{aligned}$$

where $K = 1 - \prod_{\theta_i \in \Theta}(1 - pl_X(x|\theta_i))$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Outline

1 Rationale
2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

The total belief theorem

Generalising total probability to belief functions

Theorem

Suppose Θ and Ω are two frames of discernment, and $\rho : 2^\Omega \rightarrow 2^\Theta$ the unique refining between them. Let Bel_0 be a belief function defined over $\Omega = \{\omega_1, ..., \omega_{|\Omega|}\}$. Suppose there exists a collection of belief functions $Bel_i : 2^{\Pi_i} \rightarrow [0,1]$, where $\Pi = \{\Pi_1, ..., \Pi_{|\Omega|}\}$, $\Pi_i = \rho(\{\omega_i\})$, is the partition of Θ induced by its coarsening Ω.
Then, there exists a belief function $Bel : 2^\Theta \rightarrow [0,1]$ such that:

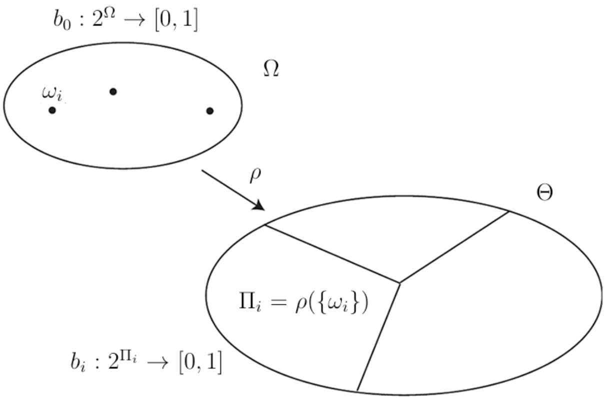
1. *Bel_0 is the restriction of Bel to Ω*
2. *$Bel \oplus Bel_{\Pi_i} = Bel_i \; \forall i = 1, ..., |\Omega|$, where Bel_{Π_i} is the logical belief function with mass*

$$m_{\Pi_i}(A) = 1 \;\; A = \Pi_i, \;\; 0 \; otherwise$$

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

The total belief theorem

Visual representation



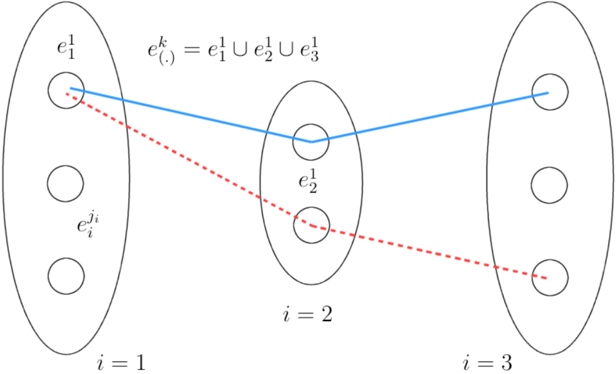
- pictorial representation of the total belief theorem

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

Structure of the focal elements

of the total belief function

- **restricted total belief theorem**: Bel_0 has only disjoint FEs
- pictorial representation of the structure of the FEs of a total BF *Bel* lying in the image of a focal element of Bel_0 of cardinality 3



Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem
Computation
Propagation
Decisions

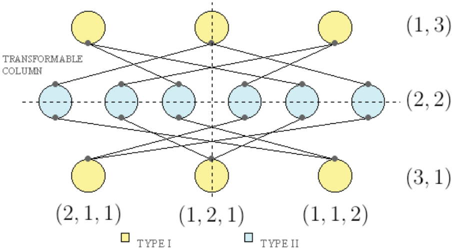
Graph of solutions

- potential solutions correspond to square linear systems, and form a graph whose nodes are linked by linear transformations of columns

$$e \mapsto e' = -e + \sum_{i \in \mathcal{C}} e_i - \sum_{j \in \mathcal{S}} e_j$$

where $\mathcal{C}$ is a covering set for e (i.e., every component of e is covered by at least one of them), $\mathcal{S}$ a set of selection columns

- at each transformation, the most negative component decreases



- general solution based on simplex-like optimisation?

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Probability transformation
Possibility transformation
Monte-Carlo methods
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Probability transformation
Possibility transformation
Monte-Carlo methods
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Efficient computation

with belief functions

- as belief functions are set functions, their complexity is exponential on the size n of the domain they are defined on
- combinining belief functions via Dempster's rule is also exponential: $(2^n)^N$, where N is the number of BFs involved
- efficient approaches based on approximating the original evidence - in particular, approaches that **transform a belief function into a less complex uncertainty measure**
  - probability (Bayesian) transformation
  - possibility (consonant) transformation
- approaches based on the **local propagation** of evidence
  - Barnett's approximation
  - hierarchical evidence
  - Cano's propagation on DAGs
  - Shafer-Shenoy architecture
- **Monte-Carlo methods** [Wilson and Moral]

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Belief functions for the working scientist

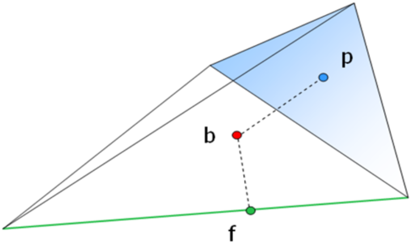
F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Probability transformation
Possibility transformation
Monte-Carlo methods
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Probability transformations

Mapping belief functions to probabilities

- **probability transform** of belief functions: an operator $pt : \mathcal{B} \rightarrow \mathcal{P}$, $b \mapsto pt[b]$ mapping belief measures onto probability distributions
- (not necessarily an element of the corresponding credal set)
- a number of transforms proposed, either as efficient implementations of ToE or tools for decision making
  - **pignistic** transform, central in the TBM [Smets]
  - **plausibility** and belief transform [Voorbraak, Cobb & Shenoy]
  - **orthogonal projection** and **intersection probability** [Cuzzolin]



Transformation in the TBM: the pignistic transform

Probability transformation

- Smets' Transferable Belief Model -> decisions made via "pignistic transform" ..
- .. resulting in a **pignistic probability**:

$$BetP[b](x) = \sum_{A \supseteq \{x\}} \frac{m_b(A)}{|A|},$$

- its purpose is to allow decision making at the level of probabilities, typically in an expected utility framework
- the result of a redistribution process in which the mass of each focal element A is re-assigned to all its elements $x \in A$ on an equal basis
- it commutes with affine combination and is the **center of mass of the credal set of consistent probabilities**

Plausibility and belief transform

Probability transformation

- **plausibility transform** [Voorbraak 89]: maps a belief function to the **relative belief of singletons**:

$$\tilde{pl}(x) = \frac{pl(x)}{\sum_{y \in \Theta} pl(y)}$$

- relative plausibility of singletons $\tilde{pl}$ is a perfect representative of *Bel* when combined with other probabilities by Dempster's rule $\oplus$
- meets a number of properties w.r.t. $\oplus$ [Coob&Shenoy 03]
- the **relative belief transform** maps each belief function to the corresponding *relative belief of singletons*:

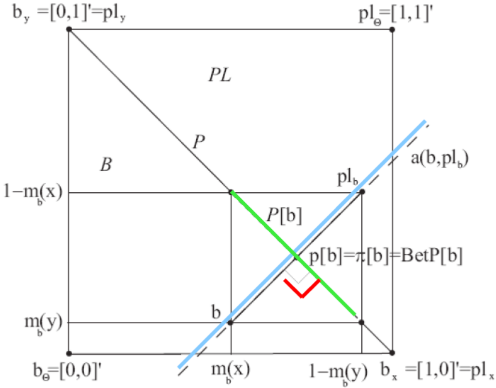
$$\tilde{bel}(x) = \frac{Bel(\{x\})}{\sum_{y \in \Theta} Bel(\{y\})}$$

- first proposed by Daniel in 2006, its geometry and that of plausibility transform analyzed in [Cuzzolin 2010 AMAI]

Geometric transformations

Probability transformation

- the probability transformation problem can be posed in geometric terms [IEEE SMC-B07]
  - **orthogonal projection** $\pi[b]$ -> minimises the L2 distance from $\mathcal{P}$
  - **intersection probability** $p[b]$ -> intersection with $\mathcal{P}$ of the belief-plausibility line



Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Probability transformation
Possibility transformation
Monte-Carlo methods
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
   - Imprecise probability
   - Fuzzy sets and possibility
   - p-Boxes
11. Advances
   - Matrix representation
   - Geometry
   - Distances
   - Combinatorics
   - Uncertainty measures
12. Toolbox
   - Classification
   - Ranking aggregation
13. Applications
   - A brief survey
   - Climate change
   - Pose estimation
14. Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Probability transformation
Possibility transformation
Monte-Carlo methods
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Possibility transformations

Outer approximations

- necessity measures have as counterparts in the ToE **consonant** belief functions, whose focal elements are nested: $A_1 \subset \cdots \subset A_m$, $A_i \subseteq \Theta$
- **outer consonant approximations** [Dubois&Prade 90]: consonant BFs *co* which are dominated by the original BF on all events:

$$co(A) \leq Bel(A) \quad \forall A \subseteq \Theta$$

- for each possible maximal chain $A_1 \subset \cdots \subset A_n$, $|A_i| = i$ of focal elements the **maximal outer consonant** approximation has mass

$$m_{\max}(A_i) = Bel(A_i) - Bel(A_{i-1})$$

- mirrors the behavior of the vertices of the credal set of probabilities dominating a belief function [Chateauneuf, Miranda& Grabish]

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Probability transformation
Possibility transformation
Monte-Carlo methods
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Possibility transformation

Isopignistic approximation

- completely different approximation in Smets' Transferable Belief Model [Smets94,05]
- **isopignistic" approximation**: the unique consonant belief function whose pignistic probability *BetP* is identical to that of *Bel* [Dubois, Aregui]
- its contour function is:

$$pl_{iso}(x) = \sum_{x' \in \Theta} \min \left\{ BetP(x), BetP(x') \right\}$$

- mass assignment:

$$m_{iso}(A_i) = i \cdot (BetP(x_i) - BetP(x_{i+1}))$$

where $\{x_i\} = A_i \setminus A_{i-1}$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Probability transformation
Possibility transformation
Monte-Carlo methods
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

A simple Monte-Carlo approach

to Dempster's combination - Wilson, 1989

- we seek $Bel = Bel_1 \oplus ... \oplus Bel_m$ on Ω, where the evidence is induced by probability distributions P_i on $\mathcal{C}_i$ via $\Gamma_i : \mathcal{C}_i \rightarrow 2^\Omega$
- Monte-Carlo algorithm **simulates the random set interpretation** of belief functions: $Bel(A) = P(\Gamma(c) \subseteq A | \Gamma(c) \neq \emptyset)$

```
for a large number of trials n = 1 : N do
    randomly pick c ∈ C such that Γ(c) ≠ ∅
    for i = 1 : m do
        randomly pick an element c_i of C_i with probability P_i(c_i)
    end for
    let c = (c_1, ..., c_m)
    if Γ(c) = ∅ then
        restart trial
    end if
    if Γ(c) ⊆ A then
        trial succeeds, T = 1
    end if
end for
```

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Probability transformation
Possibility transformation
Monte-Carlo methods
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox

A Monte-Carlo approach

Wilson, 1989

- the proportion of trials which succeed converges to $Bel(A)$: $E[\bar{T}] = Bel(A)$, $Var[\bar{T}] \leq \frac{1}{4N}$
- we say algorithms has accuracy k if $3\sigma[\bar{T}] \leq k$
- picking $c \in \mathcal{C}$ involves m random numers so it takes $A \cdot m$, A constant
- testing if $x_j \in \Gamma(c)$ takes less then Bm, constant B
- expected time of the algorithm is
$$\frac{N}{1-\kappa} m \cdot (A + B|\Omega|)$$
where κ is Shafer's conflict measure
- expected time to achieve accuracy k is then $\frac{9}{4(1-\kappa)\kappa^2} m \cdot (A + C|\Omega|)$ for constant C, better for simple support functions
- conclusion: **unless κ is close to 1 (highly conflicting evidence) Dempster's combination is feasible for large values of m (number of BFs to combine) and large** Ω (hypothesis space)

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Probability transformation
Possibility transformation
Monte-Carlo methods
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Markov-Chain Monte-Carlo

Wilson and Moral, 1996

- trials are not independent but form a Markov chain
- non-deterministic $OPERATION_i$: changes at most the i-th coordinate $c'(i)$ of c' to y, with chance $P_i(y)$

$$Pr(OPERATION_i(c') = c) \propto P_i(c(i)) \text{ if } c(i) = c'(i), \; 0 \text{ otherwise}$$

- MCMC algorithm which returns a value $BEL^N(c_0)$ which is the proportion of time in which $\Gamma(c_c) \subseteq X$

$c_c = c_0$
$S = 0$
for $n = 1 : N$ **do**
  for $i = 1 : m$ **do**
    $c_c = OPERATION_i(c_c)$
    if $\Gamma(c_c) \subseteq X$ **then**
      $S = S + 1$
    end if
  end for
end for
return $\frac{S}{Nm}$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Probability transformation
Possibility transformation
Monte-Carlo methods
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Importance sampling

Wilson and Moral, 1996

Theorem

If $\mathcal{C}$ is connected (i.e., any c, c' are linked by a chain of $OPERATION_i$) then given ϵ, δ there exist K', N' s.t. for all $K \geq K'$ and $N \geq N'$ and c_0:

$$Pr(|BEL_K^N(c_0)| < \epsilon) \geq 1 - \delta$$

- further step: **importance sampling** -> pick samples $c^1, ..., c^N$ according to an "easy to handle" probability distribution P^*
- assign to each sample a weight $w_i = \frac{P(c)}{P^*(c)}$
- if $P(c) > 0$ implies $P^*(c) > 0$ then the average $\frac{\sum_{\Gamma(c^i)\subseteq X} w_i}{N}$ is an unbiased estimator of $Bel(X)$
- try to use P^* as close as possible to the real one
- strategies are proposed to compute $P(\mathcal{C}) = \sum_c P(c)$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Barnett's method
Diagnostic trees
Shafer-Shenoy architecture
Directed Evidential Networks
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Barnett's method
Diagnostic trees
Shafer-Shenoy architecture
Directed Evidential Networks
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Graphical models for belief functions

