Epistemic random fuzzy sets

Theory and Application to Machine Learning

Thierry Denœux

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

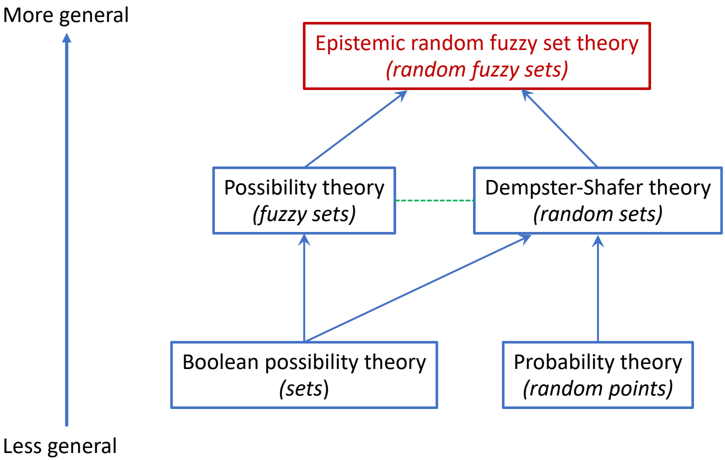
https://www.hds.utc.fr/~tdenoeux

6th School on Belief Functions and their Applications
Ishikawa, Japan, October 31, 2023

A general model of uncertainty

- Modeling uncertainty: a fundamental problem in Artificial/Computational Intelligence
  - Representation of uncertain/imperfect knowledge
  - Reasoning and decision-making with uncertainty
  - Quantification of prediction uncertainty in machine learning, etc.
- As probability theory proved to be too limited, two alternative models were introduced in the late 1970's:
  - Dempster-Shafer (DS) theory = belief functions + Dempster's rule (based on random sets, generalizes Bayesian probability theory)
  - Possibility theory = possibility measures + triangular norms (based on fuzzy sets)
- Each of these two models can be more suitable/practical than the other, depending on the available information (unreliable/uncertain vs. vague/fuzzy).
- The purpose of this lecture is to introduce a more general theoretical framework: Epistemic Random Fuzzy Sets, which unifies the two previous approaches and gives more flexibility in applications.

General picture



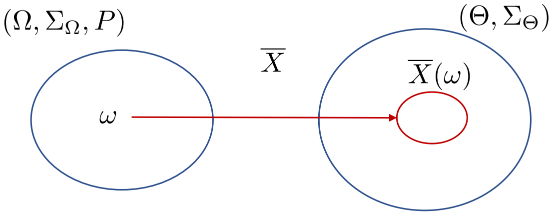
Outline

1. Classical frameworks
   - Random sets and DS theory
   - Fuzzy sets and possibility theory
2. Random fuzzy sets
   - Definitions
   - Gaussian random fuzzy numbers
   - Gaussian random fuzzy vectors
   - Extensions
3. Application to Machine Learning
   - Neural network model
   - Learning
   - Experimental results

Outline

Outline

Random set



Definition (Random Set)

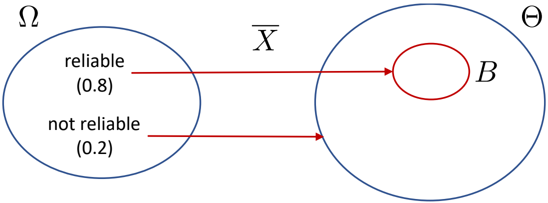
Let $(\Omega, \Sigma_\Omega, P)$ be a probability space, (Θ, Σ_Θ) a measurable space, and $\overline{X} : \Omega \to 2^\Theta$. The 6-tuple $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \overline{X})$ is a random set (RS) iff $\overline{X}$ verifies the following measurability condition:

$$\forall B \in \Sigma_\Theta, \quad \{\omega \in \Omega : \overline{X}(\omega) \cap B \neq \emptyset\} \in \Sigma_\Omega.$$

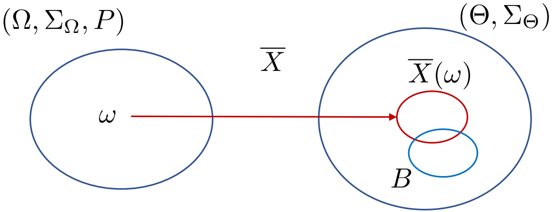
The images $\overline{X}(\omega)$ are called the focal sets of $\overline{X}$.

Interpretation and example

- In DS theory, a RS represents a piece of evidence about a variable X taking values in set Θ (called the frame of discernment):
  - Ω is a set of interpretations of the evidence
  - If interpretation $\omega \in \Omega$ holds, we know that $X \in \overline{X}(\omega)$, and nothing more
  - For any $A \in \Sigma_\Omega$, $P(A)$ is the (subjective) probability that the true interpretation is in A
- Example: unreliable sensor



Belief and plausibility functions



- For any $B \in \Sigma_\Theta$, we can compute
  - The probability that proposition "$X \in B$" is supported by the evidence:

$$Bel_{\overline{X}}(B) = P(\{\omega \in \Omega : \emptyset \neq \overline{X}(\omega) \subseteq B\})$$

  - The probability that proposition "$X \in B$" is consistent with the evidence:

$$Pl_{\overline{X}}(B) = P(\{\omega \in \Omega : \overline{X}(\omega) \cap B \neq \emptyset\})$$
$$= 1 - Bel_{\overline{X}}(B^c)$$

- Mappings $Bel_{\overline{X}} : \Sigma_\Theta \rightarrow [0,1]$ and $Pl_{\overline{X}} : \Sigma_\Theta \rightarrow [0,1]$ are called respectively, belief and plausibility functions.

Interpretation

- In DS theory, $Bel_{\overline{X}}(B)$ and $Pl_{\overline{X}}(B)$ are interpreted, respectively, as a degree of support for B, and a degree of lack of support for B^c, based on some evidence. This model is more flexible than probability theory.
- Examples:

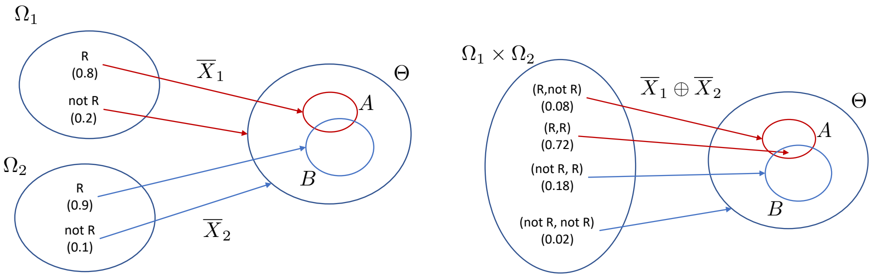
	$Bel(B)$	$Bel(B^c)$	$Pl(B)$	$Pl(B^c)$
evidence for B	0.9	0	1	0.1
mixed evidence for B and B^c	0.6	0.2	0.8	0.4
complete ignorance	0	0	1	1
probabilistic evidence	0.4	0.6	0.4	0.6

Special cases

- Precise but uncertain information: if for all $\omega \in \Omega$, $|\overline{X}(\omega)| = 1$, RS $\overline{X}$ is said to be Bayesian. $Bel_{\overline{X}}$ is then a probability measure, and $Pl_{\overline{X}} = Bel_{\overline{X}}$
- Certain but imprecise information: let $B \subseteq \Theta$; the constant RS $\overline{X}_B$ such that for all $\omega \in \Omega$, $\overline{X}(\omega) = B$ corresponds to set-valued information (we know for sure that $X \in B$, and nothing more).
- In particular, if $\overline{X}_0$ is a RS such that for all $\omega \in \Omega$, $\overline{X}_0(\omega) = \Theta$, $\overline{X}_0$ is said to be vacuous: it represents complete ignorance.

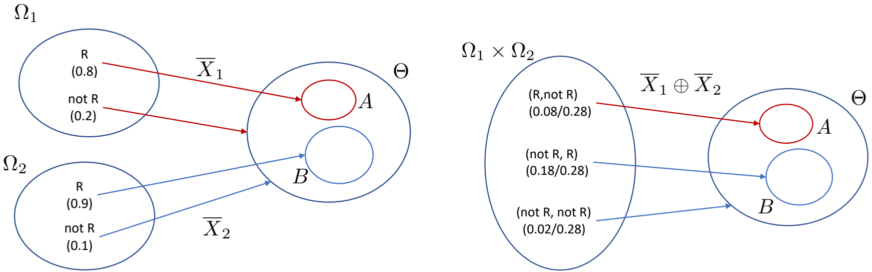
Combination of independent pieces of evidence

Case 1: no conflict

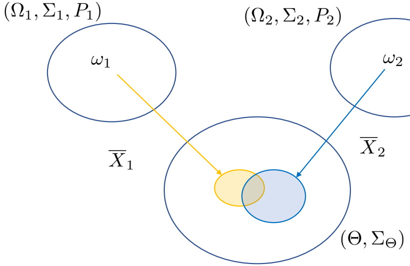


Combination of independent pieces of evidence

Case 2: conflict



Dempster's rule of combination



Definition (Dempster's rule)

