Belief Functions on the Real Line defined by Transformed Gaussian Random Fuzzy Numbers

Thierry Denœux
Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, Compiègne, France
Institut Universitaire de France, Paris, France
ORCID: 0000-0002-0660-5436

***Abstract*—The recently introduced theory of epistemic random fuzzy sets extends both Dempster-Shafer and possibility theories, by allowing the representation of partially reliable and fuzzy evidence. Within this formalism, we study transformations of random fuzzy sets by one-to-one mappings, and show that such transformations commute with combination. We apply this result to define parameterized models of random fuzzy numbers, which generalize Gaussian random fuzzy numbers and allow us to construct easily combinable belief functions on a real interval. We apply this idea to the prediction of proportions.**

***Index Terms*—Random fuzzy sets, evidence theory, regression, machine learning, uncertainty quantification.**

I. Introduction

The Dempster-Shafer (DS) theory of belief functions [1]–[3] and possibility theory [4]–[6] are two powerful frameworks for representing partial information and reasoning with uncertainty. Whereas DS theory makes it possible to represent partially reliable evidence, possibility theory allows us to express uncertainty based on vague information such as conveyed by fuzzy sets. In [7], [8], we have argued that DS and possibility theories can be viewed as two specializations of a more general theory of "epistemic random fuzzy sets". A random fuzzy set (RFS) maps each element of a probability space to a fuzzy subset of a set Θ. It is, thus, a model of evidence that can be both uncertain and fuzzy. In this framework, a possibility distribution can be viewed as a constant RFS, while a random set (a notion underlying DS theory) corresponds to the special case where all images are crisp. Random fuzzy sets induced by independent pieces of evidence can be combined by the generalized product-intersection (GPI) operator, which generalizes both Dempster's rule of combination and the normalized product intersection of possibility distributions.

Whereas the theory of belief functions has been defined from the start in a very general setting [9], most applications have used only belief functions on finite spaces. This limitation was mainly due to the absence of general enough parametric families of belief functions in continuous spaces that could easily be defined and combined by Dempster's rule of combination. In [8], we have proposed Gaussian Random Fuzzy Numbers (GRFNs) as a model for defining belief functions on the real line. A GRFN can be seen either as a Gaussian possibility distribution with random mode, or as a Gaussian random variable with fuzzy mean. The family of GRFNs is closed under the GPI operation, which makes it suitable for evidential reasoning with continuous variables. An application to machine learning was presented in [10].

Practical as it may be, the GRFN model is quite restricted. 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 $[a, b]$. Furthermore, the model is symmetric about the mean μ, i.e., intervals of the form $[\mu - r, \mu]$ and $[\mu, \mu + r]$, where $r > 0$, have the same degree of belief, a property that may not always reflect an agent's actual beliefs. It is thus of interest to define other parameterized families of random fuzzy numbers with different supports, while maintaining the closure property under the GPI rule. In this paper, this is achieved by transforming GRFNs via a bijective mapping from the real line to itself, or to real intervals. More specifically, the contributions presented in this paper are the following:

1) The notion of transformation of a RFS by a one-to-one mapping is introduced, and expressions for the induced belief and plausibility functions are provided;
2) We prove that bijective transformations commute with the GPI rule, i.e., the image of the product intersection of two RFSs by a one-to-one mapping is equal to the product intersection of their images;
3) We introduce two new parametric families of random fuzzy numbers with bounded or half-bounded supports, closed under the GPI operation;
4) We demonstrate an application of this model to the prediction of proportions.

The definitions related to RFSs and GRFNs are first recalled in Section II. Transformations of RFSs are then studied in Section III, where several models based on GRFNs are introduced. Finally, the application to the prediction of proportions is discussed in Section IV, and Section V concludes the paper.

II. Random Fuzzy Sets

The RFS setting and its relation with belief functions will first be recalled in Section II-A. The GRFN model will then be introduced in Section II-B.

A. General definitions and main properties

a) Definition: Let us consider a probability space $(\Omega, \Sigma_\Omega, P)$, a measurable space (Θ, Σ_Θ), and a mapping $\widetilde{X}$ from Ω to the set $[0,1]^\Theta$ of fuzzy subsets of Θ. For any $\alpha \in [0,1]$, let ${}^\alpha\widetilde{X}$ be the mapping from Ω to 2^Θ such

that ${}^{\alpha}\widetilde{X}(\omega) = {}^{\alpha}[\widetilde{X}(\omega)]$, where ${}^{\alpha}[\widetilde{X}(\omega)]$ is the weak α-cut of $\widetilde{X}(\omega)$. If, for any $\alpha \in [0,1]$, ${}^{\alpha}\widetilde{X}$ is $\Sigma_\Omega - \Sigma_\Theta$ strongly measurable [11], the tuple $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \widetilde{X})$ is said to be a *random fuzzy set* (also called a *fuzzy random variable*) [12]. We define the *support* of $\widetilde{X}$ as the union of the supports of its images, i.e.,

