Independence and Combination of Belief Functions

Marco Cattaneo
Department of Statistics, LMU Munich
cattaneo@stat.uni-muenchen.de

2 June 2010

random sets

Let x be a categorical variable taking values in the *finite* set $\mathcal{X} \neq \varnothing$.

random sets

Let x be a categorical variable taking values in the *finite* set $\mathcal{X} \neq \varnothing$.

In (Bayesian) *probability theory*, information about the uncertain value of x is described by the probability distribution of a random variable X taking values in $\mathcal{X}$.

random sets

Let x be a categorical variable taking values in the *finite* set $\mathcal{X} \neq \varnothing$.

In (Bayesian) *probability theory*, information about the uncertain value of x is described by the probability distribution of a random variable X taking values in $\mathcal{X}$.

In (Dempster-Shafer) *belief functions theory*, information about the uncertain value of x is described by the probability distribution of a **random subset** S of $\mathcal{X}$, with $S \neq \varnothing$ a.s.

random sets

Let x be a categorical variable taking values in the *finite* set $\mathcal{X} \neq \varnothing$.

In (Bayesian) *probability theory*, information about the uncertain value of x is described by the probability distribution of a random variable X taking values in $\mathcal{X}$.

In (Dempster-Shafer) *belief functions theory*, information about the uncertain value of x is described by the probability distribution of a **random subset** S of $\mathcal{X}$, with $S \neq \varnothing$ a.s.

Each value $A \subseteq \mathcal{X}$ of S is interpreted as "$x \in A$" (without any additional information about the value of x); random variables thus correspond to the case with $|S| = 1$ a.s.

random sets

Let x be a categorical variable taking values in the *finite* set $\mathcal{X} \neq \varnothing$.

In (Bayesian) *probability theory*, information about the uncertain value of x is described by the probability distribution of a random variable X taking values in $\mathcal{X}$.

In (Dempster-Shafer) *belief functions theory*, information about the uncertain value of x is described by the probability distribution of a **random subset** S of $\mathcal{X}$, with $S \neq \varnothing$ a.s.

Each value $A \subseteq \mathcal{X}$ of S is interpreted as "$x \in A$" (without any additional information about the value of x); random variables thus correspond to the case with $|S| = 1$ a.s.

The interpretation of the probability distribution of S varies from author to author, but it is usually an *epistemic* interpretation.

belief and plausibility functions

Given a set $B \subseteq \mathcal{X}$, each value $A \subseteq \mathcal{X}$ of S falls into one of the following 3 categories:

- $A \subseteq B$ ($S = A$ supports "$x \in B$"),
- $A \nsubseteq B$ and $A \nsubseteq B^{\mathsf{C}}$ ($S = A$ supports neither "$x \in B$" nor "$x \notin B$"),
- $A \subseteq B^{\mathsf{C}}$ ($S = A$ supports "$x \notin B$").

belief and plausibility functions

Given a set $B \subseteq \mathcal{X}$, each value $A \subseteq \mathcal{X}$ of S falls into one of the following 3 categories:

- $A \subseteq B$ ($S = A$ supports "$x \in B$"),
- $A \not\subseteq B$ and $A \not\subseteq B^{\mathsf{C}}$ ($S = A$ supports neither "$x \in B$" nor "$x \notin B$"),
- $A \subseteq B^{\mathsf{C}}$ ($S = A$ supports "$x \notin B$").

$Bel(B) = P\{S \subseteq B\}$ is the probability that S supports "$x \in B$".

belief and plausibility functions

Given a set $B \subseteq \mathcal{X}$, each value $A \subseteq \mathcal{X}$ of S falls into one of the following 3 categories:

- $A \subseteq B$ ($S = A$ supports "$x \in B$"),
- $A \not\subseteq B$ and $A \not\subseteq B^{\mathsf{C}}$ ($S = A$ supports neither "$x \in B$" nor "$x \notin B$"),
- $A \subseteq B^{\mathsf{C}}$ ($S = A$ supports "$x \notin B$").

$Bel(B) = P\{S \subseteq B\}$ is the probability that S supports "$x \in B$".