Let $(\Omega_i, \Sigma_i, P_i, \Theta, \Sigma_\Theta, \overline{X}_i)$, $i = 1, 2$ be two RSs representing independent pieces of evidence. Their orthogonal sum is the RS

$$(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, P_{12}, \Theta, \Sigma_\Theta, \overline{X}_1 \oplus \overline{X}_2)$$

where $(\overline{X}_1 \oplus \overline{X}_2)(\omega_1, \omega_2) = \overline{X}_1(\omega_1) \cap \overline{X}_2(\omega_2)$ and P_{12} is the product measure $P_1 \times P_2$ conditioned on the set $\Theta^* = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \overline{X}_1(\omega_1) \cap \overline{X}_2(\omega_2) \neq \emptyset\}$

Properties

- Commutativity:
$$\overline{X}_1 \oplus \overline{X}_2 = \overline{X}_2 \oplus \overline{X}_1$$
- Associativity:
$$(\overline{X}_1 \oplus \overline{X}_2) \oplus \overline{X}_3 = \overline{X}_1 \oplus (\overline{X}_2 \oplus \overline{X}_3)$$
- Neutral element: if $\overline{X}_0$ is vacuous,
$$\overline{X}_0 \oplus \overline{X} = \overline{X}$$
- Let $pl_{\overline{X}} : \Theta \to [0,1]$ be the contour function defined by $pl_{\overline{X}}(\theta) = Pl_{\overline{X}}(\{\theta\})$ for all $\theta \in \Theta$. We have
$$pl_{\overline{X}_1 \oplus \overline{X}_2} \propto pl_{\overline{X}_1} pl_{\overline{X}_2}$$
- Generalization of Bayesian conditioning: if $\overline{X}$ is a Bayesian RS and $\overline{X}_B$ is a constant RS with focal set B, then $\overline{X} \oplus \overline{X}_B$ is a Bayesian RS, and
$$Bel_{\overline{X} \oplus \overline{X}_B} = Bel_{\overline{X}}(\cdot \mid B)$$

Outline

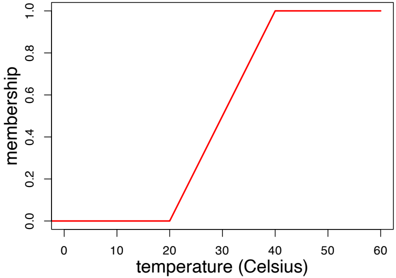
1. Classical frameworks
   - Random sets and DS theory
   - Fuzzy sets and possibility theory

2. Random fuzzy sets
   - Definitions
   - Gaussian random fuzzy numbers
   - Gaussian random fuzzy vectors
   - Extensions

3. Application to Machine Learning
   - Neural network model
   - Learning
   - Experimental results

Fuzzy set

- A fuzzy subset of a set Θ is a mapping $\widetilde{F} : \Theta \mapsto [0, 1]$.
- It represents a generalized subset of Θ with unsharp boundaries: $\widetilde{F}(\theta)$ is the degree of membership of θ to the fuzzy set $\widetilde{F}$.
- Example: if $\Theta = [-60, 60]$ is the range of outside air temperatures, the notion of "hot temperature" can be represented by the fuzzy subset



Additional definitions

- The height of $\widetilde{F}$ is

$$\text{hgt}(\widetilde{F}) = \sup_{\theta \in \Theta} \widetilde{F}(\theta)$$

- $\widetilde{F}$ is normal if $\text{hgt}(\widetilde{F}) = 1$
- For any $\alpha \in [0, 1]$, the α-cut of $\widetilde{F}$ is the set

$$^{\alpha}\widetilde{F} = \{\theta \in \Theta : \widetilde{F}(\theta) \geq \alpha\}$$

Possibility and necessity

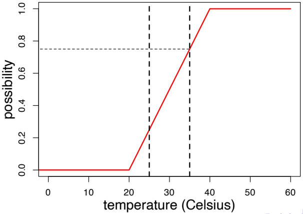
- Let X be a variable taking values in Θ. Assume that we receive a piece of evidence telling us that "X is $\widetilde{F}$", where $\widetilde{F}$ is a normal fuzzy subset of Θ.
- Such evidence can be seen as a flexible constraint on the true value of X. We define
  - The possibility distribution of X as $\pi_{\widetilde{F}} = \widetilde{F}$
  - The degree of possibility that $X \in B$ for $B \subseteq \Theta$ as

$$\Pi_{\widetilde{F}}(B) = \sup_{\theta \in B} \pi_{\widetilde{F}}(\theta)$$

  - The degree of necessity that $X \in B$ as

$$N_{\widetilde{F}}(B) = 1 - \Pi_{\widetilde{F}}(B^c)$$

- Example:



Possibility and necessity measures

- The mapping $\Pi_{\widetilde{F}} : 2^\Theta \mapsto [0,1]$ is called a possibility measure, and $N_{\widetilde{F}} : 2^\Theta \mapsto [0,1]$ is the dual necessity measure.
- Properties: for any $A, B \subseteq \Theta$,

$$\Pi_{\widetilde{F}}(A \cup B) = \max(\Pi_{\widetilde{F}}(A), \Pi_{\widetilde{F}}(B))$$

$$N_{\widetilde{F}}(A \cap B) = \min(N_{\widetilde{F}}(A), N_{\widetilde{F}}(B))$$

- $N_{\widetilde{F}}$ is a belief function, and $\Pi_{\widetilde{F}}$ is the dual plausibility function. For this reason, it has been claimed that possibility theory is a special case of DS theory. However, the two theories have different mechanisms for combining information.

Combination of possibility distributions

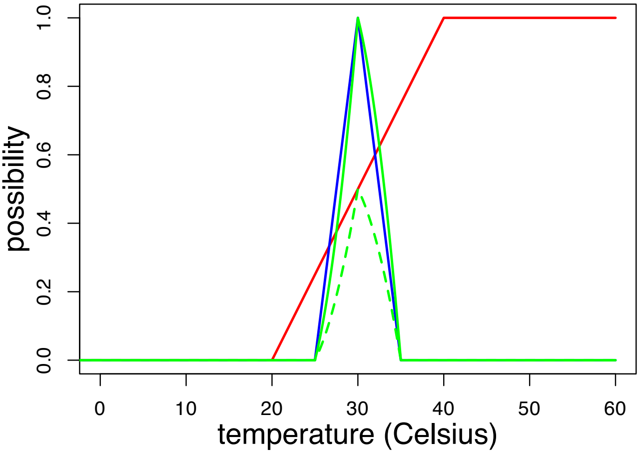
- Assume that we receive two independent pieces of information telling us that "X is $\widetilde{F}$" and "X is $\widetilde{G}$", where $\widetilde{F}$ and $\widetilde{G}$ are two fuzzy subsets of Θ.
- We can deduce that "X is $\widetilde{F} \cap_\top \widetilde{G}$", where $\cap_\top$ is a fuzzy set intersection operator based on a t-norm $\top$. The most common choices for $\top$ are the minimum and product t-norms.
- The intersection of two normal fuzzy sets is generally not normal. We define the normalized $\top$-intersection as

$$(\widetilde{F} \cap_\top^* \widetilde{G})(\theta) = \frac{\widetilde{F}(\theta) \top \widetilde{G}(\theta)}{\text{hgt}(\widetilde{F} \cap_\top \widetilde{G})}$$

- The normalized intersection is associative iff $\top =$ product; the normalized product intersection is denoted by $\odot$.

Example

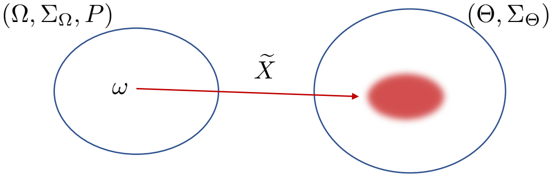
$\widetilde{F}$ = hot, $\widetilde{G}$ = around 30



Outline

Outline

Random fuzzy set



Definition (Random Fuzzy Set)

Let $(\Omega, \Sigma_\Omega, P)$ be a probability space, (Θ, Σ_Θ) a measurable space, and $\widetilde{X}$ a mapping from Ω to the set $[0,1]^\Theta$ of fuzzy subsets of Θ. The 6-tuple $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \widetilde{X})$ is a random fuzzy set (RFS) iff for any $\alpha \in [0,1]$, the mapping

$$
\begin{aligned}
{}^\alpha\widetilde{X} &: \Omega \to 2^\Theta \\
\omega &\mapsto {}^\alpha[\widetilde{X}(\omega)] = \{\theta \in \Theta : \widetilde{X}(\omega)(\theta) \geq \alpha\}
\end{aligned}
$$

is a random set.

