Random fuzzy sets: a general model of uncertainty

Thierry Denœux

Université de technologie de Compiègne, Compiègne, France
Institut universitaire de France, Paris, France

https://www.hds.utc.fr/~tdenoeux

Virtual Conference on Epistemic Uncertainty in Engineering
February 15, 2021

Uncertainty

Definition (Uncertainty)

"A situation in which something is not known" (Cambridge dictionary)

- Uncertainty always refers to lack of knowledge: it is always "epistemic".
- In engineering, "epistemic uncertainty" is often used as the opposite of "aleatory uncertainty".

Aleatory uncertainty

Definition (Aleatory uncertainty)

Uncertainty about the characteristic of an individual that is only known to belong to a population with known distribution.

- Examples:
  - Drawing a ball from an urn containing balls of different colors with known proportions
  - John has Covid. He lives in the UK. We know that 3.5% of Covid patients in the UK need hospitalization. Will John need to be hospitalized?
- An ideal situation, not often encountered in practice
  - Usually, the exact frequencies are not known
  - Sometimes, there is only one element in the population (next presidential election in France)

But it is useful as an approximation.

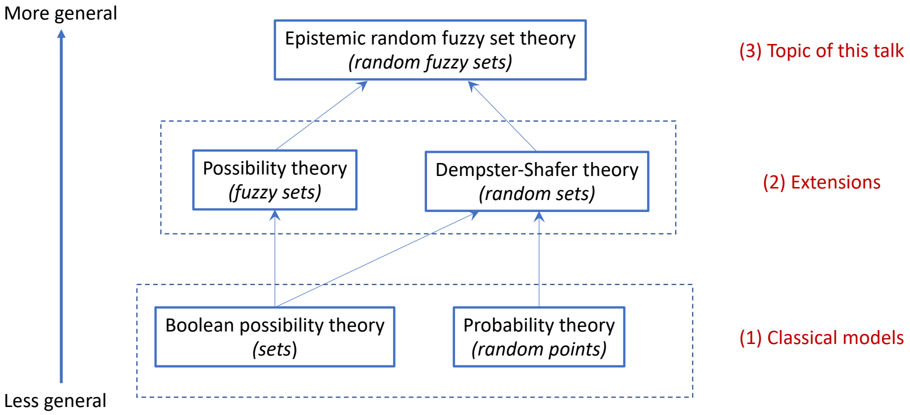
Models of uncertainty

- To reason and to make decisions under uncertainty, we need a mathematical model.
- Several models have been proposed in different fields: statistics, computer science, engineering, economics.
- These models are just tools.

  All models are wrong, but some are useful (George Box).

- Models must be defended based on
  - Their ability to support and justify sound conclusions and
  - Their ease of use in practical applications (important from an engineering perspective)

Overview of some models of uncertainty



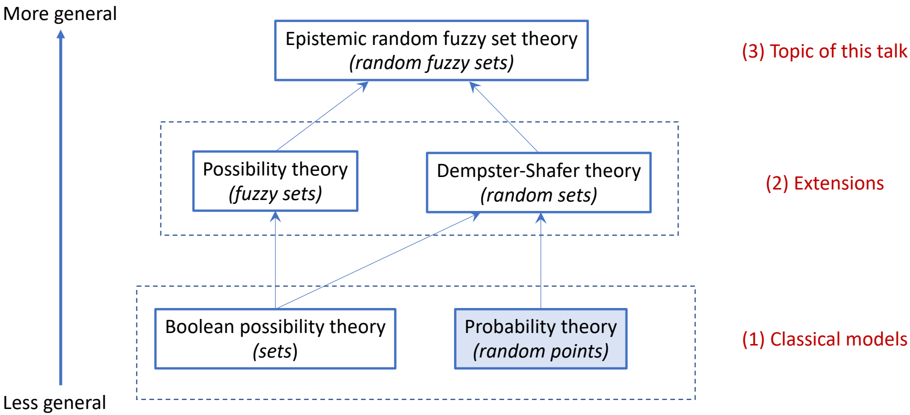
Outline

1. Classical models
   - Probability theory
   - Set-based approach

2. Extensions
   - Fuzzy sets and possibility theory
   - Random sets and Dempster-Shafer theory

3. Random fuzzy sets
   - Definitions
   - Application to statistical prediction

Probability theory



The probabilistic formalism

- We consider a question Q of interest and the finite set $\Theta = \{\theta_1, \ldots, \theta_n\}$ of possible answers to that question, one and only one of which is true. (Finiteness is only assumed to avoid mathematical technicalities).
- Let $P(A)$ denote the degree of belief that subset $A \subseteq \Theta$ contains the true answer. By convention, $P(\emptyset) = 0$ and $P(\Theta) = 1$.
- Probability theory makes 2 main assumptions:

  Additivity The mapping $P : 2^\Theta \mapsto [0, 1]$ verifies

  $$\boxed{\forall A, B \subseteq \Theta, \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

  In particular, $P(A^c) = 1 - P(A)$.

