Introduction to belief functions
Belief functions on infinite spaces

T. Denœux

July 27, 2014

Until now, the presentation of belief functions has been restricted to the case where Ω is finite. The theory of belief functions on finite frames is sufficient to represent expert opinions, because any infinite frame can always be coarsened to a finite one, which is more easily conceived by an expert. However, in some applications, the restriction to finite frames does appear as a limitation. For instance, in most statistical models, the parameter space is $\mathbb{R}^d$ for some $d \geq 1$. It is thus useful to extend the theory from finite to infinite (continuous) spaces. This extension involves, in the most general case, considerably more mathematical sophistication than involved in the finite case. In the presentation below, we will try to avoid entering technical details and we will focus on the simplest models, which are sufficient for most applications, in particular to statistical inference.

In Lecture 2, we have noticed the formal connection between belief functions and random sets. This connection remains valid in the infinite case and, as the theory of random sets is well developed mathematically [3], it will provide a solid foundation for a theory of belief functions in infinite spaces.

1 General definitions and results

In the finite case, we derived the notion of belief function from that of mass function, and we later showed the equivalence with the complete monotonicity condition. In the infinite case, there may not be a mass function associated with a completely monotone function, so that we have to define a belief function axiomatically from its properties (the most important one being complete monotonicity).

1.1 Definitions

Let $(\Omega, \mathcal{B})$ be a measurable space (i.e., $\mathcal{B}$ is a sigma-field, that is a non-empty subset of 2^Ω closed under complementation and countable union). A belief function on $\mathcal{B}$ is a function $Bel : \mathcal{B} \to [0,1]$ verifying the following three conditions:

1. $Bel(\emptyset) = 0$;
2. $Bel(\Omega) = 1$;
3. For any $k \geq 2$ and any collection $B_1, \ldots, B_k$ of elements of $\mathcal{B}$,

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

Furthermore, a belief function Bel on $(\Omega, \mathcal{B})$ is continuous [4] if for any decreasing sequence $B_1 \supset B_2 \supset B_3 \supset \ldots$ of elements of $\mathcal{B}$,

$$\lim_{i \to +\infty} Bel(B_i) = Bel\left(\bigcap_{i} B_i\right). \tag{2}$$

Similarly, a plausibility function can be defined as a function $Pl : \mathcal{B} \to [0,1]$ such that:

1. $Pl(\emptyset) = 0$;
2. $Pl(\Omega) = 1$;
3. For any $k \geq 2$ and any collection $B_1, \ldots, B_k$ of elements of $\mathcal{B}$,

$$Pl\left(\bigcap_{i=1}^{k} B_i\right) \leq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} Pl\left(\bigcup_{i \in I} B_i\right), \tag{3}$$

and it is continuous if, for any increasing sequence $B_1 \subset B_2 \subset B_3 \subset \ldots$ of elements of $\mathcal{B}$,

$$\lim_{i \to +\infty} Pl(B_i) = Pl\left(\bigcup_{i} B_i\right). \tag{4}$$

It is clear that, whenever Bel is a belief function, Pl defined by $Pl(A) = 1 - Bel(\overline{A})$ is a plausibility function.

1.2 Belief function induced by a source

A very convenient way to create a belief function is to define a source, i.e., a multivalued mapping from a probability space to $\mathcal{B}$. More precisely, let $(U, \mathcal{A}, \mu)$ be a probability space and let $\Gamma : U \to 2^{\Omega}$ be a multi-valued mapping. We can define two inverses of Γ:

1. The lower inverse

$$\Gamma_*(B) = \{u \in U | \Gamma(u) \neq \emptyset, \Gamma(u) \subseteq B\}; \tag{5}$$

2. The upper inverse

$$\Gamma^*(B) = \{u \in U | \Gamma(u) \cap B \neq \emptyset\}, \tag{6}$$

for all $B \in \mathcal{B}$. We say that Γ is *strongly measurable* with respect to $\mathcal{A}$ and $\mathcal{B}$ iff, for all $B \in \mathcal{B}$, $\Gamma^*(B) \in \mathcal{A}$.

We then have the following important theorem.