- to tackle complexity, a number of local computation schemes have been proposed
  - **Barnett's computational scheme**
  - Gordon and Shortliffe's **diagnostic trees**
  - Shafer and Logan's hierarchical evidence
  - **Shafer-Shenoy architecture**
- later on, these developed into graphical models for reasoning with conditional belief functions:
  - Cano et al - propagating uncertainty in directed acyclic networks
  - Xu and Smets - Evidential networks with conditional belief functions
  - Shenoy - graphical representation of valuation-based systems (VBS), called valuation networks
  - Ben Yaghlane and Mellouli - **Directed Evidential Networks**

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Barnett's method
Diagnostic trees
Shafer-Shenoy architecture
Directed Evidential Networks
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Barnett's scheme, 1981

- computations **linear in the size of** Ω if all BFs to combine are simple support focused on singletons or their complements
- simple support function -> as focal elements only A or Ω
- assume we have a Bel_ω with as FEs only $\{\omega, \bar{\omega}, \Omega\}$ for all ω, and we want to combine them
- uses the fact that the plausibility of the combined BF is a function of their input BFs' **commonalities** $Q(A) = \sum_{B \supseteq A} m(B)$:

$$Pl(A) = \sum_{B \subseteq A, B \neq \emptyset} (-1)^{|B|+1} \prod_{\omega \in \Omega} Q_\omega(B)$$

- we get that $Pl(A) = K \left(1 + \sum_{\omega \in A} \frac{Bel_\omega(\omega)}{1 - Bel_\omega(\omega)} - \prod_{\omega \in A} \frac{Bel_\omega(\bar{\omega})}{1 - Bel_\omega(\omega)} \right)$
- the computation of a specific plausibility value $Pl(A)$ is linear in the size of Ω (only elements of A and not subsets are involved)
- however, the number of events A themselves is still exponential

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

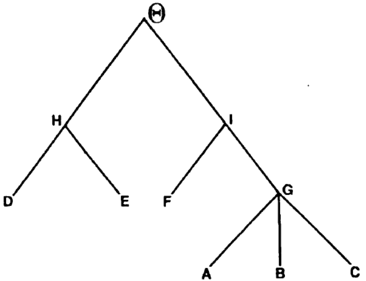
13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Gordon and Shortliffe's scheme

based on diagnostic trees

- they are interested in computing degrees of belief **only for events forming a hierarchy (diagnostic tree)**
- (in some applications certain events are not relevant, e.g. classes of diseases)



- combine simple support functions focused on or against the nodes
- produces good approximations, unless evidence is highly conflicting

Gordon and Shortliffe's scheme

based on diagnostic trees

- however, intersection of complements produces FEs not in the tree
- approximated algorithm:
  1. first we combine all simple functions focussing on the node events (by Dempster's rule)
  2. then, we successively (working down the tree) combine those focused on the complements of the nodes
  3. tricky bit: when we do that, we replace each intersection of FEs with the smallest node in the tree that contains it
- results depends on the order of the combination in phase 2
- again approximation can be poor, also no degrees of belief are assigned to complements of nodes
- therefore, we cannot compute their plausibilities!

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Barnett's method
Diagnostic trees
Shafer-Shenoy architecture
Directed Evidential Networks
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Shafer-Shenoy architecture

Qualitative Conditional Independence

- uses **qualitative Markov trees**, which generalise both diagnostic trees and causal trees (Pearl). Extend Pearl's idea to BFs
- partitions $\Psi_1, ..., \Psi_n$ of a frame are **qualitatively conditionally independent** (QCI) given the partition Ψ if

$$P \cap P_1 \cap ... \cap P_n \neq \emptyset$$

  whenever $P \in \Psi$, $P_i \in \Psi_i$ and $P \cap P_i \neq \emptyset$ for all i
- example: $\{\theta_1\} \times \{\theta_2\} \times \Theta_3$ and $\Theta_1 \times \{\theta_2\} \times \{\theta_3\}$ are QCI on $\Theta_1 \times \Theta_2 \times \Theta_3$ given $\Theta_1 \times \{\theta_2\} \times \Theta_3$ for all $\theta_i \in \Theta_i$
- does not involve probability, but only logical independence
- stochastic conditional independence does imply the above
- if two BFs Bel_1 and Bel_2 are carried by partitions Ψ_1, Ψ_2 which are QCI given Ψ then

$$(Bel_1 \oplus Bel_2)_\Psi = (Bel_1)_\Psi \oplus (Bel_2)_\Psi$$

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Barnett's method
Diagnostic trees
Shafer-Shenoy architecture
Directed Evidential Networks
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Qualitative Markov trees

Shafer-Shenoy architecture

- given a tree, deleting a node and all incident edges yields a forest - denote the collection of nodes of the j-th subtree by $\alpha_m(j)$

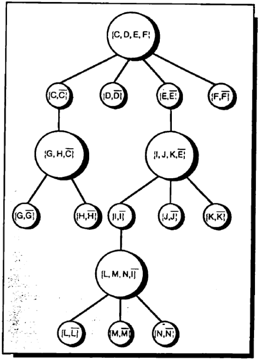


Figure 8. The enlarged qualitative Markov tree for the car that won't start.

- a **qualitative Markov tree** is a tree of partitions s.t. for every node i the minimal refinements of partitions in $\alpha_m(j)$ for $j = 1, ..., k$ are QCI given Ψ_i
- a Bayesian causal tree becomes a qualitative Markov tree whenever we associate each node B with the partition Ψ_B associated with the random variable v_B
- a QMT remains such if we insert between parent and child their common refinement
- can also be constructed from a diagnostic tree (left), same interpolation property holds

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Barnett's method
Diagnostic trees
Shafer-Shenoy architecture
Directed Evidential Networks
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Propagating belief functions

on qualitative Markov trees

- each BF to combine has to be carried by a partition in the tree
- idea: **replace Dempster's combination over the whole frame with multiple implementations over partitions**
- a **processor located at each node** Ψ_i combines BFs using Ψ_i as a frame and projects BFs to its neighbours

  1. send Bel_i to its neighbours
  2. whenever it gets a new input, computes $(Bel^T)_{\Psi_i} \leftarrow (\oplus\{(Bel_x)_{\Psi_i} : x \in N(i)\} \oplus Bel_i)_{\Psi_i}$
  3. computes $Bel_{i,y} \leftarrow (\oplus\{(Bel_x)_{\Psi_i} : x \in N(i) \setminus \{y\}\} \oplus Bel_i)_{\Psi_y}$ and sends it to its neighbour y, for each neighbour

- inputting new BFs in the tree can take place asynchronously
- final result of each processor: **coarsening to that partition of the combination of all inputted BFs**: $(\oplus_{j \in J} Bel_j)_{\Psi_i}$

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Barnett's method
Diagnostic trees
Shafer-Shenoy architecture
Directed Evidential Networks
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Graphical representation

Shafer-Shenoy architecture

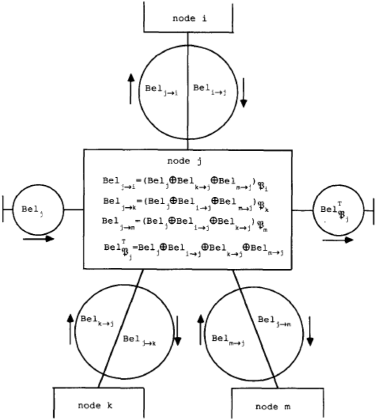


Figure 9. A Typical Processor (with Three Neighbors)

- total time to reach equilibrium is proportional to the tree's diameter

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. **Propagation**
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - **Directed Evidential Networks**
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Barnett's method
Diagnostic trees
Shafer-Shenoy architecture
Directed Evidential Networks
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Directed evidential networks

Ben Yaghlane and Mellouli, 2008

- Evidential networks with conditional belief functions (ENC) were originally proposed by Xu and Smets for the propagation of beliefs
  - (Dempster's conditioning is adopted)
- ENCs contain a directed acyclic graph with conditional beliefs defined in a different manner from conditional probabilities in Bayesian networks (BNs)
  - edges represent the existence of a conditional BF (no form of independence assumed)
  - initially defined only for binary (conditional) relationships
- Ben Yaghlane and Mellouli generalised ENCs to any number of nodes - directed evidential network (DEVN)
  - a directed acyclic graph (DAG) in which directed arcs describe the conditional dependence relations expressed by conditional BFs for each node given its parents
  - new observations introduced in the network are represented by belief functions allocated to some nodes

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Barnett's method
Diagnostic trees
Shafer-Shenoy architecture
Directed Evidential Networks
Decisions
BFs on reals
Sister theories
Advances
Toolbox

Directed evidential networks

Ben Yaghlane and Mellouli, 2008

- problem: given n BFs $Bel_1, ..., Bel_n$ over $X_1, ..., X_n$ we seek the marginal on X_i of their joint belief function
- uses the generalised Bayesian theorem (GBT) to compute the posterior $Bel(x|y)$ given the conditional $Bel(y|x)$
- the marginal is computed for each node by combining all the messages received from its neighbors and its own prior belief:

$$Bel^X = Bel_0^X \oplus Bel_{Y \to X}, \quad Bel_{Y \to X}(x) = \sum_{y \subseteq \Theta_Y} m_0(y) Bel(x|y)$$

  where $Bel(x|y)$ is given by GBT
- another application of the message-passing idea to belief functions
- propose a simplified scheme for simply directed networks
- extension to DEVNs by first transforming them to binary join trees

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Decision making with belief functions

An overview

- natural application of belief function representation of uncertainty
- problem: **selecting an act f from an available list $\mathcal{F}$** (making a "'decision'), which optimises a certain objective function
- various approaches to decision making
  - decision making in the TBM is based on expected utility via pignistic transform
  - Strat has proposed something similar in his "cloaked carnival wheel" scenario
  - **generalised expected utility** [Gilboa] based on classical expected utility theory [Savage,von Neumann]
- a lot of interest in **multicriteria decision making** (based on a number of attributes)

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Expected utility approach

Decision making under uncertainty

- a decision problem can be formalized by defining:
  - a set Ω of **states of the world**;
  - a set $\mathcal{X}$ of **consequences**;
  - a set $\mathcal{F}$ of **acts**, where an act is a function $f : \Omega \to \mathcal{X}$
- let $\succcurlyeq$ be a **preference relation** on $\mathcal{F}$, such that $f \succcurlyeq g$ means that f is at least as desirable as g
- Savage (1954) has showed that $\succcurlyeq$ verifies some rationality requirements iff there exists a **probability measure** P on Ω and a **utility function** $u : \mathcal{X} \to \mathbb{R}$ s.t.

$$\forall f, g \in \mathcal{F}, \quad f \succcurlyeq g \Leftrightarrow \mathbb{E}_P(u \circ f) \geq \mathbb{E}_P(u \circ g)$$

  where $\mathbb{E}_P$ denotes the expectation w.r.t. P
- P and u are unique up to a positive affine transformation
- does that mean that basing decisions on belief functions is irrational?

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Outline

1 Rationale
2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- **Decision making in the TBM**
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Decision making in the TBM

Expected utility using the pignistic probability

- in the TBM, decision making is done by **maximising the expected utility of actions based on the pignistic transform**
- (as opposed to computing upper and lower expected utilities directly from (*Bel*, *Pl*) via Choquet integral, as we will see later)
- the set of possible actions $\mathcal{F}$ and the set Ω of possible outcomes are distinct, and the utility function is defined on $\mathcal{F} \times \Omega$
- Smets proves the necessity of the pignistic transform by maximizing

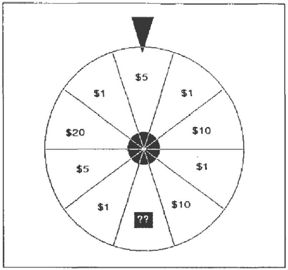
$$E[u] = \sum_{\omega \in \Omega} u(f, \omega) Pign(\omega)$$

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Strat's decision apparatus

[UAI 1990]



- Strat's decision apparatus is based on computing intervals of expected values
- assumes that the decision frame Ω is itself a set of scalar values (e.g. dollar values, see left) - does not distinguish between utilities and elements of Ω (returns)

- .. so that an **expected value interval** can be computed: $E(\Omega) = [E_*(\Omega), E^*(\Omega)]$, where

$$E_*(\Omega) \doteq \sum_{A \subseteq \Omega} \inf(A) m(A), \;\; E^*(\Omega) \doteq \sum_{A \subseteq \Omega} \sup(A) m(A)$$

- not good enough to make a decision, e.g.: should we pay a 6\$ ticket when the expected interval is $[5\$, 8\$]$?

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Strat's decision apparatus

A probability of favourable outcome

- Strat identifies ρ as the probability that the value assigned to the hidden sector is the one the player would choose
- $1 - \rho$ is the probability that the sector is chosen by the carnival hawker

Theorem

The expected value of the mass function of the wheel is
$E(\Omega) = E_*(\Omega) + \rho(E^*(\Omega) - E_*(\Omega))$

- to decide whether to play the game we only need to assess ρ
- basically, this amounts to a specific probability transform (like the pignistic one)
- Lesh, 1986 had also proposed a similar approach

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Savage's axioms

- Savage has proposed 7 axioms, 4 of which are considered as meaningful (the others are rather technical)
- let us examine the first two axioms:
- Axiom 1: $\succcurlyeq$ is a total preorder (complete, reflexive and transitive)
- Axiom 2 [**Sure Thing Principle**]. Given $f, h \in \mathcal{F}$ and $E \subseteq \Omega$, let fEh denote the act defined by

$$(fEh)(\omega) = \begin{cases} f(\omega) & \text{if } \omega \in E \\ h(\omega) & \text{if } \omega \notin E \end{cases}$$

- then the Sure Thing Principle states that $\forall E, \forall f, g, h, h'$,

$$fEh \succcurlyeq gEh \Rightarrow fEh' \succcurlyeq gEh'$$

- this axiom seems reasonable, but it is **not verified empirically**!