  Bayes' rule of conditioning: If we learn that the truth is in $B \subset \Theta$ and $P(B) > 0$, P is updated as

  $$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

The case of pure "aleatory" uncertainty

- Example: we know that a ball was taken from an urn in which 40% of the balls are black. What is our belief that this particular ball is black?
- Here, we have a question Q that pertains to some characteristic of an individual that is only known to belong to a population in which θ_k is present in proportion p_k.
- Equating degrees of belief with proportions in the population, we get

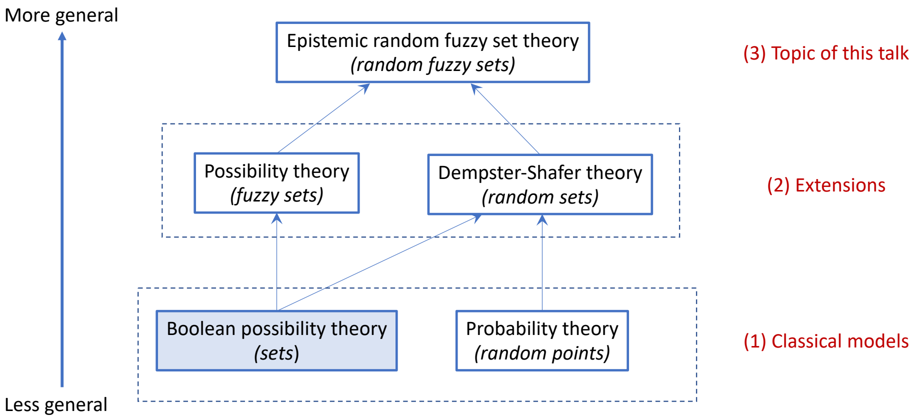
$$P(A) = \sum_{\theta_k \in A} p_k \quad \text{for all } A \subseteq \Theta,$$

a probability measure.

Principle of indifference

- It has been argued that the probabilistic model is sufficient to describe any situation of uncertainty. However, difficulties arise when describing states of complete or near ignorance with probabilities.
- Laplace's Principle of Indifference (PI): *"In the absence of any relevant evidence, agents should distribute their degrees of belief equally among all the possibilities under consideration"*.
- Example: Sirius paradox (Shafer). "Is there life around Sirius?"
  - There are two possibilities: life (L) or no life (NL).
  - Since we have no evidence, the PI gives us $P(L) = P(NL) = 1/2$.
  - But 'NL' can be split in two possibilities: there is no planet (NP), or there is a planet, but no life (PNL).
  - So we now have three possibilities: L, NP, PNL, and the PI gives us $P(L) = P(NP) = P(PNL) = 1/3$.
- Such paradoxes appear when trying to "squeeze" set-based information into the probabilistic formalism. There is no necessity to do so.

Set-based approach



Set-based approach

- When the evidence points to a set $F \subseteq \Theta$ of possibilities, an alternative is to keep that set as a representation of the evidence.
- Knowing that the truth is in F, a proposition A is possible iff $F \cap A \neq \emptyset$, and it is certain iff $F \subseteq A$.
- We can thus define two mappings from 2^Θ to $\{0,1\}$: a Boolean possibility measure

$$\Pi_F(A) = \begin{cases} 1 & \text{if } F \cap A \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

and a Boolean necessity measure:

$$N_F(A) = \begin{cases} 1 & \text{if } F \subseteq A \\ 0 & \text{otherwise} \end{cases}$$

Properties, updating

- For all $A, B \subseteq \Theta$,

$$\Pi_F(A) = 1 - N_F(A^c)$$

$$\Pi_F(A \cup B) = \max\left[\Pi_F(A), \Pi_F(B)\right]$$

$$N_F(A \cap B) = \min\left[N_F(A), N_F(B)\right]$$

- Instead of a single additive measure P, with have a maxitive measure Π_F and a dual measure N_F, both taking values in $\{0, 1\}$.
- Updating: if we receive two pieces of evidence pointing to two non-disjoint subsets F and G, then we can conclude that the truth is $F \cap G$. Here, the combination rule is set intersection.

Interval analysis

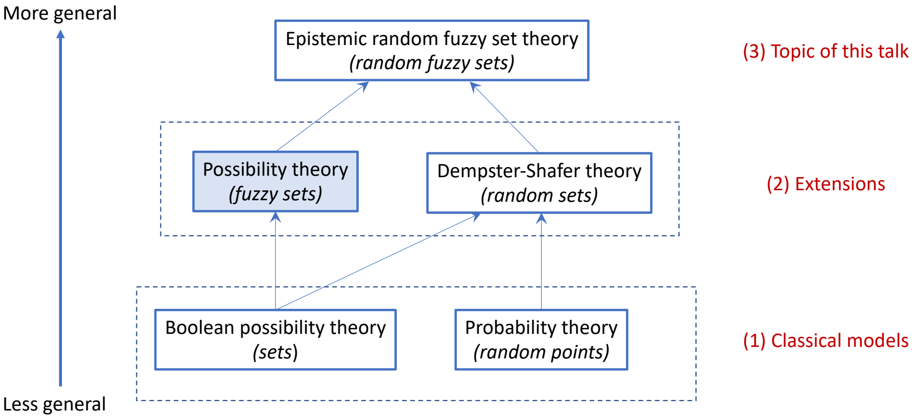
- In engineering, an example of the set-based approach is interval analysis.
- Principle: put bounds on all observations and model parameters, and propagate these bounds through numerical equations.
- Advantages:
  - Simplicity: propagating intervals is often simpler than propagating probability distribution. For instance, if $x \in [a, b]$ and $y \in [c, d]$, we can state that $x + y \in [a + c, b + d]$ and $x - y \in [a - d, b - c]$.
  - Guaranteed results: provided the initial bounds are correct, interval analysis guarantees reliable, mathematically correct results.

