

# 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>

FUZZ-IEEE 2023  
Songdo Incheon, Korea  
August 13, 2023

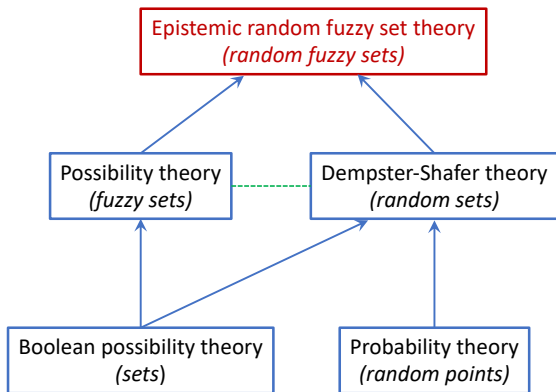
# A general model of uncertainty

- 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

More general

Less general



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

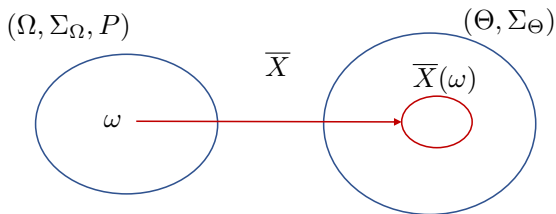
# Outline

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

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

# Random set



## Definition (Random Set)

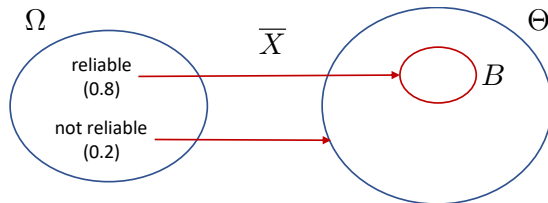
Let  $(\Omega, \Sigma_\Omega, P)$  be a probability space,  $(\Theta, \Sigma_\Theta)$  a measurable space, and  $\bar{X} : \Omega \rightarrow 2^\Theta$ . The 6-tuple  $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \bar{X})$  is a **random set (RS)** iff  $\bar{X}$  verifies the following measurability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

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

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

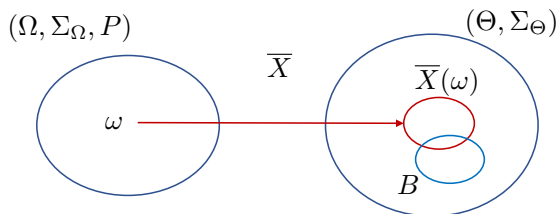
# Interpretation and example

- In DS theory, a RS represents a **piece of evidence** about a variable  $X$  taking values in set  $\Theta$  (called the **frame of discernment**):
  - $\Omega$  is a set of interpretations of the evidence
  - If interpretation  $\omega \in \Omega$  holds, we know that  $X \in \bar{X}(\omega)$ , 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:

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

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

$$\begin{aligned} Pl_{\bar{X}}(B) &= P(\{\omega \in \Omega : \bar{X}(\omega) \cap B \neq \emptyset\}) \\ &= 1 - Bel_{\bar{X}}(B^c) \end{aligned}$$

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

# Mathematical characterization

## Proposition

A mapping  $Bel : \Sigma_{\Theta} \mapsto [0, 1]$  is a *belief function* (for some RS  $\bar{X}$ ) iff it verifies the following properties:

- ①  $Bel(\emptyset) = 0$
- ②  $Bel(\Theta) = 1$
- ③ For any  $k \geq 2$  and any collection  $B_1, \dots, B_k$  of elements of  $\Sigma_{\Theta}$ ,

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

[Complete monotonicity]

# Interpretation

- In DS theory,  $Bel_{\bar{X}}(B)$  and  $Pl_{\bar{X}}(B)$  are interpreted, respectively, as a **degree of belief that  $X \in B$** , and a **degree of lack of belief in  $X \notin B$** , based on some evidence. This model is more flexible than probability theory.
- Examples:

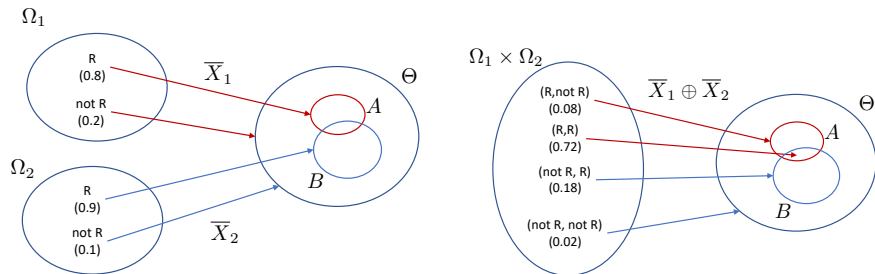
	$Bel(B)$	$Bel(B^c)$	$Pl(B)$	$Pl(B^c)$
evidence for $B$	0.9	0	1	0.1
mixed evidence for $B$ and $B^c$	0.6	0.2	0.8	0.4
complete ignorance	0	0	1	1
probabilistic evidence	0.4	0.6	0.4	0.6