$$\text{supp}(\widetilde{X}) = \bigcup_{\omega \in \Omega} \{\theta \in \Theta : \widetilde{X}(\omega)(\theta) > 0\}.$$

b) Interpretation: In epistemic random fuzzy set (ERFS) theory, RFSs are used to represent unreliable and fuzzy evidence: the set Ω is then seen as a *set of interpretations* of a piece of evidence about a variable $\boldsymbol{\theta}$ taking values in Θ. If interpretation $\omega \in \Omega$ holds, we know that "$\boldsymbol{\theta}$ is $\widetilde{X}(\omega)$", i.e., $\boldsymbol{\theta}$ is constrained by the possibility distribution defined by fuzzy set $\widetilde{X}(\omega)$. Such RFSs encode a *state of knowledge* about some variable $\boldsymbol{\theta}$, hence the adjective "epistemic". This model should not be confused with alternative interpretations of RFSs as describing a fuzzy data generation mechanism [13], [14], or as imprecise information about a "true" random variable [15], [16].

c) Belief and plausibility functions: Just as a random set, a RFS induces a belief function that can be seen as quantifying one's beliefs based on the available evidence. From now on, we will assume any RFS $\widetilde{X}$ to verify the following normalization conditions: (1) For all $\omega \in \Omega$, $\widetilde{X}(\omega)$ is either the empty set, or a normal fuzzy set, and (2) the image $\widetilde{X}(\omega)$ is almost surely nonempty, i.e., $P(\{\omega \in \Omega : \widetilde{X}(\omega) = \emptyset\}) = 0$.

For any $\omega \in \Omega$, let $\Pi_{\widetilde{X}(\omega)}$ be the possibility measure on Θ quantifying our beliefs on $\boldsymbol{\theta}$ given that interpretation ω holds; it is defined for any $B \in \Sigma_\Theta$ as

$$\Pi_{\widetilde{X}(\omega)}(B) = \sup_{\theta \in B} \widetilde{X}(\omega)(\theta). \tag{1a}$$

The dual necessity measure $N_{\widetilde{X}(\omega)}$ is

$$N_{\widetilde{X}(\omega)}(B) = \begin{cases} 1 - \Pi_{\widetilde{X}(\omega)}(B^c) & \text{if } \widetilde{X}(\omega) \neq \emptyset \\ 0 & \text{otherwise,} \end{cases} \tag{1b}$$

where B^c denotes the complement of B. For any $B \in \Sigma_\Theta$, let $Bel_{\widetilde{X}}(B)$ and $Pl_{\widetilde{X}}(B)$ denote, respectively, the *expected necessity* and the *expected possibility* of B:

$$Bel_{\widetilde{X}}(B) = \int_\Omega N_{\widetilde{X}(\omega)}(B) dP(\omega), \tag{2a}$$

$$Pl_{\widetilde{X}}(B) = \int_\Omega \Pi_{\widetilde{X}(\omega)}(B) dP(\omega) = 1 - Bel_{\widetilde{X}}(B^c). \tag{2b}$$

The mappings $B \mapsto Bel_{\widetilde{X}}(B)$ and $B \mapsto Pl_{\widetilde{X}}(B)$, are, respectively, belief and plausibility functions [12], [17].

d) Combination: The combination of independent pieces of evidence by Dempster's rule [2] is a key component of DS theory. In possibility theory, conjunctive combination operators are based on t-norms [18]. In ERFS theory, the GPI rule introduced in [7], [8] extends these operators to the general case where evidence is represented by RFSs.

Let $(\Omega_i, \Sigma_i, P_i, \Theta, \Sigma_\Theta, \widetilde{X}_i)$, $i = 1, 2$, be two RFSs encoding independent pieces of evidence. The independence assumption means here that the relevant probability measure on the joint measurable space $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2)$ is the product measure $P_1 \times P_2$. If interpretations $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$ both hold, $\boldsymbol{\theta}$ is constrained by both $\widetilde{X}_1(\omega_1)$ and $\widetilde{X}_2(\omega_2)$. It is then natural to combine these two fuzzy sets by an intersection operator. As argued in [7], [18], the product t-norm is the most suitable for combining fuzzy information from independent sources. Furthermore, the normalized product intersection operation is associative.

