Representation of evidence

Workshop on belief functions

Thierry Denœux

April, 2016

This chapter

- In this chapter, we define some of the main concepts of Dempster-Shafer theory in the finite case.
- These notions are sufficient to cope with a large number of applications.
- The extension to infinite spaces involves some mathematical intricacies and is technically more difficult, except in some simple (and practically important) cases; it will be addressed later.

Overview

Overview

- Mass function
  - Definitions
  - Semantics
- Belief and plausibility functions
  - Definitions
  - Properties
- Special cases and related theories
  - Bayesian mass functions
  - Consonant mass functions
  - Relation with imprecise probabilities

Frame of discernment

- Let Ω be a finite set of possible answers to some question Q, one and only one of which is true.
- The true answer will be denoted by $\boldsymbol{\omega}$, and an arbitrary element of Ω by ω.
- Shafer (1976) calls such a space a frame of discernment, to emphasize the fact that it is not a set of "states of nature" objectively given, but a subjective construction based on our state of knowledge.
- For instance, if Q relates to a person's state of health, Ω might contain only the diseases known at a certain time. This set could be later refined or extended if new knowledge became available.

Mass function

- A piece of evidence about Q will be represented by a mass function, defined as a mapping m from the power set 2^Ω to the interval $[0, 1]$, such that $m(\emptyset) = 0$ and

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

- Each number $m(A)$ represents the probability that the evidence supports exactly the proposition $\omega \in A$, and no more specific proposition.
- Any subset A of Ω such that $m(A) > 0$ is called a focal set of m. The union of the focal sets of a mass function is called its core.

Special cases

1. If m has only one focal set, it is said to be logical. Logical mass functions are in one-to-one correspondence with subsets of Ω: consequently, general mass functions can be viewed as generalized sets. A particular logical mass function plays a special role in the theory; it is the vacuous mass function $m_?$ defined by $m_?(\Omega) = 1$; such a mass function corresponds to a totally uninformative piece of evidence.
2. If all focal sets are singletons (i.e., sets of cardinality one), m is said to be Bayesian. To each Bayesian mass function can be associated a probability distribution $p : \Omega \to [0, 1]$ such that $p(\omega) = m(\{\omega\})$ for all $\omega \in \Omega$.

Example

- Consider the mass on functions on $\Omega = \{a, b, c\}$ shown in below. Mass function m_1 is Bayesian, m_2 is logical, $m_?$ is vacuous, and m_3 has no special form.

A	$\emptyset$	$\{a\}$	$\{b\}$	$\{a, b\}$	$\{c\}$	$\{a, c\}$	$\{b, c\}$	$\{a, b, c\}$
$m_1(A)$	0	0.2	0.5	0	0.3	0	0	0
$m_2(A)$	0	0	0	1	0	0	0	0
$m_?(A)$	0	0	0	0	0	0	0	1
$m_3(A)$	0	0.1	0.05	0.2	0.15	0.3	0.1	0.1

Overview

The murder example

- A murder has been committed and there are three suspects: Peter, John and Mary.
- The question Q of interest is the identity of the murderer and the frame of discernment is $\Omega = \{\text{Peter}, \text{John}, \text{Mary}\}$.
- The piece of evidence under study is a testimony: a witness saw the murderer. However, this witness is short-sighted and he can only report that he saw a man.
- Unfortunately, this testimony is also not fully reliable, because we know that the witness is drunk 20 % of the time.
- How can such a piece of evidence be encoded in the language of mass functions?

Formalization

- We can see here that what the testimony tells us about Q depends on the answer to another question Q': Was the witness drunk at the time of the murder?
- If he was not drunk, we know that the murderer is Peter or John. Otherwise, we know nothing.
- Since there is 80% chance that the former hypothesis holds, we may assign a 0.8 mass to the set $\{\text{Peter}, \text{John}\}$, and 0.2 to Ω:

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

A message with random meaning

- In the above example, we receive a message (a testimony) about Q, whose meaning depends on the answer to a related question Q' for which we have a chance model (a probability distribution).
- We can compare our evidence to a canonical example where we know that the outcomes of a random experiment are s_1 and s_2 with corresponding chances $p_1 = 0.8$ and $p_2 = 0.2$, and the message can only be interpreted with knowledge of the outcome.
- If the outcome is s_1, then the meaning is $\omega \in \{\text{Peter}, \text{John}\}$, otherwise the meaning is $\omega \in \Omega$, i.e., the message is totally uninformative.