# 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**.

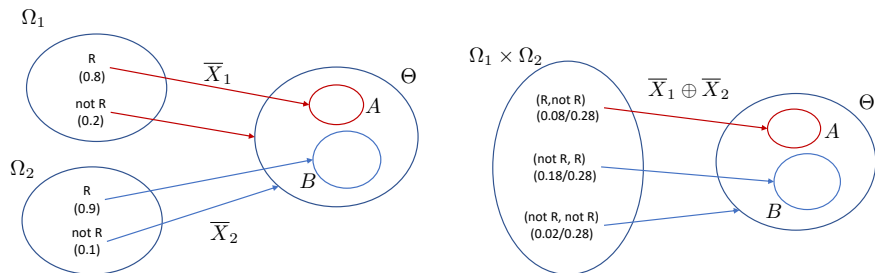
# Combination of independent pieces of evidence

Case 1: no conflict

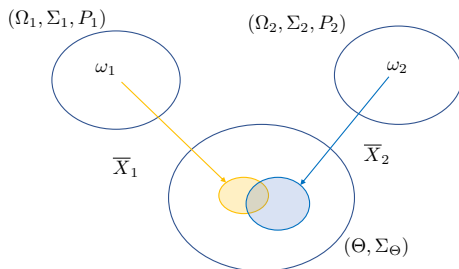


# Combination of independent pieces of evidence

Case 2: conflict



# Dempster's rule of combination



## Definition (Dempster's rule)

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

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

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

# Properties

- Commutativity:

$$\bar{X}_1 \oplus \bar{X}_2 = \bar{X}_2 \oplus \bar{X}_1$$

- Associativity:

$$(\bar{X}_1 \oplus \bar{X}_2) \oplus \bar{X}_3 = \bar{X}_1 \oplus (\bar{X}_2 \oplus \bar{X}_3)$$

- Neutral element: if  $\bar{X}_0$  is vacuous,

$$\bar{X}_0 \oplus \bar{X} = \bar{X}$$

- Let  $pl_{\bar{X}} : \theta \rightarrow [0, 1]$  be the **contour function** defined by  $pl_{\bar{X}}(\theta) = Pl_{\bar{X}}(\{\theta\})$  for all  $\theta \in \Theta$ . We have

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

- Generalization of **Bayesian conditioning**: if  $\bar{X}$  is a Bayesian RS and  $\bar{X}_B$  is a constant RS with focal set  $B$ , then  $\bar{X} \oplus \bar{X}_B$  is a Bayesian RS, and

$$Bel_{\bar{X} \oplus \bar{X}_B} = Bel_{\bar{X}}(\cdot \mid B)$$

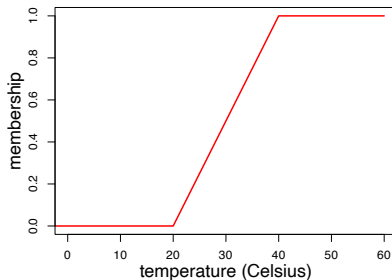


# Outline

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

# Fuzzy set

- A **fuzzy subset** of a set  $\Theta$  is a mapping  $\tilde{F} : \Theta \mapsto [0, 1]$ .
- It represents a generalized subset of  $\Theta$  with unsharp boundaries:  $\tilde{F}(\theta)$  is the **degree of membership** of  $\theta$  to the fuzzy set  $\tilde{F}$ .
- 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  $\tilde{F}$  is

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

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

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

# Possibility and necessity

- Let  $X$  be a variable taking values in  $\Theta$ . Assume that we receive a piece of evidence telling us that “ $X$  is  $\tilde{F}$ ”, where  $\tilde{F}$  is a normal fuzzy subset of  $\Theta$ .
- Such evidence can be seen as a **flexible constraint** on the true value of  $X$ . We define

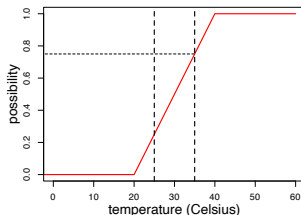
- The **possibility distribution** of  $X$  as  $\pi_{\tilde{F}} = \tilde{F}$
- The degree of possibility that  $X \in B$  for  $B \subseteq \Theta$  as

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

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

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

- Example:



# Possibility and necessity measures

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

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

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

- $N_{\tilde{F}}$  is a **belief function**, and  $\Pi_{\tilde{F}}$  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