Interpretation

- We use RFSs as a model of unreliable and fuzzy evidence[1]:
  - Θ is the domain of an uncertain variable/quantity X
  - Ω is a set of interpretations of a piece of evidence about X
  - $\forall A \in \Sigma_\Omega$, $P(A)$ is the probability that the true interpretation lies in A
  - If $\omega \in \Omega$ holds, we know that "X is $\widetilde{X}(\omega)$", i.e., X is constrained by the possibility distribution $\widetilde{X}(\omega)$.
- Such RFSs are called "epistemic" to stress that they represent a state of knowledge.
- Example: a witness tells us that "the temperature was hot on Monday", and this witness is 50% reliable
  - $\Omega = \{\text{rel}, \neg\text{rel}\}$, $p(\text{rel}) = 0.5$
  - X = temperature on Monday in Celsius, $\Theta = [-60, 60]$
  - $\widetilde{X}(\text{rel})$ = hot (a fuzzy subset of Θ), $\widetilde{X}(\neg\text{rel}) = \Theta$

[1] This interpretation is different from previous interpretations of RFSs as a model of random mechanism for generating fuzzy data (Puri & Ralescu, Gil), or as imperfect knowledge of a random variable (Kruse & Meyer, Couso & Sánchez)

Belief and plausibility functions

- If interpretation $\omega \in \Omega$ holds, the degrees of possibility and necessity that X belongs to $B \in \Sigma_\Theta$ are

$$\Pi_{\widetilde{X}(\omega)}(B) = \sup_{\theta \in B} \widetilde{X}(\omega)(\theta), \quad N_{\widetilde{X}(\omega)}(B) = 1 - \Pi_{\widetilde{X}(\omega)}(B^c)$$

- The expected necessity and possibility degrees (Zadeh, 1979) are

$$Bel_{\widetilde{X}}(B) = \int_\Omega N_{\widetilde{X}(\omega)}(B) dP(\omega), \quad Pl_{\widetilde{X}}(B) = \int_\Omega \Pi_{\widetilde{X}(\omega)}(B) dP(\omega).$$

Proposition (Zadeh, 1979; Couso & Sánchez, 2011)

Function $Bel_{\widetilde{X}}$ is a completely monotone capacity (a belief function), and $Pl_{\widetilde{X}}$ is the dual plausibility function .

A RFS is thus (like a random set) a way of specifying a belief function. The RFS model is more flexible.

Example

- Continuing the previous example, what are the degrees of belief and plausibility that $X \in B = [25, 35]$?
- We have

$$\Pi_{\widetilde{X}(\text{rel})}(B) = 0.75, \quad \Pi_{\widetilde{X}(\neg\text{rel})}(B) = 1$$

so

$$Pl_{\widetilde{X}}(B) = 0.5 \times 0.75 + 0.5 \times 1 = 0.875$$

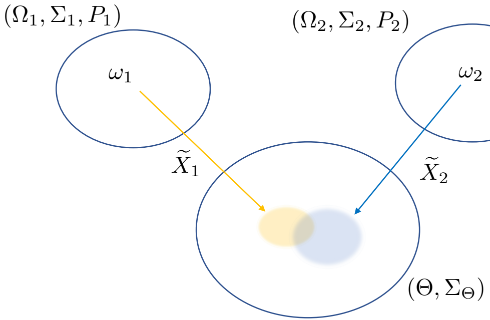
- Now,

$$N_{\widetilde{X}(\text{rel})}(B) = 0, \quad N_{\widetilde{X}(\neg\text{rel})}(B) = 0$$

so

$$Bel_{\widetilde{X}}(B) = 0$$

Combination of independent RFSs



- We consider two RFSs $\widetilde{X}_1 : \Omega_1 \rightarrow [0,1]^\Theta$ and $\widetilde{X}_2 : \Omega_2 \rightarrow [0,1]^\Theta$ representing independent pieces of evidence.
- if $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$ both hold, we can deduce "X is $\widetilde{X}_1(\omega_1) \cap \widetilde{X}_2(\omega_2)$", where $\cap$ denotes fuzzy intersection.
- We need (1) a definition of fuzzy intersection and (2) a way to handle possible conflict (inconsistency) between the two sources.

Definition of intersection and conflict

- Fuzzy intersection: as mentioned before, the normalized product intersection is suitable for combining fuzzy information from independent sources, and it is associative.
- With fuzzy sets, conflict is a matter of degree. We define the fuzzy set of consistent pairs of interpretations as

$$\widetilde{\Theta}^*(\omega_1, \omega_2) = \sup_{\Theta} \left(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2) \right)$$

- The product measure $P_1 \times P_2$ is conditioned on fuzzy event $\widetilde{\Theta}^*$:

$$\widetilde{P}_{12}(B) = \frac{(P_1 \times P_2)(B \cap \widetilde{\Theta}^*)}{(P_1 \times P_2)(\widetilde{\Theta}^*)} = \frac{\int_{\Omega_1} \int_{\Omega_2} B(\omega_1, \omega_2) \widetilde{\Theta}^*(\omega_1, \omega_2) dP_2(\omega_2) dP_1(\omega_1)}{\int_{\Omega_1} \int_{\Omega_2} \widetilde{\Theta}^*(\omega_1, \omega_2) dP_2(\omega_2) dP_1(\omega_1)}$$

where $B(\cdot, \cdot)$ denotes the indicator function of B. This process is called soft normalization.

Product-intersection rule[2]

Definition (Product-intersection rule)

The orthogonal sum of $\widetilde{X}_1$ and $\widetilde{X}_2$ is the RFS

$$(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \widetilde{P}_{12}, \Theta, \Sigma_\Theta, \widetilde{X}_1 \oplus \widetilde{X}_2)$$

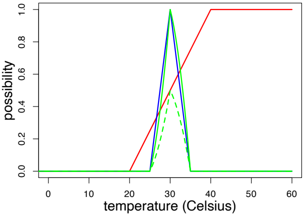
where

$$(\widetilde{X}_1 \oplus \widetilde{X}_2)(\omega_1, \omega_2) = \widetilde{X}_1(\omega_1) \odot \widetilde{X}_2(\omega_2)$$

and $\widetilde{P}_{12}$ is the product measure $P_1 \times P_2$ conditioned on the fuzzy set $\widetilde{\Theta}^*(\omega_1, \omega_2)$. This operation is called the product intersection of $\widetilde{X}_1$ and $\widetilde{X}_2$ (with soft normalization).

[2] T. Denœux. Reasoning with fuzzy and uncertain evidence using epistemic random fuzzy sets: general framework and practical models. *Fuzzy Sets and Systems* 453:1–36, 2023

Example



- As before, let $\Theta = [-60, +60]$, $\widetilde{F}$ = hot, $\widetilde{G}$ = around 30.
- Evidence 1: $\Omega_1 = \{\text{rel}, \neg\text{rel}\}$, $p_1(\text{rel}) = 0.5$, $\widetilde{X}_1(\text{rel}) = \widetilde{F}$, $\widetilde{X}_1(\neg\text{rel}) = \Theta$.
- Evidence 2: $\Omega_2 = \{\text{rel}, \neg\text{rel}\}$, $p_2(\text{rel}) = 0.7$, $\widetilde{X}_2(\text{rel}) = \widetilde{G}$, $\widetilde{X}_2(\neg\text{rel}) = \Theta$.

- $\widetilde{\Theta}^*(\text{rel}, \text{rel}) = 0.5$, $\widetilde{\Theta}^*(\text{rel}, \neg\text{rel}) = \widetilde{\Theta}^*(\neg\text{rel}, \text{rel}) = \widetilde{\Theta}^*(\neg\text{rel}, \neg\text{rel}) = 1$
- $(P_1 \times P_2)(\widetilde{\Theta}^*) = 0.35 \times 0.5 + 0.15 \times 1 + 0.35 \times 1 + 0.15 \times 1 = 0.825$
- $\widetilde{p}_{12}(\text{rel}, \text{rel}) = 0.35 \times 0.5/0.825$, $\widetilde{p}_{12}(\neg\text{rel}, \text{rel}) = 0.35/0.825$, $\widetilde{p}_{12}(\text{rel}, \neg\text{rel}) = 0.15/0.825$, $\widetilde{p}_{12}(\neg\text{rel}, \neg\text{rel}) = 0.15/0.825$
- $(\widetilde{X}_1 \oplus \widetilde{X}_2)(\text{rel}, \text{rel}) = \widetilde{F} \odot \widetilde{G}$, $(\widetilde{X}_1 \oplus \widetilde{X}_2)(\text{rel}, \neg\text{rel}) = \widetilde{F}$, $(\widetilde{X}_1 \oplus \widetilde{X}_2)(\text{rel}, \neg\text{rel}) = \widetilde{G}$, $(\widetilde{X}_1 \oplus \widetilde{X}_2)(\neg\text{rel}, \neg\text{rel}) = \Theta$.

Properties

1. Commutativity, associativity
2. Generalization of Dempster's rule and the normalized product intersection of possibility distributions
3. Multiplication of contour functions

$$pl_{\widetilde{X}_1 \oplus \widetilde{X}_2} \propto pl_{\widetilde{X}_1} pl_{\widetilde{X}_2}$$

4. Generalization of conditioning of a probability measure by a fuzzy event: if $\overline{X}$ is a Bayesian RS and $\widetilde{X}_{\widetilde{B}}$ is a constant RF with fuzzy focal set $\widetilde{B}$, then $\overline{X} \oplus \widetilde{X}_{\widetilde{B}}$ is a Bayesian RS, and

$$Bel_{\overline{X} \oplus \widetilde{X}_{\widetilde{B}}} = Bel_{\overline{X}}(\cdot | \widetilde{B})$$

i.e.