Theorem 1 *Let $(U, , \mu)$ be a probability space, $(\Omega, \mathcal{B})$ a measurable space and Γ a strongly measurable mapping w.r.t. $\mathcal{A}$ and $\mathcal{B}$ such that $\mu(\Gamma^*(\Omega)) = 1$. Let the lower and upper probability measures be defined as follows:*

$$\mu_*(B) = \mu[\Gamma_*(B)], \tag{7}$$

$$\mu^*(B) = \mu[\Gamma^*(B)] = 1 - \mu_*(\overline{B}), \tag{8}$$

for all $B \in \mathcal{B}$. Then, μ_* is a continuous belief function and μ^* is the dual plausibility function.

Proof. We can remark that

$$\Gamma_*\left(\bigcap_i B_i\right) = \bigcap_i \Gamma_*(B_i) \tag{9}$$

and

$$\Gamma^*\left(\bigcup_i B_i\right) \supseteq \bigcup_i \Gamma_*(B_i). \tag{10}$$

Consequently, for any k and any collection $B_1, \ldots, B_k$ of elements of $\mathcal{B}$,

$$\sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \mu_*\left(\bigcap_{i \in I} B_i\right) =$$
$$\sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \mu\left(\bigcap_{i \in I} \Gamma_*(B_i)\right) = \mu\left(\bigcup_{i \in I} \Gamma_*(B_i)\right)$$
$$\leq \mu\left[\Gamma_*\left(\bigcup_i B_i\right)\right] = \mu_*\left(\bigcup_i B_i\right). \tag{11}$$

Now, for any decreasing sequence $B_1 \supset B_2 \supset B_3 \supset \ldots$ of elements of $\mathcal{B}$, $\Gamma_*(B_1) \supset \Gamma_*(B_2) \supset \Gamma_*(B_3) \supset \ldots$ is a decreasing sequence of elements of $\mathcal{A}$.

Consequently,

$$\lim_{i\to+\infty}\mu_*(B_i) = \lim_{i\to+\infty}\mu\left(\Gamma_*(B_i)\right) =$$
$$\mu\left(\bigcap_i \Gamma_*(B_i)\right) = \mu\left[\Gamma_*\left(\bigcap_i B_i\right)\right] = \mu_*\left(\bigcap_i B_i\right). \quad (12)$$

Thus, to define a belief function on $(\Omega, \mathcal{B})$, it suffices to define a probability space $(U, \mathcal{A}, \mu)$ and a strongly measurable mapping Γ from U to $\mathcal{A}$. By analogy with the finite case, the sets $\Gamma(u)$ for $u \in U$ can be called the focal sets of Bel.

We can remark that, to insure the existence of a commonality function Q, we have to impose that, for any $B \in \mathcal{B}$, $\{u \in U | \Gamma(u) \supseteq B\}$ is in $\mathcal{A}$. We can then define the commonality function as

$$Q(B) = \mu\left(\{u \in U | \Gamma(u) \supseteq B\}\right). \quad (13)$$

As shown by Shafer [4], any belief function Bel on $(\Omega, \mathcal{B})$ can be extended to $(\Omega, 2^\Omega)$ as

$$\widetilde{Bel}(A) = \sup\{Bel(B) | B \in \mathcal{B}, B \subseteq A\}, \quad (14)$$

for all $A \subseteq \Omega$.

1.3 Relationship with random sets

As remarked by Nguyen [2], any belief function constructed as described in the previous section is the probability distribution of a random set.

To define a random set, we need to define a σ-field $\widehat{\mathcal{B}}$ on 2^Ω. This can be done as follows: for any $\widehat{T} \subseteq 2^\Omega$,

$$\widehat{T} \in \widehat{\mathcal{B}} \Leftrightarrow \Gamma^{-1}(\widehat{T}) \in \mathcal{A}. \quad (15)$$

It is clear that the mapping Γ is $\mathcal{A}$-$\widehat{\mathcal{B}}$ measurable. Let $\widehat{\mu}$ be the probability measure on $(2^\Omega, \widehat{\mathcal{B}})$ defined by

$$\widehat{\mu}(\widehat{T}) = \mu[\Gamma^{-1}(\widehat{T})]. \quad (16)$$

For all $B \in \mathcal{B}$, let $I(B) = \{C \subseteq \Omega | C \subseteq B\}$. It is easy to see that

$$\widehat{\mu}[I(B)] = \mu_*(B). \quad (17)$$

Hence, μ_* is the distribution function of a random set.