$Pl(B) = P\{S \subseteq B\} + P\{S \not\subseteq B \text{ and } S \not\subseteq B^{\mathsf{C}}\} = 1 - Bel(B^{\mathsf{C}})$
is the probability that S does not support "$x \notin B$".

belief and plausibility functions

Given a set $B \subseteq \mathcal{X}$, each value $A \subseteq \mathcal{X}$ of S falls into one of the following 3 categories:

- $A \subseteq B$ ($S = A$ supports "$x \in B$"),
- $A \not\subseteq B$ and $A \not\subseteq B^{\mathsf{C}}$ ($S = A$ supports neither "$x \in B$" nor "$x \notin B$"),
- $A \subseteq B^{\mathsf{C}}$ ($S = A$ supports "$x \notin B$").

$Bel(B) = P\{S \subseteq B\}$ is the probability that S supports "$x \in B$".

$Pl(B) = P\{S \subseteq B\} + P\{S \not\subseteq B \text{ and } S \not\subseteq B^{\mathsf{C}}\} = 1 - Bel(B^{\mathsf{C}})$
is the probability that S does not support "$x \notin B$".

$Bel, Pl : 2^{\mathcal{X}} \to [0,1]$ are **dual, monotonic set functions** with $Bel \leq Pl$.

belief and plausibility functions

Given a set $B \subseteq \mathcal{X}$, each value $A \subseteq \mathcal{X}$ of S falls into one of the following 3 categories:

- $A \subseteq B$ ($S = A$ supports "$x \in B$"),
- $A \not\subseteq B$ and $A \not\subseteq B^{\mathsf{C}}$ ($S = A$ supports neither "$x \in B$" nor "$x \notin B$"),
- $A \subseteq B^{\mathsf{C}}$ ($S = A$ supports "$x \notin B$").

$Bel(B) = P\{S \subseteq B\}$ is the probability that S supports "$x \in B$".

$Pl(B) = P\{S \subseteq B\} + P\{S \not\subseteq B \text{ and } S \not\subseteq B^{\mathsf{C}}\} = 1 - Bel(B^{\mathsf{C}})$
is the probability that S does not support "$x \notin B$".

$Bel, Pl : 2^{\mathcal{X}} \to [0, 1]$ are **dual, monotonic set functions** with $Bel \leq Pl$.

$Bel = Pl \quad \Leftrightarrow \quad Bel$ is *additive* $\quad \Leftrightarrow \quad Pl$ is *additive* $\quad \Leftrightarrow \quad |S| = 1$ a.s.

imprecise probabilities

Bel and *Pl* correspond to coherent *lower and upper probabilities*, respectively, in the theory of Walley (1991).

imprecise probabilities

Bel and *Pl* correspond to coherent *lower and upper probabilities*, respectively, in the theory of Walley (1991).

However, the connection with imprecise probabilities can be misleading: for example, if $\mathcal{X} = \{e, \neg e\}$, and on the basis of completely different approaches two experts assign the probabilities 0.8 and 0.9, respectively, to the event $x = e$, then

imprecise probabilities

Bel and *Pl* correspond to coherent *lower and upper probabilities*, respectively, in the theory of Walley (1991).

However, the connection with imprecise probabilities can be misleading: for example, if $\mathcal{X} = \{e, \neg e\}$, and on the basis of completely different approaches two experts assign the probabilities 0.8 and 0.9, respectively, to the event $x = e$, then

- the combined (precise or imprecise) probability of $x = e$ will be in or around the interval $[0.8, 0.9]$,

imprecise probabilities

Bel and *Pl* correspond to coherent *lower and upper probabilities*, respectively, in the theory of Walley (1991).

However, the connection with imprecise probabilities can be misleading: for example, if $\mathcal{X} = \{e, \neg e\}$, and on the basis of completely different approaches two experts assign the probabilities 0.8 and 0.9, respectively, to the event $x = e$, then

- the combined (precise or imprecise) probability of $x = e$ will be in or around the interval $[0.8, 0.9]$,
- while the combined belief in $x = e$ will be $0.\overline{972}$ (using *Dempster's rule of combination*).

imprecise probabilities

Bel and *Pl* correspond to coherent *lower and upper probabilities*, respectively, in the theory of Walley (1991).

However, the connection with imprecise probabilities can be misleading: for example, if $\mathcal{X} = \{e, \neg e\}$, and on the basis of completely different approaches two experts assign the probabilities 0.8 and 0.9, respectively, to the event $x = e$, then

- the combined (precise or imprecise) probability of $x = e$ will be in or around the interval $[0.8, 0.9]$,
- while the combined belief in $x = e$ will be $0.\overline{972}$ (using *Dempster's rule of combination*).