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Ellsberg's paradox

- suppose you have an urn containing 30 red balls and 60 balls, either black or yellow. Consider the following gambles:
  - f_1: you receive 100 euros if you draw a **red ball**
  - f_2: you receive 100 euros if you draw a **black ball**
  - f_3: you receive 100 euros if you draw a **red or yellow ball**
  - f_4: you receive 100 euros if you draw a **black or yellow ball**
- in this example $\Omega = \{R, B, Y\}$, $f_i : \Omega \to \mathbb{R}$ and $\mathcal{X} = \mathbb{R}$
- the four acts are the mappings in the left table
- empirically it is observed that most people strictly prefer f_1 to f_2, but they strictly prefer f_4 to f_3

	R	B	Y
f_1	100	0	0
f_2	0	100	0
f_3	100	0	100
f_4	0	100	100

Now, pick $E = \{R, B\}$: by definition

$$f_1\{R,B\}0 = f_1, \qquad f_2\{R,B\}0 = f_2$$
$$f_1\{R,B\}100 = f_3, \qquad f_2\{R,B\}100 = f_4$$

- since $f_1 \succcurlyeq f_2$, i.e. $f_1\{R,B\}0 \succcurlyeq f_2\{R,B\}0$ the Sure Thing principle would imply $f_1\{R,B\}100 \succcurlyeq f_2\{R,B\}100$, i.e., $f_3 \succcurlyeq f_4$
- empirically **the Sure Thing Principle is violated!**

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Gilboa's theorem

- Gilboa (1987) proposed a modification of Savage's axioms with, in particular, a **weaker form of Axiom 2**
- a preference relation $\succcurlyeq$ meets these weaker requirements iff there exists a **(non necessarily additive) measure** μ and a **utility function** $u : \mathcal{X} \to \mathbb{R}$ such that

$$\forall f, g \in \mathcal{F}, \quad f \succcurlyeq g \Leftrightarrow C_\mu(u \circ f) \geq C_\mu(u \circ g),$$

where C_μ is the **Choquet integral**, defined for $X : \Omega \to \mathbb{R}$ as

$$C_\mu(X) = \int_0^{+\infty} \mu(X(\omega) \geq t)dt + \int_{-\infty}^0 [\mu(X(\omega) \geq t) - 1]dt.$$

- given a belief function *Bel* on Ω and a utility function u, this theorem supports **making decisions based on the Choquet integral** of u with respect to *Bel* or *Pl*

Lower and upper expected utilities

- for finite Ω, it can be shown that

$$C_{Bel}(u \circ f) = \sum_{B \subseteq \Omega} m(B) \min_{\omega \in B} u(f(\omega))$$

$$C_{Pl}(u \circ f) = \sum_{B \subseteq \Omega} m(B) \max_{\omega \in B} u(f(\omega))$$

- let $\mathcal{P}(Bel)$ as usual be the set of probability measures P compatible with Bel, i.e., such that $Bel \leq P$. Then, it can be shown that

$$C_{Bel}(u \circ f) = \min_{P \in \mathcal{P}(Bel)} \mathbb{E}_P(u \circ f) = \underline{\mathbb{E}}(u \circ f)$$

$$C_{Pl}(u \circ f) = \max_{P \in \mathcal{P}(Bel)} \mathbb{E}_P(u \circ f) = \overline{\mathbb{E}}(u \circ f)$$

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Decision making

Strategies

- for each act f we have two expected utilities $\underline{\mathbb{E}}(f)$ and $\overline{\mathbb{E}}(f)$. How do we make a decision?
- possible decision criteria based on interval dominance:
  1. $f \succcurlyeq g$ iff $\underline{\mathbb{E}}(u \circ f) \geq \overline{\mathbb{E}}(u \circ g)$ (**conservative** strategy)
  2. $f \succcurlyeq g$ iff $\underline{\mathbb{E}}(u \circ f) \geq \underline{\mathbb{E}}(u \circ g)$ (**pessimistic** strategy)
  3. $f \succcurlyeq g$ iff $\overline{\mathbb{E}}(u \circ f) \geq \overline{\mathbb{E}}(u \circ g)$ (**optimistic** strategy)
  4. $f \succcurlyeq g$ iff

     $$\alpha\underline{\mathbb{E}}(u \circ f) + (1-\alpha)\overline{\mathbb{E}}(u \circ f) \geq \alpha\underline{\mathbb{E}}(u \circ g) + (1-\alpha)\overline{\mathbb{E}}(u \circ g)$$

     for some $\alpha \in [0, 1]$ called a pessimism index (**Hurwicz criterion**)
- the conservative strategy yields only a partial preorder: f and g are not comparable if $\underline{\mathbb{E}}(u \circ f) < \overline{\mathbb{E}}(u \circ g)$ and $\underline{\mathbb{E}}(u \circ g) < \overline{\mathbb{E}}(u \circ f)$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities
BFs on reals
Sister theories
Advances
Toolbox

Ellsberg's paradox revisited

- going back to the example, the evidence naturally translates into a belief function
- we have $m(\{R\}) = 1/3$, $m(\{B, Y\}) = 2/3$
- we can then compute lower and upper expected utilities for each action:

	R	B	Y	$\underline{\mathbb{E}}(u \circ f)$	$\overline{\mathbb{E}}(u \circ f)$
f_1	100	0	0	u(100)/3	u(100)/3
f_2	0	100	0	u(0)	u(200)/3
f_3	100	0	100	u(100)/3	u(100)
f_4	0	100	100	u(200)/3	u(200)/3

- the observed behavior ($f_1 \succcurlyeq f_2$ and $f_4 \succcurlyeq f_3$) is explained by the pessimistic strategy

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Allocations of probabilitity
Random sets
Random closed intervals
Sister theories
Advances
Toolbox

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Allocations of probabilitity
Random sets
Random closed intervals
Sister theories
Advances
Toolbox

Continuous formulations

of the theory of belief functions

- in the original formulation by Shafer [1976], belief functions are defined on finite sets only
- need for generalising this to arbitrary domains has been recognised at an early stage
- main approaches to continuous formulation presented here:
  - Shafer's **allocations of probability** [1982]
  - belief functions as **random sets** [Nguyen]
  - **belief functions on Borel intervals** of the real line [Strat,Smets]
- other approaches, with limited (so far) impact
  - generalised evidence theory
  - MV algebras
  - several others

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Allocations of probability

Shafer, 1979

- every belief function can be represented as an **allocation of probability**, i.e., $\cap$-homomorphisms into positive and completely additive probability algebra (deduced from the integral representation due to Choquet)
  - for every belief function *Bel* defined on a class of events $\mathcal{E} \subseteq 2^\Omega$ there exists a complete Boolean algebra $\mathcal{M}$, a positive measure μ and an allocation of probability ρ between $\mathcal{E}$ and $\mathcal{M}$ such that $Bel = \mu \circ \rho$
- two regularity conditions for a belief function over an infinite domain are considered: **continuity** and **condensability**
- **canonical continuous extensions** of belief functions to arbitrary power sets can be introduced by allocation of probability
- the approach shows significant resemblance with the notions of inner measure and extension of capacities [Honda]

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Allocations of probabilitity
Random sets
Random closed intervals
Sister theories
Advances
Toolbox

Continuity and condensability

Shafer's allocations of probability

- $\mathcal{E} \subset 2^\Theta$ is a multiplicative subclass of 2^Θ if $A \cap B \in \mathcal{E}$ for all $A, B \in \mathcal{E}$
- a function $Bel : \mathcal{E} \to [0,1]$ such that $Bel(\emptyset) = 0$, $Bel(\Theta) = 1$ and Bel is monotone of order ∞ is a belief function
    - equally, an upper probability (plausibility) function is alternating of order ∞ ($\geq$ is exchanged with $\leq$)
- a BF on 2^Θ is **continuous** if $Bel(\cap_i A_i) = \lim_{i\to\infty} Bel(A_i)$ for every decreasing sequence of A_is. A BF on a multiplicative subclass $\mathcal{E}$ is continuous if it can be extended to a continuous one on 2^Θ
    - continuity arises from partial beliefs on 'objective' probabilities
- a BF on 2^Θ is **condensable** if $Bel(\cap \mathcal{A}) = \inf_{A \in \mathcal{A}} Bel(A)$ for every downward net $\mathcal{A}$ in 2^Θ. A BF on a multiplicative subclass $\mathcal{E}$ is condensable if it can be extended to a condensable one on 2^Θ
    - a downward net is such that given two elements there is always an element subset of their intersection
- condensability is restrictive, but related to Dempster's rule

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Allocations of probabilitity
Random sets
Random closed intervals
Sister theories
Advances
Toolbox

Choquet's representation

Shafer's allocations of probability

- Choquet's integral representation says that every belief function can be represented by allocation of probability
- $r : \mathcal{E} \rightarrow \mathcal{F}$ is a $\cap$-homomorphism if it preserves $\cap$

Choquet's theorem

For every BF *Bel* on a multiplicative subclass $\mathcal{E}$ of 2^Θ, $\exists$ a set $\mathcal{X}$ and an algebra $\mathcal{F}$ of its subsets, a finitely additive probability measure μ on $\mathcal{F}$, and a $\cap$-homomorphism $r : \mathcal{E} \rightarrow \mathcal{F}$ such that $Bel = \mu \circ r$.

- if we replace the measure space $(\mathcal{X}, \mathcal{F}, \mu)$ with a probability algebra (a complete Boolean algebra $\mathcal{M}$ with a completely additive prob measure μ) we get

Allocation of probability

For every BF *Bel* on a multiplicative subclass $\mathcal{E}$ of 2^Θ, $\exists$ an allocation of probability $\rho : \mathcal{E} \rightarrow \mathcal{M}$ such that $Bel = \mu \circ \rho$.

- non-zero elements of $\mathcal{M}$ can be thought of as focal elements

Canonical extension

Shafer's allocations of probability

Theorem

*a BF on a multiplicative subclass $\mathcal{E}$ can always be extended to a belief function on 2^Θ by **canonical extension***

$$\overline{Bel}(A) \doteq \sup_{n\geq 1, A_1,\ldots,A_n\in\mathcal{E}} \sum \left\{ (-1)^{|I|+1} Bel(\cap_{i\in I} A_i) | \emptyset \neq I \subset \{1,...,n\} \right\}$$

- proof is based on the existence of an allocation for the extension
- note the similarity with the superadditivity axiom
- also related to **inner measures**, which provide approximate belief values for subsets not in a sigma-algebra
- $\overline{Bel}$ is the minimal such extension
- what about evidence combination? **condensability** ensures that the Boolean algebra $\mathcal{M}$ represents intersection properly for arbitrary (not just finite) collections $\mathcal{B}$ of subsets:

$$\rho(\cap\mathcal{B}) = \bigwedge_{B\in\mathcal{B}} \rho(B) \quad \forall\mathcal{B} \subset 2^\Omega$$

- allows us to imagine Dempster's combinations of infinitely many belief functions

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Allocations of probabilitity
Random sets
Random closed intervals
Sister theories
Advances
Toolbox

Random sets

to extend belief functions to arbitrary domains

- the notion of condensability has been studied by Nguyen for upper probabilities generated by random sets too [Nguyen 1978]
- efforts directed at a general theory on arbitrary domains
- for finite random sets (i.e. with a finite number of focal elements), under independence of variables **Dempster's rule can be applied**:

$$(\mathcal{F}, m) = \left\{ A_{i_1,\ldots,i_d} = \times_{j=1}^{d} A_{i_j},\ m_{i_1,\ldots,i_d} = m_{i_1} \cdot \cdots \cdot m_{i_d} \right\}$$

- for dependent sources Fetz and Oberguggenberger have proposed an "unknown interaction" model
- for infinite random sets Alvarez (see p-boxes later) a **Monte-Carlo sampling** method

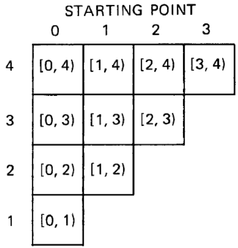
Outline

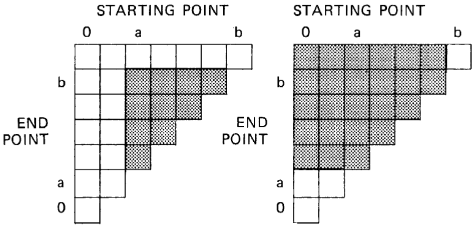
1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Continuous belief functions

Strat's approach

- idea: take a real interval I and split it into N bits
- take as frame of discernment the **set of possible intervals** with these extreme: $[0,1)$, $[0,2)$, $[1,4]$ etc
- a belief function there has $\sim N^2/2$ possible focal elements, so that its mass lives on a triangle (left), and one can compute **belief and plausibility by integration** (right)





(a) $Spt([a,b))$ (b) $Pls([a,b))$

Belief functions for the working scientist

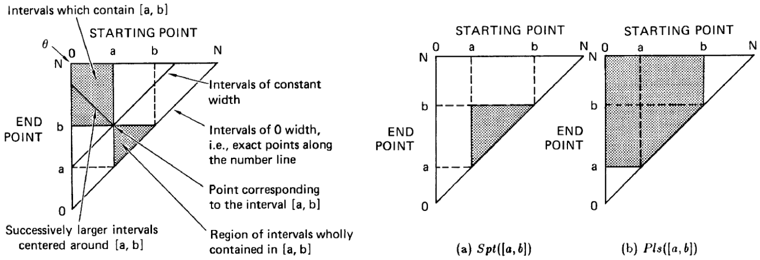
F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Allocations of probabilitity
Random sets
Random closed intervals
Sister theories
Advances
Toolbox

Continuous belief functions

Strat's approach

- this trivially generalises to **all arbitrary intervals of** I (below)



$$Bel([a,b]) = \int_a^b \int_x^b m(x,y)dydx, \quad Pl([a,b]) = \int_0^b \int_{\max(a,x)}^N m(x,y)dydx$$

- **Dempster's rule generalises** as $Bel_1 \oplus Bel_2([a,b]) =$

$$\frac{1}{K}\int_0^a \int_b^N \Big[m_1(x,b)m_2(a,y) + m_2(x,b)m_1(a,y) + m_1(a,b)m_2(x,y) + m_2(a,b)m_1(x,y)\Big]dydx$$