Limitations of the set-based approach

- The set-based approach (including interval analysis) does not allow one to express doubt and, as a consequence, it is often too conservative.
- The available information is sometimes too imprecise or vague to allow us to define precise bounds or sharp boundaries.
    - For instance, information conveyed in natural language, using words such as “small”, “large”, “old”, “young”, “roughly”, “about”, etc.

Outline

Fuzzy sets and possibility theory



Fuzzy sets

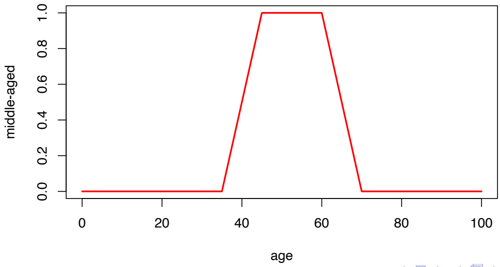
Definition

"A set with an unsharp boundary" (Zadeh, 1965). Formally, a fuzzy subset of Θ *is a mapping*

$$\widetilde{F} : \Theta \mapsto [0, 1],$$

that assigns a membership degree $\widetilde{F}(\theta)$ *to each* $\theta \in \Theta$*. The fuzzy set* $\widetilde{F}$ *is normal if that* $\widetilde{F}(\theta) = 1$ *for some* $\theta \in \Theta$*.*

Example: "middle-aged"



Possibility and necessity measures

- Assume we receive “vague” evidence described by a normal fuzzy subset $\widetilde{F}$ of possibilities acting as a flexible constraint. (For instance, a testimony that tells us that “John is middle-aged”).
- Given this evidence,
  - $A \subseteq \Theta$ is all the more possible that there is some $\theta \in A$ that satisfies the constraint, and
  - A is all the more certain that all $\theta \notin A$ fail to satisfy the constraint.
- We can thus define the degrees of possibility and necessity of $A \subseteq \Theta$ as

$$\boxed{\Pi_{\widetilde{F}}(A) = \max_{\theta \in A} \widetilde{F}(\theta)}$$

and

$$\boxed{N_{\widetilde{F}}(A) = \min_{\theta \notin A} \left[1 - \widetilde{F}(\theta)\right] = 1 - \Pi_{\widetilde{F}}(A^c)}$$

both taking values in $[0, 1]$.

Properties

- Generalization of the Boolean case.
- For all $A, B \subseteq \Theta$,

$$\Pi_{\widetilde{F}}(A \cup B) = \max\left[\Pi_{\widetilde{F}}(A), \Pi_{\widetilde{F}}(B)\right]$$
$$N_{\widetilde{F}}(A \cap B) = \min\left[N_{\widetilde{F}}(A), N_{\widetilde{F}}(B)\right]$$

Definition (Possibility distribution)

The mapping $\pi_{\widetilde{F}} : \theta \mapsto \Pi_{\widetilde{F}}(\{\theta\})$ is called a possibility distribution. We have $\pi_{\widetilde{F}} = \widetilde{F}$.

Combination of possibility distributions

- Assume that we receive two vague pieces of evidence represented by fuzzy sets $\widetilde{F}$ and $\widetilde{G}$. How to combine them?
- Assuming both pieces of evidence to be reliable, it makes sense to describe the combined evidence by the intersection of $\widetilde{F}$ and $\widetilde{G}$. We also need a normalization step to get a normal fuzzy set.
- There are several definitions for the intersection of fuzzy sets. The following operator has good properties:

$$(\widetilde{F} \odot \widetilde{G})(\theta) = \frac{(\widetilde{F} \cdot \widetilde{G})(\theta)}{h(\widetilde{F} \cdot \widetilde{G})}$$

  where $h(\cdot)$ denotes the height of a fuzzy set (the maximum membership degree).
- $\odot$ is associative.

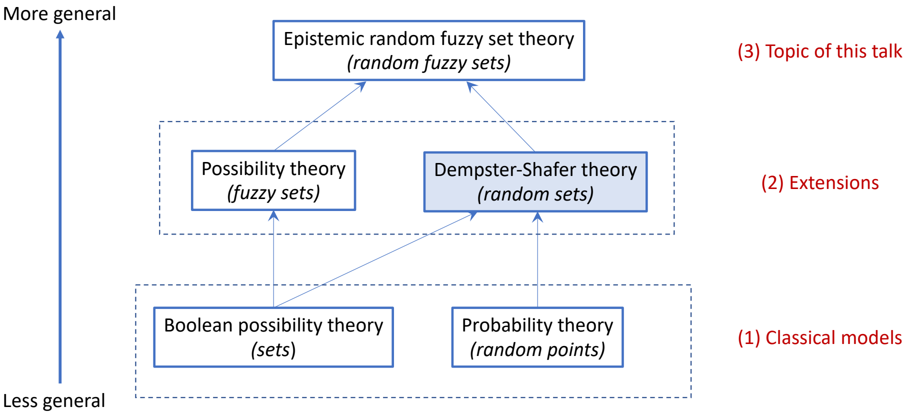
Likelihood

- Possibility theory is suitable for representing not only evidence expressed using natural language, but also statistical evidence.
- Let $X \in \mathcal{X}$ be random data generated according to a probability mass or density function $f(x;\theta)$, where $\theta \in \Theta$ is a parameter.
- After observing a realization x of X, we define the relative likelihood function as the mapping from Θ to $[0,1]$ defined by

