Introduction to belief functions
Lecture 2: Representation of evidence

T. Denœux

July 18, 2014

In this lecture, 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 is postponed to a following lecture.

1 Mass function

1.1 Definitions

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 [2] 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. We call back later to the important issue of defining and modifying the frame of discernment.

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

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

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*.

Before discussing the semantics of a mass function, it is interesting to point out two 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 $m_?(\Omega) = 1$; as we will see later, such a mass function corresponds to an 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 $\omega \in \Omega$.

A belief function may thus be viewed both as a generalized set and as a non-additive measure. As we will see in following lectures, 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).

1.2 Semantics

The following example will show how the formalism of mass functions can be used to represent a piece of evidence. It will also serve as an illustration of the semantics of mass functions.

Example 1 *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 $\Omega = \{Peter, John, Mary\}$. 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 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?*

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 $\{Peter, John\}$, and 0.2 to Ω:

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

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 o_1 and o_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 o_1, then the meaning is $\boldsymbol{\omega} \in \{Peter, John\}$, otherwise the meaning is the meaning is $\boldsymbol{\omega} \in \Omega$, i.e., the message is totally uninformative.

As remarked by Shafer [3], 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.

More precisely, two scenarios are specially useful to construct canonical examples for mass functions.

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 $\boldsymbol{\omega} \in A$ 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.

The above story is simple and very useful to model situations in which a partially reliable source of information provides a simple statement of the form $\boldsymbol{\omega} \in A$ and we can assess the probability of the source to be reliable. How, it is not general enough to cover all kinds of evidence. In [3], Shafer introduced a more sophisticated scenario that is general enough to produce canonical examples for arbitrary mass functions. In this scenario, a source holds some *true information* of the form $\boldsymbol{\omega} \in A^*$ 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 $U = \{c_1, \ldots, c_r\}$, according to some known probability measure μ (Figure 1). We know the set of codes as well as the chances of each code to be selected. If we decode the message using code c, we get a decoded message of the form $\boldsymbol{\omega} \in \Gamma(c)$ for some subset $\Gamma(c)$ of Ω. Then,

$$m(A) = \mu(\{c \in U | \Gamma(c) = A\}) \tag{2}$$

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

In the above framework, the mapping $\Gamma : U \to 2^\Omega \setminus \{\emptyset\}$ is called a multi-valued mapping and the triple (U, μ, Γ) is called a source. We can observe that a source corresponds formally to a random set [1]. 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

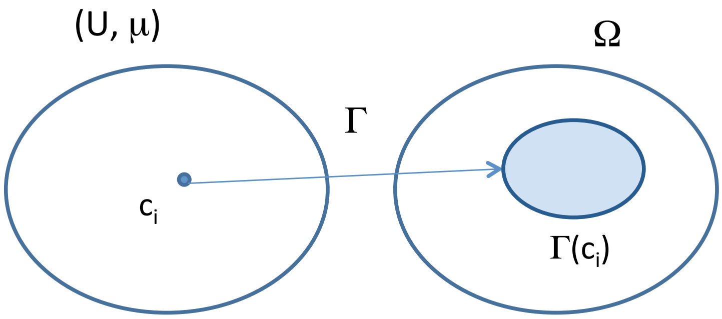

Figure 1: Random code setup.

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.

It is clear that a source (U, μ, Γ) always induces a mass function from (1) and, conversely, any mass function can be seen as generated by a source. However, as will shall see, the concept of a source is more general than that of mass function, because a source can be used in the infinite case to general belief functions even when a mass function does not exist.

2 Belief and plausibility functions

2.1 Definitions

Let us assume the available evidence to be encoded by a mass function m on Ω generated by a source (U, μ, Γ). For any $A \subseteq \Omega$, the uncertainty of the proposition $\boldsymbol{\omega} \in A$ can be quantified by two numbers:

1. The probability that the evidence implies A, defined by

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

2. The probability that the evidence does not contradict A, given by

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

Clearly, $Bel(\emptyset) = Pl(\emptyset) = 0$, $Bel(\Omega) = Pl(\Omega) = 1$, $Bel(A) \leq Pl(A)$ and $Pl(A) = 1 - Bel(\overline{A})$, where $\overline{A}$ denotes the complement of A. 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.* In contrast, $Pl(A)$ 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.*

2.2 Properties

Theorem 1 *A function $Bel : 2^\Omega \to [0,1]$ 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,*

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

Proof: See [2, page 51].

In general, a function satisfying (5) for a given k is said to be monotone of order k. It is clear that monotonicity of order k implies monotonicity of order k' for all $k' < k$. A function that is monotone for any k is said to be monotone of order infinite, or completely monotone. Furthermore, properties 1 and 2 above imply that Bel is increasing. To see this, let A and B be two subsets of Ω such that $A \subseteq B$ and let $C = B \setminus A$. We have $B = A \cup C$ and $A \cap C = \emptyset$. From (5) with $k = 2$, we have

$$\begin{aligned} Bel(A \cup C) = Bel(B) \geq Bel(A) + Bel(C) - Bel(A \cap C) \\ = Bel(A) + Bel(C) \geq Bel(A). \end{aligned} \tag{6}$$

Theorem 1 tells us that a completely monotone set function such that $Bel(\emptyset) = 0$ and $Bel(\Omega) = 1$ is induced by some mass function m using (3b). We may wonder whether there exists a unique m generating a belief function Bel. Indeed, (3b) for $A \in 2^\Omega \setminus \{\emptyset, \Omega\}$ provide $2^{|\Omega|} - 2$ equations and there are $2^{|\Omega|} - 2$ free mass numbers (taking into account constraint (1)). Consequently, one must be able to recover m from Bel in a unique way. The following theorem states that m is actually the Möbius inverse of Bel, a notion from Combinatorial theory [2].