Continuous belief functions

on the Borel algebra of intervals

- a pretty much identical approach is followed by Smets
- allows us to define a **continuous pignistic PDF** as

$$Bet(a) \doteq \lim_{\epsilon \to 0} \int_0^a dx \int_{a+\epsilon}^1 \frac{m(x,y)}{y-x} dy$$

- can be easily extended to the real line, by considering **belief functions defined on the Borel σ-algebra** of subsets of $\mathbb{R}$ generated by the collection $\mathcal{I}$ of closed intervals
- the theory provides a way of **building a continuous belief function from a pignistic density**, by applying the least commitment principle and assuming unimodal pignistic PDFs

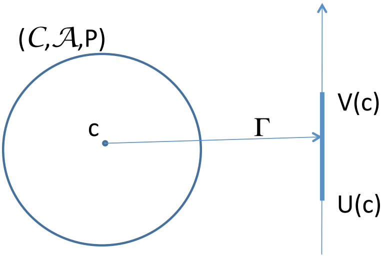
$$Bel(s) = -(s - \bar{s}) \frac{dBet(s)}{ds}$$

where $\bar{s}$ is such that $Bet(s) = Bet(\bar{s})$
- example: $Bet(x) = \mathcal{N}(x, \mu, \sigma)$ is normal $\rightarrow$ $Bel(y) = \frac{2y}{\sqrt{2\pi}} e^{-y^2}$, where $y = (x-\mu)/\sigma$

Continuous belief functions

induced by random closed intervals

- formal setting:
- let (U, V) be a two-dimensional random variable from $(\mathcal{C}, \mathcal{A}, P)$ to $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ such that $P(U \leq V) = 1$ and $\Gamma(c) = [U(c), V(c)] \subseteq \mathbb{R}$

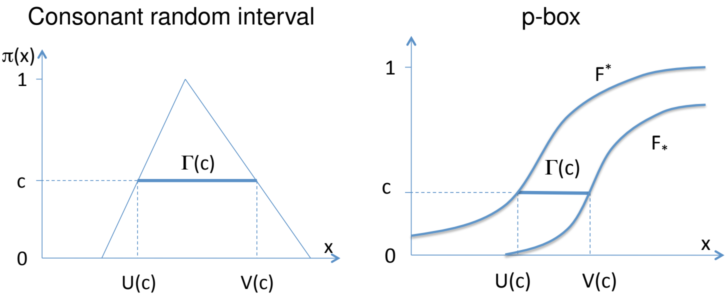


- this setting defines a **random closed interval**, which induces a belief function on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by

$$Bel(A) = P([U, V] \subseteq A), \quad \forall A \in \mathcal{B}(\mathbb{R})$$

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Allocations of probabilitity
Random sets
Random closed intervals
Sister theories
Advances
Toolbox

Special cases
of random closed intervals



- special cases
- a fuzzy set on the real line induces a mapping to a collection of nested intervals, parameterised by the level c
- a p-box, i.e, upper and lower bounds to a cumulative distribution function (see later) also induces a family of intervals

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

Relationships with other theories of uncertainty

- belief functions have meaningful relationships with a number of other theories of uncertainty
- here we briefly recall the most significant ones:
  - **imprecise probabilities** [Walley]
  - **credal sets** [Levi]
  - **possibility theory** [Zadeh, Dubois & Prade]
  - belief functions **on fuzzy sets** [Zadeh & others]
  - **p-boxes** [Ferson]
- others we will not touch here for lack of time:
  - probability intervals [Moral]
  - monotone capacities
  - fuzzy measures
  - rough sets [Pawlak]
  - probabilistic and modal logic

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

Outline

1 Rationale
2 Basic notions
 Belief functions
 Dempster's rule
 Families of frames
3 Semantics
 Lower probabilities
 Credal sets
 Set functions
 Generalised probabilities
 Random sets
4 Inference
 Dempster's approach
 Likelihood-based
 From preferences
5 Conditioning
 Dempster's conditioning
 Lower conditional envelopes
 Unnormalised vs geometric conditioning
 Conditional events as equivalence classes
 Generalised Bayes theorem
 The total belief theorem
6 Computation
 Probability transformation
 Possibility transformation
 Monte-Carlo methods
7 Propagation
 Barnett's method
 Diagnostic trees
 Shafer-Shenoy architecture
 Directed Evidential Networks
8 Decisions
 Decision making in the TBM
 Strat's decision apparatus
 Upper and lower expected utilities
9 BFs on reals
 Allocations of probabilitity
 Random sets
 Random closed intervals
10 Sister theories
 Imprecise probability
 Fuzzy sets and possibility
 p-Boxes
11 Advances
 Matrix representation
 Geometry
 Distances
 Combinatorics
 Uncertainty measures
12 Toolbox
 Classification
 Ranking aggregation
13 Applications
 A brief survey
 Climate change
 Pose estimation
14 Future trends

Lower probabilities and credal sets

in Imprecise Probability

- a **lower probability** $\underline{P}$ is a function from 2^Ω, the power set of Ω, to the unit interval $[0, 1]$
- with any lower probability $\underline{P}$ is associated a dual upper probability function $\overline{P}$, defined for any $A \subseteq \Omega$ as $\overline{P}(A) = 1 - \underline{P}(A^c)$
- with any lower probability $\underline{P}$ we can associate a closed convex set (**credal set** [Levi])

$$\mathcal{P}(\underline{P}) = \Big\{ P : P(A) \geq \underline{P}(A), \forall A \subseteq \Omega \Big\}$$

  of probability measures P which dominate $\underline{P}$
- note that not all convex sets of probabilities can be described by merely focusing on events [Walley]

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

Coherent lower probabilities

in Imprecise Probability

- a lower probability $\underline{P}$ is called 'consistent' if $\mathcal{P}(\underline{P}) \neq \emptyset$ and 'tight' if
$$\inf_{p \in \mathcal{P}(\underline{P})} P(A) = \underline{P}(A)$$
- (respectively $\underline{P}$ 'avoids sure loss" and $\underline{P}$ is 'coherent' in Walley's terminology)
- consistency means that the lower bound constraints $\underline{P}(A)$ can indeed be satisfied by some probability measure
- tightness indicates that $\underline{P}$ is the lower envelope on subsets of $\mathcal{P}(\underline{P})$
- belief functions are indeed a **special type of coherent lower probabilities**, which in turn can be seen as a special class of *lower previsions*
- having said that, the two approaches depart on the fundamental epistemic representation of evidence

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

Outline

1 Rationale
2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- **Fuzzy sets and possibility**
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

Possibility theory

and consonant belief functions

- a **possibility measure** on a domain Ω is a function $Pos : 2^\Omega \to [0,1]$ such that $Pos(\emptyset) = 0$, $Pos(\Omega) = 1$ and

$$Pos\left(\bigcup_i A_i\right) = \sup_i Pos(A_i)$$

  for any family $\{A_i | A_i \in 2^\Omega, i \in I\}$ where I is an arbitrary set index
- it is uniquely characterized by a **membership function** or "possibility distribution" $\pi(x) \doteq Pos(\{x\})$, as $Pos(A) = \sup_{x\in A} \pi(x)$
  - $Nec(A) = 1 - Pos(A^c)$ is called **necessity measure**
- call "plausibility assignment" pl the restriction of the plausibility function to singletons $pl(x) = Pl(\{x\})$ - then [Shafer]:
- ***Bel*** **is a necessity measure iff** ***Bel*** **is consonant**
- the membership function coincides with the plausibility assignment
  - according to Shafer, the difference between possibilities and consonant BFs is just in the language used

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

Belief functions on fuzzy sets

1. a finite fuzzy set is equivalent to a consonant belief function
2. **generalisations of belief functions defined on fuzzy sets** have been proposed [Zadeh]

- basic idea: belief measures generalised on fuzzy sets as follows:

$$Bel(X) = \sum_{A \in \mathcal{M}} I(A \subseteq X) m(A)$$

  where X is a fuzzy set defined on Ω, m is a mass function defined on the collection of fuzzy sets on Ω
- $I(A \subseteq X)$ is a measure of how much the fuzzy set A is included in the fuzzy set X
- various measures of inclusion in $[0, 1]$ can be proposed:
  - Lukasiewicz: $I(x, y) = \min\{1, 1 - x - y\}$ [Ishizuka]
  - Kleene-Dienes: $I(x, y) = \max\{1 - x, y\}$ [Yager]
- from which one can get: $\mathcal{I}(A, B) = \bigwedge_{x \in \Theta} I(A(x), B(y))$ [Wu 2009]

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
   - Imprecise probability
   - Fuzzy sets and possibility
   - p-Boxes
11. Advances
   - Matrix representation
   - Geometry
   - Distances
   - Combinatorics
   - Uncertainty measures
12. Toolbox
   - Classification
   - Ranking aggregation
13. Applications
   - A brief survey
   - Climate change
   - Pose estimation
14. Future trends

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

Probability-boxes (p-Boxes)

Classes of cumulative distribution functions

- a **probability box** or **p-box** [Ferson and Hajagos] $\langle \underline{F}, \overline{F} \rangle$ is a class of cumulative distribution functions (CDFs)

$$\langle \underline{F}, \overline{F} \rangle = \left\{ F \ CDF : \underline{F} \leq F \leq \overline{F} \right\}$$

  delimited by upper and lower CDF bounds $\underline{F}$ and $\overline{F}$
- represents the epistemic uncertainty about the CDF of a random variable
- every RS generates a unique p-box whose CDFs are all those consistent with the evidence:

$$\underline{F}(x) = Bel((-\infty, x]), \quad \overline{F}(x) = Pl((-\infty, x])$$

- every p-box generates an infinite RS with as focal elements the following infinite collection of intervals of $\mathbb{R}$:

$$\left\{ [\overline{F}^{-1}(\alpha), \underline{F}^{-1}(\alpha)] \ \forall \alpha \in [0, 1] \right\}$$

  where $\overline{F}^{-1}(\alpha) \doteq \inf\{\overline{F}(x) \geq \alpha\}$, $\underline{F}^{-1}(\alpha) \doteq \inf\{\underline{F}(x) \geq \alpha\}$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

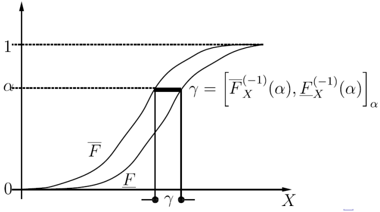
Approximate computations

for random sets

- in an infinite RS the computation of the integral $Bel(A) = \int_{c \in \mathcal{C}} I[\Gamma(c) \subset A] dP(c)$ (or those for $Pl(A)$, etc) is not trivial
- we can use the representation of infinite RSs provided by p-boxes, with set of focal elements

$$\mathcal{F} = \left\{ \gamma = [\overline{F}^{-1}(\alpha), \underline{F}^{-1}(\alpha)] \ \forall \alpha \in [0,1] \right\}$$

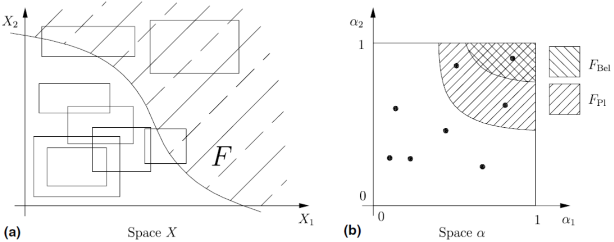
- if α has its own CDF F_α, we can sample from it
- after sampling FEs from the RS, we can compute belief and plausibility integrals



Approximate combination of random sets

α-representation

- we can also calculate the combination of the sampled FEs
- if d random sets to combine, FEs are vectors of indices from all constituing RS: $\alpha = [\alpha_1, ..., \alpha_d] \in (0, 1]^d$



- suppose a copula C is defined on the unit hypercube (i.e. a prob distribution whose marginals are uniform)..

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes
Advances
Toolbox
Applications

Approximate combination of random sets

Monte-Carlo approach

- ..we can use it to compute the desired integrals, i.e.

$$P_\Gamma(G) = \int_{\alpha \in G} dC(\alpha)$$

- if input RS are independent, these integrals decompose as, e.g.

$$\mathrm{Bel}_{(\mathscr{F},P_\Gamma)}(F) = \underbrace{\int_{0^+}^{1} \cdots \int_{0^+}^{1}}_{d\text{-times}} I[[\alpha_1, \ldots, \alpha_d] \in F_{\mathrm{Bel}}] \, \mathrm{d}C(\alpha_1, \ldots, \alpha_d)$$

- Monte-Carlo approach [Alvarez 2006] - for $j = 1, \ldots, n$:
  1. randomly extract a sample α_j from the copula C
  2. form the corresponding focal element $A_j = \times_{i=1,\ldots,d} \gamma_i^d$
  3. assign to it mass $m(A_j) = \frac{1}{n}$
- can prove that such approximation converges as $n \to +\infty$ almost surely to the actual random set

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

The mathematics of belief functions

- belief functions are rather complex mathematical objects, therefore:
  - have links with a number of fields of (applied) mathematics
  - lead to interesting generalisations of standard results of classical probability (e.g. Bayes' theorem, total probability)
- **matrix representation**
- **geometric approach** to uncertainty [Cuzzolin]
- measuring **distances** [Jousselme et al]
- algebra of frames [Kohlas]
- abstract independence, Boolean algebras and matroids [Cuzzolin]
- **Moebius transforms**
- entropy and other **measures of uncertainty** [Yager, Klir, Harmanec]

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. **Advances**
    - **Matrix representation**
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

Matrix representation

Linear algebra

- given an ordering of the subsets of Ω mass, belief, and plausibility functions can be represented as vectors $\mathbf{m}$, $\mathbf{bel}$ and $\mathbf{pl}$
- various operations with belief functions can be expressed via vectors and matrices
- negation ($\overline{m}(A) = m(\overline{A})$): $\overline{\mathbf{m}} = J\mathbf{m}$ where J is the matrix whose inverse diagonal is made of 1s
- belief value: $\mathbf{b} = BfrM\mathbf{m}$, where $BfrM(A, B) = 1$ iff $B \subseteq A$ and 0 otherwise
- $BfrM_{i+1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \otimes BfrM_i$ where $\otimes$ is the Kronecker product
- other transformation matrices for Moebius inversion can be defined:

$$MfrB = BfrM^{-1}, \ QfrM = JBfrMJ, \ MfrQ = JBfrM^{-1}J$$

- normalised BFs and plausibilities: $\mathbf{Bel} = \mathbf{b} - b(\emptyset)\mathbf{1}$, $\mathbf{pl} = \mathbf{1} - J\mathbf{b}$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

Fast Moebius Transform

- to efficiently compute Moebius transforms (e.g. from *Bel* to *m*)

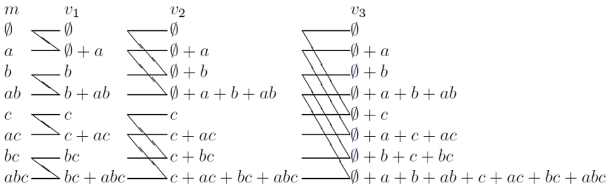


Figure 1: Detail of the FMT when $\Omega = \{a, b, c\}$. The symbol a denotes $m(a)$, ab denotes $m(a, b)$ etc...