Bel and *Pl* are descriptions of the *support provided by the available evidence*, while a (precise or imprecise) probability distribution is the description of an *equilibrium*.

imprecise probabilities

Bel and *Pl* correspond to coherent *lower and upper probabilities*, respectively, in the theory of Walley (1991).

However, the connection with imprecise probabilities can be misleading: for example, if $\mathcal{X} = \{e, \neg e\}$, and on the basis of completely different approaches two experts assign the probabilities 0.8 and 0.9, respectively, to the event $x = e$, then

- the combined (precise or imprecise) probability of $x = e$ will be in or around the interval $[0.8, 0.9]$,
- while the combined belief in $x = e$ will be $0.\overline{972}$ (using *Dempster's rule of combination*).

Bel and *Pl* are descriptions of the *support provided by the available evidence*, while a (precise or imprecise) probability distribution is the description of an *equilibrium*.

In the above example, the probability ratios are multiplied (as if they were likelihood ratios): $\frac{0.8}{1-0.8} \times \frac{0.9}{1-0.9} = 36 = \frac{0.\overline{972}}{1-0.\overline{972}}$. In fact, *Bel* and *Pl* were rather interpreted as generalizations of *likelihood functions* or *fiducial probabilities* by Dempster and Shafer: see also Wiencierz (2009).

information fusion

If the probability distributions of the random subsets $S_1, \ldots, S_n$ of $\mathcal{X}$ describe the information (about the uncertain value of x) obtained from n different sources, respectively, then the **combined information** is described by the probability distribution of $S_1 \cap \cdots \cap S_n$, which depends on the *joint probability distribution* of $S_1, \ldots, S_n$.

information fusion

If the probability distributions of the random subsets $S_1, \ldots, S_n$ of $\mathcal{X}$ describe the information (about the uncertain value of x) obtained from n different sources, respectively, then the **combined information** is described by the probability distribution of $S_1 \cap \cdots \cap S_n$, which depends on the *joint probability distribution* of $S_1, \ldots, S_n$.

The *independence* of $S_1, \ldots, S_n$ is often assumed, but is in general **incompatible** with the condition that $S_1 \cap \cdots \cap S_n \neq \varnothing$ a.s.

information fusion

If the probability distributions of the random subsets $S_1, \ldots, S_n$ of $\mathcal{X}$ describe the information (about the uncertain value of x) obtained from n different sources, respectively, then the **combined information** is described by the probability distribution of $S_1 \cap \cdots \cap S_n$, which depends on the *joint probability distribution* of $S_1, \ldots, S_n$.

The *independence* of $S_1, \ldots, S_n$ is often assumed, but is in general **incompatible** with the condition that $S_1 \cap \cdots \cap S_n \neq \varnothing$ a.s.

Dempster's rule of combination consists in assuming the independence of $S_1, \ldots, S_n$ and **then** conditioning on $\{S_1 \cap \cdots \cap S_n \neq \varnothing\}$ (if possible).

However, in general the conditional joint probability distribution neither has the right marginal distributions for $S_1, \ldots, S_n$, nor describes their independence.

information fusion

If the probability distributions of the random subsets $S_1, \ldots, S_n$ of $\mathcal{X}$ describe the information (about the uncertain value of x) obtained from n different sources, respectively, then the **combined information** is described by the probability distribution of $S_1 \cap \cdots \cap S_n$, which depends on the *joint probability distribution* of $S_1, \ldots, S_n$.

The *independence* of $S_1, \ldots, S_n$ is often assumed, but is in general **incompatible** with the condition that $S_1 \cap \cdots \cap S_n \neq \varnothing$ a.s.

Dempster's rule of combination consists in assuming the independence of $S_1, \ldots, S_n$ and **then** conditioning on $\{S_1 \cap \cdots \cap S_n \neq \varnothing\}$ (if possible).

However, in general the conditional joint probability distribution neither has the right marginal distributions for $S_1, \ldots, S_n$, nor describes their independence.