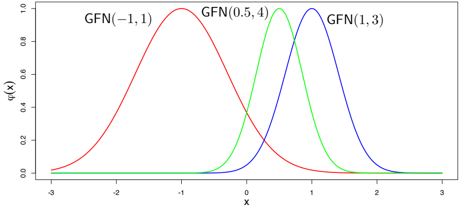
$$\forall A \in \Sigma_\Theta, \quad Bel_{\overline{X} \oplus \widetilde{X}_{\widetilde{B}}}(A) = \frac{\int_A \widetilde{B}(\theta) dBel_{\overline{X}}(\theta)}{\int_\Theta \widetilde{B}(\theta) dBel_{\overline{X}}(\theta)}$$

Outline

Motivation

- In probability theory and statistics, the Gaussian probability distribution is widely used because it allows for simple calculations and easy manipulation (conditioning, marginalization, etc.)
- Until recently, a similar workable model had been missing in DS theory to represent uncertainty on continuous variables (possibility distributions or p-boxes are not closed under Dempster's rule)
- Gaussian random fuzzy numbers (GRFNs) and extensions are simple models of RFSs making it possible to define families of belief functions on $\mathbb{R}$, $\mathbb{R}^p$, $[a, b]$, etc., which can be easily combined by the product-intersection operator $\oplus$.

Gaussian fuzzy numbers



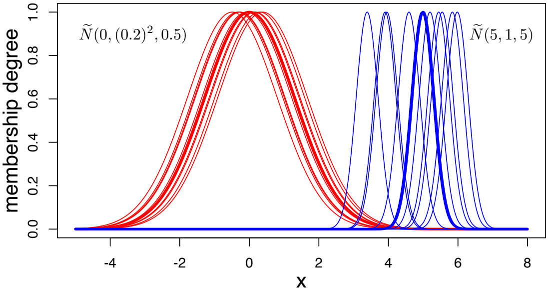
Definition (Gaussian fuzzy number)

A Gaussian fuzzy number (GFN) with mode $m \in \mathbb{R}$ and precision $h \geq 0$ is a fuzzy subset of $\mathbb{R}$ with membership function $\varphi(x; m, h) = \exp\left(-\frac{h}{2}(x-m)^2\right)$. It is denoted by $\mathrm{GFN}(m, h)$.

Proposition

$GFN(m_1, h_1) \odot GFN(m_2, h_2) = GFN(m_{12}, h_1 + h_2)$ with $m_{12} = \frac{h_1 m_1 + h_2 m_2}{h_1 + h_2}$

Gaussian random fuzzy numbers



Definition (Gaussian random fuzzy number)

A Gaussian random fuzzy number (GRFN) $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$ with mean μ, variance σ^2 and precision $h \geq 0$ is a Gaussian fuzzy number GFN(M, h) whose mode is a Gaussian random variable: $M \sim N(\mu, \sigma^2)$. Formally, it is a mapping $\widetilde{X} : \Omega \mapsto [0,1]^{\mathbb{R}}$ such that $\widetilde{X}(\omega) = \text{GFN}(M(\omega), h)$ with $M \sim N(\mu, \sigma^2)$.

Special cases

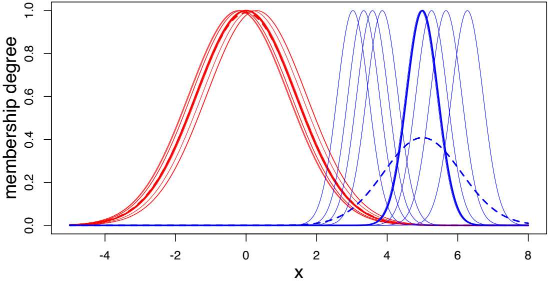
- If $h = 0$, $\widetilde{X}(\omega) = \mathbb{R}$ for all ω: $\widetilde{X}$ induces the vacuous belief function on $\mathbb{R}$; it represents complete ignorance
- If $h = +\infty$, $\widetilde{X}$ is equivalent to a GRV with mean μ and variance σ^2:

$$\widetilde{N}(\mu, \sigma^2, +\infty) = N(\mu, \sigma^2)$$

- If $\sigma^2 = 0$, $\widetilde{X}$ is equivalent to a Gaussian possibility distribution:

$$\widetilde{N}(\mu, 0, h) = GFN(\mu, h)$$

Contour function



- The contour function of $\widetilde{X}$ is

$$pl_{\widetilde{X}}(x) = \frac{1}{\sqrt{1+h\sigma^2}} \exp\left(-\frac{h(x-\mu)^2}{2(1+h\sigma^2)}\right)$$

- Remarks: (1) for all x, $pl_{\widetilde{X}}(x) \to 0$ when $\sigma^2 \neq 0$ and $h \to \infty$; (2) when $\sigma^2 = 0$, $pl_{\widetilde{X}}$ is the possibility distribution of $\widetilde{X} \sim GFN(\mu, h)$.

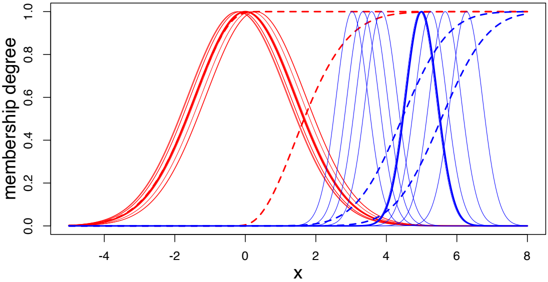
Belief and plausibility of intervals

$$
\begin{aligned}
Bel_{\widetilde{X}}([x,y]) = \Phi\left(\frac{y-\mu}{\sigma}\right) - \Phi\left(\frac{x-\mu}{\sigma}\right) - \\
pl_{\widetilde{X}}(x)\left[\Phi\left(\frac{(x+y)/2-\mu}{\sigma\sqrt{h\sigma^2+1}}\right) - \Phi\left(\frac{x-\mu}{\sigma\sqrt{h\sigma^2+1}}\right)\right] - \\
pl_{\widetilde{X}}(y)\left[\Phi\left(\frac{y-\mu}{\sigma\sqrt{h\sigma^2+1}}\right) - \Phi\left(\frac{(x+y)/2-\mu}{\sigma\sqrt{h\sigma^2+1}}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
Pl_{\widetilde{X}}([x,y]) = \Phi\left(\frac{y-\mu}{\sigma}\right) - \Phi\left(\frac{x-\mu}{\sigma}\right) + pl_{\widetilde{X}}(x)\Phi\left(\frac{x-\mu}{\sigma\sqrt{h\sigma^2+1}}\right) + \\
pl_{\widetilde{X}}(y)\left[1 - \Phi\left(\frac{y-\mu}{\sigma\sqrt{h\sigma^2+1}}\right)\right]
\end{aligned}
$$

where Φ is the normal cumulative distribution function (cdf).

Lower and upper distribution functions



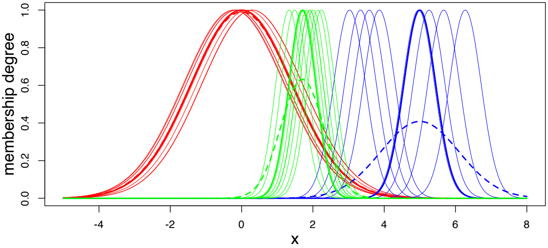
In particular, the lower and upper cdfs of $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$ are

$$Bel_{\widetilde{X}}((-\infty, y]) = \Phi\left(\frac{y-\mu}{\sigma}\right) - pl_{\widetilde{X}}(y)\Phi\left(\frac{y-\mu}{\sigma\sqrt{h\sigma^2+1}}\right)$$

and

$$Pl_{\widetilde{X}}((-\infty, y]) = \Phi\left(\frac{y-\mu}{\sigma}\right) + pl_{\widetilde{X}}(y)\left[1 - \Phi\left(\frac{y-\mu}{\sigma\sqrt{h\sigma^2+1}}\right)\right]$$

Combination of GRFNs



Theorem (Product-intersection of GRFNs)

Given two GRFNs $\widetilde{X}_1 \sim \widetilde{N}(\mu_1, \sigma_1^2, h_1)$ *and* $\widetilde{X}_2 \sim \widetilde{N}(\mu_2, \sigma_2^2, h_2)$, *we have*

$$\boxed{\widetilde{X}_1 \oplus \widetilde{X}_2 \sim \widetilde{N}(\widetilde{\mu}_{12}, \widetilde{\sigma}_{12}^2, h_1 + h_2)}$$

(Expressions of $\widetilde{\mu}_{12}$ *and* $\widetilde{\sigma}_{12}^2$ *on next slide)*