- can also be computed in matrix form as $BfrM = M_3 \cdot M_2 \cdot M_1$, where

$$\mathbf{M_1} = \begin{bmatrix} 1 & . & . & . & . & . & . & . \\ 1 & 1 & . & . & . & . & . & . \\ . & . & 1 & . & . & . & . & . \\ . & . & 1 & 1 & . & . & . & . \\ . & . & . & . & 1 & . & . & . \\ . & . & . & . & 1 & 1 & . & . \\ . & . & . & . & . & . & 1 & . \\ . & . & . & . & . & . & 1 & 1 \end{bmatrix} \mathbf{M_2} = \begin{bmatrix} 1 & . & . & . & . & . & . & . \\ . & 1 & . & . & . & . & . & . \\ 1 & . & 1 & . & . & . & . & . \\ . & 1 & . & 1 & . & . & . & . \\ . & . & . & . & 1 & . & . & . \\ . & . & . & . & . & 1 & . & . \\ . & . & . & . & 1 & . & 1 & . \\ . & . & . & . & . & 1 & . & 1 \end{bmatrix} \mathbf{M_3} = \begin{bmatrix} 1 & . & . & . & . & . & . & . \\ . & 1 & . & . & . & . & . & . \\ . & . & 1 & . & . & . & . & . \\ . & . & . & 1 & . & . & . & . \\ 1 & . & . & . & 1 & . & . & . \\ . & 1 & . & . & . & 1 & . & . \\ . & . & 1 & . & . & . & 1 & . \\ . & . & . & 1 & . & . & . & 1 \end{bmatrix}$$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

Outline

1 Rationale

2 Basic notions
Belief functions
Dempster's rule
Families of frames

3 Semantics
Lower probabilities
Credal sets
Set functions
Generalised probabilities
Random sets

4 Inference
Dempster's approach
Likelihood-based
From preferences

5 Conditioning
Dempster's conditioning
Lower conditional envelopes
Unnormalised vs geometric conditioning
Conditional events as equivalence classes
Generalised Bayes theorem
The total belief theorem

6 Computation
Probability transformation
Possibility transformation
Monte-Carlo methods

7 Propagation
Barnett's method
Diagnostic trees
Shafer-Shenoy architecture
Directed Evidential Networks

8 Decisions
Decision making in the TBM
Strat's decision apparatus
Upper and lower expected utilities

9 BFs on reals
Allocations of probabilitity
Random sets
Random closed intervals

10 Sister theories
Imprecise probability
Fuzzy sets and possibility
p-Boxes

11 Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

12 Toolbox
Classification
Ranking aggregation

13 Applications
A brief survey
Climate change
Pose estimation

14 Future trends

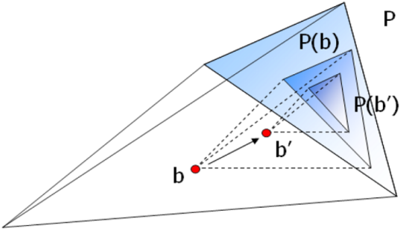
A geometric approach

to the theory of evidence

- the collection $\mathcal{B}$ of all the vectors $\mathbf{b} = [Bel(A), \emptyset \subsetneq A \subsetneq \Omega]'$ representing a belief function on Ω is a “simplex" (in rough words a higher-dimensional triangle), the **belief space**

$$\mathcal{B} = Cl(\mathbf{b}_A, \emptyset \subsetneq A \subseteq \Omega)$$

which is the convex closure of (the vectors of) all “logical" BFs $\mathbf{b}_A$



- alternatively we can adopt mass vectors $\mathbf{m}_b = [m_b(A), \emptyset \subsetneq A \subseteq \Omega]'$, living in a **mass space**: $\mathcal{M} = Cl(\mathbf{m}_A, \emptyset \subsetneq A \subseteq \Omega)$

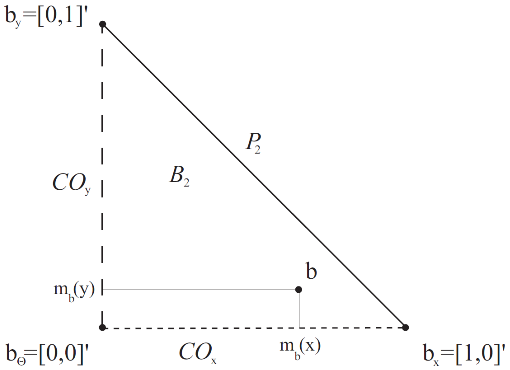
Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

Binary example

The simplex of BFs on a frame of size 2



- belief/mass space $\mathcal{B}_2 = \mathcal{M}_2$ for a binary frame
- set of probabilities is a face of the simplex (triangle)
- region of consonant BFs is a "simplicial complex" $\mathcal{CO} = \bigcup_{x \in \Omega} Cl(\mathbf{b}_A, A \ni x)$

Geometry of Dempster's rule

Conditional subspaces

- Dempster's rule behavior w.r.t. affine combination

$$\mathbf{b} \oplus \sum_i \alpha_i \mathbf{b}_i = \sum_i \beta_i (\mathbf{b} \oplus \mathbf{b}_i), \quad \beta_i = \frac{\alpha_i \kappa(b, b_i)}{\sum_{j=1}^n \alpha_j k(b, b_j)}$$

  where $\kappa(b, b_i)$ is the usual Dempster's conflict
- convex closure (Cl) and $\oplus$ **commute** in the belief space

$$\mathbf{b} \oplus Cl(\mathbf{b}_1, \cdots, \mathbf{b}_n) = Cl(\mathbf{b} \oplus \mathbf{b}_1, \cdots, \mathbf{b} \oplus \mathbf{b}_n)$$

- the **conditional subspace** $\langle \mathbf{b} \rangle$ - the set of all BFs (Dempster-) conditioned by $\mathbf{b}$:

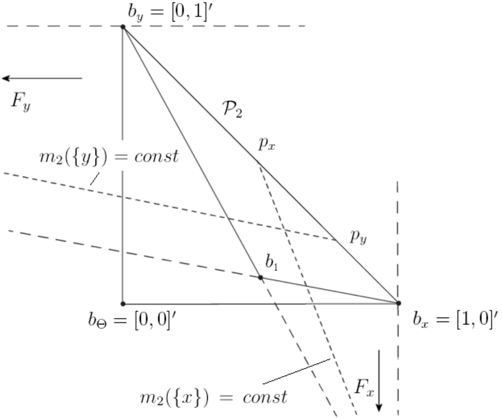
$$\langle \mathbf{b} \rangle \doteq \left\{ \mathbf{b} \oplus \mathbf{b}', \forall \mathbf{b}' \in \mathcal{B} \; s.t. \; \exists \; \mathbf{b} \oplus \mathbf{b}' \right\}$$

  is the convex closure

$$\langle \mathbf{b} \rangle = Cl(\mathbf{b} \oplus \mathbf{b}_A, \forall A \subseteq \mathcal{C}_{\mathbf{b}})$$

Geometry of Dempster's rule

Geometric construction



- the pointwise behavior of $\oplus$ depends on the notions of **constant mass locus** [Cuzzolin, 2004] and of **foci** $\{F_x, x \in \Omega\}$ of a conditional subspace

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

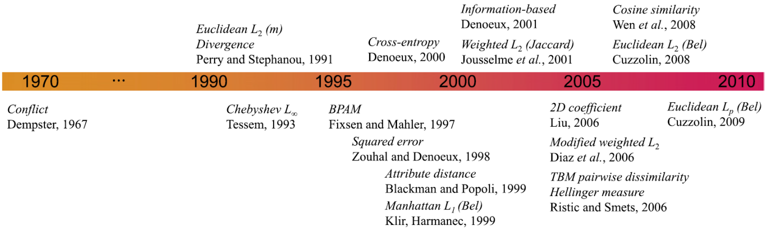
13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Dissimilarity measures

between belief functions - an overview

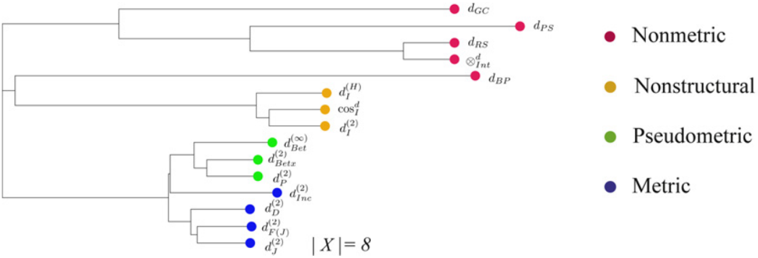
- a number of norms can be introduced for belief functions
- e.g., generalizations to belief functions of the classical Kullback-Leibler divergence of two probability distributions P, Q: $D_{KL}(P|Q) = \int_{-\infty}^{\infty} p(x) \log(\frac{p(x)}{q(x)}) dx$
- measures based on information theory such as fidelity and entropy-based norms [Jousselme IJAR'11]
- many others have been proposed [diaz,jiang,khatibi,shi], exhaustive analysis huge task!



Families of distances

between belief functions

- experimental tests on randomly generated BFs lead to the emergence of four families
  - metric (i.e. proper distance functions)
  - pseudo-metric (dissimilarities)
  - non-structural (do not account for structure of focal elements)
  - non-metric



Jousselme's distance

- most popular and cited measure of dissimilarity
- was proposed as a "measure of performance" of algorithms (e.g. object identification) where successive evidence combination leads to convergence to the "true" solution
- based on the geometric representation of mass functions m

$$d_J(m_1, m_2) \doteq \sqrt{\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T D(\mathbf{m}_1 - \mathbf{m}_2)}$$

where $D(A, B) = \frac{|A \cap B|}{|A \cup B|}$ for all $A, B \in 2^\Theta$

- D so defined:
  - is definite positive, therefore it defines a metric distance
  - takes into account the similarity among subsets
  - is such that $d(A, B) < d(A, C)$ is C is "closer" to A than B
- this notion remains not well specified though

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Moebius inverses

of plausibilities and commonalities?

- belief function are sum functions: $Bel(A) = \sum_{B \subseteq A} m(B)$
- analogous of integral in calculus, derivative = Moebius inversion
- plausibilities and commonalities **have Moebius inverses**
- only, b.pl.a.s can be negative; b.comm.a.s are not even normalised

belief function $\quad m_b(A) = \sum_{B \subseteq A} (-1)^{|A-B|} b(B) \quad$ b.b.a.

plausibility function $\quad \mu_b(A) \doteq \sum_{B \subseteq A} (-1)^{|A-B|} pl_b(B) \quad$ b.pl.a.

commonality function $\quad q_b(B) = \sum_{\emptyset \subseteq A \subseteq B} (-1)^{|B \setminus A|} Q_b(A) \quad$ b.comm.a.

- plausibilities and commonalities **live in simplices congruent with the belief space $\mathcal{B}$**

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

Measures of uncertainty

for belief functions

- various measures have been proposed - see for instance the (rather outdated) survey by Nikhil Pal, 1992
- **Yager's entropy measure** (350+ citations):

$$E(m) = -\sum_{A \in \mathcal{F}} m(A) \log Pl(A)$$

  - Yager's entropy is 0 for consonant or **consistent** BFs ($A_i \cap A_j \neq \emptyset$ for all FEs)
  - is maximal for disjoint focal elements with equal mass
- Hohle's measure of confusion: $C(m) = -\sum_{A \in \mathcal{F}} m(A) \log Bel(A)$
- **specificity** of belief measures: $N(m) = \sum_{A \in \mathcal{F}} \frac{m(A)}{|A|}$
  - measure the dispersion of the evidence
  - clearly related to pignistic function
- Klir's non-specificity (extended by Dubois & Prade):

$$I(m) = \sum_{A \in \mathcal{F}} m(A) \log |A|$$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

Measures of uncertainty

Global measures

- composite measures: Lamata and Moral: $E(m) + I(m)$
- $E(m)$ was criticised by Klir & Ramer for it expresses conflict as $A \cap B = \emptyset$ rather than $B \not\subseteq A$
- $C(m)$ was criticised for it does not measure to what extent two focal elements disagree (size of $A \cap B$)
- Klir & Ramer's **global uncertainty measure**: $D(m) + I(m)$, where

$$D(m) = -\sum_{A \in \mathcal{F}} m(A) \log \left[\sum_{B \in \mathcal{F}} m(B) \frac{|A \cap B|}{|B|} \right]$$

- Pal argues that none of them is really satisfactory: none of the composite measures have a unique maximum
- there is no sounding rationale for simply adding conflct and non-specificity measures together to get a “total” one
- some are computationally very expensive

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Matrix representation
Geometry
Distances
Combinatorics
Uncertainty measures

Aggregated Uncertainty vs Ambiguity Measure

- Harmanec's **Aggregated Uncertainty** (AU) as the maximal Shannon entropy of all consistent probabilities
  - obviously assumes a credal set interpretation
  - it is the minimal measure meeting eight requirements: symmetry, continuity, expansibility, subadditivity, additivity, monotonicity, normalisation
- criticised by Klir and Smith for being insensitive to changes in evidence
- replaced by a linear combination of AU and nonspecificity $I(m)$
- still high computational complexity
- Jousselme et al, 2006: **Ambiguity Measure** (AM), basically classical entropy of pignistic function