Conflict needs to be handled at two levels. First, the product-intersection of fuzzy sets $\widetilde{X}_1(\omega_1)$ and $\widetilde{X}_2(\omega_2)$ has to be normalized to obtain a normal fuzzy set, or the empty set in case of extreme conflict. Denoting by $\odot$ the normalized product intersection, we thus consider the mapping $\widetilde{X}_\odot(\omega_1, \omega_2) = \widetilde{X}_1(\omega_1) \odot \widetilde{X}_2(\omega_2)$, which will be assumed to be $\Sigma_1 \otimes \Sigma_2$-$\Sigma_\Theta$ strongly measurable. Secondly, the product probability measure $P_1 \times P_2$ needs to be conditioned to eliminate pairs of fully inconsistent interpretations $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ such that $\text{hgt}(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2)) = 0$ (where $\text{hgt}(\cdot)$ denotes the height of a fuzzy set), but also to downweigh pairs of partially inconsistent pairs such that $0 < \text{hgt}(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2)) < 1$. This is achieved by *soft normalization* proposed in [7], [8], which consists in conditioning the product probability $P_1 \times P_2$ by the fuzzy subset $\widetilde{\Theta}^*$ of consistent pairs of interpretations, with membership function

$$\widetilde{\Theta}^*(\omega_1, \omega_2) = \text{hgt}\left(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2)\right). \tag{3}$$

The product-intersection operator with soft normalization is denoted by $\oplus$, and the combined RFS $\widetilde{X}_1 \oplus \widetilde{X}_2$ is called the *product intersection*, or the *orthogonal sum* of $\widetilde{X}_1$ and $\widetilde{X}_2$. The operator $\oplus$ is commutative and associative; it generalizes both Dempster's rule and the normalized product intersection of possibility distributions.

B. Gaussian Random Fuzzy Numbers

The important role played by the Gaussian distribution in probability theory and statistics is partly due to the fact that it is amenable to easy calculation. Until recently, such a practical model was missing in DS theory, which hindered its application to uncertain reasoning with real variables. The GRFN model fills this gap by blending Gaussian possibility distributions and Gaussian random variables.

Let us start by recalling the definition of a *Gaussian Fuzzy Number* (GFN) as a fuzzy subset of $\mathbb{R}$ with membership function

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

where $m \in \mathbb{R}$ is the *mode* and $h \in [0, +\infty]$ is the *precision*. Such a fuzzy number will be denoted by $\text{GFN}(m, h)$. GFNs are easily combined by the normalized product-intersection operator, as the following property holds: $\text{GFN}(m_1, h_1) \odot \text{GFN}(m_2, h_2) = \text{GFN}(m_{12}, h_{12})$, with $m_{12} = (h_1 m_1 + h_2 m_2)/(h_1 + h_2)$ and $h_{12} = h_1 + h_2$.

Let us now consider a probability space $(\Omega, \Sigma_\Omega, P)$ and a Gaussian random variable (GRV) $M : \Omega \to \mathbb{R}$ with mean

μ and variance σ^2. The random fuzzy set $\widetilde{X} : \Omega \to [0,1]^{\mathbb{R}}$ defined as

$$\widetilde{X}(\omega) = \text{GFN}(M(\omega), h)$$

is called a *Gaussian random fuzzy number* (GRFN) with mean μ, variance σ^2 and precision h, which we write $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$. A GRFN can, thus, be seen as a GFN whose mode is uncertain and described by a Gaussian probability distribution. It is defined by a location parameter μ, and two parameters h and σ^2 corresponding, respectively, to possibilistic and probabilistic uncertainty. In the special case where the precision is infinite, $\widetilde{X}$ becomes equivalent to a GRV with mean μ and variance σ^2, which we can write: $\widetilde{N}(\mu, \sigma^2, +\infty) = N(\mu, \sigma^2)$. If $\sigma^2 = 0$, M is constant and $\widetilde{X}$ is equivalent to possibility distribution GFN(μ, h), i.e., $\widetilde{N}(\mu, 0, h) = \text{GFN}(\mu, h)$. Finally, when $h = 0$, we have $\widetilde{X}(\omega)(x) = 1$ for all $\omega \in \Omega$ and all $x \in \mathbb{R}$: such a RFS represents total ignorance and the corresponding belief function is said to be *vacuous*.

Formulas to compute the plausibility and belief degrees of any real interval $[x, y]$ induced by a GRFN $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$ are given in [8]. In particular, the contour function of $\widetilde{X}$ is given by

$$pl_{\widetilde{X}}(x) = \frac{1}{\sqrt{1 + h\sigma^2}} \exp\left(-\frac{h(x-\mu)^2}{2(1+h\sigma^2)}\right). \tag{4}$$

The lower and upper cumulative distribution functions (cdfs) defined, respectively, as the mappings $x \mapsto Bel(-\infty, x]$ and $x \mapsto Pl(-\infty, x]$ have the following expressions:

$$Bel_{\widetilde{X}}((-\infty, x]) = \Phi\left(\frac{x-\mu}{\sigma}\right) - pl_{\widetilde{X}}(x)\Phi\left(\frac{x-\mu}{\sigma\sqrt{h\sigma^2+1}}\right), \tag{5a}$$

$$Pl_{\widetilde{X}}((-\infty, x]) = Bel_{\widetilde{X}}((-\infty, x]) + pl_{\widetilde{X}}(x). \tag{5b}$$