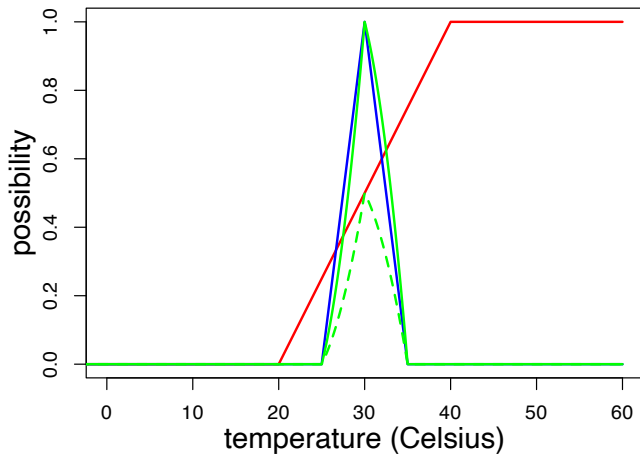
- Assume that we receive two independent pieces of information telling us that “ $X$  is  $\tilde{F}$ ” and “ $X$  is  $\tilde{G}$ ”, where  $\tilde{F}$  and  $\tilde{G}$  are two fuzzy subsets of  $\Theta$ .
- We can deduce that “ $X$  is  $\tilde{F} \cap_{\top} \tilde{G}$ ”, where  $\cap_{\top}$  is 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

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

- When  $\top = \text{product}$ , the normalized intersection is associative and is denoted by  $\odot$ . **Product intersection** has a reinforcement effect that is appropriate when the information sources are assumed to be independent.

# Example

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



# Outline

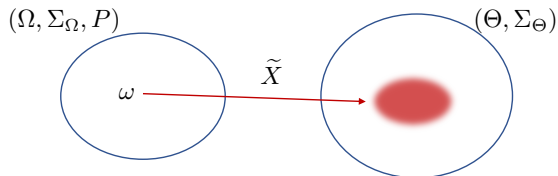
- 1 Classical frameworks
  - Random sets and DS theory
  - Fuzzy sets and possibility theory
- 2 Random fuzzy sets
  - Definitions
  - Gaussian random fuzzy numbers
  - Gaussian random fuzzy vectors
- 3 Application to Machine Learning
  - Neural network model
  - Learning
  - Experimental results



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

# Random fuzzy set



## Definition (Random Fuzzy Set)

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

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

is a random set.

# Interpretation

- We use RFSs as a model of **unreliable and fuzzy evidence**<sup>1</sup>:
  - $\Theta$  is the domain of an uncertain variable/quantity  $X$
  - $\Omega$  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  $\tilde{X}(\omega)$ ”, i.e.,  $X$  is constrained by the possibility distribution  $\tilde{X}(\omega)$ .
- 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
  - $\Omega = \{\text{rel}, \neg\text{rel}\}$ ,  $p(\text{rel}) = 0.5$
  - $X = \text{temperature on Monday in Celsius}$ ,  $\Theta = [-60, 60]$
  - $\tilde{X}(\text{rel}) = \text{hot}$  (a fuzzy subset of  $\Theta$ ),  $\tilde{X}(\neg\text{rel}) = \Theta$

<sup>1</sup>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

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

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

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

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

Function  $Bel_{\tilde{X}}$  is a completely monotone capacity (a **belief function**), and  $Pl_{\tilde{X}}$  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

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

so

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

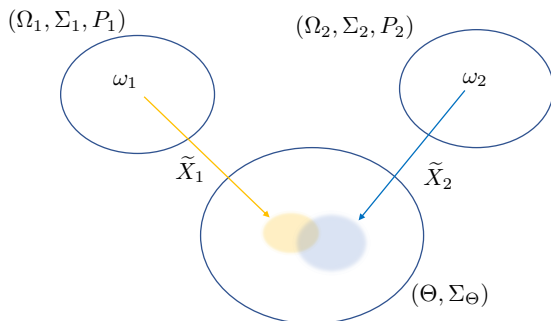
- Now,

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

so

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

# Combination of independent RFSs



- We consider two RFSs  $\tilde{X}_1 : \Omega_1 \rightarrow [0, 1]^\Theta$  and  $\tilde{X}_2 : \Omega_2 \rightarrow [0, 1]^\Theta$  representing **independent pieces of evidence**.
- if  $\omega_1 \in \Omega_1$  and  $\omega_2 \in \Omega_2$  both hold, we can deduce “ $X$  is  $\tilde{X}_1(\omega_1) \cap \tilde{X}_2(\omega_2)$ ”, 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

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

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

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

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

# Product-intersection rule<sup>2</sup>

## Definition (Product-intersection rule)

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

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

where

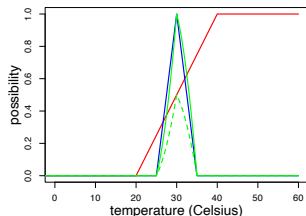
$$(\tilde{X}_1 \oplus \tilde{X}_2)(\omega_1, \omega_2) = \tilde{X}_1(\omega_1) \odot \tilde{X}_2(\omega_2)$$

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

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

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

# Properties

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

$$p|_{\tilde{X}_1 \oplus \tilde{X}_2} \propto p|_{\tilde{X}_1} p|_{\tilde{X}_2}$$

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

$$Bel_{\bar{X} \oplus \tilde{X}_{\tilde{B}}} = Bel_{\bar{X}}(\cdot | \tilde{B})$$

i.e.