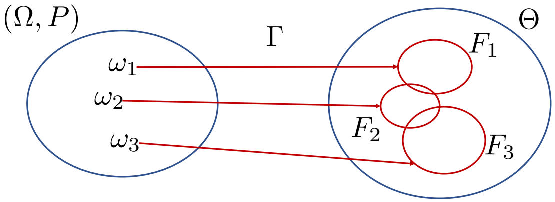
$$\boxed{\widetilde{L}_x(\theta) = \frac{f(x;\theta)}{\max_{\theta \in \theta} f(x;\theta)}}$$

- Several authors (Smets, Dubois) have noticed that $\widetilde{L}_x$ can be interpreted as a possibility distribution. Equivalently, $\widetilde{L}_x$ can be seen as the fuzzy set of likely values of the parameter, after observing x.
- If x and x' are independent samples, $\widetilde{L}_{x,x'} = \widetilde{L}_x \odot \widetilde{L}_{x'}$.

Random sets and Dempster-Shafer theory



Random set



- A random set is a tuple $(\Omega, 2^\Omega, P, \Gamma)$ where $(\Omega, 2^\Omega, P)$ is a probability space and Γ is a mapping from Ω to $2^\Theta \setminus \{\emptyset\}$.
- Meaning: Ω is a set of interpretations of a piece of evidence, with known probabilities. If $\omega \in \Omega$ holds, the evidence tells us that the truth is in a nonempty subset $\Gamma(\omega) \subseteq \Theta$, and nothing more.
- For instance, a sensor tells us that the parameter of interest is in F, but this sensor is 80% reliable. We thus have $\Omega = \{\text{rel}, \neg\text{rel}\}$, $P(\{\text{rel}\}) = 0.8$, $P(\{\neg\text{rel}\}) = 0.2$, $\Gamma(\text{rel}) = F$, $\Gamma(\neg\text{rel}) = \Theta$.

Mass function

- The multi-valued mapping carries probabilities from Ω to 2^Θ. We obtain a probability mass function on 2^Θ, called a mass function and defined as

$$m(F) = P(\{\omega \in \Omega : \Gamma(\omega) = F\})$$

for all $F \subseteq \Theta$.

- Properties:

$$\sum_{F \subseteq \Theta} m(F) = 1, \quad m(\emptyset) = 0.$$

- $m(F)$ is the probability of knowing that the truth is in F, given the evidence.
- If $m(F) > 0$, F is called a focal set.
- Special cases:
  - If m has only one focal set, it is said to be logical (it is equivalent to a set).
  - If the focal sets are singletons, m is said to be Bayesian (it is equivalent to a probability mass function).

Belief and plausibility

- If the truth is in F, the (Boolean) necessity and possibility of A are, respectively, $N_F(A)$ and $\Pi_F(A)$.
- The expected necessity of A is

$$Bel_m(A) = \sum_{F \subseteq \Theta} m(F) \underbrace{N_F(A)}_{I(F \subseteq A)}$$

$Bel_m(A)$ measures the degree to which the evidence supports A.

- Similarly, the expected possibility of A is then

$$Pl_m(A) = \sum_{F \subseteq \Theta} m(F) \underbrace{\Pi_F(A)}_{I(F \cap A \neq \emptyset)} = 1 - Bel_m(A^c)$$

$Pl_m(A)$ measures the degree to which the evidence fails to support A^c.

Belief and plausibility functions

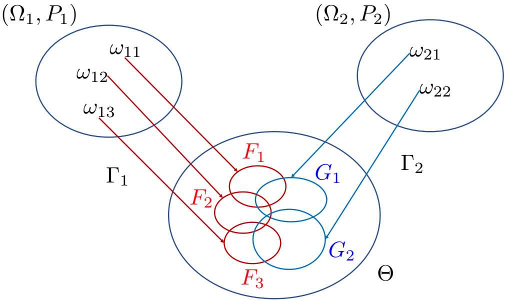
- Mappings $Bel_m : 2^\Theta \to [0,1]$ and $Pl_m : 2^\Theta \to [0,1]$ are called, respectively, belief and the plausibility functions.
- Characterization of belief functions: a mapping $F : 2^\Theta \to [0,1]$ is a belief function iff $F(\emptyset) = 0$, $F(\Theta) = 1$, and F is completely monotone; i.e., for any $k \geq 2$ and any collection $A_1, \ldots, A_k$ of subsets of Θ,

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

- The mapping $m \to Bel_m$ is one-to-one.
- Special cases:
  - If m is Bayesian, Bel_m is a probability measure and $Pl_m = Bel_m$
  - If m is logical with focal set F, $Bel_m = N_F$ and $Pl_m = \Pi_F$

Dempster's rule

- Assume we have two independent pieces of evidence represented by random sets $(\Omega_1, 2^{\Omega_1}, P_1, \Gamma_1)$ and $(\Omega_2, 2^{\Omega_2}, P_2, \Gamma_2)$.
- If the pair of interpretations $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ holds, we know that the truth is in $\Gamma_\cap(\omega_1, \omega_2) = \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2)$, provided this intersection is nonempty.