In the experts' example, after the conditioning, $S_1 = S_2$ a.s. with
$P\{S_i = \{e\}\} = 0.\overline{972}$ and $P\{S_i = \{\neg e\}\} = 0.\overline{027}$.

information fusion

If the probability distributions of the random subsets $S_1, \dots, S_n$ of $\mathcal{X}$ describe the information (about the uncertain value of x) obtained from n different sources, respectively, then the **combined information** is described by the probability distribution of $S_1 \cap \dots \cap S_n$, which depends on the *joint probability distribution* of $S_1, \dots, S_n$.

The *independence* of $S_1, \dots, S_n$ is often assumed, but is in general **incompatible** with the condition that $S_1 \cap \dots \cap S_n \neq \varnothing$ a.s.

Dempster's rule of combination consists in assuming the independence of $S_1, \dots, S_n$ and **then** conditioning on $\{S_1 \cap \dots \cap S_n \neq \varnothing\}$ (if possible).

However, in general the conditional joint probability distribution neither has the right marginal distributions for $S_1, \dots, S_n$, nor describes their independence.

In the experts' example, after the conditioning, $S_1 = S_2$ a.s. with $P\{S_i = \{e\}\} = 0.\overline{972}$ and $P\{S_i = \{\neg e\}\} = 0.\overline{027}$.

Hence, Dempster's rule of combination can at best be considered as corresponding to an **approximation** of independence.

combination without the assumption of independence

When **no dependence structure is assumed** for $S_1, \ldots, S_n$, there are in general *many possible probability distributions* for $S_1 \cap \cdots \cap S_n$.

combination without the assumption of independence

When **no dependence structure is assumed** for $S_1, \ldots, S_n$, there are in general *many possible probability distributions* for $S_1 \cap \cdots \cap S_n$.

A typical solution in theories dealing with uncertainty is to select the *least precise* description of information (for instance by *entropy maximization*).

combination without the assumption of independence

When **no dependence structure is assumed** for $S_1, \ldots, S_n$, there are in general *many possible probability distributions* for $S_1 \cap \cdots \cap S_n$.

A typical solution in theories dealing with uncertainty is to select the *least precise* description of information (for instance by *entropy maximization*).

However, this approach has several problems, such as:

- there are *many different definitions* of "least precise" belief function,

combination without the assumption of independence

When **no dependence structure is assumed** for $S_1, \ldots, S_n$, there are in general *many possible probability distributions* for $S_1 \cap \cdots \cap S_n$.

A typical solution in theories dealing with uncertainty is to select the *least precise* description of information (for instance by *entropy maximization*).

However, this approach has several problems, such as:

- there are *many different definitions* of "least precise" belief function,
- for each of them the least precise belief function is in general *not unique*,

combination without the assumption of independence

When **no dependence structure is assumed** for $S_1, \ldots, S_n$, there are in general *many possible probability distributions* for $S_1 \cap \cdots \cap S_n$.

A typical solution in theories dealing with uncertainty is to select the *least precise* description of information (for instance by *entropy maximization*).

However, this approach has several problems, such as:

- there are *many different definitions* of "least precise" belief function,
- for each of them the least precise belief function is in general *not unique*,
- the selection of a whole belief function can be *computationally too demanding*,

combination without the assumption of independence

When **no dependence structure is assumed** for $S_1, \ldots, S_n$, there are in general *many possible probability distributions* for $S_1 \cap \cdots \cap S_n$.

A typical solution in theories dealing with uncertainty is to select the *least precise* description of information (for instance by *entropy maximization*).

However, this approach has several problems, such as:

- there are *many different definitions* of "least precise" belief function,
- for each of them the least precise belief function is in general *not unique*,
- the selection of a whole belief function can be *computationally too demanding*,
- in general the condition that $S_1 \cap \cdots \cap S_n \neq \varnothing$ a.s. *cannot be satisfied*.

combination without the assumption of independence

When **no dependence structure is assumed** for $S_1, \ldots, S_n$, there are in general *many possible probability distributions* for $S_1 \cap \cdots \cap S_n$.

A typical solution in theories dealing with uncertainty is to select the *least precise* description of information (for instance by *entropy maximization*).

However, this approach has several problems, such as:

- there are *many different definitions* of "least precise" belief function,
- for each of them the least precise belief function is in general *not unique*,
- the selection of a whole belief function can be *computationally too demanding*,
- in general the condition that $S_1 \cap \cdots \cap S_n \neq \varnothing$ a.s. *cannot be satisfied*.

