Random fuzzy sets and belief functions

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

SIPTA Seminar
May 24, 2023

Two models of uncertainty

- Modeling uncertainty: a fundamental problem in AI
  - Representation of uncertain/imperfect knowledge
  - Reasoning and decision-making with uncertainty
  - Quantification of prediction uncertainty in machine learning
  - Etc.
- Two of the most widely used models:
  - Dempster-Shafer (DS) theory = belief functions + Dempster's rule (based on random sets, generalizes Bayesian probability theory)
  - Possibility theory = possibility distributions + triangular norms (based on fuzzy sets)
- Claims:
  - These two theories are distinct
  - They are needed simultaneously in some applications
  - $\rightarrow$ We need to embed them into a more general framework

The case of statistical inference

- Shafer (1976) proposed to interpret the relative likelihood as defining as a consonant belief function (mathematically equivalent to a possibility distribution).
- Combining the likelihood-based belief function with a Bayesian prior by Dempster's rule yields the Bayesian posterior.
- However, relative likelihood functions from independent samples must be combined by the normalized product t-norm (possibility theory).
- To make this approach consistent, we need a more general mathematical framework encompassing both possibility and DS theories.

Contents of this talk

1. Introduction of a new model of uncertainty based on random fuzzy sets (RFSs) + a new combination rule generalizing both Dempster's rule and the normalized product intersection of possibility theory
2. Definition of practical and easily combinable RFS models allowing us to represent uncertainty on continuous variables (in $\mathbb{R}$, $[a, b]$, $\mathbb{R}^p$, probability simplex, etc.)
3. Application to machine learning: neural network model for regression quantifying prediction uncertainty by RFSs

Outline

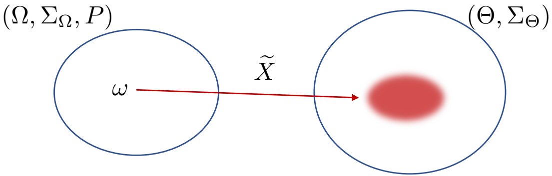
1. Random fuzzy sets
   - Basic definitions
   - Combination
2. Practical models
   - Gaussian random fuzzy numbers
   - Extensions: transformations and mixtures
3. Application to regression
   - Neural network model
   - Learning
   - Experimental results

Outline

Outline

1. Random fuzzy sets
   - Basic definitions
   - Combination
2. Practical models
   - Gaussian random fuzzy numbers
   - Extensions: transformations and mixtures
3. Application to regression
   - Neural network model
   - Learning
   - Experimental results

Random fuzzy set



A random fuzzy set (RFS) is a mapping $\widetilde{X}$ from Ω to the set $[0,1]^\Theta$ of fuzzy subsets of Θ, such that for any $\alpha \in [0,1]$, the mapping ${}^\alpha\widetilde{X}$ from Ω to 2^Θ defined as

$${}^\alpha\widetilde{X}(\omega) = {}^\alpha[\widetilde{X}(\omega)] = \{\theta \in \Theta : \widetilde{X}(\omega)(\theta) \geq \alpha\}$$

is Σ_Ω-Σ_Θ strongly measurable (i.e., $\forall B \in \Sigma_\Theta, \{\omega \in \Omega : {}^\alpha[\widetilde{X}(\omega)] \cap B \neq \emptyset\} \in \Sigma_\Omega$). (Couso & Sánchez, 2011)

Epistemic random fuzzy sets

- We use RFSs as a model of unreliable and fuzzy evidence[1]:
  - Θ is the domain of an uncertain variable/quantity $\boldsymbol{\theta}$
  - Ω is a set of interpretations of a piece of evidence about $\boldsymbol{\theta}$
  - If $\omega \in \Omega$ holds, we know that "$\boldsymbol{\theta}$ is $\widetilde{X}(\omega)$", i.e., $\boldsymbol{\theta}$ is constrained by the possibility distribution $\widetilde{X}(\omega)$.
- Example: a witness tells us that "John is tall", and this witness is 50% reliable
  - $\Omega = \{\text{rel}, \neg\text{rel}\}$, $p(\text{rel}) = 0.5$
  - $\boldsymbol{\theta}$ = John's height in meters, $\Theta = [0, 2.5]$
  - $\widetilde{X}(\text{rel})$ = tall (a fuzzy subset of Θ), $\widetilde{X}(\neg\text{rel}) = \Theta$
- This interpretation is different from previous interpretations of RFSs as
  - A model of random mechanism for generating fuzzy data (Puri & Ralescu, Gil)
  - Imperfect knowledge of a random variable (Kruse & Meyer, Couso & Sánchez)