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

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

for all $A \subseteq \Omega$.

Proof: See [2, page 52].

Using the identity $Pl(A) = 1 - Bel(\overline{A})$ for any $A \subseteq \Omega$, it is easy to obtain the following theorem, which is a counterpart of Theorem 1:

Theorem 3 *A function $Pl : 2^\Omega \to [0, 1]$ 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,*

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

A set function verifying the third property in Theorem 3 is said to be alternating of order infinite, or completely alternating. A plausibility function is a completely alternating set function Pl such that $Pl(\emptyset) = 0$ and $Pl(\Omega) = 1$.

From the above result, 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.

3 Special cases and relation with other theories

3.1 Bayesian mass functions

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,

$$Bel(A \cup B) = \sum_{\omega \in A \cup B} m(\{\omega\}) = \\ \sum_{\omega \in A} m(\{\omega\}) + \sum_{\omega \in B} m(\{\omega\}) = Bel(A) + Bel(B). \tag{8}$$

Consequently, belief functions induced by Bayesian mass functions are probability measures and are equal to their dual plausibility functions. 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 $\omega \in \Omega$.

In other terms, the set of probability measures is 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 later, 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.

3.2 Consonant mass functions

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 $A_1 \subset \ldots \subset A_r$. In that case, functions Bel and Pl satisfy the following properties.

For any $A, B \subseteq \Omega$, let i_1 and i_2 be the largest indices such that $A_i \subseteq A$ and $A_i \subseteq B$, respectively. Then, $A_i \subseteq A \cap B$ iff $i \leq \min(i_1, i_2)$ and

$$
\begin{aligned}
Bel(A \cap B) &= \sum_{i=1}^{\min(i_1,i_2)} m(A_i) && \text{(9a)}\\
&= \min\left(\sum_{i=1}^{i_1} m(A_i), \sum_{i=1}^{i_2} m(A_i)\right) && \text{(9b)}\\
&= \min(Bel(A), Bel(B)). && \text{(9c)}
\end{aligned}
$$

Now, it is easy to deduce, from De Morgan laws, that

$$
Pl(A \cup B) = \max(Pl(A), Pl(B)). \tag{10}
$$

Properties (9c) and (10) characterize, respectively, possibility and necessity measures, which form the basis of Possibility theory introduced by Zadeh in [5]. In this theory, $Pl(A)$ is the degree to which proposition A is possible, and $Bel(A)$ is the degree to which A is certain, i.e., the degree to which $\overline{A}$ is impossible. As possibility measures are special plausibility functions (induced by consonant mass functions), the theory of belief functions can be considered as more expressive than Possibility theory. However, as we shall see, the two theories depart in the way different pieces of information are combined: in the belief function approach, a mass function resulting from the combination of two consonant mass functions will generally not be consonant.

An important consequence of (10) 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 $A \subseteq \Omega$,

$$
Pl(A) = \max_{\omega \in A} pl(\omega). \tag{11}
$$

We note that the condition $Pl(\Omega) = 1$ implies that $\max_{\omega \in \Omega} pl(\omega) = 1$.

We have seen that the plausibility function induced by a consonant mass function is a possibility measure. Conversely, a possibility measure is always a plausibility function for some consonant mass function, which can be recovered from the contour function as follows. Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be the frame of discernment, with elements arranged by decreasing order of plausibility, i.e.,

$$1 = pl(\omega_1) \geq pl(\omega_2) \geq \ldots \geq pl(\omega_n).$$

Then, the corresponding mass function is obtained by the following formula:

$$\begin{aligned}
m(\emptyset) &= 0 \\
m(\{\omega_1\}) &= pl(\omega_1) - pl(\omega_2) \\
&\vdots \\
m(\{\omega_1, \ldots, \omega_i\}) &= pl(\omega_i) - pl(\omega_{i+1}) \\
&\vdots \\
m(\Omega) &= pl(\omega_n).
\end{aligned}$$

Possibility theory has a strong connection with the theory of Fuzzy Sets. More precisely, if we receive evidence of the form "$\boldsymbol{\omega}$ is F", where F is a fuzzy subset of Ω with membership function μ_F, then this piece of evidence may be represented by a consonant belief function with contour function $pl = \mu_F$.

3.3 Relation with imprecise probabilities

To each belief function Bel can be associated 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

$$Bel(A) \leq P(A) \leq Pl(A), \quad \forall A \subseteq \Omega. \tag{12}$$

Any probability measure P verifying (12) 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. Any compatible probability can be obtained by distributing each mass $m(A)$ among the elements of A.

A belief function can thus be seen the lower envelope of a non-empty set of probability measure. Such a function is called a coherent lower probability [4]. However, a coherent lower probability is not always a belief function: in some sense, the notion of coherent lower probability is thus more general. Anyway, the definition of the credal set associated with a belief function is purely formal, as these probabilities have no particular interpretation. The theory of belief functions is not a theory of imprecise probabilities.

References

[1] H.T. Nguyen. *An Introduction to Random Sets*. Chapman and Hall/CRC Press, Boca Raton, Florida, 2006.

[2] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.

[3] G. Shafer. Constructive probability. *Synthese*, 48(1):1–60, 1981.

[4] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[5] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.