Dempster's rule (continued)

- We have the combined random set

$$\boxed{\left(\Omega_1 \times \Omega_2, 2^{\Omega_1 \times \Omega_2}, (P_1 \otimes P_2)(\cdot|\Theta^*), \Gamma_\cap\right)}$$

where $P_1 \otimes P_2$ is the product measure (independence assumption) and

$$\Theta^* = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) \neq \emptyset\}$$

is the subset of non-contradictory pairs of interpretations.

- The corresponding combined mass function is

$$(m_1 \oplus m_2)(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \frac{\sum_{F \cap G = A} m_1(F) m_2(G)}{\sum_{F \cap G \neq \emptyset} m_1(F) m_2(G)} & \text{otherwise} \end{cases}$$

Properties of Dempster's rule

- $\oplus$ is commutative and associative.
- Generalization of set intersection: If m_F and m_G are two logical mass functions with focal sets F and G such that $F \cap G \neq \emptyset$, then

$$m_F \oplus m_G = m_{F\cap G}$$

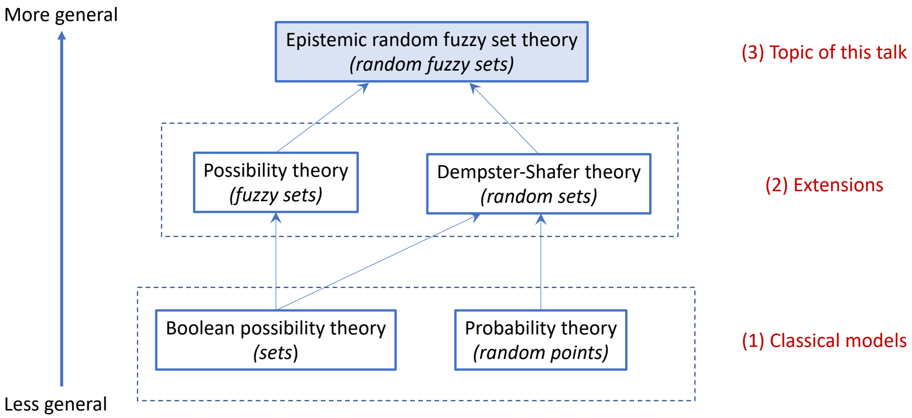
- Generalization of Bayesian conditioning: If m is a Bayesian mass function and F is a set such that $Bel_m(F) > 0$, then $m \oplus m_F$ is a Bayesian mass function, and the corresponding belief function is the conditional probability measure

$$Bel_{m\oplus m_F}(A) = \frac{Bel_m(A\cap F)}{Bel_m(F)}$$

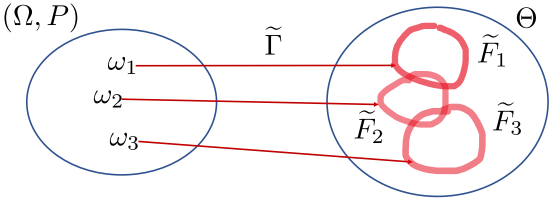
Outline

1. Classical models
   - Probability theory
   - Set-based approach
2. Extensions
   - Fuzzy sets and possibility theory
   - Random sets and Dempster-Shafer theory
3. Random fuzzy sets
   - Definitions
   - Application to statistical prediction

Random fuzzy sets



Random fuzzy sets



- As before, we consider a set Ω of interpretations of a given body of evidence. If $\omega \in \Omega$ holds, the truth is constrained by a normal fuzzy subset $\widetilde{\Gamma}(\omega)$ of Θ.
- For instance, a witness tells us that "John is middle-aged", but this witness is only 80% reliable. If "middle-aged" is represented by fuzzy set $\widetilde{F}$, we have $\Omega = \{\text{rel}, \neg\text{rel}\}$, $P(\{\text{rel}\}) = 0.8$, $P(\{\neg\text{rel}\}) = 0.2$, $\widetilde{\Gamma}(\text{rel}) = \widetilde{F}$, $\widetilde{\Gamma}(\neg\text{rel}) = \Theta$.
- The tuple $(\Omega, 2^\Omega, P, \widetilde{\Gamma})$ is a random fuzzy set.

Fuzzy mass function

- We define the corresponding fuzzy mass function $\widetilde{m}$ from $\mathcal{F}^*(\Theta)$ to $[0,1]$ as

$$\widetilde{m}(\widetilde{F}) = P(\{\omega \in \Omega : \widetilde{\Gamma}(\omega) = \widetilde{F}\}) \quad \text{for all } \widetilde{F} \in \mathcal{F}^*(\Theta)$$

- If $\widetilde{m}(\widetilde{F}) > 0$, $\widetilde{F}$ is called a (fuzzy) focal set. The collection of focal sets will be denoted as $\mathbb{F}(\widetilde{m}) = \{\widetilde{F}_1, \ldots, \widetilde{F}_f\}$.
- $\widetilde{m}(\widetilde{F})$ is interpreted as the probability that the truth is constrained by fuzzy set $\widetilde{F}$.