[1] 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

Belief and plausibility functions

- If interpretation $\omega \in \Omega$ holds, the degrees of possibility and necessity that $\boldsymbol{\theta}$ 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

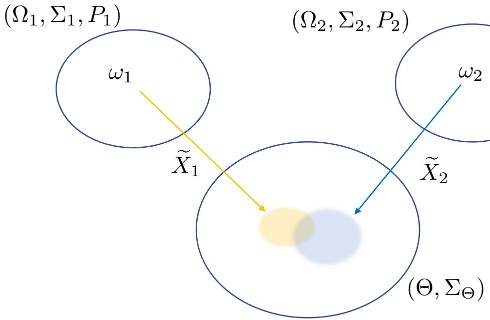
$$Bel_{\widetilde{X}}(B) = \int_\Omega N_{\widetilde{X}(\omega)}(B) dP(\omega), \quad Pl_{\widetilde{X}}(B) = \int_\Omega \Pi_{\widetilde{X}(\omega)}(B) dP(\omega).$$

- Function $Bel_{\widetilde{X}}$ is a completely monotone capacity (a belief function), and $Pl_{\widetilde{X}}$ is the dual plausibility function (Zadeh, 1979; Couso & Sánchez, 2011).
- A RFS is thus (like a random set) a way of specifying a belief function. The RFS model is more flexible.

Outline

1. Random fuzzy sets
   - Basic definitions
   - Combination
2. Practical models
   - Gaussian random fuzzy numbers
   - Extensions: transformations and mixtures
3. Application to regression
   - Neural network model
   - Learning
   - Experimental results

Combination of independent RFSs



- We consider two RFSs $\widetilde{X}_1 : \Omega_1 \to [0,1]^\Theta$ and $\widetilde{X}_2 : \Omega_2 \to [0,1]^\Theta$ representing independent pieces of evidence.
- if $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$ both hold, we can deduce "$\boldsymbol{\theta}$ 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.

Combination of independent RFSs (continued)

- Fuzzy intersection: the product t-norm is the most suitable for combining fuzzy information from independent sources. The normalized product intersection of fuzzy sets/possibility distributions (is defined as

$$(\widetilde{F} \odot \widetilde{G})(\theta) = \begin{cases} \dfrac{\widetilde{F}(\theta)\widetilde{G}(\theta)}{\sup_{\Theta}(\widetilde{F} \cdot \widetilde{G})} & \text{if } \sup_{\Theta}(\widetilde{F} \cdot \widetilde{G}) > 0 \\ 0 & \text{otherwise} \end{cases}$$

  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 probability measure on $\Omega_1 \times \Omega_2$ is then obtained by conditioning $P_1 \times P_2$ on the fuzzy event $\widetilde{\Theta}^*$. This process is called soft normalization.

Product-intersection rule

- The combined RFS

$$\begin{array}{rccl} \widetilde{X}_{12}: & \Omega_1 \times \Omega_2 & \to & [0,1]^\Theta \\ & (\omega_1, \omega_2) & \mapsto & \widetilde{X}_1(\omega_1) \odot \widetilde{X}_2(\omega_2) \end{array}$$

  associated with the probability measure $(P_1 \times P_2)(\cdot|\widetilde{\Theta}^*)$ is called the product intersection[2] of $\widetilde{X}_1$ and $\widetilde{X}_2$ (with soft normalization). We write $\widetilde{X}_{12} = \widetilde{X}_1 \oplus \widetilde{X}_2$.