Combination of GRFNs

Expressions of $\widetilde{\mu}_{12}$ and $\widetilde{\sigma}^2_{12}$

$$\widetilde{\mu}_{12} = \frac{h_1\widetilde{\mu}_1 + h_2\widetilde{\mu}_2}{h_1 + h_2}, \quad \widetilde{\sigma}^2_{12} = \frac{h_1^2\widetilde{\sigma}_1^2 + h_2^2\widetilde{\sigma}_2^2 + 2\rho h_1 h_2 \widetilde{\sigma}_1\widetilde{\sigma}_2}{(h_1 + h_2)^2}$$

with

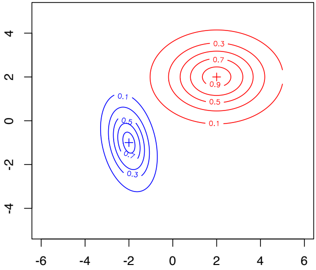
$$\widetilde{\mu}_1 = \frac{\mu_1(1 + \overline{h}\sigma_2^2) + \mu_2\overline{h}\sigma_1^2}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}, \quad \widetilde{\mu}_2 = \frac{\mu_2(1 + \overline{h}\sigma_1^2) + \mu_1\overline{h}\sigma_2^2}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}$$

$$\widetilde{\sigma}_1^2 = \frac{\sigma_1^2(1 + \overline{h}\sigma_2^2)}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}, \quad \widetilde{\sigma}_2^2 = \frac{\sigma_2^2(1 + \overline{h}\sigma_1^2)}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}$$

$$\rho = \frac{\overline{h}\sigma_1\sigma_2}{\sqrt{(1 + \overline{h}\sigma_1^2)(1 + \overline{h}\sigma_2^2)}} \quad \text{and} \quad \overline{h} = \frac{h_1 h_2}{h_1 + h_2}$$

Outline

Gaussian fuzzy vectors

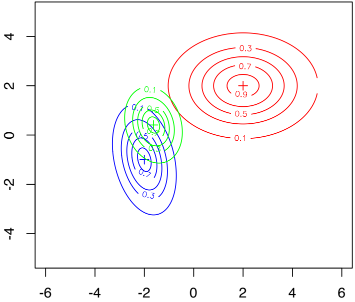


Definition (Gaussian fuzzy vector)

A p-dimensional Gaussian fuzzy vector (GFV) with mode $\boldsymbol{m} \in \mathbb{R}^p$ and symmetric and positive semidefinite precision matrix $\boldsymbol{H} \in \mathbb{R}^{p \times p}$, denoted by GFV($\boldsymbol{m}, \boldsymbol{H}$), is a fuzzy subset of $\mathbb{R}^p$ with membership function

$$\varphi(\boldsymbol{x}; \boldsymbol{m}, \boldsymbol{H}) = \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m})^T \boldsymbol{H}(\boldsymbol{x} - \boldsymbol{m})\right).$$

Product intersection of GFVs



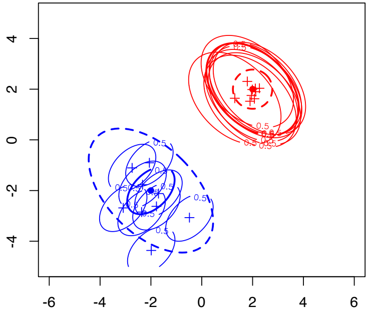
Proposition

$$GFV(\boldsymbol{m}_1, \boldsymbol{H}_1) \odot GFV(\boldsymbol{m}_2, \boldsymbol{H}_2) = GFV(\boldsymbol{m}_{12}, \boldsymbol{H}_{12}),$$

with

$$\boldsymbol{m}_{12} = (\boldsymbol{H}_1 + \boldsymbol{H}_2)^{-1}(\boldsymbol{H}_1\boldsymbol{m}_1 + \boldsymbol{H}_2\boldsymbol{m}_2) \quad \textit{and} \quad \boldsymbol{H}_{12} = \boldsymbol{H}_1 + \boldsymbol{H}_2.$$

Gaussian random fuzzy vectors

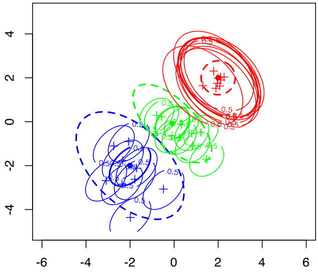


Definition (Gaussian random fuzzy vector)

A Gaussian random fuzzy vector (GRFV) $\widetilde{X} \sim \widetilde{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{H})$ with covariance matrix $\boldsymbol{\Sigma}$ and precision matrix $\boldsymbol{H}$ is random fuzzy set $\widetilde{X} : \Omega \to [0,1]^{\mathbb{R}^p}$ such that

$$\widetilde{X}(\omega) = \text{GFV}(\boldsymbol{M}(\omega), \boldsymbol{H}) \quad \text{with} \quad \boldsymbol{M} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Combination of GRFVs



Theorem (Product-intersection of GRFVs)

Let $\widetilde{X}_1 \sim \widetilde{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{H}_1)$ and $\widetilde{X}_2 \sim \widetilde{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{H}_2)$ be two independent GRFVs such that matrices $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$, $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are all positive definite. We have

$$\boxed{\widetilde{X}_1 \oplus \widetilde{X}_2 \sim \widetilde{N}(\widetilde{\boldsymbol{\mu}}_{12}, \widetilde{\boldsymbol{\Sigma}}_{12}, \boldsymbol{H}_1 + \boldsymbol{H}_2)}$$

(Expressions of $\widetilde{\boldsymbol{\mu}}_{12}$ and $\widetilde{\boldsymbol{\Sigma}}_{12}$ on next slide)

Combination of GRFVs

Expressions of $\widetilde{\mu}_{12}$ and $\widetilde{\Sigma}_{12}$

$$\widetilde{\mu}_{12} = A\widetilde{\mu} \quad \text{and} \quad \widetilde{\Sigma}_{12} = A\widetilde{\Sigma}A^T$$

where A is the constant $p \times 2p$ matrix defined as

$$A = H_{12}^{-1} \begin{pmatrix} H_1 & H_2 \end{pmatrix}$$

$$\widetilde{\Sigma} = \begin{pmatrix} \Sigma_1^{-1} + \overline{H} & -\overline{H} \\ -\overline{H} & \Sigma_2^{-1} + \overline{H} \end{pmatrix}^{-1}$$

$$\widetilde{\mu} = \begin{pmatrix} \overline{H}^{-1}\Sigma_1^{-1} + I_p & -I_p \\ -I_p & \overline{H}^{-1}\Sigma_2^{-1} + I_p \end{pmatrix}^{-1} \begin{pmatrix} \overline{H}^{-1}\Sigma_1^{-1} & 0 \\ 0 & \overline{H}^{-1}\Sigma_2^{-1} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

and

$$\overline{H} = (H_1^{-1} + H_2^{-1})^{-1}.$$

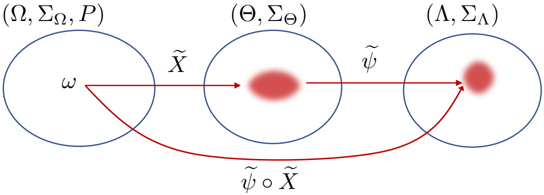
Outline

Limitations of the GRFN model

- The domain of a GRFN is the whole real line, making the model unsuitable for representing belief functions on a real interval such as $(0, +\infty)$ or $[a, b]$.
- A GRFN is unimodal and symmetric about the mean μ; these properties may not always reflect an agent's actual beliefs.
- We need more flexible parameterized families of random fuzzy numbers and vectors with different supports and different "shapes", while maintaining the closure property under the product-intersection rule.
- This can be achieved by composing a RFS $\widetilde{X} : \Omega \to [0, 1]^{\Theta}$ with a one-to-one mapping[3] from Θ to another space Λ, to obtain a a RFS $\widetilde{Y} : \Omega \to [0, 1]^{\Lambda}$.

[3] T. Denœux. Parametric families of continuous belief functions based on generalized Gaussian random fuzzy numbers. *Fuzzy Sets and Systems*, 471:108679, 2023.

Transformation of a RFS



- Let ψ be a one-to-one mapping from Θ to some set Λ.
- Zadeh's extension principle allows us to extend ψ to fuzzy subsets of Θ; the extended mapping $\widetilde{\psi} : [0,1]^\Theta \rightarrow [0,1]^\Lambda$ is defined as

$$\forall \widetilde{F} \in [0,1]^\Theta, \quad \widetilde{\psi}(\widetilde{F})(\lambda) = \sup_{\lambda=\psi(\theta)} \widetilde{F}(\theta) = \widetilde{F}(\psi^{-1}(\lambda)).$$

Proposition

If $\widetilde{X} : \Omega \mapsto [0,1]^\Theta$ is a RFS, the composed mapping $\widetilde{\psi} \circ \widetilde{X} : \Omega \mapsto [0,1]^\Lambda$, such that $(\widetilde{\psi} \circ \widetilde{X})(\omega) = \widetilde{\psi}[\widetilde{X}(\omega)]$, is a RFS.