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

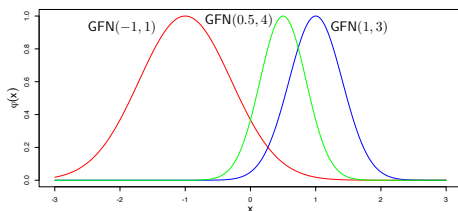
# Outline

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

# 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



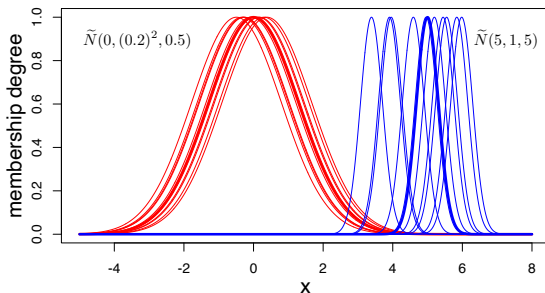
## Definition (Gaussian fuzzy number)

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  $\text{GFN}(m, h)$ .

## Proposition

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

# Gaussian random fuzzy numbers



## Definition (Gaussian random fuzzy number)

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

# Special cases

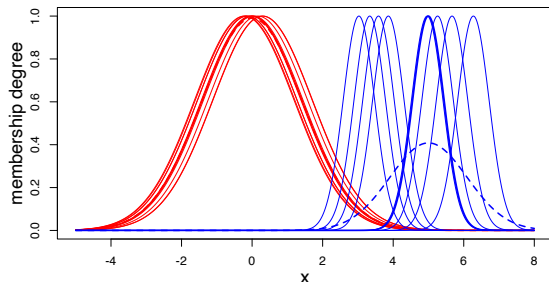
- If  $h = 0$ ,  $\tilde{X}(\omega) = \mathbb{R}$  for all  $\omega$ :  $\tilde{X}$  induces the **vacuous belief function** on  $\mathbb{R}$ ; it represents complete ignorance
- If  $h = +\infty$ ,  $\tilde{X}$  is equivalent to a GRV with mean  $\mu$  and variance  $\sigma^2$ :

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

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

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

# Contour function



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

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

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



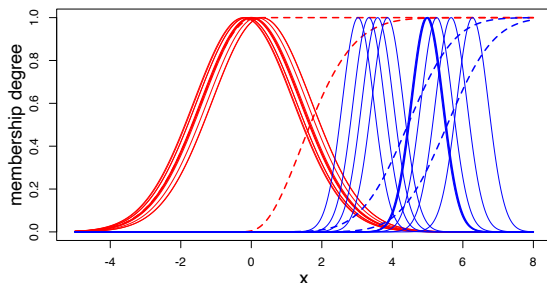
# Belief and plausibility of intervals

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

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

where  $\Phi$  is the normal cumulative distribution function (cdf).

# Lower and upper distribution functions



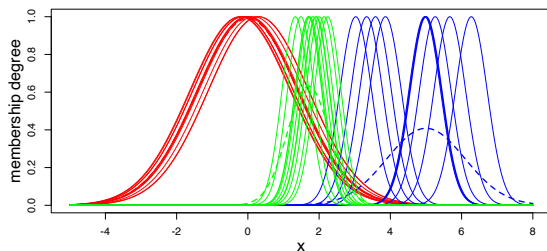
In particular, the lower and upper cdfs of  $\tilde{X} \sim \tilde{N}(\mu, \sigma^2, h)$  are

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

and

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

# Combination of GRFNs



## Theorem (Product-intersection of GRFNs)

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

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

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

# Combination of GRFNs

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

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

with

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

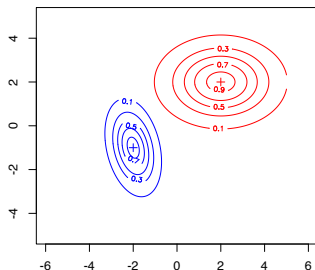
$$\tilde{\sigma}_1^2 = \frac{\sigma_1^2(1 + \bar{h}\sigma_2^2)}{1 + \bar{h}(\sigma_1^2 + \sigma_2^2)}, \quad \tilde{\sigma}_2^2 = \frac{\sigma_2^2(1 + \bar{h}\sigma_1^2)}{1 + \bar{h}(\sigma_1^2 + \sigma_2^2)}$$

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

# 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**
- 3 Application to Machine Learning
  - Neural network model
  - Learning
  - Experimental results