Canonical examples

- We have seen that, in the constructive approach, probability judgements can be made by comparing the available evidence to some canonical example involving a chance setup.
- In the Bayesian theory, we compare our evidence to a situation where the truth is governed by chance (e.g., by thinking of the murderer as having been selected at random).
- In the belief function approach, the canonical example describes a situation where the meaning of the evidence is governed by chance.
- Two scenarios are specially useful to construct canonical examples for mass functions.

The unreliable machine

- The first scenario involves a machine that has two modes of operation, normal and faulty. We know that in the normal mode it broadcasts true messages, but we are completely unable to predict what it does in the faulty mode.
- We further assume that the operating mode of the machine is random and there a chance p that it is in the normal mode.
- It is then natural to say that a message $\omega \in A$ produced by the machine has a chance p of meaning what it says and a chance $1 - p$ of meaning nothing.
- This leads to the mass function $m(A) = p$ and $m(\Omega) = 1 - p$.
- Such a mass function, with two focal sets including Ω, is called a simple mass function.

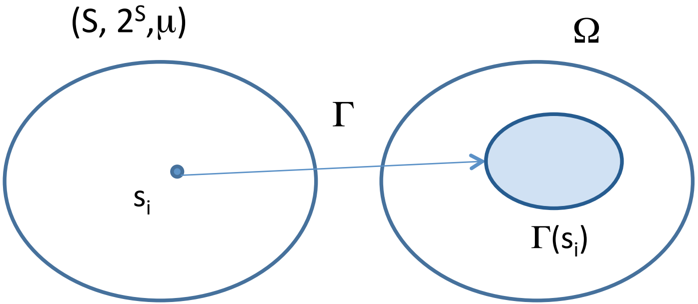
Random code

- The above story is not general enough to cover all kinds of evidence.
- More sophisticated scenario: a source holds some true information of the form $\omega \in A^*$ for some $A^* \subseteq \Omega$.
- It sends us this information as an encoded message using a code chosen at random from a set of codes $S = \{s_1, \ldots, s_r\}$, according to some known probability measure μ.
- We know the set of codes as well as the chances of each code to be selected. If we decode the message using code s, we get a decoded message of the form $\omega \in \Gamma(s)$ for some subset $\Gamma(s)$ of Ω. Then,

$$m(A) = \mu(\{s \in S | \Gamma(s) = A\}) \tag{2}$$

is the chance that the original message was "$\omega \in A$", i.e., the probability of knowing that $\omega \in A$, and nothing more.

The random code – continued



Random set

- In the above framework, the mapping $\Gamma : S \to 2^\Omega \setminus \{\emptyset\}$ is called a multi-valued mapping and the 4-tuple $(S, 2^S, \mu, \Gamma)$ is called a source.
- We can observe that a source corresponds formally to a random set.
- However, the term "random set" may be misleading here, because we are not interested in situations where a set is selected at random (such as, e.g., drawing a handful of marbles from a bag).
- Here, the true answer to the question of interest is a single element of Ω and it is not assumed to have been selected at random. Instead, chances are introduced when comparing our evidence to a situation where the meaning of a message depends on the result of a random experiment.

Relation between random sets and mass functions

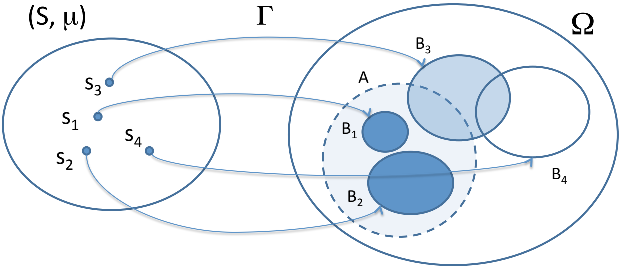
- It is clear that a source $(S, 2^S, \mu, \Gamma)$ always induces a mass function.
- Conversely, any mass function can be seen as generated by a source. For instance, if $A_1, \ldots, A_n$ are the focal sets of a mass function m, we may set $S = \{1, \ldots, n\}$ and $\mu(\{i\}) = m(A_i)$ for $1 \leq i \leq n$.
- However, as we will see later, the concept of a source is more general than that of mass function, because a source can be used in the infinite case to generate a belief function even when a mass function does not exist.