Belief and plausibility functions

- If the truth is constrained by $\widetilde{F}$, the possibility and necessity of a subset $A \subseteq \Theta$ are, respectively, $\Pi_{\widetilde{F}}(A)$ and $N_{\widetilde{F}}(A)$.
- The expected possibility and the expected necessity of A are, thus,

$$Pl_{\widetilde{m}}(A) = \sum_{\widetilde{F} \in \mathbb{F}(\widetilde{m})} m(\widetilde{F}) \underbrace{\Pi_{\widetilde{F}}(A)}_{\max_{\theta \in A} \widetilde{F}(\theta)}$$

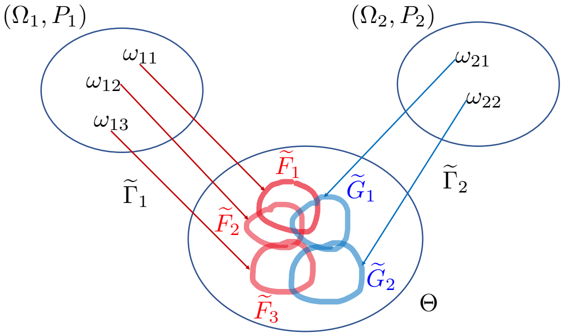
and

$$Bel_{\widetilde{m}}(A) = \sum_{\widetilde{F} \in \mathbb{F}(\widetilde{m})} m(\widetilde{F}) \underbrace{N_{\widetilde{F}}(A)}_{\min_{\theta \not\in A}[1 - \widetilde{F}(\theta)]} = 1 - Pl_{\widetilde{m}}(A^c)$$

- $Bel_{\widetilde{m}}$ is a belief function, and $Pl_{\widetilde{m}}$ is the dual plausibility function.
- The mapping $\widetilde{m} \rightarrow Bel_{\widetilde{m}}$ is many-to-one.

Generalized Dempster's rule

- Assume we have two independent and reliable pieces of evidence represented by random fuzzy sets $(\Omega_1, 2^{\Omega_1}, P_1, \widetilde{\Gamma}_1)$ and $(\Omega_2, 2^{\Omega_2}, P_2, \widetilde{\Gamma}_2)$.
- If the pair of interpretations $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ holds, we can deduce that the truth is in $\widetilde{\Gamma}_\cap(\omega_1, \omega_2) = \widetilde{\Gamma}_1(\omega_1) \cap \widetilde{\Gamma}_2(\omega_2)$, where $\cap$ is a normalized fuzzy intersection.



Generalized Dempster's rule (continued)

- To ensure associativity, we consider the normalized product intersection:

$$\widetilde{\Gamma}_{\odot}(\omega_1,\omega_2) = \widetilde{\Gamma}_1(\omega_1) \odot \widetilde{\Gamma}_2(\omega_2).$$

- The combined random fuzzy set is then

$$\boxed{\left(\Omega_1 \times \Omega_2, 2^{\Omega_1\times\Omega_2}, (P_1 \otimes P_2)(\cdot|\widetilde{\Theta}^*), \widetilde{\Gamma}_{\odot}\right)}$$

where $\widetilde{\Theta}^*$ is the fuzzy subset of non-contradictory pairs of interpretations

$$\widetilde{\Theta}^*(\omega_1,\omega_2) = h\left(\widetilde{\Gamma}_1(\omega_1) \cdot \widetilde{\Gamma}_2(\omega_2)\right).$$

- The corresponding combined fuzzy mass function is

$$(\widetilde{m}_1 \oplus \widetilde{m}_2)(\widetilde{F}) = \frac{\sum_{\widetilde{G}\odot\widetilde{H}=\widetilde{F}} h(\widetilde{G}\cdot\widetilde{H})\widetilde{m}_1(\widetilde{G})\widetilde{m}_2(\widetilde{H})}{\sum_{(\widetilde{G},\widetilde{H})\in\mathbb{F}(\widetilde{m}_1)\times\mathbb{F}(\widetilde{m}_2)} h(\widetilde{G}\cdot\widetilde{H})\widetilde{m}_1(\widetilde{G})\widetilde{m}_2(\widetilde{H})}$$

Properties

- Commutativity, associativity
- Generalization of Dempster's rule
- Generalization of the conjunctive combination of possibility distributions: if $\widetilde{m}_{\widetilde{F}}$ and $\widetilde{m}_{\widetilde{G}}$ are two logical fuzzy mass functions with focal sets $\widetilde{F}$ and $\widetilde{G}$, then
$$\widetilde{m}_{\widetilde{F}} \oplus \widetilde{m}_{\widetilde{G}} = \widetilde{m}_{\widetilde{F} \odot \widetilde{G}}$$
- Generalization of conditioning by a fuzzy event (Zadeh, 1968): Let
  - m be a Bayesian mass function with corresponding probability measure P and
  - $\widetilde{m}_{\widetilde{F}}$ a logical fuzzy mass function with focal set $\widetilde{F}$.