# Gaussian fuzzy vectors

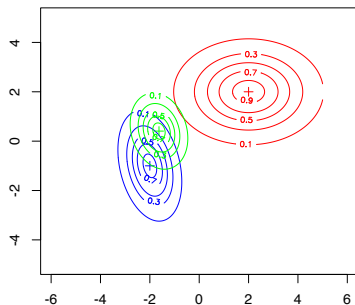


## Definition (Gaussian fuzzy vector)

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

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

# Product intersection of GFVs



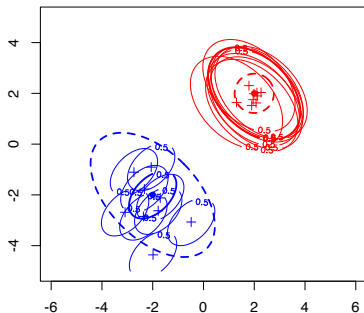
## Proposition

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

with

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

# Gaussian random fuzzy vectors



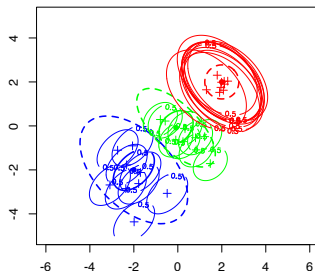
## Definition (Gaussian random fuzzy vector)

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

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



# Combination of GRFVs



## Theorem (Product-intersection of GRFVs)

Let  $\tilde{X}_1 \sim \tilde{N}(\mu_1, \Sigma_1, H_1)$  and  $\tilde{X}_2 \sim \tilde{N}(\mu_2, \Sigma_2, H_2)$  be two independent GRFVs such that matrices  $\Sigma_1$ ,  $\Sigma_2$ ,  $H_1$  and  $H_2$  are all positive definite. We have

$$\tilde{X}_1 \oplus \tilde{X}_2 \sim \tilde{N}(\tilde{\mu}_{12}, \tilde{\Sigma}_{12}, H_1 + H_2)$$

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

# Combination of GRFVs

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

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

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

$$\mathbf{A} = \mathbf{H}_{12}^{-1} (\mathbf{H}_1 \quad \mathbf{H}_2)$$

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma_1^{-1} + \bar{\mathbf{H}} & -\bar{\mathbf{H}} \\ -\bar{\mathbf{H}} & \Sigma_2^{-1} + \bar{\mathbf{H}} \end{pmatrix}^{-1}$$

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

and

$$\bar{\mathbf{H}} = (\mathbf{H}_1^{-1} + \mathbf{H}_2^{-1})^{-1}.$$

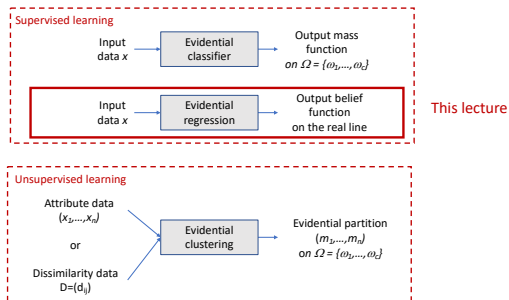
# Extension of the GRFN model

- The GRFN model can be extended to allow the definition of random fuzzy numbers and vectors with
  - Different supports ( $[a, b]$ ,  $[a, +\infty)$ , probability simplex, etc.)
  - Different “shapes” (skewed, heavy-tailed etc.)while maintaining the closure property under the product-intersection rule.
- This can be achieved by composing a RFS  $\tilde{X} : \Omega \rightarrow [0, 1]^\Theta$  with a **one-to-one mapping** from  $\Theta$  to another space  $\Lambda$ , to obtain a a RFS  $\tilde{Y} : \Omega \rightarrow [0, 1]^\Lambda$ .
- More details in my paper “Belief Functions on the Real Line defined by Transformed Gaussian Random Fuzzy Numbers” to be presented on Tuesday, August 15, session “Fuzzy Machine Learning”, 8:00-10:00, Room #113.

# 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
- 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  $\mathbf{X} = (X_1, \dots, X_p)$ , based on a learning set  $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ .
- We have proposed a **neural network model**<sup>3</sup> (ENNreg), which quantifies uncertainty about the response  $Y$  given input vector  $\mathbf{X} = \mathbf{x}$  by a **GRFN**  $\tilde{Y}(\mathbf{x})$  with associated belief function  $Bel_{\tilde{Y}(\mathbf{x})}$ .
- 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.

<sup>3</sup>T. Denœux. Quantifying Prediction Uncertainty in Regression using Random Fuzzy Sets: the ENNreg model. *IEEE Transactions on Fuzzy Systems*, 2023. 