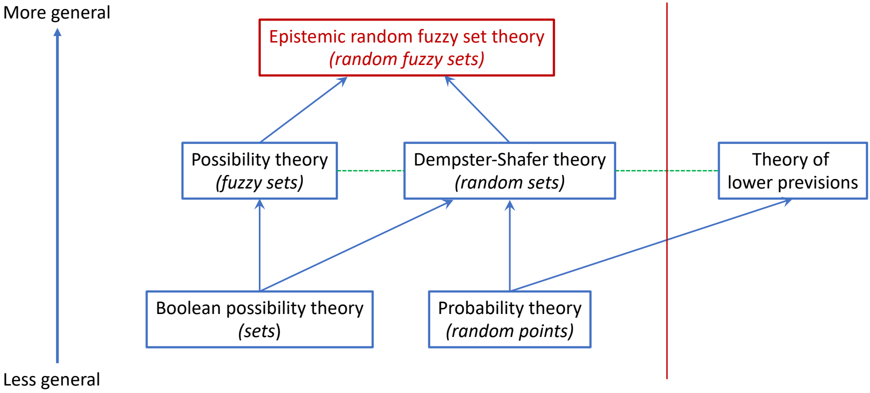
- Properties:
  1. Commutativity, associativity
  2. Generalization of Dempster's rule and the normalized product intersection of possibility distributions
  3. For all $\theta \in \Theta$,

     $$Pl_{\widetilde{X}_1 \oplus \widetilde{X}_2}(\{\theta\}) = c\, Pl_{\widetilde{X}_1}(\{\theta\}) Pl_{\widetilde{X}_2}(\{\theta\})$$

     where c does not depend on θ.

[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

General picture



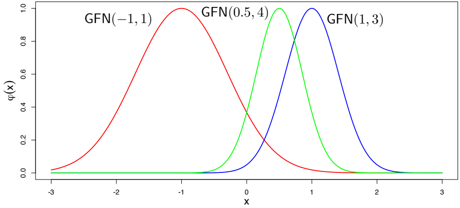
Outline

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 now, a similar workable model has 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 $\oplus$.

Outline

Gaussian fuzzy numbers



- A Gaussian fuzzy number (GFN) is a normal fuzzy subset of $\mathbb{R}$ with membership function

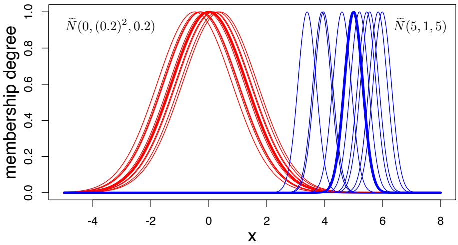
$$\varphi(x; m, h) = \exp\left(-\frac{h}{2}(x-m)^2\right),$$

where $m \in \mathbb{R}$ is the mode and $h \in [0, +\infty]$ is the precision. It is denoted by $\mathsf{GFN}(m, h)$.

- Property: $\mathsf{GFN}(m_1, h_1) \odot \mathsf{GFN}(m_2, h_2) = \mathsf{GFN}(m_{12}, h_{12})$ with

$$m_{12} = \frac{h_1 m_1 + h_2 m_2}{h_1 + h_2} \quad \text{and} \quad h_{12} = h_1 + h_2.$$

Gaussian random fuzzy numbers



- A Gaussian random fuzzy number (GRFN)[3] is a GFN whose mode is a Gaussian random variable (GRV): it can be seen as an uncertain GFN or as a fuzzy GRV.
- Formally: a GRFN with mean μ, variance σ^2 and precision h is a RFS $\widetilde{X} : \Omega \rightarrow [0,1]^{\mathbb{R}}$ defined as $\widetilde{X}(\omega) = \text{GFN}(M(\omega), h)$ where $M \sim N(\mu, \sigma^2)$. We write $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$.