Main results

Proposition

Let Σ_Λ be the image of Σ_Θ by ψ. For any $C \in \Sigma_\Lambda$,

$$Bel_{\widetilde{\psi}\circ\widetilde{X}}(C) = Bel_{\widetilde{X}}(\psi^{-1}(C))$$

and

$$Pl_{\widetilde{\psi}\circ\widetilde{X}}(C) = Pl_{\widetilde{X}}(\psi^{-1}(C))$$

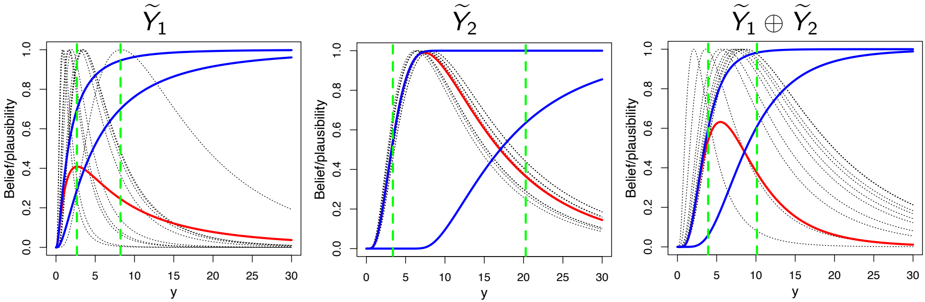
Theorem

Let $\widetilde{X}_1 : \Omega_1 \to [0,1]^\Theta$ and $\widetilde{X}_2 : \Omega_2 \to [0,1]^\Theta$ be two RFSs representing independent evidence. We have

$$\boxed{\widetilde{\psi} \circ (\widetilde{X}_1 \oplus \widetilde{X}_2) = (\widetilde{\psi} \circ \widetilde{X}_1) \oplus (\widetilde{\psi} \circ \widetilde{X}_2)}$$

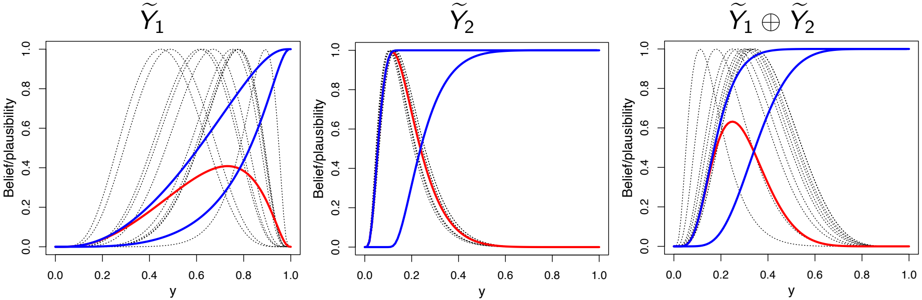
Lognormal RFNs

- Let $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$ and $\psi = \exp$.
- The RFN $\widetilde{Y} = \widetilde{\psi} \circ \widetilde{X}$ with support equal to $(0, +\infty)$ is called a lognormal RFN; we write $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \log)$.



Logit-normal RFNs

- Let $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$ and $\psi(x) = [1 + \exp(-x)]^{-1}$.
- The RFN $\widetilde{Y} = \widetilde{\psi} \circ \widetilde{X}$ with support equal to $(0, 1)$ is called a logit-normal RFN; we write $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \text{logit})$, where $\text{logit}(y) = \log \frac{y}{1-y}$.



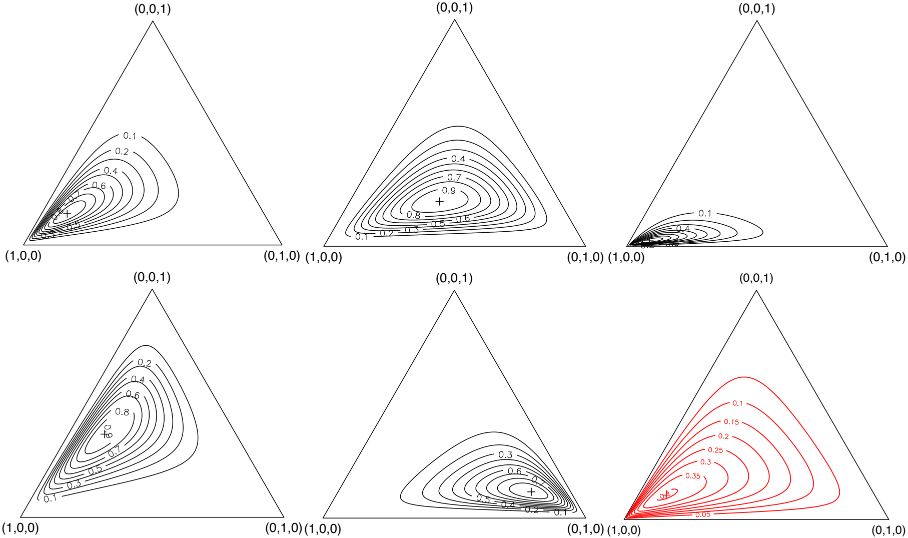
Logistic-normal RFVs

- Let $\widetilde{X} \sim \widetilde{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{H})$ be a $p-1$ dimensional GRFV and ψ_S the softmax transformation from $\mathbb{R}^{p-1}$ to the simplex $\mathcal{S}_p$ of p-dimensional probability vectors:

$$\psi_S(\boldsymbol{x}) = \left[\frac{\exp(x_1)}{1+\sum_{j=1}^{p}\exp(x_j)}, \ldots, \frac{\exp(x_{p-1})}{1+\sum_{j=1}^{p}\exp(x_j)}, \frac{1}{1+\sum_{j=1}^{p}\exp(x_j)}\right]^T$$

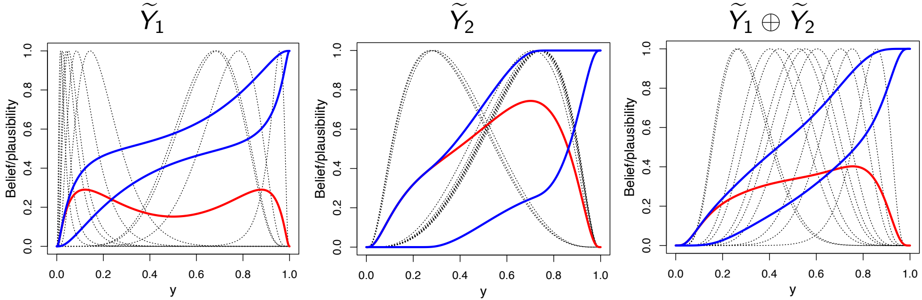
- The random fuzzy vector $\widetilde{Y} = \widetilde{\psi}_S \circ \widetilde{X}$ is a logistic-normal RFV; we write $\widetilde{Y} \sim T\widetilde{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{H}, \psi_S^{-1})$. Its support is the simplex $\mathcal{S}_p$.

Logistic-normal RFVs: Example



Mixtures of (transformed) GRFNs

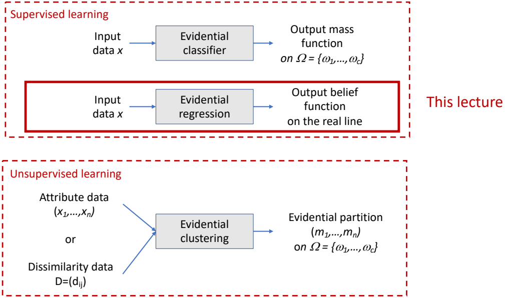
- Mixtures of GRFNs = a GFN whose mode is a mixture of GRVs.
- Can be transformed by a one-to-one mappings.
- Defines new families of RFNs closed under the product-intersection rule.
- Example: $\widetilde{Y}_1 \sim 0.5 T\widetilde{N}(2, 1, 2, \text{logit}) + 0.5 T\widetilde{N}(-2, 1, 2, \text{logit})$, $\widetilde{Y}_2 \sim 0.3 T\widetilde{N}(-1, 0.1^2, 1, \text{logit}) + 0.7 T\widetilde{N}(1, 0.1^2, 1, \text{logit})$



Outline

1. Classical frameworks
   - Random sets and DS theory
   - Fuzzy sets and possibility theory
2. Random fuzzy sets
   - Definitions
   - Gaussian random fuzzy numbers
   - Gaussian random fuzzy vectors
   - Extensions
3. Application to Machine Learning
   - Neural network model
   - Learning
   - Experimental results

Evidential Machine Learning



- Evidential Machine Learning (ML): an approach to ML in which uncertainty is quantified by belief functions.
- Previous work has mainly focussed on clustering and classification because these learning tasks only require belief functions on finite frames.
- With models for defining and combining belief functions on continuous frames, it is now possible to tackle other learning tasks, such as regression.

The ENNreg model

- We consider a regression problem: the task is to predict a continuous random response variable Y from p input variables $\boldsymbol{X} = (X_1, \ldots, X_p)$, based on a learning set $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$.
- We have proposed a neural network model[4] (ENNreg), which quantifies uncertainty about the response Y given input vector $\boldsymbol{X} = \boldsymbol{x}$ by a GRFN $\widetilde{Y}(\boldsymbol{x})$ with associated belief function $Bel_{\widetilde{Y}(\boldsymbol{x})}$.
- ENNreg is based on prototypes. The distances to the prototypes are treated as independent pieces of evidence about the response and are combined by the product-intersection rule.

[4] T. Denœux. Quantifying Prediction Uncertainty in Regression using Random Fuzzy Sets: the ENNreg model. *IEEE Transactions on Fuzzy Systems*, 31(10):3690–3699, 2023.

Outline

Propagation equations (1/2)

- Let $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K$ denote K vectors in the p-dimensional input space, called prototypes.
- The similarity between input vector $\boldsymbol{x}$ and prototype $\boldsymbol{w}_k$ is measured by

$$\boxed{s_k(\boldsymbol{x}) = \exp(-\gamma_k^2 \|\boldsymbol{x} - \boldsymbol{w}_k\|^2)}$$

where $\gamma_k > 0$ is a scale parameter.