# 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
- 3 Application to Machine Learning
  - **Neural network model**
  - Learning
  - Experimental results

# Propagation equations (1/2)

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

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

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

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

$$\tilde{Y}_k(\mathbf{x}) \sim \tilde{N}(\mu_k(\mathbf{x}), \sigma_k^2, s_k(\mathbf{x})h_k)$$

where  $\sigma_k^2$  and  $h_k$  are variance and precision parameters, and

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

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



# Propagation equations (2/2)

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

$$\tilde{Y}(\mathbf{x}) = \tilde{Y}_1(\mathbf{x}) \boxplus \dots \boxplus \tilde{Y}_K(\mathbf{x})$$

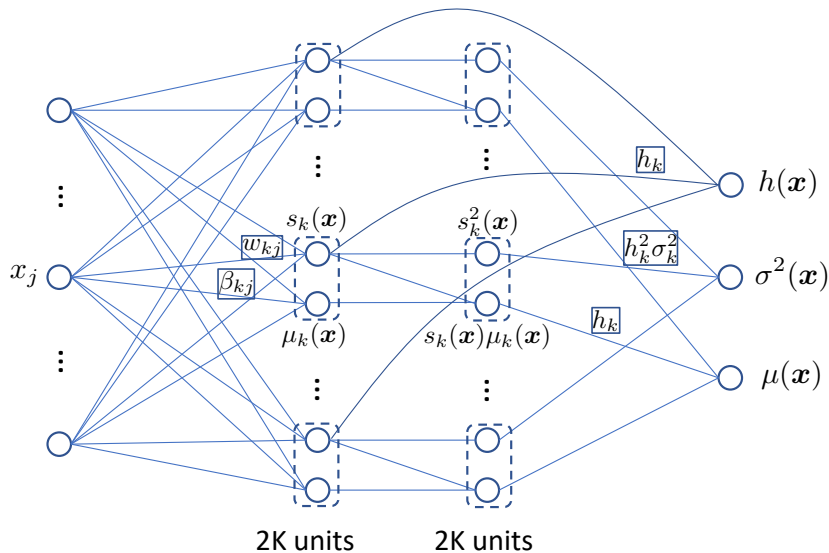
where  $\boxplus$  denotes product intersection without soft conditioning (to simplify calculations).

- We have  $\tilde{Y}(\mathbf{x}) \sim \tilde{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x}), h(\mathbf{x}))$ , with

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

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

# Neural network architecture



# 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
- 3 Application to Machine Learning
  - Neural network model
  - **Learning**
  - Experimental results

# Negative log-likelihood loss (probabilistic forecasts)

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

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

- 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

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

So,  $\mathcal{L}(y, \hat{f}) = -\ln \hat{P}([y]_\epsilon) + \text{cst.}$

- Generalization in the case of prediction in the form of a belief function?

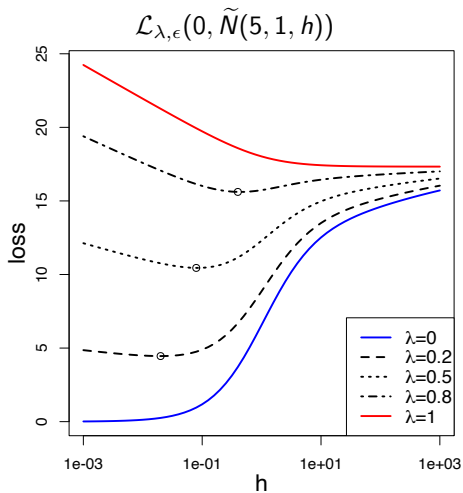
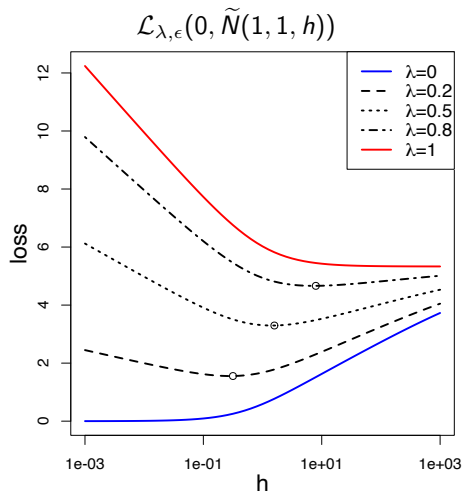
# Extension

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

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

with  $\lambda \in [0, 1]$  and  $\epsilon > 0$ .

- Smaller values of  $\lambda$  correspond to more cautious predictions.

Influence of  $\lambda$ 

# Training

- The network is trained by minimizing the **regularized average loss**

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

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:  $\lambda = 0.9$ ,  $\epsilon = 0.01\hat{\sigma}_Y$ ,  $\xi$  and  $\rho$  tuned using a validation set or cross-validation.