[3] T. Denœux. *Fuzzy Sets and Systems* 453:1–36, 2023

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:

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

Mathematical formulas

Contour function:

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

Belief and plausibility of an interval $[x, y]$:

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

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

Combination 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

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

with

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

where

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

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

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

Gaussian random fuzzy vectors

- Multidimensional generalization of GRFNs.
- A p-dimensional Gaussian fuzzy vector (GFV) with mode $\boldsymbol{m} \in \mathbb{R}^p$ and symmetric and positive semidefinite precision matrix $\boldsymbol{H} \in \mathbb{R}^{p \times p}$ is defined as the fuzzy subset of $\mathbb{R}^p$ with membership function

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

  It is denoted as $\text{GFV}(\boldsymbol{m}, \boldsymbol{H})$.
- A Gaussian random fuzzy vector (GRFV) $\widetilde{X} \sim \widetilde{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{H})$ is random fuzzy set $\widetilde{X} : \Omega \to [0,1]^{\mathbb{R}^p}$ defined as

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

- The product intersection of two GRFVs is a GRFV.

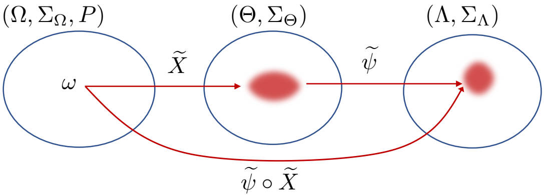
Outline

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 in two complementary ways[4]:
  1. Compose a RFS $\widetilde{X} : \Omega \to [0,1]^\Theta$ with a one-to-one mapping from Θ to another space Λ, to obtain a a RFS $\widetilde{Y} : \Omega \to [0,1]^\Lambda$
  2. Define mixtures of RFSs

[4] T. Denœux. Parametric families of continuous belief functions based on generalized Gaussian random fuzzy numbers. Preprint hal-04060251, 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; specifically, we define a mapping $\widetilde{\psi} : [0,1]^\Theta \rightarrow [0,1]^\Lambda$ such that

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

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

Properties

1. For any $C \in \Sigma_\Lambda = \{\psi(B) : B \in \Sigma_\Theta\}$,

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

and

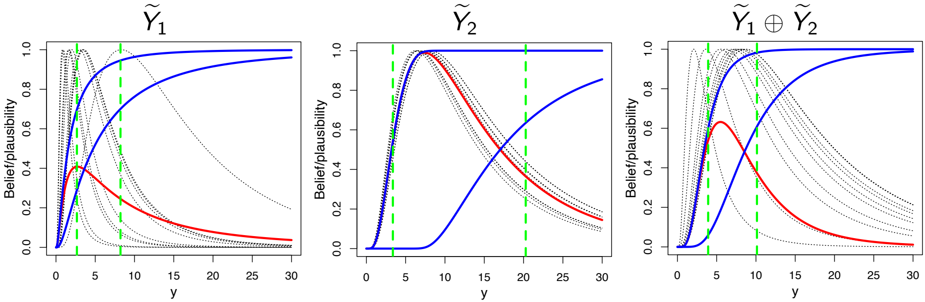
$$Pl_{\widetilde{\psi}\circ\widetilde{X}}(C) = Pl_{\widetilde{X}}(\psi^{-1}(C))$$

2. 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

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

Example: 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 $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \log)$.



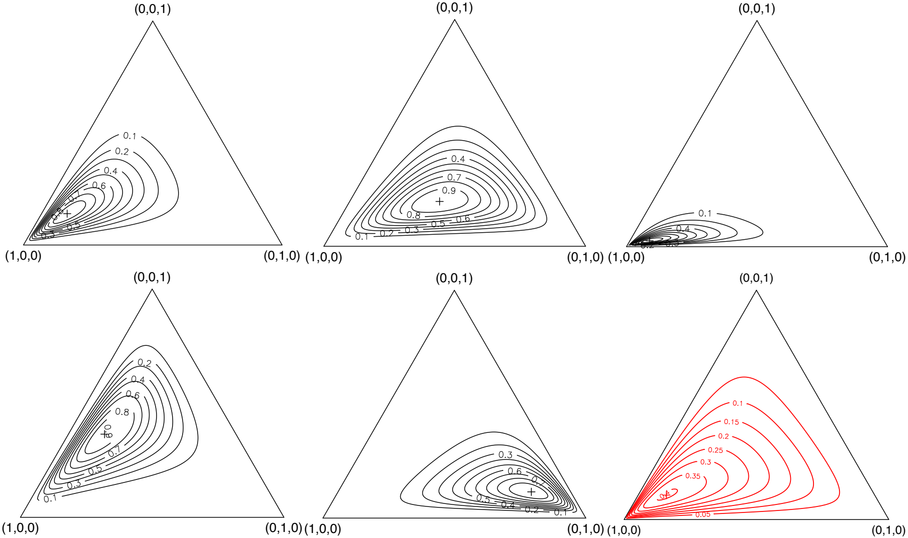
Logistic-normal RFVs

- 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:

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

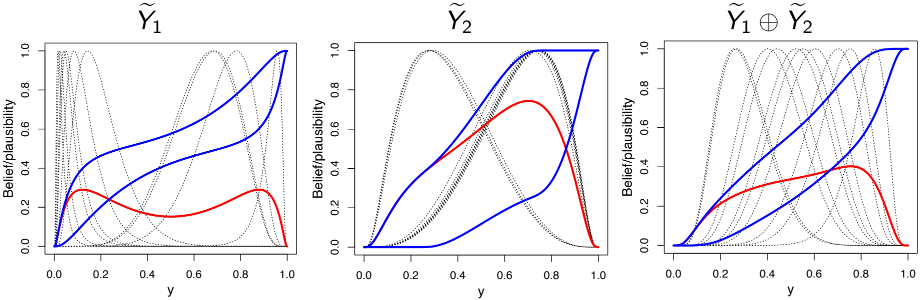
- The random fuzzy vector $\widetilde{Y} = \widetilde{\psi}_S \circ \widetilde{X}$ is a logistic-normal RFV; we write $\widetilde{Y} \sim T\widetilde{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{H}, \psi_S^{-1})$. Its support is the simplex $\mathcal{S}_p$.

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

Evidential Machine Learning

- Evidential Machine Learning (ML): an approach to ML in which uncertainty is quantified by belief functions.
- Existing methods mainly address clustering (ECM, EVCLUS, etc.) and classification (EKNN, ENN, etc.), because these learning tasks only require belief functions on finite frames.
- The availability of models for defining and combining belief functions on continuous frames now makes it 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 propose a neural network model[5] (ENNreg), which for an observed input vector $\boldsymbol{X} = \boldsymbol{x}$ computes a GRFN $\widetilde{Y}(\boldsymbol{x})$ with associated belief function $Bel_{\widetilde{Y}(\boldsymbol{x})}$ representing uncertainty about Y.
- 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

[5] T. Denœux. Quantifying Prediction Uncertainty in Regression using Random Fuzzy Sets: the ENNreg model. *IEEE Transactions on Fuzzy Systems*, 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

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

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

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

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

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

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

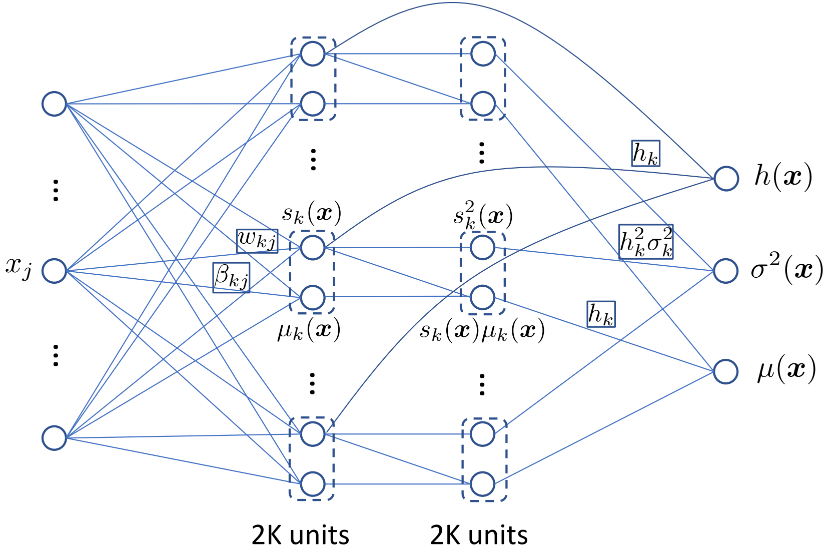
where $\boxplus$ denotes product intersection without the normalization step (to simplify calculations).