Most importantly, as shown in [8], the family of GRFNs is closed under the GPI combination operation $\oplus$: 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 $\widetilde{X}_1 \oplus \widetilde{X}_2 \sim \widetilde{N}(\widetilde{\mu}_{12}, \widetilde{\sigma}_{12}^2, h_{12})$, with $h_{12} = h_1 + h_2$,

$$\widetilde{\mu}_{12} = \frac{h_1\widetilde{\mu}_1 + h_2\widetilde{\mu}_2}{h_1 + h_2}, \quad \widetilde{\sigma}_{12}^2 = \frac{h_1^2\widetilde{\sigma}_1^2 + h_2^2\widetilde{\sigma}_2^2 + 2\rho h_1 h_2 \widetilde{\sigma}_1\widetilde{\sigma}_2}{(h_1+h_2)^2}, \tag{6a}$$

where

$$\widetilde{\mu}_1 = \frac{\mu_1(1 + \overline{h}\sigma_2^2) + \mu_2\overline{h}\sigma_1^2}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}, \tag{6b}$$

$$\widetilde{\mu}_2 = \frac{\mu_2(1 + \overline{h}\sigma_1^2) + \mu_1\overline{h}\sigma_2^2}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}, \tag{6c}$$

$$\widetilde{\sigma}_1^2 = \frac{\sigma_1^2(1 + \overline{h}\sigma_2^2)}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}, \quad \widetilde{\sigma}_2^2 = \frac{\sigma_2^2(1 + \overline{h}\sigma_1^2)}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}, \tag{6d}$$

$$\rho = \frac{\overline{h}\sigma_1\sigma_2}{\sqrt{(1 + \overline{h}\sigma_1^2)(1 + \overline{h}\sigma_2^2)}}, \tag{6e}$$

and $\overline{h} = h_1 h_2/(h_1 + h_2)$.

III. Transformation of a Random Fuzzy Set

As mentioned in Section I, the GRFN model is very convenient for uncertain reasoning with real variables due to its closure property with respect to the $\oplus$ operator, but it also has several limitations. In particular, its support is the whole real line, making it unsuitable to represent evidence about variables taking values in a strict subset of $\mathbb{R}$. In this section, we overcome this limitation by considering bijective transformations of RFSs. The main result, stated in Section III-A, is that the image of the orthogonal sum of two RFSs under a bijective mapping is the orthogonal sum of the images. Two useful transformations of GRFNs are studied in Section III-B.

A. Definitions and main result

a) Definition: Let $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \widetilde{X})$ be a RFS, and $\psi : \Theta \to \Lambda$ be a one-to-one mapping from Θ to some set Λ. Zadeh's extension principle [19] allows us to extend mapping ψ to fuzzy subsets of Θ; specifically, we can define a mapping $\widetilde{\psi} : [0,1]^\Theta \to [0,1]^\Lambda$ as follows:

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

Let Σ_Λ be the set containing the images of all elements of Σ_Θ by ψ:

$$\Sigma_\Lambda = \{\psi(B) : B \in \Sigma_\Theta\},$$

and consider the mapping $\widetilde{\psi} \circ \widetilde{X} : \Omega \to [0,1]^\Lambda$ such that $(\widetilde{\psi} \circ \widetilde{X})(\omega) = \widetilde{\psi}(\widetilde{X}(\omega))$ for all $\omega \in \Omega$. It is easy to show that Σ_Λ is a σ-algebra on Λ, and that $(\Omega, \Sigma_\Omega, P, \Lambda, \Sigma_\Lambda, \widetilde{\psi} \circ \widetilde{X})$ is a RFS; we say that $\widetilde{\psi} \circ \widetilde{X}$ is the result of the *transformation* of $\widetilde{X}$ by ψ.

b) Belief and plausibility: From (1), for all $C \in \Sigma_\Lambda$,

$$\begin{aligned}\Pi_{(\widetilde{\psi}\circ\widetilde{X})(\omega)}(C) &= \sup_{\lambda \in C} (\widetilde{\psi} \circ \widetilde{X})(\omega)(\lambda) \\ &= \sup_{\lambda \in C} \widetilde{X}(\omega)(\psi^{-1}(\lambda)) \\ &= \sup_{\theta \in \psi^{-1}(C)} \widetilde{X}(\omega)(\theta) = \Pi_{\widetilde{X}(\omega)}(\psi^{-1}(C)),\end{aligned}$$

and, similarly,

$$N_{(\widetilde{\psi}\circ\widetilde{X})(\omega)}(C) = N_{\widetilde{X}(\omega)}(\psi^{-1}(C)).$$

Consequently, from (2),

$$Bel_{\widetilde{\psi}\circ\widetilde{X}}(C) = Bel_{\widetilde{X}}(\psi^{-1}(C)), \tag{7a}$$

and

$$Pl_{\widetilde{\psi}\circ\widetilde{X}}(C) = Pl_{\widetilde{X}}(\psi^{-1}(C)). \tag{7b}$$

Eq. (7) provides simple expressions for the belief and plausibility functions associated to a transformed RFS.

c) Combination: Let now consider the combination of two transformed RFSs $\psi \circ \widetilde{X}_1$ and $\psi \circ \widetilde{X}_2$ with the same transformation ψ. The following lemma states that the image of the product intersection of two fuzzy subsets of Θ is equal to the product intersection of their images, and the *degree of conflict* (defined as the height of the product intersection before normalization) of the fuzzy subsets equals that of their images.