  Then $m \oplus \widetilde{m}_{\widetilde{F}}$ is a Bayesian mass function and the corresponding belief function is
$$Bel_{m \oplus \widetilde{m}_{\widetilde{F}}} = P(\cdot \mid \widetilde{F})$$

Application: statistical prediction

- Let X be the lifetime of a piece of equipment supposed to have an exponential distribution $\mathcal{E}(\theta)$ with $\theta > 0$.
- We have observed the lifetimes $\boldsymbol{x} = (x_1, \ldots, x_n)$ of n machines, assumed to be a realization of an iid sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ with parent distribution $\mathcal{E}(\theta)$.
- What can we say about the lifetime Y of a new piece of equipment?

Formalization

- General method: express Y as a function of θ and a pivotal random variable W with known distribution.
- Here, the cdf of Y is $F_Y(y) = 1 - \exp(-\theta y)$
- We know that $W = F_Y(Y) \sim \mathcal{U}(0,1)$, hence Y can be written as

$$\boxed{Y = -\frac{\ln(1-W)}{\theta}} \quad (1)$$

where

- $W \sim \mathcal{U}(0,1)$
- θ is unknown but constrained by the normalized likelihood function (possibility distribution)

$$\widetilde{L}_{\boldsymbol{x}}(\theta) = \left(\frac{\theta}{\widehat{\theta}}\right)^n \exp\left[n\left(1 - \frac{\theta}{\widehat{\theta}}\right)\right] \quad (2)$$

where $\widehat{\theta} = 1/\overline{x}$ is the MLE of θ.

Random fuzzy set

- From (1) and (2), we get the conditional possibility distribution on Y given $W = w$:

$$\boxed{\widetilde{Y}(y|w) = \left(\frac{\widehat{y}}{y}\right)^n \exp\left[n\left(1 - \frac{\widehat{y}}{y}\right)\right]}$$

with mode

$$\widehat{y} = -\frac{\ln(1-w)}{\widehat{\theta}}.$$

- Let $\widetilde{\Gamma}$ be the mapping from $[0,1]$ to $\mathcal{F}^*(\mathbb{R})$ defined by

$$\widetilde{\Gamma} : w \mapsto \widetilde{Y}(\cdot|w).$$

- Knowledge about Y is represented by the (continuous) random fuzzy set

$$\boxed{([0,1], \mathcal{B}([0,1]), P_W, \widetilde{\Gamma})}$$

Predictive belief and plausibility functions

- This random fuzzy set induces predictive plausibility and belief functions defined as

$$\boxed{Pl_Y(A) = \mathbb{E}_W\left[\sup_{y\in A} \widetilde{Y}(y|W)\right]}$$

$$\boxed{Bel_Y(A) = 1 - Pl_Y(A^c)}$$

for all $A \in \mathcal{B}(\mathbb{R})$.

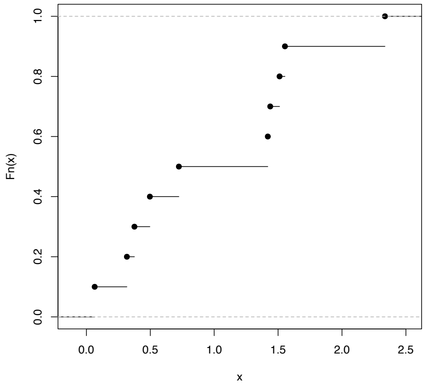
- In particular, we can define lower and upper predictive cdfs as

$$F_*(y) = Bel_Y((-\infty, y]) \quad \text{and} \quad F^*(y) = Pl_Y((-\infty, y])$$

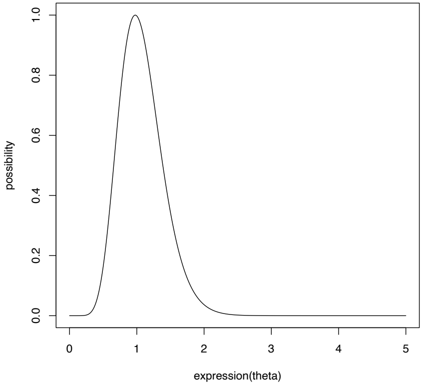
- These functions can be approximated by Monte Carlo simulation.

Example

$n = 10$ observations from $\mathcal{E}(1)$

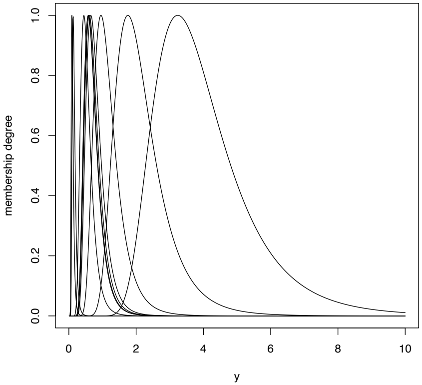


Relative likelihood

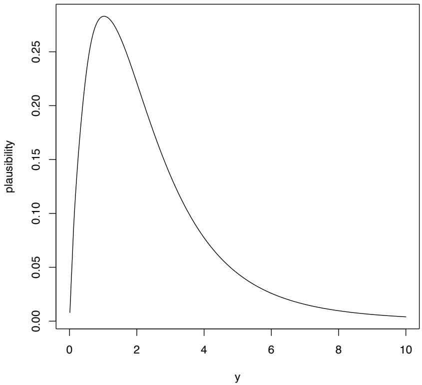


Example (continued)