# Calibration

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

$$\forall \alpha \in (0, 1], \quad P_{\mathbf{X}, Y}(Y \in \mathcal{B}_\alpha(\mathbf{X})) \geq \alpha \quad (1)$$

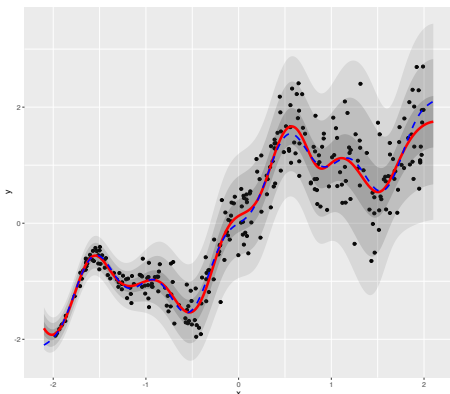
- As in the probabilistic case, the calibration of evidential predictions can be checked graphically using a **calibration plot** (see infra).
- The precision output  $h(\mathbf{x})$  can be multiplied by a constant  $c > 0$  to ensure (1) with predictions as precise as possible.



# Example

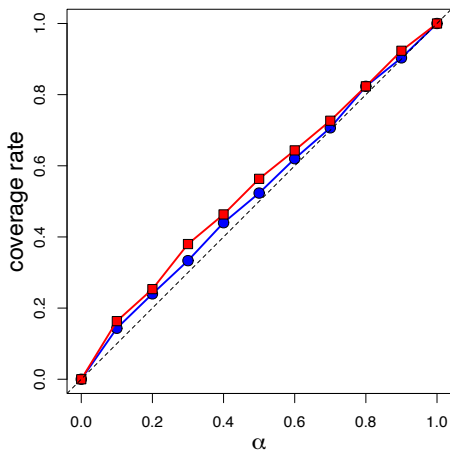
We consider iid data with one-dimensional input  $X \sim \text{Unif}(-2, 2)$  and

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



- Learning and validation sets of size  $n = 300$ .
- Network with  $K = 30$  prototypes initialized by the k-means algorithm.
- $\xi$  and  $\rho$  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

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

# 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)

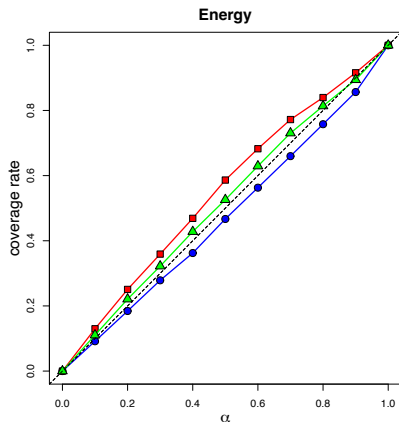
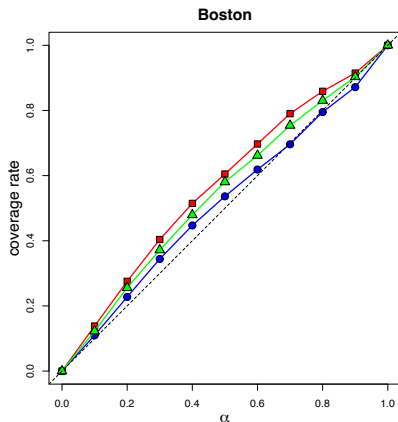
	ENNreg	RBF	RVM	SVM	GP	RF	MLP
Boston	<b>2.87</b> ± <b>0.14</b>	3.31 ± 0.19	3.42 ± 0.17	3.17 ± 0.15	3.70 ± 0.22	<b>3.11</b> ± <b>0.14</b>	<b>3.14</b> ± <b>0.14</b>
Energy	<b>1.06</b> ± <b>0.05</b>	2.06 ± 0.08	1.79 ± 0.05	1.39 ± 0.06	2.58 ± 0.07	1.75 ± 0.06	<b>0.95</b> ± <b>0.16</b>
Concr.	5.10 ± 0.12	6.30 ± 0.19	6.38 ± 0.16	5.62 ± 0.13	6.93 ± 0.13	<b>4.64</b> ± <b>0.12</b>	<b>4.82</b> ± <b>0.16</b>
Yacht	<b>0.44</b> ± <b>0.04</b>	2.00 ± 0.20	1.88 ± 0.20	1.93 ± 0.11	6.12 ± 0.31	0.96 ± 0.08	<b>0.50</b> ± <b>0.05</b>
Wine	0.63 ± 0.01	0.63 ± 0.01	0.80 ± 0.02	0.61 ± 0.01	0.61 ± 0.01	<b>0.56</b> ± <b>0.01</b>	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	<b>0.07</b> ± <b>0.00</b>
Crime	<b>0.14</b> ± <b>0.00</b>	<b>0.14</b> ± <b>0.00</b>	<b>0.14</b> ± <b>0.00</b>	<b>0.14</b> ± <b>0.00</b>	<b>0.14</b> ± <b>0.00</b>	<b>0.14</b> ± <b>0.00</b>	<b>0.14</b> ± <b>0.00</b>
Resid.	<b>0.11</b> ± <b>0.01</b>	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	<b>1.46</b> ± <b>0.03</b>	1.70 ± 0.04	2.58 ± 0.04	2.37 ± 0.04	2.49 ± 0.04	<b>1.44</b> ± <b>0.04</b>	1.53 ± 0.04
Bike	<b>6.59</b> ± <b>0.19</b>	<b>6.49</b> ± <b>0.15</b>	<b>6.64</b> ± <b>0.14</b>	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	<b>2.87 ± 0.14</b>	<b>3.01 ± 0.18</b>	<b>2.97 ± 0.19</b>	<b>3.28 ± 1.00</b>	<b>3.06 ± 0.16</b>
Energy	<b>1.06 ± 0.05</b>	1.80 ± 0.05	1.66 ± 0.04	2.09 ± 0.29	2.06 ± 0.10
Concr.	<b>5.10 ± 0.12</b>	5.67 ± 0.09	<b>5.23 ± 0.12</b>	6.03 ± 0.58	5.85 ± 0.15
Yacht	<b>0.44 ± 0.04</b>	1.02 ± 0.05	1.11 ± 0.09	1.58 ± 0.48	1.57 ± 0.56
Wine	<b>0.63 ± 0.01</b>	<b>0.64 ± 0.01</b>	<b>0.62 ± 0.01</b>	<b>0.64 ± 0.04</b>	<b>0.61 ± 0.02</b>
kin8nm	<b>0.08 ± 0.00</b>	0.10 ± 0.00	0.10 ± 0.00	0.09 ± 0.00	0.09 ± 0.00