Lemma 1. *Let $\widetilde{F}$ and $\widetilde{G}$ be two fuzzy subsets of Θ. We have*

$$\widetilde{\psi}(\widetilde{F} \odot \widetilde{G}) = \widetilde{\psi}(\widetilde{F}) \odot \widetilde{\psi}(\widetilde{G})$$

and

$$hgt(\widetilde{\psi}(\widetilde{F}) \cdot \widetilde{\psi}(\widetilde{G})) = hgt(\widetilde{F} \cdot \widetilde{G}).$$

Proof. For any $\lambda \in \Lambda$,

$$\begin{aligned}\widetilde{\psi}(\widetilde{F} \odot \widetilde{G})(\lambda) &= (\widetilde{F} \odot \widetilde{G})[\psi^{-1}(\lambda)] \\ &= \frac{\widetilde{F}[\psi^{-1}(\lambda)]\widetilde{G}[\psi^{-1}(\lambda)]}{\sup_{\lambda'} \widetilde{F}[\psi^{-1}(\lambda')]\widetilde{G}[\psi^{-1}(\lambda')]} \\ &= \frac{\widetilde{\psi}(\widetilde{F})(\lambda)\widetilde{\psi}(\widetilde{G})(\lambda)}{\sup_{\lambda'} \widetilde{\psi}(\widetilde{F})(\lambda')\widetilde{\psi}(\widetilde{G})(\lambda')} \\ &= (\widetilde{\psi}(\widetilde{F}) \odot \widetilde{\psi}(\widetilde{G}))(\lambda).\end{aligned} \tag{8}$$

Now, the degree of conflict between $\widetilde{\psi}(\widetilde{F})$ and $\widetilde{\psi}(\widetilde{G})$ is the denominator on the right-hand side of (8). It is equal to

$$\sup_{\lambda} \widetilde{F}[\psi^{-1}(\lambda)]\widetilde{G}[\psi^{-1}(\lambda)] = \sup_{\theta} \widetilde{F}(\theta)\widetilde{G}(\theta).$$

□

We can now state the main result of this section.

Theorem 1. *Let $(\Omega_i, \Sigma_i, P_i, \Theta, \Sigma_\Theta, \widetilde{X}_i)$, $i = 1, 2$, be two RFSs representing independent evidence. We have*

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

Proof. As recalled in Section II-A, the orthogonal sum of $\widetilde{\psi} \circ \widetilde{X}_1$ and $\widetilde{\psi} \circ \widetilde{X}_2$ is defined by mapping

$$(\omega_1, \omega_2) \mapsto (\widetilde{\psi} \circ \widetilde{X}_1)(\omega_1) \odot (\widetilde{\psi} \circ \widetilde{X}_2)(\omega_2),$$

and the joint probability measure $P_1 \times P_2$ conditioned by the fuzzy subset of $\Omega_1 \times \Omega_2$ with membership function

$$\Theta^*(\omega_1, \omega_2) = \text{hgt}\left((\widetilde{\psi} \circ \widetilde{X}_1)(\omega_1) \cdot (\widetilde{\psi} \circ \widetilde{X}_2)(\omega_2)\right).$$

Now, from Lemma 1,

$$(\widetilde{\psi} \circ \widetilde{X}_1)(\omega_1) \odot (\widetilde{\psi} \circ \widetilde{X}_2)(\omega_2) = \widetilde{\psi}\left[\widetilde{X}_1(\omega_1) \odot \widetilde{X}_2(\omega_2)\right]$$

and

$$\text{hgt}((\widetilde{\psi} \circ \widetilde{X}_1)(\omega_1) \cdot (\widetilde{\psi} \circ \widetilde{X}_2)(\omega_2)) = \text{hgt}(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2)),$$

from which the result directly follows. □

B. Transformed Gaussian Random Fuzzy Numbers

Applying the idea developed in Section III-A to GRFNs makes it possible to define a wide variety of parametric families of random fuzzy numbers and associated belief functions on the real line. Let $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$ be a GRFN, and ψ a one-to-one mapping from $\mathbb{R}$ to $\Lambda \subseteq \mathbb{R}$. Let $\widetilde{\psi} \circ \widetilde{X}$ the result of the transformation of $\widetilde{X}$ by ψ. We will say that $\widetilde{\psi} \circ \widetilde{X}$ is a transformed GRFN and we will write $\widetilde{\psi} \circ \widetilde{X} \sim T\widetilde{N}(\mu, \sigma^2, h, \psi^{-1})$. For any random fuzzy number $\widetilde{Y}$, it is clear that

$$\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \psi^{-1}) \Leftrightarrow \widetilde{\psi}^{-1} \circ \widetilde{Y} \sim \widetilde{N}(\mu, \sigma^2, h). \tag{9}$$