Overview

Overview

Example

- Assume the available evidence to be encoded by a mass function m on Ω generated by a source $(S, 2^S, \mu, \Gamma)$.
- How to quantify the uncertainty on the proposition "$\omega \in A$"?



Belief function

- For any $A \subseteq \Omega$, the probability that the evidence supports (implies) A is

$$\begin{aligned} Bel(A) &= \mu(\{s \in S | \Gamma(s) \subseteq A\}) && (3a) \\ &= \sum_{B \subseteq A} m(B); && (3b) \end{aligned}$$

- The quantity $Bel(A)$ can be interpreted as a degree of support for proposition A, or as a degree of belief. The function $Bel : 2^{\Omega} \to [0, 1]$ is called a belief function.

Plausibility function

- The probability that the evidence does not contradict A is

$$\begin{aligned} Pl(A) &= \mu(\{s \in S | \Gamma(s) \cap A \neq \emptyset\}) & \text{(4a)} \\ &= \sum_{B \cap A \neq \emptyset} m(B). & \text{(4b)} \end{aligned}$$

- $Pl(A)$ can be seen as the degree to which one fails to doubt A; this number is called the plausibility of A and the function $Pl : 2^{\Omega} \to [0, 1]$ is called a plausibility function.

Relation between belief and plausibility

- The uncertainty pertaining to the proposition $\omega \in A$ is quantified by two numbers: $Bel(A)$ and $Pl(A)$.
- We have

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

$$Pl(A) = 1 - Bel(\overline{A})$$

- Special case: if $m(\Omega) = 1$, then

$$Bel(A) = 0, \quad \forall A \subset \Omega$$

$$Pl(A) = 1, \quad \forall A \subseteq \Omega, A \neq \emptyset$$

Overview

Complete monotonicity

Theorem

A function $Bel : 2^{\Omega} \rightarrow [0,1]$ *is a belief function iff it satisfies the following conditions:*

1. $Bel(\emptyset) = 0$;
2. $Bel(\Omega) = 1$;
3. *For any* $k \geq 2$ *and any collection* $A_1, \ldots, A_k$ *of subsets of* Ω,

$$Bel\left(\bigcup_{i=1}^{k} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} A_i\right). \quad (5)$$

A function verifying (5) for given k is said to be monotone of order k. A belief function is monotone of order k for any k (monotone of order infinite).

Möbius transform

Theorem

Let $Bel : 2^{\Omega} \to [0,1]$ *be a belief function induced by a mass function* m*. Then*

$$m(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|} Bel(B), \qquad (6)$$

for all $A \subseteq \Omega$*.*

Completely alternating capacity

Theorem

A function $Pl : 2^\Omega \rightarrow [0,1]$ *is a plausibility function iff it satisfies the following conditions:*

1. $Pl(\emptyset) = 0$;
2. $Pl(\Omega) = 1$;
3. *For any* $k \geq 2$ *and any collection* $A_1, \ldots, A_k$ *of subsets of* Ω,

$$Pl\left(\bigcap_{i=1}^{k} A_i\right) \leq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} Pl\left(\bigcup_{i \in I} A_i\right). \quad (7)$$

A set function verifying (7) is said to be alternating of order infinite, or completely alternating. A plausibility function is thus a completely alternating set function Pl such that $Pl(\emptyset) = 0$ and $Pl(\Omega) = 1$.

Recovering m from Pl

Theorem

Let $Pl : 2^\Omega \to [0, 1]$ be a plausibility function induced by a mass function m. Then

$$m(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|+1} Pl(\overline{B}), \tag{8}$$

for all $A \subseteq \Omega$.