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Classification
Ranking aggregation
Applications

Outline

1 Rationale
2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

A set of tools for the working scientist

using belief functions

- scientists face on a daily basis problems such as:
    - **making decisions** based on the available data (we already covered this)
    - **estimating** a quantity of interest give the available data (which can be missing, incomplete,conflicting,partially specified)
    - **classifying** data-points into bins
        - extending k-NN classification approaches
        - fusing the results of multiple classifiers
    - **clustering** clouds of data to make sense of them
    - learning a mapping from measurements to a domain of interest (**regression**)
    - ranking objects
- belief functions can provide useful approaches to all these problems when in the presence of (heavy) uncertainty

Belief functions in Machine Learning

- the theory of belief functions has great potential to help solve **complex machine learning (ML) problems**, particularly those involving:
  - **weak information** (partially labeled data, unreliable sensor data, etc.);
  - **multiple sources** of information (classifier or clustering ensembles) [Quost et al., 2007; Masson & Denoeux, 2011]
- other recent ML applications of belief functions:
  - regression [Petit-Renaud & Denoeux, 2004]
  - multi-label classification [Denoeux et al. 2010]
  - clustering [Masson & Denoeux, 2008; Antoine et al., 2012]

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

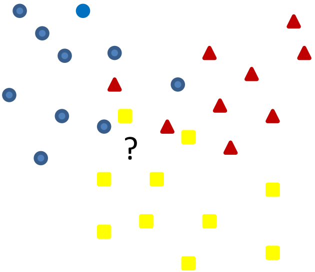
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Classification
Ranking aggregation
Applications

Outline

1 Rationale
2 Basic notions
 - Belief functions
 - Dempster's rule
 - Families of frames
3 Semantics
 - Lower probabilities
 - Credal sets
 - Set functions
 - Generalised probabilities
 - Random sets
4 Inference
 - Dempster's approach
 - Likelihood-based
 - From preferences
5 Conditioning
 - Dempster's conditioning
 - Lower conditional envelopes
 - Unnormalised vs geometric conditioning
 - Conditional events as equivalence classes
 - Generalised Bayes theorem
 - The total belief theorem
6 Computation
 - Probability transformation
 - Possibility transformation
 - Monte-Carlo methods
7 Propagation
 - Barnett's method
 - Diagnostic trees
 - Shafer-Shenoy architecture
 - Directed Evidential Networks
8 Decisions
 - Decision making in the TBM
 - Strat's decision apparatus
 - Upper and lower expected utilities
9 BFs on reals
 - Allocations of probabilitity
 - Random sets
 - Random closed intervals
10 Sister theories
 - Imprecise probability
 - Fuzzy sets and possibility
 - p-Boxes
11 Advances
 - Matrix representation
 - Geometry
 - Distances
 - Combinatorics
 - Uncertainty measures
12 Toolbox
 - Classification
 - Ranking aggregation
13 Applications
 - A brief survey
 - Climate change
 - Pose estimation
14 Future trends

Classification problems

in machine learning



- a population is assumed to be partitioned in c groups or classes
- let $\Omega = \{\omega_1, \ldots, \omega_c\}$ denote the set of classes
- each instance is described by
  - a feature vector $\mathbf{x} \in \mathbb{R}^p$
  - a class label $y \in \Omega$

- problem: given a **learning set** $\mathcal{L} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, **predict the class** of a new instance described by $\mathbf{x}$

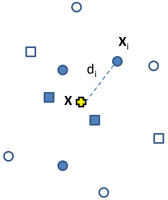
Main approaches

Classification with belief functions

1. Approach 1 (**ensemble classification**): Convert the outputs from standard classifiers into belief functions and combine them using Dempster's rule or any other alternative rule (e.g., Quost al., *IJAR*, 2011)
2. Approach 2: Develop **evidence-theoretic classifiers** directly providing belief functions as outputs:
    - **Generalized Bayes theorem**, extends the Bayesian classifier when class densities and priors are ill-known [Appriou, 1991; Denœux & Smets, 2008]
    - **Distance-based approach**: evidential *k*-NN rule [Denœux, 1995], evidential neural network classifier [Denœux, 2000]

- today we will just see the evidential k-NN rule (for complete and partially specified data)

Evidential K-NN



- let Ω be the set of classes
- let $\mathcal{N}_k(\mathbf{x}) \subset \mathcal{L}$ denote the set of the k nearest neighbors of $\mathbf{x}$ in $\mathcal{L}$, based on some distance measure d
- each $\mathbf{x}_i \in \mathcal{N}_k(\mathbf{x})$ can be considered as a **piece of evidence** regarding the class of $\mathbf{x}$ represented by a mass function m_i on Ω:

$$m_i(\{y_i\}) = \varphi(d_i), \quad m_i(\Omega) = 1 - \varphi(d_i)$$

- the **strength of this evidence decreases with the distance** d_i between $\mathbf{x}$ and $\mathbf{x}_i$ - φ is a **decreasing function** such that $\lim_{d \to +\infty} \varphi(d) = 0$
- pooling of evidence: $m = \bigoplus_{\mathbf{x}_i \in \mathcal{N}_k(\mathbf{x})} m_i$
- the function φ can be fixed heuristically or selected among a family $\{\varphi_\theta | \theta \in \Theta\}$ using, e.g., cross-validation
- decision: select the class with the **highest plausibility**

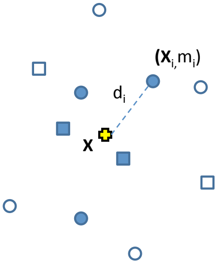
Classification with partially specified data

- in some applications, learning instances are labeled by experts or indirect methods (no ground truth)
- **class labels of learning data are then uncertain**: partially supervised learning problem
- formalization of the learning set: $\mathcal{L} = \{(\mathbf{x}_i, m_i), i = 1, \ldots, n\}$, where
  - $\mathbf{x}_i$ is the attribute vector for instance i, and
  - m_i is a mass function representing **uncertain expert knowledge** about the class y_i of instance i
- special cases:
  - $m_i(\{\omega_k\}) = 1$ for all i: **supervised learning**
  - $m_i(\Omega) = 1$ for all i: **unsupervised learning**
- the evidential k-NN rule can easily be adapted to handle such uncertain learning data

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Classification
Ranking aggregation
Applications

Evidential k-NN rule

for partially supervised data



- Ω is again the collection of classes
- each mass function m_i is **discounted** with a rate depending on the distance d_i:

$$m_i'(A) = \varphi(d_i)\, m_i(A), \quad \forall A \subset \Omega$$

$$m_i'(\Omega) = 1 - \sum_{A \subset \Omega} m_i'(A)$$

- the k mass functions m_i' are combined using **Dempster's rule**:

$$m = \bigoplus_{\mathbf{x}_i \in \mathcal{N}_k(\mathbf{x})} m_i'$$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Classification
Ranking aggregation
Applications
Future trends

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

13 Applications
- A brief survey
- Climate change
- Pose estimation

14 Future trends

Ranking aggregation

Problem definition

- we consider a **set of alternatives** $O = \{o_1, o_2, ..., o_n\}$ and an **unknown linear order** (transitive, antisymmetric and complete relation) on O
- typically, this linear order corresponds to **preferences** held by an agent or a group of agents, so that $o_i \succ o_j$ is interpreted as "alternative o_i is preferred to alternative o_j"
  - (compare inference from qualitative data)
- a source of information (elicitation procedure, classifier) provides us with $n(n-1)/2$ **paired comparisons**, affected by uncertainty
- problem: derive the **most plausible linear order** from this uncertain (and possibly conflicting) information

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Classification
Ranking aggregation
Applications

Example

Tritchler & Lockwood, 1991

- consider four scenarios $O = \{A, B, C, D\}$ describing ethical dilemmas in health care
- suppose two experts gave their preference for all six possible scenario pairs with confidence degrees described below

- assuming the existence of a unique **consensus linear ordering** L^* and seeing the expert assessments as sources of information, what can we say about L^*?

Combining pairwise masses

on the space of linear orders

- the frame of discernment is the set $\mathcal{L}$ of **linear orders over O**
- comparing each pair of objects (o_i, o_j) yields a **pairwise mass function** $m^{\Theta_{ij}}$ on a coarsening $\Theta_{ij} = \{o_i \succ o_j, o_j \succ o_i\}$ with:

$$m^{\Theta_{ij}}(o_i \succ o_j) = \alpha_{ij}, \quad m^{\Theta_{ij}}(o_j \succ o_i) = \beta_{ij}, \quad m^{\Theta_{ij}}(\Theta_{ij}) = 1 - \alpha_{ij} - \beta_{ij}$$

- $m^{\Theta_{ij}}$ may come from a single expert (e.g., an evidential classifier) or from the combination of the evaluations of several experts
- let $\mathcal{L}_{ij} = \{L \in \mathcal{L} | (o_i, o_j) \in L\}$. **Vacuously extending $m^{\Theta_{ij}}$ in $\mathcal{L}$** yields

$$m^{\Theta_{ij}\uparrow\mathcal{L}}(\mathcal{L}_{ij}) = \alpha_{ij}, \quad m^{\Theta_{ij}\uparrow\mathcal{L}}(\overline{\mathcal{L}_{ij}}) = \beta_{ij}, \quad m^{\Theta_{ij}\uparrow\mathcal{L}}(\mathcal{L}) = 1 - \alpha_{ij} - \beta_{ij}$$

- **combining the pairwise mass functions** using Dempster's rule:

$$m^{\mathcal{L}} = \bigoplus_{i<j} m^{\Theta_{ij}\uparrow\mathcal{L}}$$

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Classification
Ranking aggregation
Applications

Plausibility of a linear order

An integer programming problem

- the plausibility of the combination $m^{\mathcal{L}}$ is:

$$pl(L) = \frac{1}{1-\kappa} \prod_{i<j} (1-\beta_{ij})^{\ell_{ij}} (1-\alpha_{ij})^{1-\ell_{ij}},$$

  where $\ell_{ij} = 1$ if $(o_i, o_j) \in L$ and 0 otherwise
- its logarithm $pl(L)$ can be maximized by solving the following **binary integer programming** problem:

$$\max_{\ell_{ij} \in \{0,1\}} \sum_{i<j} \ell_{ij} \ln \left(\frac{1-\beta_{ij}}{1-\alpha_{ij}} \right)$$

  subject to:

$$\begin{cases} \ell_{ij} + \ell_{jk} - 1 \leq \ell_{ik}, & \forall i<j<k \quad (1) \\ \ell_{ik} \leq \ell_{ij} + \ell_{jk}, & \forall i<j<k \quad (2) \end{cases}$$

- constraint (1) ensures that $\ell_{ij} = 1$ and $\ell_{jk} = 1 \Rightarrow \ell_{ik} = 1$, and (2) ensures that $\ell_{ij} = 0$ and $\ell_{jk} = 0 \Rightarrow \ell_{ik} = 0$.

On preference aggregation

A summary

- the framework of belief functions allows us to **model uncertainty in paired comparisons**
- the **most plausible linear order** can be computed efficiently using a binary linear programming approach
- the approach has been applied to **label ranking**, in which the task is to **learn a "ranker"** that maps *p*-dimensional feature vectors *x* describing an agent to a linear order over a finite set of alternatives, describing the agent's preferences [Denœux and Masson, 2012]
- the method can easily be extended to the elicitation of preference relations with **indifference** and/or **incomparability** between alternatives [Denœux and Masson. *AOR* 195(1):135-161, 2012]

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. **Applications**
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

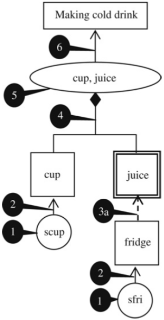
Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

New applications

of belief functions



(a) Making cold drink.

- sensor fusion has always been a stronghold of belief calculus
- mainly about merging different sensors using Dempster's rule
- typical applications: tracking and data association, reliability in engineering, image processing, robotics, medical imaging and diagnosis, business and finance (audit)
- a new wave of applications, on:
  - geographical information systems
  - communication networks and security
  - earth sciences
- here we present one (or two!) in more detail:
  - **climate change**
  - **motion capture** in computer vision

Outline

1 Rationale

2 Basic notions
- Belief functions
- Dempster's rule
- Families of frames

3 Semantics
- Lower probabilities
- Credal sets
- Set functions
- Generalised probabilities
- Random sets

4 Inference
- Dempster's approach
- Likelihood-based
- From preferences

5 Conditioning
- Dempster's conditioning
- Lower conditional envelopes
- Unnormalised vs geometric conditioning
- Conditional events as equivalence classes
- Generalised Bayes theorem
- The total belief theorem

6 Computation
- Probability transformation
- Possibility transformation
- Monte-Carlo methods

7 Propagation
- Barnett's method
- Diagnostic trees
- Shafer-Shenoy architecture
- Directed Evidential Networks

8 Decisions
- Decision making in the TBM
- Strat's decision apparatus
- Upper and lower expected utilities

9 BFs on reals
- Allocations of probabilitity
- Random sets
- Random closed intervals

10 Sister theories
- Imprecise probability
- Fuzzy sets and possibility
- p-Boxes

11 Advances
- Matrix representation
- Geometry
- Distances
- Combinatorics
- Uncertainty measures

12 Toolbox
- Classification
- Ranking aggregation

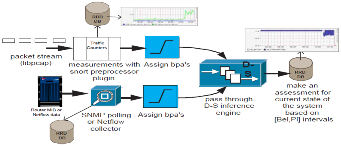
13 Applications
- A brief survey
- Climate change
- Pose estimation

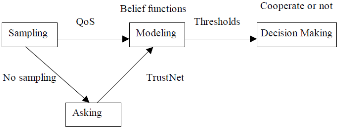
14 Future trends

Most popular applications

of belief functions

- information quality in **financial accounting** [A conceptual framework and belief-function approach to assessing overall information quality (158)]
- **auditing** [The Bayesian and belief-function formalisms: A general perspective for auditing (148)]
- reputation and **trust management in telecoms** [An evidential model of distributed reputation management (615)]
- security [An information systems security risk assessment model under the DS theory of belief functions (137)]
- **DoS** [Towards multisensor data fusion for DoS detection (137)]





Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Most popular applications

of belief functions

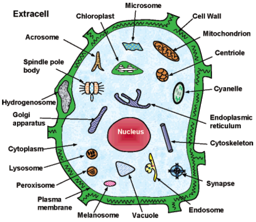


Figure 1. Schematic illustration to show the 22 subcellular locations of eukaryotic proteins: (1) acrosome, (2) cell wall, (3)

- **molecular biology** [Ensemble classifier for protein fold pattern recognition (257)], [**Predicting eukaryotic protein subcellular location by fusing optimized evidence-theoretic K-nearest neighbor classifiers** (205)]
- **medical imaging** [Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account (218)]
- **earth sciences** and ecology [Integration of geophysical and geological data using evidential belief function (106)], [Decision support system for the sustainable forest management (131)]
- **context-aware HCI** and sensing [Sensor fusion using Dempster-Shafer theory (238)], [Evidential fusion of sensor data for activity recognition in smart homes (158)]

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Most popular applications

of belief functions

- measurement theory [Measurement uncertainty: An approach via the mathematical theory of evidence (77)]
- **reliability** [Engine fault diagnosis based on multi-sensor information fusion using Dempster-Shafer evidence theory (243)]
- **engineering** [Modelling global risk factors affecting construction cost performance (303)]
- semantic web and **information retrieval** [Dempster-Shafer's theory of evidence applied to structured documents: modelling uncertainty (154)], [EDM: a general framework for data mining based on evidence theory (109)]
- reputation management in e-commerce [Distributed reputation management for electronic commerce (257)]
- **climate change** [Utilizing belief functions for the estimation of future climate change (86)]
- chemistry [Application of belief theory to similarity data fusion for use in analog searching and lead hopping (99)]

Most popular applications

of belief functions

- **robotics and navigation** [An evidential approach to map-building for autonomous vehicles (229)], [Dempster-Shafer theory for sensor fusion in autonomous mobile robots (192)]
- **tracking and data association** [Shafer-Dempster reasoning with applications to multisensor target identification systems (317)]
- **image processing and computer vision** [Image annotations by combining multiple evidence Wordnet (231)], [Evidence-based recognition of 3-D objects (176)]
- **biometrics** [Image quality assessment for iris biometric (160)]

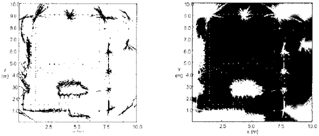
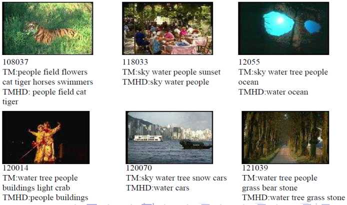


Fig. 10. Occupancy maps generated using the Bayes' fusion rule, values above and below 0.5 separated into full and empty map res. and normalized into the interval [0 1] for comparison to the D-S method.



Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. Future trends

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Climate change

Adaptation of flood defense structures

- climate change is expected to have **enormous economic impact**, including threats to infrastructure assets through
  - damage or destruction from extreme events;
  - coastal flooding and inundation from sea level rise, etc.
- **adaptation of infrastructure** to climate change is a major issue
- engineering design processes and standards are based on **analysis of historical climate data** (using, e.g. Extreme Value Theory), with the assumption of a stable climate
- commonly, flood defenses in coastal areas are designed to withstand at least **100 years return period events**. However, due to climate change, they will be subject during their life time to higher loads than the design estimations
- the main impact is related to the **increase of the mean sea level**, which affects the frequency and intensity of surges
- for adaptation purposes, statistics of extreme sea levels derived from historical data should be combined with projections of the future sea level rise (SLR)

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Assumptions and approach

- the **annual maximum sea level** Z at a given location is often assumed to have a Gumbel distribution

$$P(Z \leq z) = \exp\left[-\exp\left(-\frac{z-\mu}{\sigma}\right)\right]$$

  with mode μ and scale parameter σ
- current design procedures are based on the **return level** z_T associated with a return period T, defined as the quantile at level $1 - 1/T$: $z_T = \mu - \sigma \log\left[-\log\left(1 - \frac{1}{T}\right)\right]$
- because of climate change, it is assumed that the distribution of annual maximum sea level at the end of the century will be **shifted to the right**, with shift equal to the SLR : $z'_T = z_T + SLR$
- **approach**:
  1. represent the evidence on z_T by a likelihood-based belief function using past sea level measurements;
  2. represent the evidence on SLR by a belief function describing expert opinions;
  3. combine these two items of evidence to get a belief function on $z'_T = z_T + SLR$.

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Statistical evidence on z_T

- let $z_1, \ldots, z_n$ be n i.i.d. observations of Z. The likelihood function is:

$$L(z_T, \mu; z_1, \ldots, z_n) = \prod_{i=1}^{n} f(z_i; z_T, \mu),$$

where the pdf of Z has been reparametrized as a function of z_T and μ

- the corresponding contour function is thus:

$$pl(z_T, \mu; z_1, \ldots, z_n) = \frac{L(z_T, \mu; z_1, \ldots, z_n)}{\sup_{z_T, \mu} L(z_T, \mu; z_1, \ldots, z_n)}$$

and the marginal contour function of z_T is

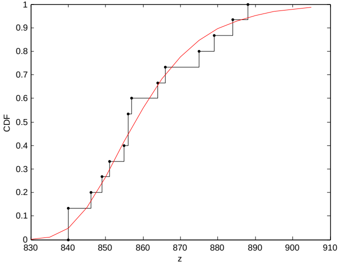
$$pl(z_T; z_1, \ldots, z_n) = \sup_{\mu} pl(z_T, \mu; z_1, \ldots, z_n)$$

Belief functions for the working scientist

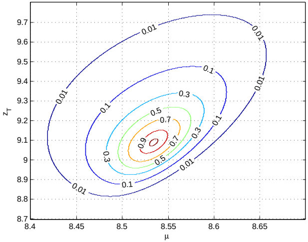
F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

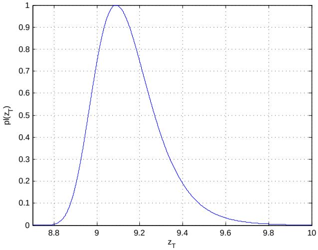
Example of inference



- 15 years of sea level data at Le Havre, France
- corresponding contour function (left) and marginal contour function (right)



$pl(z_{100}, \mu)$



$pl(z_{100})$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Expert evidence on SLR

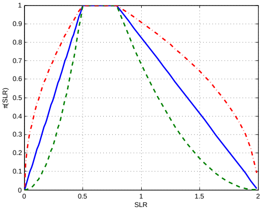
- future SLR projections provided by the IPCC last Assessment Report (2007) give **[0.18 m, 0.79 m]** as a likely range of values for SLR over the 1990-2095 period
- however, it is indicated that **higher values cannot be excluded**
- other recent SLR assessments based on semi-empirical models have been undertaken. For example, based on a simple statistical model, Rahmstorf (2007) suggests **[0.5m, 1.4 m]** as a likely range
- recent studies indicate that **the threshold of 2 m cannot be exceeded** by the end of this century due to physical constraints

Representation

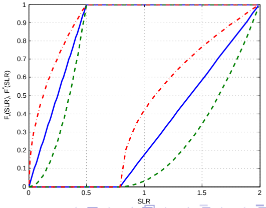
of expert evidence

- the interval $[0.5, 0.79] = [0.18, 0.79] \cap [0.5, 1.4]$ seems to be fully supported, as considered highly plausible by all three sources
- while values outside the interval $[0, 2]$ are considered as impossible
- three representations of expert evidence:
  - **consonant random intervals** with core $[0.5, 0.79]$, support $[0, 2]$ and different contour functions π;
  - **p-boxes** with same cumulative *Bel* and *Pl* as above;
  - random sets $[U, V]$ with **independent** U **and** V and same cumulative belief and plausibility functions as above

Contour functions



Cumulative Bel and Pl



Combination

Principle

- let $[U_{z_T}, V_{z_T}]$ and $[U_{SLR}, V_{SLR}]$ be the **independent random intervals** representing evidence on z_T and SLR, respectively
- the random interval for $z'_T = z_T + SLR$ is

$$[U_{z_T}, V_{z_T}] + [U_{SLR}, V_{SLR}] = [U_{z_T} + U_{SLR}, V_{z_T} + V_{SLR}]$$

- the corresponding belief and plausibility functions are

$$\begin{aligned} Bel(A) &= P([U_{z_T} + U_{SLR}, V_{z_T} + V_{SLR}] \subseteq A) \\ Pl(A) &= P([U_{z_T} + U_{SLR}, V_{z_T} + V_{SLR}] \cap A \neq \emptyset) \end{aligned}$$

  for all $A \in \mathcal{B}(\mathbb{R})$
- $Bel(A)$ and $Pl(A)$ can be estimated by **Monte Carlo simulation** (see Computation)

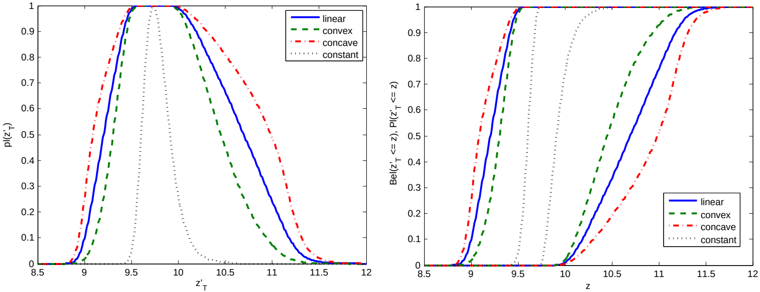
Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Results

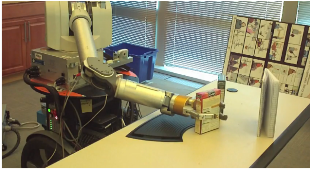
of combining expert and historical belief functions



Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. **Applications**
    - A brief survey
    - Climate change
    - **Pose estimation**
14. Future trends

Pose estimation



- estimating the position and orientation of an object, along with its internal configuration or pose
- **“model-based"**: explicitly known parametric body model
- **“learning-based"**: exploit the fact that typical (human) motions involve a far smaller set of poses
- directly recover pose estimates **from observable image quantities (features)**

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Example-based estimation

- explicitly store a **set of training examples** whose 3D poses are known
- estimate pose by searching for training image(s) similar to the given input image and interpolating from their poses
- **no prior structure** of the pose space is incorporated
- typical architecture:
  - features are extracted from individual images
  - a **map from the features space to the pose space is learned from a training set of examples**
  - the likely pose of the object is then predicted by feeding this feature vector to the learnt map
- approaches: Relevant Vector Machines (RVMs), shape context matching, Local Weighted Regression, BoostMap, Bayesian Mixture of Experts

Scenario

- the available evidence comes in the form of a **training set of images** containing sample poses of an unspecified object
- configuration: a vector $q \in \mathcal{Q} \subset \mathbb{R}^D$
- an **"oracle"** provides for each training image I_k the configuration q_k of the object portrayed in the image
- object location within each training image is known in the form of a **bounding box**
- in **training**, the object explores its range of possible configurations and both samples poses $\tilde{\mathcal{Q}} \doteq \left\{ q_k, k = 1, ..., T \right\}$ and N features $\tilde{\mathcal{Y}} \doteq \left\{ y_i(k), k = 1, ..., T \right\}, i = 1, ..., N$ are collected
- source of ground truth: motion capture system
- in **testing**, a supervised localization algorithm is employed to locate the object within the test image
- such features are exploited to produce an estimate of the object's configuration

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Pose estimation via belief calculus: why

- let us assume features and poses are described by probability distributions
- as **feature-to-pose maps are typically multi-valued** ..
- ..they induce belief functions on the space of poses
- also, in pose estimation training sets are of limited size
- as in the credal interpretation belief functions amount to a **set of linear constraints on the actual conditional pose distribution (given the features)** ..
- .. they **encode the uncertainty induced by the size of the training set**
- finally, multiple features defined as belief functions on different (feature) domains can be moved to a common refinement (the pose space) and there combined

Training session

- learn from the training data an approximation $\tilde{\rho}$ of the unknown mapping between each feature space $\mathcal{Y}_i$ and the pose space $\mathcal{Q}$
- we apply **EM clustering** to the N training sequences of feature values $\{y_i(k), k = 1, ..., T\}, i = 1, ..., N$ ..
- ..obtaining a obtain a Mixture of Gaussians (MoG)

$$\left\{\Gamma_i^j, j = 1, ..., n_i\right\}, \quad \Gamma_i^j \sim \mathcal{N}(\mu_i^j, \Sigma_i^j)$$

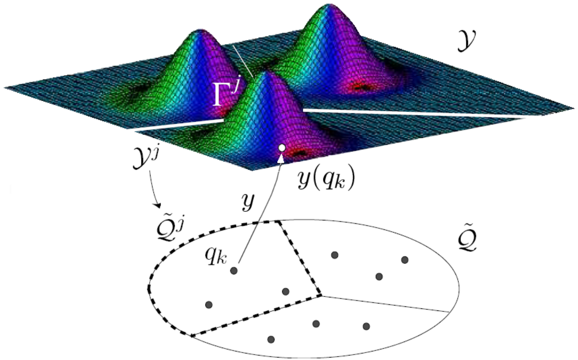
approximation of each feature space $\mathcal{Y}_i$

- .. and a **discrete approximation of the feature-pose** mapping

$$\rho_i : \mathcal{Y}_i^j \mapsto \tilde{\mathcal{Q}}_i^j \doteq \left\{q_k \in \tilde{\mathcal{Q}} : y_i(k) \in \mathcal{Y}_i^j\right\}$$

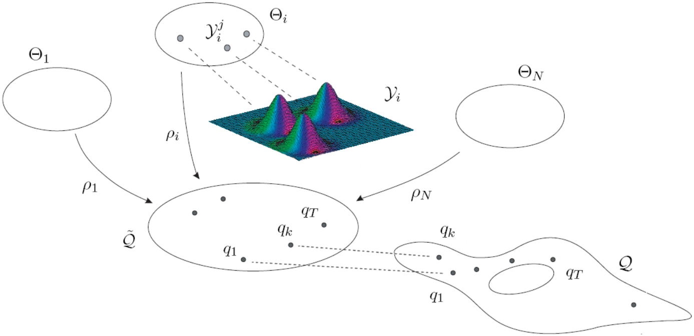
  - $\mathcal{Y}_i^j$ is the region of $\mathcal{Y}_i$ in which the j-th Gaussian dominates