From Theorem 1, given two transformed GRFNs $\widetilde{Y}_1 \sim T\widetilde{N}(\mu_1, \sigma_1^2, h_1, \psi^{-1})$ and $\widetilde{Y}_2 \sim T\widetilde{N}(\mu_2, \sigma_2^2, h_2, \psi^{-1})$, we have $Y_1 \oplus Y_2 \sim T\widetilde{N}(\widetilde{\mu}_{12}, \widetilde{\sigma}_{21}^2, h_{12}, \psi^{-1})$, where $\widetilde{\mu}_{12}$, $\widetilde{\sigma}_{21}^2$ and h_{12} are given by (6).

Hereafter, we will consider two cases for the choice of function ψ allowing us to define belief functions on the positive real line and on a closed real interval.

a) Lognormal random fuzzy numbers: Using a one-to-one mapping from $\mathbb{R}$ to $(0, +\infty)$ allows us to define a random fuzzy number with support equal to the positive real line. Choosing $\psi = \exp$, we obtain a *log-normal random fuzzy number RFN)* $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \log)$. From (9), $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \log)$ if and only if $\widetilde{\log}(\widetilde{Y}) \sim \widetilde{N}(\mu, \sigma^2, h)$. A log-normal random variable is recovered when $h = +\infty$. From (4) and (7b), the contour function of $\widetilde{Y}$ is

$$pl_{\widetilde{Y}}(y) = \frac{1}{\sqrt{1 + h\sigma^2}} \exp\left(-\frac{h(\log y - \mu)^2}{2(1 + h\sigma^2)}\right).$$

Similarly, the lower and upper cdfs of $\widetilde{Y}$ can easily be computed from (5) and (7).

Example 1. *Figure 1 displays two log-normal RFNs $\widetilde{Y}_1 \sim T\widetilde{N}(1, 1, 5, \log)$ and $\widetilde{Y}_2 \sim T\widetilde{N}(2, 0.1, 2, \log)$ as well as their orthogonal sum $\widetilde{Y}_1 \oplus \widetilde{Y}_2$. For each RFN, we plot ten realizations, the contour functions, as well as the lower and upper cdfs.*

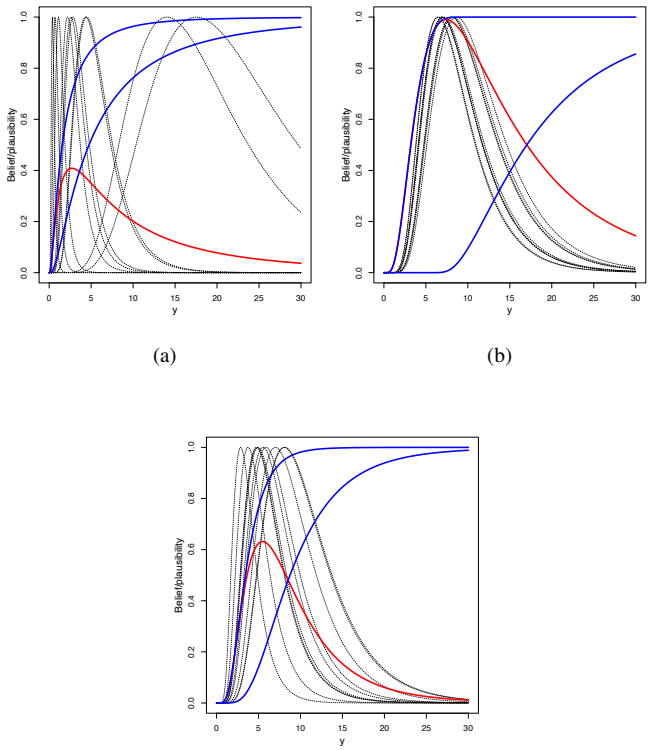
b) Logit-normal random fuzzy numbers: Any cdf F can be used to define a RFN with support equal to interval $[0, 1]$ (or more generally, using an additional affine transformation, an interval $[a, b]$ with $b > a$). A natural choice is the cdf of the standard logistic distribution, $F_L(x) = [1 + \exp(-x)]^{-1}$. The corresponding quantile function is the logit function,

$$F_L^{-1}(y) = \text{logit}(y) = \log \frac{y}{1 - y}.$$

A RFN $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma, h, \text{logit})$ will said to be *logit-normal*. A logit-normal random variable [20] is recovered when $h = +\infty$. From (4) and (7b), the contour function of $\widetilde{Y}$ is

$$pl_{\widetilde{Y}}(y) = \frac{1}{\sqrt{1 + h\sigma^2}} \exp\left(-\frac{h(\text{logit}(y) - \mu)^2}{2(1 + h\sigma^2)}\right).$$

The lower and upper cdfs of $\widetilde{Y}$ can be computed from (5) and (7) in a similar manner.