Equivalence of representations

- From the above results, it is clear that, given any of the three functions m, Bel and Pl, we can recover the other two.
- Consequently, these three functions can be seen as different facets of the same information.
- In the sequel, we will sometimes use the term "belief function" to refer to any of these functions, when there will be no risk of confusion.

Overview

Mass function
- Definitions
- Semantics

Belief and plausibility functions
- Definitions
- Properties

Special cases and related theories
- Bayesian mass functions
- Consonant mass functions
- Relation with imprecise probabilities

Overview

Bayesian belief functions are Probability measures

- If m is Bayesian, then

$$Bel(A) = Pl(A) = \sum_{\omega \in A} m(\{\omega\})$$

  for any $A \subseteq \Omega$.

- Furthermore, for any two disjoint subsets A and B of Ω,

$$Bel(A \cup B) = \sum_{\omega \in A \cup B} m(\{\omega\}) = \sum_{\omega \in A} m(\{\omega\}) + \sum_{\omega \in B} m(\{\omega\}) = Bel(A) + Bel(B). \quad (9)$$

- Consequently, belief functions induced by Bayesian mass functions are probability measures and are equal to their dual plausibility functions.

Bayesian belief function=Probability measure

- Conversely, it is clear that each probability measure P is a belief function induced by the Bayesian mass function m such that $m(\{\omega\}) = P(\{\omega\})$ for all $\omega \in \Omega$.
- The set of probability measures is thus exactly the set of belief functions induced by Bayesian mass functions.
- This results shows us that the language of belief functions is more general than that of probability theory.
- As we will see, the conditioning operation, which plays a major role in updating beliefs based on new evidence in the Bayesian framework, can also be seen as a special case of a more general operation in the belief function framework.

Overview

Consonant plausibility functions are possibility measures

- A mass function m is said to be consonant if its focal sets are nested, i.e., if they can be arranged in an increasing sequence $A_1 \subset \ldots \subset A_r$.

Theorem

Let m be a consonant mass function. Then, the belief and plausibility functions verify the following properties for all $A, B \subseteq \Omega$,

$$Bel(A \cap B) = \min(Bel(A), Bel(B)), \tag{10}$$

$$Pl(A \cup B) = \max(Pl(A), Pl(B)). \tag{11}$$

- Properties (10) and (11) characterize, respectively, possibility and necessity measures, which form the basis of Possibility theory introduced by Zadeh (1978).

Contour function

- An important consequence of (11) is that function Pl can be deduced from its restriction to singletons.
- More precisely, let $pl : \Omega \to [0, 1]$ be the contour function of m, defined by $pl(\omega) = Pl(\{\omega\})$, for all $\omega \in \Omega$.
- For all $A \subseteq \Omega$,

$$Pl(A) = \max_{\omega \in A} pl(\omega). \tag{12}$$

- The condition $Pl(\Omega) = 1$ implies that $\max_{\omega \in \Omega} pl(\omega) = 1$. The contour function pl is then the possibility distribution associated to the possibility measure Pl.

Possibility measures are consonant plausibility functions

Theorem

Let π be a possibility distribution on the frame $\Omega = \{\omega_1, \ldots, \omega_n\}$, with elements arranged by decreasing order of plausibility, i.e.,

$$1 = \pi(\omega_1) \geq \pi(\omega_2) \geq \ldots \geq \pi(\omega_n),$$

and let A_i denote the set $\{\omega_1, \ldots, \omega_i\}$, for $1 \leq i \leq n$. Then, π is the contour function for a mass function m obtained by the following formula:

$$m(A_i) = \pi(\omega_i) - \pi(\omega_{i+1}), \quad 1 \leq i \leq n-1, \tag{13}$$

$$m(\Omega) = \pi(\omega_n). \tag{14}$$

Example

- Consider, for instance, the following possibility distribution defined on the frame $\Omega = \{a, b, c, d\}$:

ω	a	b	c	d
$\pi(\omega)$	0.3	0.5	1	0.7

- The corresponding mass function is

$$\begin{aligned} m(\{c\}) &= 1 - 0.7 = 0.3 \\ m(\{c, d\}) &= 0.7 - 0.5 = 0.2 \\ m(\{c, d, b\}) &= 0.5 - 0.3 = 0.2 \\ m(\{c, d, b, a\}) &= 0.3. \end{aligned}$$