In the experts' example, $P\{S_1 \cap S_2 = \varnothing\} \in [0.1, 0.3]$ for all possible joint probability distributions of S_1, S_2.

Fréchet bounds

The new idea in Cattaneo (2010) is to *approximate* by a belief function the set function $F : 2^{\mathcal{X}} \to [0, 1]$ that is **pointwise least precise**: F assigns to each $B \subseteq \mathcal{X}$ the *minimum* of $P\{S_1 \cap \cdots \cap S_n \subseteq B\}$ over all possible joint probability distributions of $S_1, \ldots, S_n$ (that is, F is a *lower envelope*).

Fréchet bounds

The new idea in Cattaneo (2010) is to *approximate* by a belief function the set function $F : 2^{\mathcal{X}} \to [0, 1]$ that is **pointwise least precise**: F assigns to each $B \subseteq \mathcal{X}$ the *minimum* of $P\{S_1 \cap \cdots \cap S_n \subseteq B\}$ over all possible joint probability distributions of $S_1, \ldots, S_n$ (that is, F is a *lower envelope*).

In particular, the **minimal conflict** $F(\varnothing)$ is a very interesting measure of disagreement among belief functions: see also Cattaneo (2003).

Fréchet bounds

The new idea in Cattaneo (2010) is to *approximate* by a belief function the set function $F : 2^{\mathcal{X}} \to [0,1]$ that is **pointwise least precise**: F assigns to each $B \subseteq \mathcal{X}$ the *minimum* of $P\{S_1 \cap \cdots \cap S_n \subseteq B\}$ over all possible joint probability distributions of $S_1, \ldots, S_n$ (that is, F is a *lower envelope*).

In particular, the **minimal conflict** $F(\varnothing)$ is a very interesting measure of disagreement among belief functions: see also Cattaneo (2003).

For each $B \subseteq \mathcal{X}$, the quantity

$$\max_{\substack{B_1,\ldots,B_n \subseteq \mathcal{X}:\\ B_1 \cap \cdots \cap B_n \subseteq B}} \left(P\{S_1 \subseteq B_1\} + \cdots + P\{S_n \subseteq B_n\}\right) + 1 - n$$

is a simple *lower approximation* of $F(B)$, which is exact when $n \leq 2$, as follows from a result by Strassen (1965).

Fréchet bounds

The new idea in Cattaneo (2010) is to *approximate* by a belief function the set function $F : 2^{\mathcal{X}} \to [0,1]$ that is **pointwise least precise**: F assigns to each $B \subseteq \mathcal{X}$ the *minimum* of $P\{S_1 \cap \cdots \cap S_n \subseteq B\}$ over all possible joint probability distributions of $S_1, \ldots, S_n$ (that is, F is a *lower envelope*).

In particular, the **minimal conflict** $F(\varnothing)$ is a very interesting measure of disagreement among belief functions: see also Cattaneo (2003).

For each $B \subseteq \mathcal{X}$, the quantity

$$\max_{\substack{B_1,\ldots,B_n \subseteq \mathcal{X}: \\ B_1 \cap \cdots \cap B_n \subseteq B}} \left(P\{S_1 \subseteq B_1\} + \cdots + P\{S_n \subseteq B_n\}\right) + 1 - n$$

is a simple *lower approximation* of $F(B)$, which is exact when $n \leq 2$, as follows from a result by Strassen (1965).

In the experts' example, $F(\varnothing) = 0.1, \quad F(\{e\}) = 0.9, \quad F(\{\neg e\}) = 0.2,$ and $F(\mathcal{X}) = 1$. Hence, there is a joint probability distribution of S_1, S_2 with $F(B) = P\{S_1 \cap S_2 \subseteq B\}$ for all $B \subseteq \mathcal{X}$, but $P\{S_1 \cap S_2 = \varnothing\} = 0.1$.

references

Cattaneo (2003). Combining belief functions issued from dependent sources. In *ISIPTA '03*, Carleton Scientific, pp. 133–147.

Cattaneo (2010). Belief functions combination without the assumption of independence of the information sources. *To appear*.

Strassen (1965). The existence of probability measures with given marginals. *Ann. Math. Stat.* 36(2):423–439.

Walley (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.

Wiencierz (2009). *Arthur P. Dempster's Generalized Inference Theory*. Master's thesis, LMU Munich.