(a) (b)

(c)

Fig. 1: (a) and (b): Two log-normal random fuzzy numbers $\widetilde{Y}_1 \sim T\widetilde{N}(1, 1, 5, \log)$ and $\widetilde{Y}_2 \sim T\widetilde{N}(2, 0.1, 2, \log)$; (c): combined log-normal random fuzzy number $\widetilde{Y}_1 \oplus \widetilde{Y}_2$. For each RFN, we plot ten realizations (black dotted curves), the contour functions (red curve), as well as the lower and upper cdfs (blue curves).

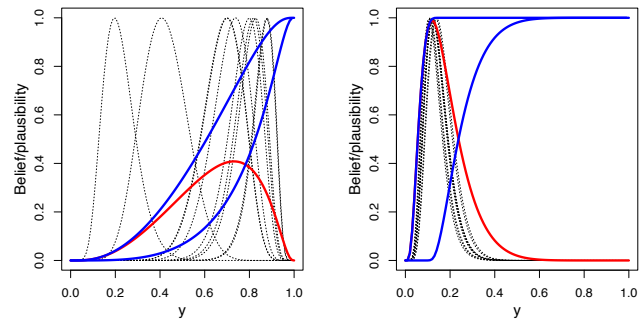
Example 2. *Figure 3 shows 10 realizations of a logit-normal RFN* $\widetilde{Y} \sim T\widetilde{N}(1, 1, 5, logit)$*, together with its contour function as well as its lower and upper cdfs.*

IV. Application to Prediction of Proportions

In [10], we have introduced the ENNreg model, a regression neural network that quantifies prediction uncertainty by GRFNs. This model is appropriate when the response variable takes values in the whole real line. The notion of transformed GRFN introduced in this paper makes it possible to apply ENNreg to learning problems in which the response is positive, or takes values in a closed interval. We will first briefly recall the ENNreg model in Section IV-A. Its application to the prediction of proportions, with uncertainty quantified by logit-normal RFNs will then be discussed in Section IV-B.

A. The ENNreg model

The Evidential Neural Network for regression (ENNreg) model solves regression tasks by comparing an input vector to prototypes. Each prototype is considered as a source of



(a) (b)

Fig. 2: Representation of logit-normal random fuzzy numbers $\widetilde{Y} \sim T\widetilde{N}(1, 1, 5, \text{logit})$ (a) and $\widetilde{Y} \sim T\widetilde{N}(-2, 0.1, 2, \text{logit})$ (b): ten realizations (black dotted curves), contour function (red curve), lower and upper cdfs (blue curves).

information, which provides evidence represented by a GRFN. The total evidence is pooled by a simplified version of the GPI rule recalled in Section II-A.

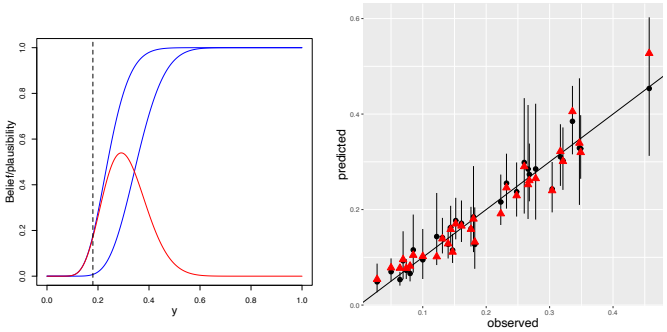
More precisely, consider a regression task in which a response variable Y taking values in the real line is to be predicted from a p-dimensional input vector $\boldsymbol{x}$. Let $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K$ denote K p-dimensional prototypes in input space. The similarity between input vector $\boldsymbol{x}$ and prototype $\boldsymbol{w}_k$ is measured by

$$s_k(\boldsymbol{x}) = \exp(-\gamma_k^2 \|\boldsymbol{x} - \boldsymbol{w}_k\|^2), \tag{10}$$

where $\gamma_k > 0$ is a scale parameter. The evidence of 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 for prototype k; the mean $\mu_k(\boldsymbol{x})$ is defined as $\mu_k(\boldsymbol{x}) = \boldsymbol{\beta}_k^T \boldsymbol{x} + \beta_{k0}$, where $\boldsymbol{\beta}_k$ is a p-dimensional vector of coefficients, and β_{k0} is a scalar parameter. The output $\widetilde{Y}(\boldsymbol{x})$ for input $\boldsymbol{x}$ is computed by combining the GRFNs $\widetilde{Y}_k(\boldsymbol{x})$, $k = 1, \ldots, K$ induced by the K prototypes. To simplify the computations, hard normalization is used instead of the soft normalization described in Section II-A. The network output is a GRFN $\widetilde{Y}(\boldsymbol{x}) \sim \widetilde{N}(\mu(\boldsymbol{x}), \sigma^2(\boldsymbol{x}), h(\boldsymbol{x}))$, with