Overview

Coherent lower and upper probabilities

- Let $\mathcal{P}$ be a non empty set of probability measures on some frame Ω.
- Its lower and upper envelopes are set functions defined as follows:

$$P_*(A) = \inf_{P \in \mathcal{P}} P(A), \tag{15a}$$

$$P^*(A) = \sup_{P \in \mathcal{P}} P(A). \tag{15b}$$

  for all subsets A of Ω.
- Functions P_* and P^* are called, respectively, coherent lower and upper probabilities.
- Clearly, $P^*(A) = 1 - P_*(\overline{A})$ for all A, which is reminiscent of the relation between belief and plausibility functions.
- What is the relation between these notions?

Credal set

- To each belief function Bel we can associate the set of probability measures P that dominate Bel, i.e., the set of probability measures such that $P(A) \geq Bel(A)$ for all subset A of Ω.
- Because of the relation $Bel(A) = 1 - Pl(\overline{A})$, we also have $P(A) \leq Pl(A)$ for all A, or

$$Bel(A) \leq P(A) \leq Pl(A), \quad \forall A \subseteq \Omega. \tag{16}$$

- Any probability measure P verifying (16) is said to be compatible with Bel, and the set $\mathcal{P}(Bel)$ of all probability measures compatible with Bel is called the credal set of Bel.

Allocation of probability

- An arbitrary element of $\mathcal{P}(Bel)$ can be obtained by distributing each mass $m(A)$ among the elements of A.
- More precisely, let us call an allocation of m any function

$$\alpha : \Omega \times 2^{\Omega} \setminus \{\emptyset\} \to [0,1] \tag{17}$$

such that, for all $A \subseteq \Omega$,

$$\sum_{\omega \in A} \alpha(\omega, A) = m(A). \tag{18}$$

- Each quantity $\alpha(\omega, A)$ can be viewed as a part of $m(A)$ allocated to the element ω of A.
- By summing up the numbers $\alpha(\omega, A)$ for each ω, we get a probability mass function on Ω,

$$p_{\alpha}(\omega) = \sum_{A \ni \omega} \alpha(\omega, A). \tag{19}$$

Belief functions are coherent lower probabilities

- It can be shown (Dempster, 1967) that the set of probability measures constructed in that way is exactly equal to the credal set $\mathcal{P}(Bel)$.
- Furthermore, the bounds in (16) are attained. A belief function is thus a coherent lower probability.
- However, a coherent lower probability is not always a belief function.

A counterexample

- Suppose a fair coin is tossed twice, in such a way that the outcome of the second toss may depend on the outcome of the first toss.
- The outcome of the experiment can be denoted by $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$.
- Let $H_1 = \{(H,H),(H,T)\}$ and $H_2 = \{(H,H),(T,H)\}$ the events that we get Heads in the first and second toss, respectively.
- Let $\mathcal{P}$ be the set of probability measures on Ω which assign $P(H_1) = P(H_2) = 1/2$ and have an arbitrary degree of dependence between tosses.
- Let P_* be the lower envelope of $\mathcal{P}$.

A counterexample – continued

- It is clear that $P_*(H_1) = 1/2$, $P_*(H_2) = 1/2$ and $P_*(H_1 \cap H_2) = 0$ (as the occurrence Heads in the first toss may never lead to getting Heads in the second toss).
- Now, in the case of complete positive dependence, $P(H_1 \cup H_2) = P(H_1) = 1/2$, hence $P_*(H_1 \cup H_2) \leq 1/2$.
- We thus have

$$P_*(H_1 \cup H_2) < P_*(H_1) + P_*(H_2) - P_*(H_1 \cap H_2), \qquad (20)$$

which violates the complete monotonicity condition (5) for $k = 2$.

Two different theories

- Mathematically, the notion of coherent lower probability is thus more general than that of belief function.
- However, the definition of the credal set associated with a belief function is purely formal, as these probabilities have no particular interpretation in our framework.
- The theory of belief functions is not a theory of imprecise probabilities.