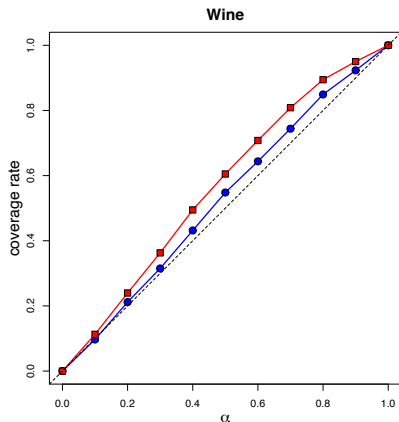
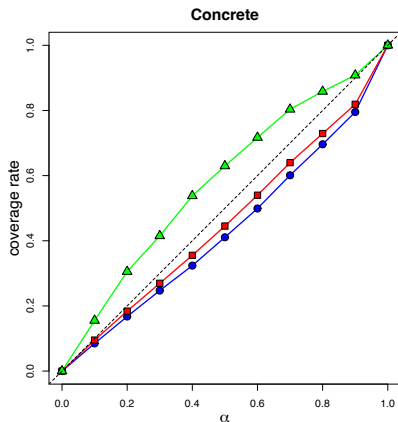
	NLL				
	ENNreg	PBP	MC-dropout	Deep ens.	Deep ev. reg.
Boston	2.53 ± 0.07	2.57 ± 0.09	<b>2.46 ± 0.06</b>	<b>2.41 ± 0.25</b>	<b>2.35 ± 0.06</b>
Energy	<b>1.14 ± 0.07</b>	2.04 ± 0.02	1.99 ± 0.02	<b>1.38 ± 0.22</b>	1.39 ± 0.06
Concr.	3.38 ± 0.13	3.16 ± 0.02	<b>3.04 ± 0.02</b>	<b>3.06 ± 0.18</b>	<b>3.01 ± 0.02</b>
Yacht	<b>0.13 ± 0.12</b>	1.63 ± 0.02	1.55 ± 0.03	1.18 ± 0.21	1.03 ± 0.19
Wine	<b>0.94 ± 0.01</b>	0.97 ± 0.01	<b>0.93 ± 0.01</b>	<b>0.94 ± 0.12</b>	<b>0.89 ± 0.05</b>
kin8nm	<b>-1.19 ± 0.00</b>	-0.90 ± 0.01	-0.95 ± 0.01	<b>-1.20 ± 0.02</b>	-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  $\mathbf{x}$  is a GRFN defined by three numbers:
  - a point prediction  $\mu(\mathbf{x})$
  - a variance  $\sigma^2(\mathbf{x})$  measuring **random** uncertainty
  - a precision  $h(\mathbf{x})$  representing **epistemic** uncertainty
- Other applications include knowledge elicitation and statistical inference<sup>4</sup>.

---

<sup>4</sup>T. Denœux. Parametric families of continuous belief functions based on generalized Gaussian random fuzzy numbers. *Fuzzy Sets and Systems* (to appear), 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* (to appear), 2023.

<https://hal.science/hal-04060251>

# 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*, 2023.

<https://doi.org/10.1109/TFUZZ.2023.3268200>



T. Denœux

evreg: Evidential Regression

R package version 1.0.2, 2023. Available:

<https://CRAN.R-project.org/package=evreg>