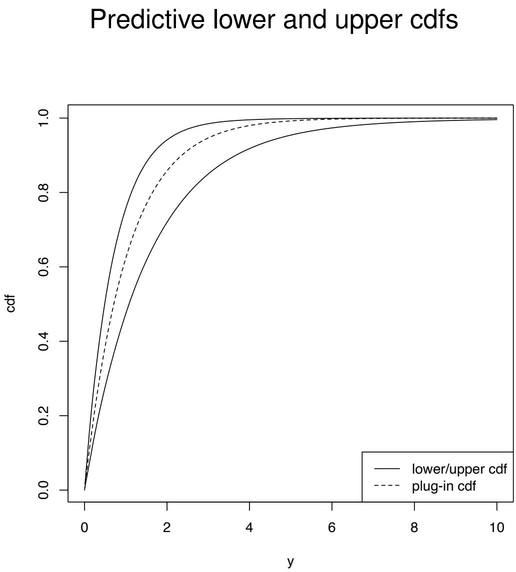
10 fuzzy sets $\widetilde{Y}(\cdot|w)$



Predictive contour function pl_Y



Example (continued)



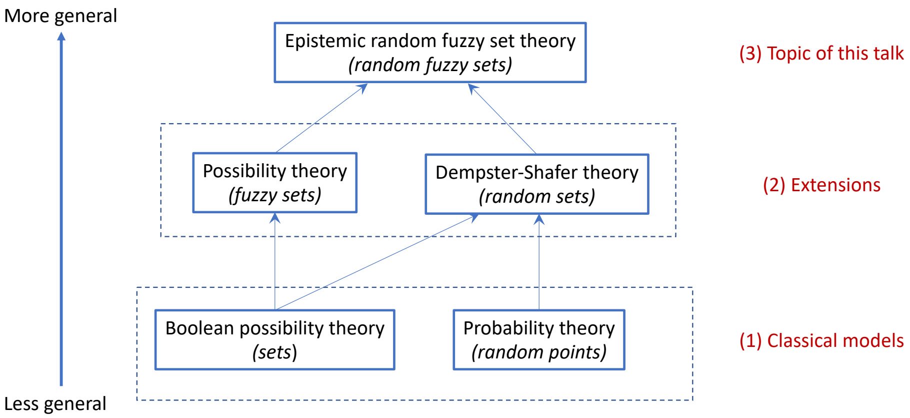
Properties

- Frequency calibration: Let $c_\alpha = \exp(-0.5\chi^2_{1;1-\alpha})$. The following inequalities

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad {}^{c_\alpha}Bel_Y(A) \leq P_Y(A) \leq {}^{c_\alpha}Pl_Y(A)$$

  hold asymptotically for at least $100(1-\alpha)\%$ of the samples.
- Compatibility with Bayesian inference: If the possibility distribution $\widetilde{L}_{\boldsymbol{x}}$ is combined with a Bayesian prior by Dempster's rule, then $Bel_Y = Pl_Y$ is equal to the Bayesian predictive distribution of Y.

Summary


More general
Less general
Epistemic random fuzzy set theory
(random fuzzy sets)
Possibility theory
(fuzzy sets)
Dempster-Shafer theory
(random sets)
Boolean possibility theory
(sets)
Probability theory
(random points)
(3) Topic of this talk
(2) Extensions
(1) Classical models

Concluding remarks

- Random fuzzy sets (aka "fuzzy random variables") appear in the literature with different interpretations:
  - As a model of a random mechanism for generating fuzzy data (Puri & Ralescu, Gil)
  - As a representation of an ill-known random variable (Kruse, Couso), etc.

  In contrast, the interpretation proposed here is purely epistemic.
- In DS theory, a belief function is not interpreted as the lower envelope of a set of probability measures, and degrees of belief are not interpreted as maximum buying prices for gambles. Walley's theory of imprecise probability is an alternative, competing theory of uncertainty.
- Propagating random (fuzzy) sets requires a combination of
  - Set-propagation techniques (interval analysis, constrained optimization) and
  - Probabilistic techniques (Monte Carlo simulation).

  Work in this direction is in progress.

References

cf. https://www.hds.utc.fr/~tdenoeux

T. Denœux.
Belief functions induced by random fuzzy sets: A general framework for representing uncertain and fuzzy evidence
Fuzzy Sets and Systems (in press, 2020)
https://doi.org/10.1016/j.fss.2020.12.004

T. Denœux and S. Li.
Frequency-Calibrated Belief Functions: Review and New Insights
International Journal of Approximate Reasoning 92:232–254, 2018.

O. Kanjanatarakul, T. Denœux and S. Sriboonchitta
Prediction of future observations using belief functions: a likelihood-based approach.
International Journal of Approximate Reasoning 72:71-94, 2016.