When the set of Ω has a topological structure, some special classes of random sets are particularly amenable to mathematical analysis. In particular, let $\mathcal{C}$ be the set of closed subsets of Ω. For any $A \subseteq \Omega$, let

$$\mathcal{C}_A = \{C \in \mathcal{C} | C \cap A \neq \emptyset\}. \quad (18)$$

Let $(U, \mathcal{A}, \mu)$ be a probability space. Then a map $\Gamma : U \to \mathcal{C}$ is a random closed set if, for any A in $\mathcal{C}$, $\Gamma^{-1}(\mathcal{C}_A) \in \mathcal{A}$.

1.4 Consonant random closed sets

A practical way of constructing a random closed set is as follows. For simplicity, we will assume that $\Omega = \mathbb{R}^d$. Let π be an upper semi-continuous map from $\mathbb{R}^d$ to $[0,1]$, i.e., for any $u \in [0,1]$, the set

$$^u\pi = \{x \in \mathbb{R}^p | \pi(x) \geq u\} \tag{19}$$

is closed. Furthermore, assume that $\pi(x) = 1$ for some x. Let $U = [0,1]$, $\mathcal{A}$ be the Borel σ-field on $[0,1]$, μ the Lebesgue measure, and Γ the mapping defined by $\Gamma(u) = {}^u\pi$. Then Γ is a random closed set [3]. We can observe that its focal sets are nested: it is said to be *consonant*. Let Bel, Pl and Q be the corresponding belief, plausibility and commonality functions. Then, for any $A \subset \mathbb{R}^d$:

$$\begin{aligned}
Pl(A) &= \mu(\{u \in U | {}^u\pi \cap A \neq \emptyset\}) && \text{(20a)}\\
&= \mu(\{u \in U | \exists x \in A, \pi(x) \geq u\}) && \text{(20b)}\\
&= \mu(\{u \in U | u \leq \sup_{x \in A} \pi(x)\}) && \text{(20c)}\\
&= \sup_{x \in A} \pi(x), && \text{(20d)}
\end{aligned}$$

$$Bel(A) = 1 - Pl(\overline{A}) = 1 - \sup_{x \notin A} \pi(x) = \inf_{x \notin A} (1 - \pi(x)) \tag{21}$$

and

$$\begin{aligned}
Q(A) &= \mu(\{u \in U | {}^u\pi \supseteq A\}) && \text{(22a)}\\
&= \mu(\{u \in U | \forall x \in A, \pi(x) \geq u\}) && \text{(22b)}\\
&= \mu(\{u \in U | u \leq \inf_{x \in A} \pi(x)\}) && \text{(22c)}\\
&= \inf_{x \in A} \pi(x). && \text{(22d)}
\end{aligned}$$

In particular, $Pl\{x\} = Q(\{x\}) = \pi(x)$ for all x.

2 Random closed intervals

In this section, we consider the case where $\Omega = \mathbb{R}$. In this case, a special class of random closed set is of special interest: random closed intervals [1].

2.1 Definition and properties

Let (U, V) be a bi-dimensional random vector from $(S, \mathcal{A}, \mathbb{P})$ to $\mathbb{R}^2$ such that

$$\mathbb{P}(\{s \in S | U(s) \leq V(s)\}) = 1. \tag{23}$$

We can define the corresponding random closed set

$$\Gamma : s \to \Gamma(s) = [U(s), V(s)], \tag{24}$$

which is called a random closed interval. Two special cases are of interest:

1. If the random vector (U, V) is discrete, with $\mathbb{P}(U = u_i; V = V_i) = m_i$, we have a discrete random interval; it is characterized by a mass function m with focal sets $I_i = [u_i, v_i]$ and masses $m(I_i) = m_i$.