- The evidence from prototype $\boldsymbol{w}_k$ is represented by a GRFN

$$\boxed{\widetilde{Y}_k(\boldsymbol{x}) \sim \widetilde{N}(\mu_k(\boldsymbol{x}), \sigma_k^2, s_k(\boldsymbol{x}) h_k)}$$

where σ_k^2 and h_k are variance and precision parameters, and

$$\mu_k(\boldsymbol{x}) = \boldsymbol{\beta}_k^T \boldsymbol{x} + \beta_{k0}$$

where $\boldsymbol{\beta}_k$ is a p-dimensional vector of coefficients, and β_{k0} is a scalar parameter.

Propagation equations (2/2)

- The output $\widetilde{Y}(\boldsymbol{x})$ for input $\boldsymbol{x}$ is computed as

$$\boxed{\widetilde{Y}(\boldsymbol{x}) = \widetilde{Y}_1(\boldsymbol{x}) \boxplus \ldots \boxplus \widetilde{Y}_K(\boldsymbol{x})}$$

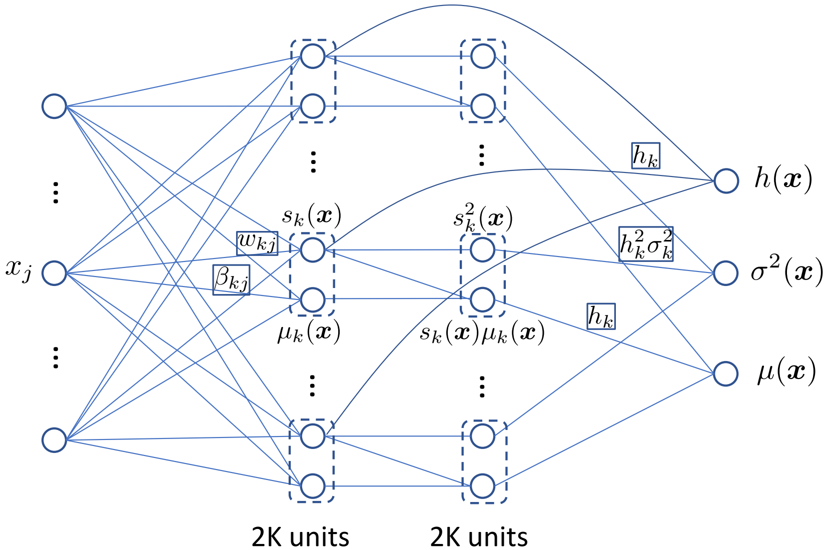
  where $\boxplus$ denotes product intersection without soft normalization (to simplify calculations).

- We have $\widetilde{Y}(\boldsymbol{x}) \sim \widetilde{N}\left(\mu(\boldsymbol{x}), \sigma^2(\boldsymbol{x}), h(\boldsymbol{x})\right)$, with

$$\mu(\boldsymbol{x}) = \frac{\sum_{k=1}^K s_k(\boldsymbol{x}) h_k \mu_k(\boldsymbol{x})}{\sum_{k=1}^K s_k(\boldsymbol{x}) h_k}$$

$$\sigma^2(\boldsymbol{x}) = \frac{\sum_{k=1}^K s_k^2(\boldsymbol{x}) h_k^2 \sigma_k^2}{\left(\sum_{k=1}^K s_k(\boldsymbol{x}) h_k\right)^2} \quad \text{and} \quad h(\boldsymbol{x}) = \sum_{k=1}^K s_k(\boldsymbol{x}) h_k$$

Neural network architecture



Outline

Negative log-likelihood loss (probabilistic forecasts)

- In the case of a probabilistic forecast with pdf $\widehat{f}$, we typically measure the prediction error (or loss) by the negative log-likelihood

$$\mathcal{L}(y, \widehat{f}) = -\ln \widehat{f}(y)$$

- We actually never observe a real number y with infinite precision, but an interval $[y]_\epsilon = [y - \epsilon, y + \epsilon]$ centered at y. The probability of that interval is

$$\widehat{P}([y]_\epsilon) = \widehat{F}(y + \epsilon) - \widehat{F}(y - \epsilon) \approx 2\widehat{f}(y)\epsilon,$$

  So, $\mathcal{L}(y, \widehat{f}) = -\ln \widehat{P}([y]_\epsilon) + \text{cst}$.
- Generalization in the case of prediction in the form of a belief function?

Extension

- $\mathcal{L}_\epsilon(y, \widetilde{Y}) = -\ln Bel_{\widetilde{Y}}([y]_\epsilon)$ does not work (does not reward imprecision).
- $\mathcal{L}_\epsilon(y, \widetilde{Y}) = -\ln Pl_{\widetilde{Y}}([y]_\epsilon)$ also does not work (minimized when $\widetilde{Y}$ is vacuous).
- Proposal:

$$\boxed{\mathcal{L}_{\lambda,\epsilon}(y, \widetilde{Y}) = -\lambda \ln Bel_{\widetilde{Y}}([y]_\epsilon) - (1-\lambda) \ln Pl_{\widetilde{Y}}([y]_\epsilon)}$$

  with $\lambda \in [0, 1]$ and $\epsilon > 0$.
- Smaller values of λ correspond to more cautious predictions.

Training

- The network is trained by minimizing the regularized average loss

$$C^{(R)}_{\lambda,\epsilon,\xi,\rho}(\Psi) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_{\lambda,\epsilon}(y_i, \widetilde{Y}(\boldsymbol{x}_i;\Psi))}_{C_{\lambda,\epsilon}(\Psi)} + \underbrace{\frac{\xi}{K}\sum_{k=1}^{K} h_k}_{R_1(\Psi)} + \underbrace{\frac{\rho}{K}\sum_{k=1}^{K}\gamma_k^2}_{R_2(\Psi)},$$

  where
  - $R_1(\Psi)$ has the effect of reducing the number of prototypes used for the prediction (setting $h_k = 0$ amounts to discarding prototype k)
  - $R_2(\Psi)$ shrinks the solution towards a linear model (setting $\gamma_k = 0$ for all k yields a linear model).
- Heuristics: $\lambda = 0.9$, $\epsilon = 0.01\widehat{\sigma}_Y$, ξ and ρ tuned using a validation set or cross-validation.

Calibration

- For any $\alpha \in (0,1]$, we define an α-level belief prediction interval (BPI) as an interval $\mathcal{B}_\alpha(\boldsymbol{x})$ centered at $\mu(\boldsymbol{x})$, such that $Bel_{\widetilde{Y}(\boldsymbol{x})}(\mathcal{B}_\alpha(\boldsymbol{x})) = \alpha$.
- The predictions are said to be calibrated if, for all $\alpha \in (0,1]$, α-level BPIs have a coverage probability at least equal to α, i.e,

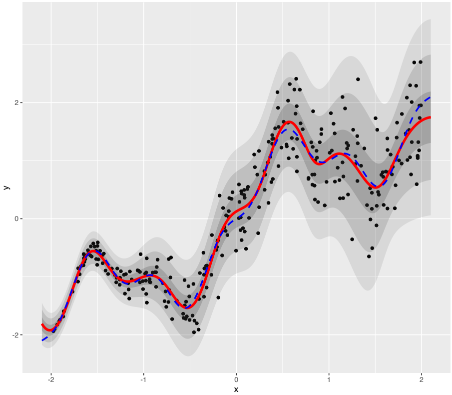
$$\boxed{\forall \alpha \in (0,1], \quad P_{\boldsymbol{X},Y}\left(Y \in \mathcal{B}_\alpha(\boldsymbol{X})\right) \geq \alpha} \tag{1}$$

- As in the probabilistic case, the calibration of evidential predictions can be checked graphically using a calibration plot (see infra).
- The precision output $h(\boldsymbol{x})$ can be multiplied by a constant $c > 0$ to ensure (1) with predictions as precise as possible.