Discrete feature-pose maps



- each element $\mathcal{Y}_i^j$ of the approximate feature space is associated with the set of training poses $q_k \in \mathcal{Q}_k$ whose i-th feature value falls in $\mathcal{Y}_i^j$

Evidential model



- as the applications ρ_i map approximate feature spaces to disjoint partitions of the set of sample poses $\tilde{\mathcal{Q}}$ they are refinings
- $\tilde{\mathcal{Q}}$ is a common refinement for the approximate feature spaces $\Theta_1, ..., \Theta_N$
- the collection of FODs $\tilde{\mathcal{Q}}, \Theta_1, ..., \Theta_N$ along with their refinings $\rho_1, ..., \rho_N$ is characteristic of object to track, features y_i, and training data
- we call it the **evidential model** of the moving object, learned by example

Testing: Dirichlet belief functions

- when one or more new test images are acquired, new visual features $y_1, ..., y_N$ are extracted
- feature values can be mapped to a collection of belief functions $Bel_1, ..., Bel_N$ on the set of sample poses $\mathcal{Q}$
- belief functions also allow to **take into account the scarcity of the training samples**..
- .. by assigning some mass $m(\Theta_i)$ to the whole feature space

$$m_i : 2^{\Theta_i} \to [0,1], \quad m_i(\mathcal{Y}_i^j) = \frac{\Gamma_i^j(y_i)}{\sum_k \Gamma_i^k(y_i)}(1 - m_i(\Theta_i))$$

  - a reasonable choice is $m_i(\Theta_i) = \frac{1}{n_i}$, as when $n_i \to \infty$ the discount factor tends to zero and the approximate feature space Θ_i tends to the real feature space $\mathcal{Y}_i$

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Computing pose estimates

Belief estimate

- the belief functions inferred from the test feature values are then mapped to the approximate pose space $\tilde{\mathcal{Q}}$ by **vacuous extension**
- where they are combined by **conjunctive combination**
- this yields the **belief estimate** of the pose, which amounts to a convex set of probabilities on the set of sample poses $\tilde{\mathcal{Q}}$. Then:

1. we can compute the **expected pose associated with each vertex** of the credal set:

$$\hat{q} = \sum_{k=1}^{T} p(q_k) q_k$$

   - degree of confidence on the accuracy of the pointwise estimate $\rightarrow$ size of the credal set

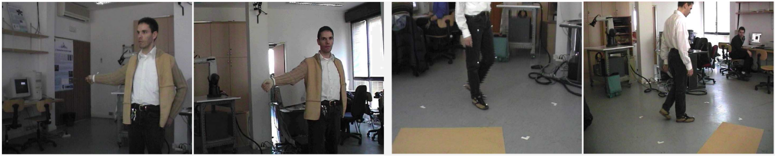
2. or, we can approximate $\hat{b}$ with a probability $\hat{p}$ on $\tilde{\mathcal{Q}}$ (e.g. the pignistic function)

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

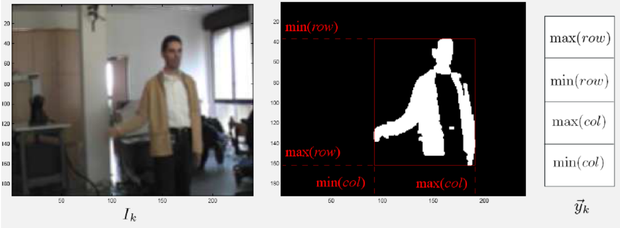
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
A brief survey
Climate change
Pose estimation

Two human pose estimation experiments



- person filmed by two uncalibrated DV cameras
- **arm experiment**: subject moves his arm, while standing in a fixed floor location
- **legs experiment**: person walking normally on the floor, training set collected by sampling a random walk on a section of the floor
- length of the training sequences: 1726 frames for the arm and 1952 frames for the legs
- quite challenging setup: background was highly non-static, with people coming in and out the scene and flickering monitors; self-occlusions

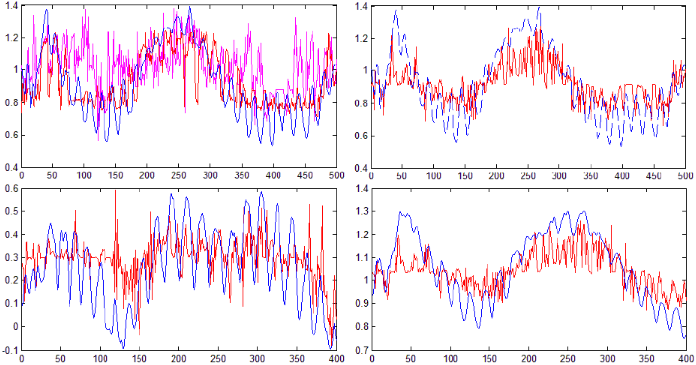
Annotation and features



- color-based segmentation to get the object of interest (to automatically generate the bounding box annotation required)
- simple feature vector directly from the bounding box: the collection max(row), min(row), max(col), min(col)
- built different evidential models with 2 features from left view, 3 features from right view, or both
- MoG with $n = 5$ components for each feature space

Empirical results

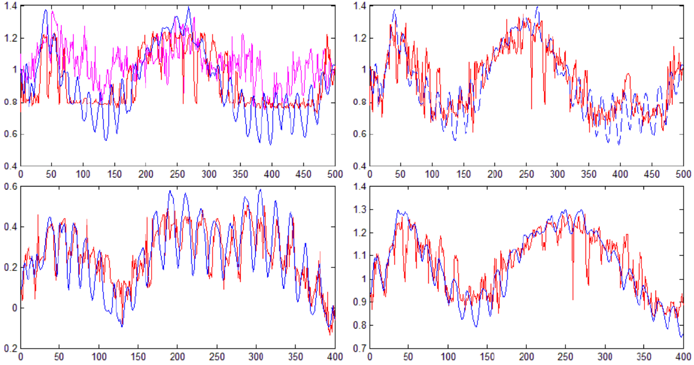
Comparison with Relevant Vector Machine



- results for component 9 on top, components 1 and 6 at bottom
- blue → ground truth, red → pignistic estimate
- average Euclidean errors: 25.0, 10.6, 18.6, and 7.0 centimeters
- our belief-theoretical approach outperforms the competitor

Empirical results

Comparison with Gaussian Process Regression



- used to build feature-pose maps in, for instance, [14] and [16]
- same components of the pose vector, same test sequence
- our BMR approach clearly outperforms a standard RVM implementation
- average Euclidean errors: 31.2, 13.6, 23.0, and 4.5 centimeters

Conclusions

on the Belief Modeling Regression approach

- presented a novel approach to example-based pose estimation
- framing the problem within belief calculus is natural
- tested in a fairly challenging human pose recovery setup
- exhibits interesting relationships with Gaussian Process approach we did not mention
- future: efficient conflict resolution mechanism
- future: testing of the framework in higher-dimensional pose ranges
- full development of an evidential tracking approach, exploiting temporal information as well via tht total belief theorem (see Conditioning)

Outline

1. Rationale
2. Basic notions
   - Belief functions
   - Dempster's rule
   - Families of frames
3. Semantics
   - Lower probabilities
   - Credal sets
   - Set functions
   - Generalised probabilities
   - Random sets
4. Inference
   - Dempster's approach
   - Likelihood-based
   - From preferences
5. Conditioning
   - Dempster's conditioning
   - Lower conditional envelopes
   - Unnormalised vs geometric conditioning
   - Conditional events as equivalence classes
   - Generalised Bayes theorem
   - The total belief theorem
6. Computation
   - Probability transformation
   - Possibility transformation
   - Monte-Carlo methods
7. Propagation
   - Barnett's method
   - Diagnostic trees
   - Shafer-Shenoy architecture
   - Directed Evidential Networks
8. Decisions
   - Decision making in the TBM
   - Strat's decision apparatus
   - Upper and lower expected utilities
9. BFs on reals
   - Allocations of probabilitity
   - Random sets
   - Random closed intervals
10. Sister theories
    - Imprecise probability
    - Fuzzy sets and possibility
    - p-Boxes
11. Advances
    - Matrix representation
    - Geometry
    - Distances
    - Combinatorics
    - Uncertainty measures
12. Toolbox
    - Classification
    - Ranking aggregation
13. Applications
    - A brief survey
    - Climate change
    - Pose estimation
14. **Future trends**

Belief functions for the working scientist
F. Cuzzolin and T. Denoeux
Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
Future trends

A summary

of what we have learned in this tutorial

- the theory of belief functions is a modeling language for **representing elementary items of evidence and combining them**, in order to form a representation of our beliefs about certain aspects of the world
- it is relatively **simple to implement** and has been successfully used in a wide range of applications
- has strong relationships with other theories of uncertainty
- belief functions have interesting mathematical properties in terms of geometry, algebra, combinatorics
- evidential reasoning can be implemented even for **very large spaces and numerous pieces of evidence**, because
  - elementary items of evidence induce **simple belief functions**, which can be combined very efficiently;
  - the **most plausible hypothesis** can be found without computing the whole combined belief function;
  - Monte-Carlo approximations are easily implementable
  - local propagation schemes allow parallelisation

A summary

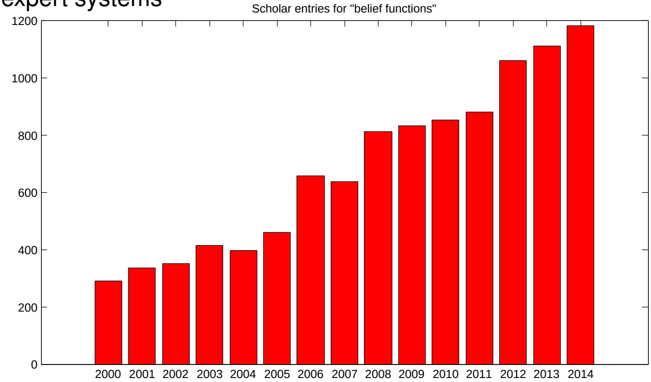
of what we have learned in this tutorial

- statistical evidence may be represented by **likelihood-based belief functions**, generalizing both likelihood-based and Bayesian inference
- inference can also be performed from qualitative data
- decision making strategies based on **intervals of expected utilities** can be formulated that are more cautious than traditional ones
- the **extension to continuous domains** can be tackled via the Borel interval representation, possibly in connection with p-Boxes
- a **toolbox of estimation, classification, regression tools** based on the theory of belief functions is available

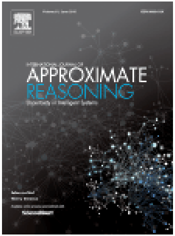
Recent trends

in the theory and application of belief functions

- in 2014 alone, almost 1200 papers were published on belief functions and their applications
- new applications are gaining ground, beyond sensor fusion or expert systems



Publications venues



- **conferences** on the theory of uncertainty:
  - BFAS's International Conference on Belief Functions (BELIEF)
  - Uncertainty in Artificial Intelligence (UAI)
  - International Conference on Information Fusion (FUSION)
  - International Symposium on Imprecise Probability - Theories and Applications (ISIPTA)
  - Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)
  - IEEE Systems, Man and Cybernetics (SMC)
  - Information Processing and Management under Uncertainty (IPMU)
- **journals** (for theoretical contributions):
  - International Journal of Approximate Reasoning (IJAR)
  - IEEE Transactions on Fuzzy Systems (I.F. 6.306)
  - IEEE Transactions on Cybernetics (I.F. 3.781)
  - Artificial Intelligence
  - Information Sciences (4.038)
  - Fuzzy Sets and Systems

Open issues

and future developments

- what still needs to be resolved:
  - competing epistemic interpretations of belief function theory
  - conditioning issue still open, a variety of approaches proposed depending on semantic adopted and revision principles
  - correct mechanism for evidence combination still debated, depend on meta-information on sources hardly accessible
  - local propagation models (e.g. Shenoy-Shafer) still assume low complexity of local cliques
- what are the next steps?
  - relationships with several fields of mathematics not completely understood
  - generalisation of the total probability theorem in full generality
  - full development of evidential graphical models (e.g. evidential HMMs)
  - tackling current machine learning trends such as transfer learning
  - can it cope with big data paradigm?

Belief functions for the working scientist

F. Cuzzolin and T. Denoeux

Rationale
Basic notions
Semantics
Inference
Conditioning
Computation
Propagation
Decisions
BFs on reals
Sister theories
Advances
Toolbox
Applications
Future trends

For Further Reading

Papers and Matlab software available at:

`https://www.hds.utc.fr/~tdenoeux`

Belief Functions Encyclopedia:

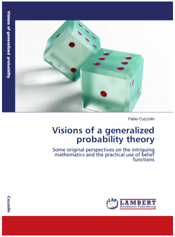
`http://cms.brookes.ac.uk/staff/FabioCuzzolin`

These slides are available online at:

`/FabioCuzzolin/uai-tutorial.pdf`

THANK YOU!

For Further Reading I

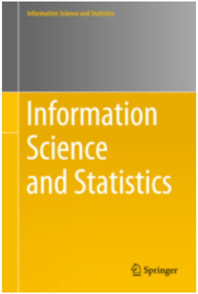


G. Shafer.
A mathematical theory of evidence.
Princeton University Press, 1976.

F. Cuzzolin.
Visions of a generalized probability theory.
Lambert Academic Publishing, 2014.

F. Cuzzolin (Ed.).
Belief functions: theory and applications.
LNCS Volume 8764, Springer, 2014.

For Further Reading I



F. Cuzzolin.
Fifty years of belief functions: a survey. Part I: Theory
International Journal of Approximate Reasoning (in preparation) 2000.

F. Cuzzolin and C. Sengul.
Fifty years of belief functions: a survey. Part II: Applications
International Journal of Approximate Reasoning (in preparation) 2000.

F. Cuzzolin.
The geometry of uncertainty.
Springer-Verlag, 2016.