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

1. Random fuzzy sets
   - Basic definitions
   - Combination

2. Practical models
   - Gaussian random fuzzy numbers
   - Extensions: transformations and mixtures

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

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?

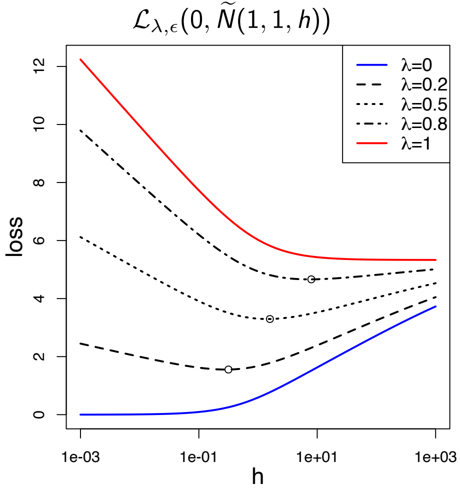
Extension

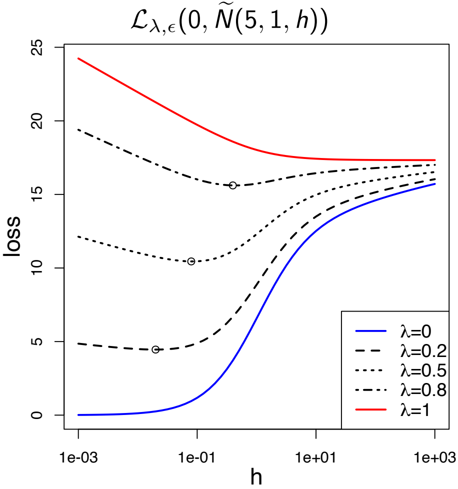
- $\mathcal{L}_\epsilon(y, \widetilde{Y}) = -\ln Bel_{\widetilde{Y}}([y]_\epsilon)$ does not work (does not reward imprecision).
- $\mathcal{L}_\epsilon(y, \widetilde{Y}) = -\ln Pl_{\widetilde{Y}}([y]_\epsilon)$ also does not work (minimized when $\widetilde{Y}$ is vacuous).
- Proposal:

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

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

Influence of λ





Training

- The network is training 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 will be 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,

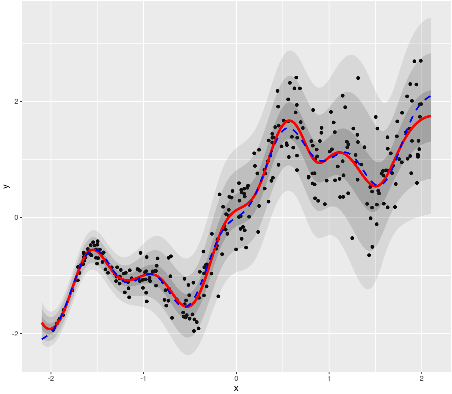
$$\boxed{\forall \alpha \in (0,1], \quad P_{\boldsymbol{X},Y}\left(Y \in \mathcal{B}_\alpha(\boldsymbol{X})\right) \geq \alpha} \tag{1}$$

- 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

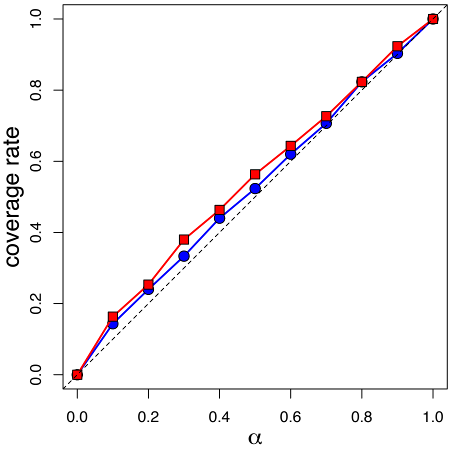
We consider iid data with one-dimensional input $X \sim \text{Unif}(-2, 2)$ and

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



- 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: 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)