Example

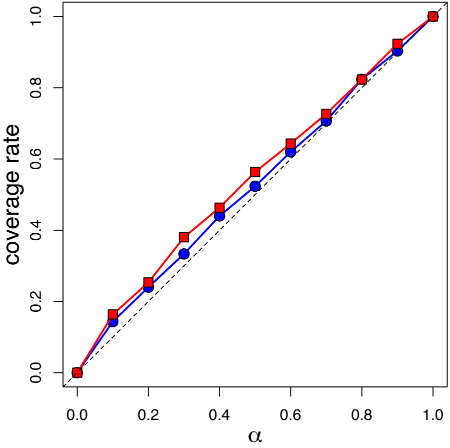
We consider iid data with one-dimensional input $X \sim \text{Unif}(-2, 2)$ and

$$Y = X + (\sin 3X)^3 + \frac{X+2}{4\sqrt{2}} U, \quad U \sim N(0,1)$$



- Learning and validation sets of size $n = 300$.
- Network with $K = 30$ prototypes initialized by the k-means algorithm.
- ξ and ρ determined by minimizing the validation MSE.
- Shown: true regression function (blue), expected values $\mu(x)$ (red) with BPIs at levels 0.5, 0.9 and 0.99

Calibration curves



Calibration curves for the probabilistic PIs $\mu(x) \pm u_{(1+\alpha)/2}\sigma(x)$ (in blue) and the BPIs (in red)

Outline

Data sets

	n	p	response
Boston	506	13	`medv`
Energy	768	8	`Y2`
Concrete	1030	8	`strength`
Yacht	308	6	`Y`
Wine	1599	11	`quality`
kin8nm	8192	8	`V9`
Crime	1994	100	`ViolentCrimesPerPop`
Residential	372	103	`V10`
Airfoil	1503	5	`Y`
Bike	731	9	`cnt`

Comparison with classical methods (RMS)

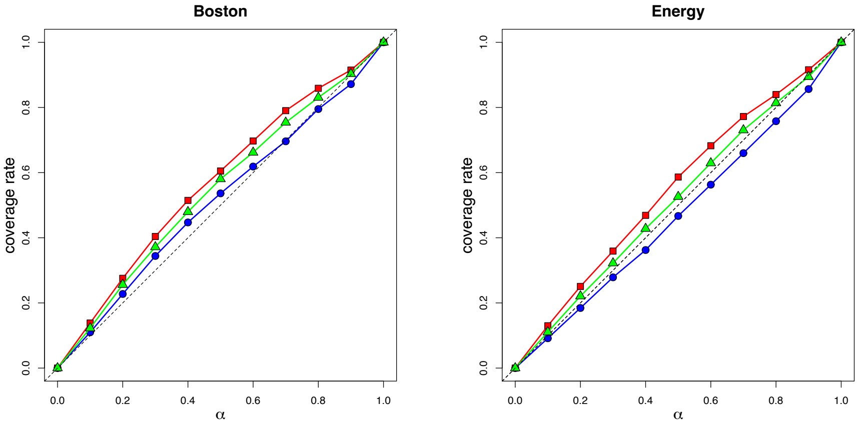
	ENNreg	RBF	RVM	SVM	GP	RF	MLP
Boston	**2.87 ± 0.14**	3.31 ± 0.19	3.42 ± 0.17	3.17 ± 0.15	3.70 ± 0.22	**3.11 ± 0.14**	**3.14 ± 0.14**
Energy	**1.06 ± 0.05**	2.06 ± 0.08	1.79 ± 0.05	1.39 ± 0.06	2.58 ± 0.07	1.75 ± 0.06	**0.95 ± 0.16**
Concr.	5.10 ± 0.12	6.30 ± 0.19	6.38 ± 0.16	5.62 ± 0.13	6.93 ± 0.13	**4.64 ± 0.12**	**4.82 ± 0.16**
Yacht	**0.44 ± 0.04**	2.00 ± 0.20	1.88 ± 0.20	1.93 ± 0.11	6.12 ± 0.31	0.96 ± 0.08	**0.50 ± 0.05**
Wine	0.63 ± 0.01	0.63 ± 0.01	0.80 ± 0.02	0.61 ± 0.01	0.61 ± 0.01	**0.56 ± 0.01**	0.77 ± 0.01
kin8nm	0.08 ± 0.00	0.11 ± 0.00	–	0.09 ± 0.00	0.08 ± 0.00	0.14 ± 0.00	**0.07 ± 0.00**
Crime	**0.14 ± 0.00**	**0.14 ± 0.00**	**0.14 ± 0.00**	**0.14 ± 0.00**	**0.14 ± 0.00**	**0.14 ± 0.00**	**0.14 ± 0.00**
Resid.	**0.11 ± 0.01**	0.16 ± 0.01	0.17 ± 0.01	0.15 ± 0.01	0.22 ± 0.01	0.16 ± 0.01	0.14 ± 0.01
Airfoil	**1.46 ± 0.03**	1.70 ± 0.04	2.58 ± 0.04	2.37 ± 0.04	2.49 ± 0.04	**1.44 ± 0.04**	1.53 ± 0.04
Bike	**6.59 ± 0.19**	**6.49 ± 0.15**	**6.64 ± 0.14**	7.11 ± 0.16	7.55 ± 0.14	6.86 ± 0.17	9.68 ± 0.20

Comparison with SOTA methods (RMS & NLL)

	RMS				
	ENNreg	PBP	MC-dropout	Deep ens.	Deep ev. reg.
Boston	**2.87** ± **0.14**	**3.01** ± **0.18**	**2.97** ± **0.19**	**3.28** ± **1.00**	**3.06** ± **0.16**
Energy	**1.06** ± **0.05**	1.80 ± 0.05	1.66 ± 0.04	2.09 ± 0.29	2.06 ± 0.10
Concr.	**5.10** ± **0.12**	5.67 ± 0.09	**5.23** ± **0.12**	6.03 ± 0.58	5.85 ± 0.15
Yacht	**0.44** ± **0.04**	1.02 ± 0.05	1.11 ± 0.09	1.58 ± 0.48	1.57 ± 0.56
Wine	**0.63** ± **0.01**	**0.64** ± **0.01**	**0.62** ± **0.01**	**0.64** ± **0.04**	**0.61** ± **0.02**
kin8nm	**0.08** ± **0.00**	0.10 ± 0.00	0.10 ± 0.00	0.09 ± 0.00	0.09 ± 0.00

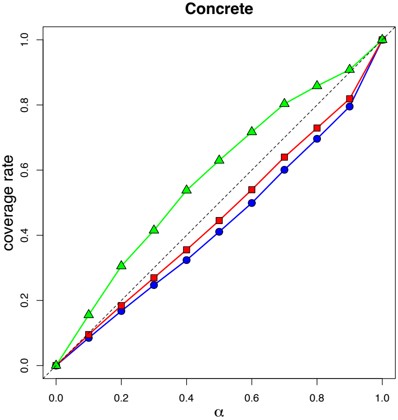
	NLL				
	ENNreg	PBP	MC-dropout	Deep ens.	Deep ev. reg.
Boston	2.53 ± 0.07	2.57 ± 0.09	**2.46** ± **0.06**	**2.41** ± **0.25**	**2.35** ± **0.06**
Energy	**1.14** ± **0.07**	2.04 ± 0.02	1.99 ± 0.02	**1.38** ± **0.22**	1.39 ± 0.06
Concr.	3.38 ± 0.13	3.16 ± 0.02	**3.04** ± **0.02**	**3.06** ± **0.18**	**3.01** ± **0.02**
Yacht	**0.13** ± **0.12**	1.63 ± 0.02	1.55 ± 0.03	1.18 ± 0.21	1.03 ±0.19
Wine	**0.94** ± **0.01**	0.97 ± 0.01	**0.93** ± **0.01**	**0.94** ± **0.12**	**0.89** ± **0.05**
kin8nm	**-1.19** ± **0.00**	-0.90 ± 0.01	-0.95 ± 0.01	**-1.20** ± **0.02**	-1.24 ± 0.01

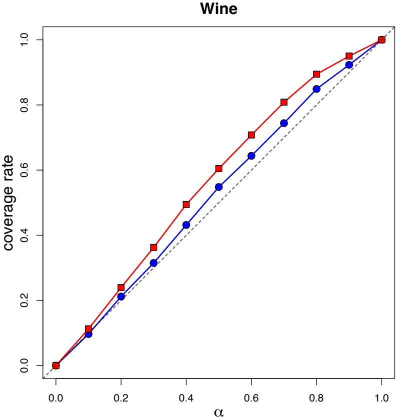
Calibration plots



Probabilistic predictions (blue), raw evidential predictions (red) and adjusted evidential predictions (green).

Calibration plots





Probabilistic predictions (blue), raw evidential predictions (red) and adjusted evidential predictions (green).

Summary

- The theory of epistemic RFSs is a very general framework, generalizing both possibility theory and DS theory. It allows one to represent and reason with uncertain, imprecise and vague information.
- Practical models of RFNs and RFVs indexed by 3 parameters (mode, variance and precision) make it possible to define belief functions on continuous frames that can be easily manipulated and combined, overcoming a limitation of DS theory.
- As an example of application, we have described the ENNreg model, a regression neural network based on the combination of GRFNs. The network output for input vector $\boldsymbol{x}$ is a GRFN defined by three numbers:
  - a point prediction $\mu(\boldsymbol{x})$
  - a variance $\sigma^2(\boldsymbol{x})$ measuring random uncertainty
  - a precision $h(\boldsymbol{x})$ representing epistemic uncertainty
- Other applications include knowledge elicitation and statistical inference[5].

[5] T. Denœux. Parametric families of continuous belief functions based on generalized Gaussian random fuzzy numbers. *Fuzzy Sets and Systems*, 471:108679, 2023.

References on epistemic RFSs

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, 424:63–91, 2021

T. Denœux
Reasoning with fuzzy and uncertain evidence using epistemic random fuzzy sets: general framework and practical models.
Fuzzy Sets and Systems, 453:1–36, 2023

T. Denœux
Parametric families of continuous belief functions based on generalized Gaussian random fuzzy numbers.
Fuzzy Sets and Systems, 471:108679, 2023.

References on the ENNreg model

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

T. Denœux
Quantifying Prediction Uncertainty in Regression using Random Fuzzy Sets: the ENNreg model.
IEEE Transactions on Fuzzy Systems, 31(10):3690–3699, 2023.

T. Denœux
evreg: Evidential Regression
R package version 1.0.2, 2023. Available:
https://CRAN.R-project.org/package=evreg