2. If (U, V) is absolutely continuous with density $f(u, v)$, we have a continuous random interval.

For all $x \in \mathbb{R}$, we have:

$$Bel((-\infty, x]) = \mathbb{P}([U, V] \subseteq (-\infty, x]) = \mathbb{P}(V \leq x) = F_V(x), \tag{25}$$

where F_V is the cumulative distribution function (cdf) of V, and

$$Pl((-\infty, x]) = \mathbb{P}([U, V] \cap (-\infty, x] \neq \emptyset) = \mathbb{P}(U \leq x) = F_U(x). \tag{26}$$

These functions are called, respectively, the lower and upper cdf of $[U, V]$. Now, for any $a \leq b$, we have

$$Bel([a, b]) = \mathbb{P}([U, V] \subseteq [a, b]) = \mathbb{P}(U \geq a; V \leq b), \tag{27}$$

$$\begin{aligned} Pl([a, b]) = \mathbb{P}([U, V] \cap [a, b] \neq \emptyset) = \\ 1 - \mathbb{P}([U, V] \cap [a, b] = \emptyset) = 1 - \mathbb{P}(V < a) - \mathbb{P}(U > b) \end{aligned} \tag{28}$$

and

$$Q([a, b]) = \mathbb{P}([U, V] \supseteq [a, b]) = \mathbb{P}(U \leq a; V \geq b). \tag{29}$$

We can observe that If $[U, V]$ is continuous, these probabilities can be computed by integrating the joint density $f(u, v)$. For instance,

$$Q([a, b]) = \int_{-\infty}^{a} \int_{b}^{+\infty} f(u, v) dv du, \tag{30}$$

$$Bel([a, b]) = \int_{a}^{b} \int_{u}^{b} f(u, v) dv du. \tag{31}$$

Conversely,

$$f(u, v) = -\frac{\partial^2 Q([a, b])}{\partial a \partial b} = -\frac{\partial^2 Bel([a, b])}{\partial a \partial b}. \tag{32}$$

2.2 Combination of random intervals

As in the finite case, random closed intervals can be combined using Dempster's rule. Let $[U_1, V_1]$ and $[U_2, V_2]$ be two random closed intervals, and let Q_1 and Q_2 be their commonality functions. We have the following equality:

$$(Q_1 \oplus Q_2)([a, b]) = \frac{1}{1 - \kappa} Q_1([a, b]) Q_2([a, b]), \tag{33}$$

where

$$\kappa = \mathbb{P}([U_1, V_2] \cap [U_2, V_2] = \emptyset) \tag{34}$$

is the degree of conflict between the two random sets. To see this, we may observe that

$$\begin{aligned}
(Q_1 \oplus Q_2)([a,b]) &= \mathbb{P}([U_1, V_1] \cap [U_2, V_2] \supseteq [a,b] | [U_1, V_1] \cap [U_2, V_2] \neq \emptyset) && \text{(35a)}\\
&= \frac{\mathbb{P}([U_1, V_1] \supseteq [a,b], [U_2, V_2] \supseteq [a,b])}{\mathbb{P}([U_1, V_1] \cap [U_2, V_2] \neq \emptyset)} && \text{(35b)}\\
&= \frac{Q_1([a,b]) Q_2([a,b])}{1-\kappa}. && \text{(35c)}
\end{aligned}$$

When $[U_1, V_1]$ and $[U_2, V_2]$ are continuous, the combination of $[U_1, V_1] \oplus [U_2, V_2]$ may be cumbersome or even intractable. We may then compute an approximation, either by discretizing the two random intervals, or by using Monte Carlo simulation. For instance, the following algorithm can be used to approximate $(Pl_1 \oplus Pl_2)(A)$ for some $A \subseteq \mathbb{R}$:

$k = 0$
for $i = 1 : N$ **do**
  Generate realizations $[u_1, v_1]$ and $[u_2, v_2]$ of $[U_1, V_1]$ and $[U_2, V_2]$
  $I = [u_1, v_1] \cap [u_2, v_2]$
  if $I \cap A \neq \emptyset$ **then**
    $k = k + 1$
  end if
end for
$(\widehat{Pl_1 \oplus Pl_2})(A) = \frac{k}{N}$

References

[1] A. P. Dempster. Upper and lower probabilities generated by a random closed interval. *Annals of Mathematical Statistics*, 39(3):957–966, 1968.

[2] H. T. Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65:531–542, 1978.

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

[4] G. Shafer. Allocations of probability. *Annals of Probability*, 7(5):827–839, 1979.