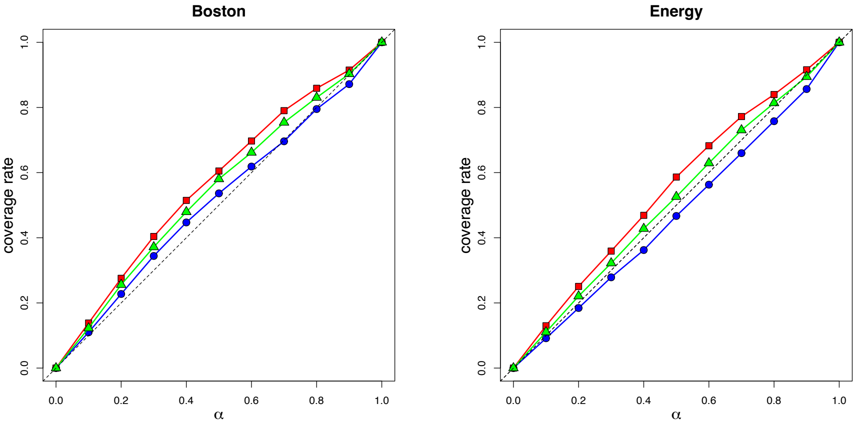
	ENNreg	RBF	RVM	SVM	GP	RF	MLP
Boston	**2.87** $\pm$ **0.14**	3.31 $\pm$ 0.19	3.42 $\pm$ 0.17	3.17 $\pm$ 0.15	3.70 $\pm$ 0.22	**3.11** $\pm$ **0.14**	**3.14** $\pm$ **0.14**
Energy	**1.06** $\pm$ **0.05**	2.06 $\pm$ 0.08	1.79 $\pm$ 0.05	1.39 $\pm$ 0.06	2.58 $\pm$ 0.07	1.75 $\pm$ 0.06	**0.95** $\pm$ **0.16**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

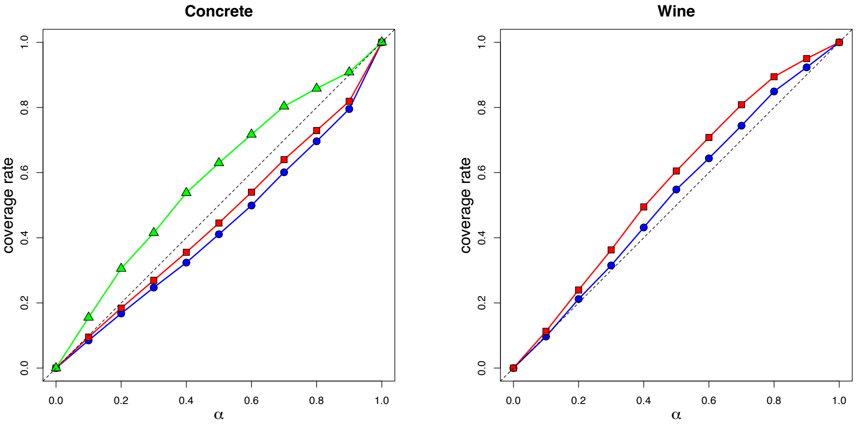
	NLL				
	ENNreg	PBP	MC-dropout	Deep ens.	Deep ev. reg.
Boston	2.53 ± 0.07	2.57 ± 0.09	**2.46** ± **0.06**	**2.41** ± **0.25**	**2.35** ± **0.06**
Energy	**1.14** ± **0.07**	2.04 ± 0.02	1.99 ± 0.02	**1.38** ± **0.22**	1.39 ± 0.06
Concr.	3.38 ± 0.13	3.16 ± 0.02	**3.04** ± **0.02**	**3.06** ± **0.18**	**3.01** ± **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 extends both possibility theory and DS theory. It allows one to represent and reason with uncertain, imprecise and vague information.
- We have defined flexible families of RFNs and RFVs indexed by 3 parameters (mode, variance and precision). They make it possible to define belief functions on continuous frames that can be easily manipulated and combined, overcoming a limitation of DS theory.
- The ENNreg model is 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
- Experimental results show that ENNreg performs as well as, or better than state-of-the-art regression methods, while providing conservative (cautious) predictions.

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.
Preprint hal-04060251, 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