$$\mu(\boldsymbol{x}) = \frac{\sum_{k=1}^K s_k(\boldsymbol{x}) h_k \mu_k(\boldsymbol{x})}{\sum_{k=1}^K s_k(\boldsymbol{x}) h_k}, \sigma^2(\boldsymbol{x}) = \frac{\sum_{k=1}^K s_k^2(\boldsymbol{x}) h_k^2 \sigma_k^2}{\left(\sum_{k=1}^K s_k(\boldsymbol{x}) h_k\right)^2},$$

and $h(\boldsymbol{x}) = \sum_{k=1}^K s_k(\boldsymbol{x}) h_k$. This GRFN quantifies prediction uncertainty: $\mu(\boldsymbol{x})$ is the most plausible value of Y given $\boldsymbol{x}$, while $\sigma^2(\boldsymbol{x})$ and $h(\boldsymbol{x})$ represent, respectively, aleatory and epistemic uncertainty. A description of the loss function and the training process can be found in [10]. This model is implemented in R package `evreg` [21].



(a) (b)

Fig. 3: Prediction of crude oil proportion for the Gasoline Yield dataset. (a): Contour function (red curve) and lower/upper cdfs (blue curves) for a predictive logit-normal RFN associated to a particular test sample; the true proportion is shown as a broken vertical line; (b): predicted vs. observed values; black dots with error bars are the ENNreg predictions with 80% belief intervals; red triangles are the predictions using Beta regression.

B. Prediction of proportions

Let us now assume that Y is a proportion, taking values in the interval $[0, 1]$. Let $Z = \text{logit}(Y)$ be the transformed proportion. Training the ENNreg model with Z as the response variable allows us to compute a GRFN $\widetilde{Z}(\boldsymbol{x}) \sim \widetilde{N}(\mu(\boldsymbol{x}), \sigma^2(\boldsymbol{x}), h(\boldsymbol{x}))$ quantifying the uncertainty on Z given $\boldsymbol{x}$. The uncertainty on Y is then described by the transformed GRFN $\widetilde{Y} \sim T\widetilde{N}(\mu(\boldsymbol{x}), \sigma^2(\boldsymbol{x}), h(\boldsymbol{x}), \text{logit})$.

Example 3. *As an example, let us consider the Gasoline Yield dataset available in R package* `betareg` *[22]. In this dataset, the response is the proportion of crude oil after distillation and fractionation. The dataset contains 32 observations of the response and of the four input variables: crude oil gravity, vapor pressure of crude oil, temperature at which 10% of crude oil has vaporized, and temperature at which all gasoline has vaporized. (The dataset has an additional input factor, which we did not use in this experiment). We trained a network with $K = 5$ prototypes. We used nested cross-validation to tune the ξ and ρ hyperparameters (see [10]) and to estimate the prediction performance. An example of a predictive logit-normal RFN is shown in Figure 3a, and a scatterplot of predicted vs. observed values with 80% belief intervals (see [10]) is displayed in Figure 3b. The root mean squared errors for ENNreg and Beta regression [22] are, respectively, 0.0241 and 0.0285.*

V. Conclusions

The GRFN model introduced in [8] makes it possible to define belief functions on the real line that can be easily combined by the GRP rule, a combination operator generalizing both Dempster's rule of DS theory and the normalized product intersection of possibility distributions. The notion of transformed RFS introduced in this paper allows us to extend the GRFN model by defining easily combinable belief functions on a bounded or half-bounded real interval. We have presented an application to the prediction of a proportion in machine learning. Even more flexible models can be defined using a parameterized family of transformations. Applications to belief elicitation and generalized Bayesian inference, as well as further applications to machine learning will be reported in future publications.

References

[1] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.

[2] G. Shafer, *A mathematical theory of evidence*. Princeton, N.J.: Princeton University Press, 1976.

[3] T. Denœux, D. Dubois, and H. Prade, "Representations of uncertainty in artificial intelligence: Beyond probability and possibility," in *A Guided Tour of Artificial Intelligence Research*, P. Marquis, O. Papini, and H. Prade, Eds. Springer Verlag, 2020, vol. 1, ch. 4, pp. 119–150.

[4] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems*, vol. 1, pp. 3–28, 1978.

[5] D. Dubois and H. Prade, *Possibility Theory: An approach to computerized processing of uncertainty*. New-York: Plenum Press, 1988.

[6] T. Denœux, D. Dubois, and H. Prade, "Representations of uncertainty in artificial intelligence: Probability and possibility," in *A Guided Tour of Artificial Intelligence Research*, P. Marquis, O. Papini, and H. Prade, Eds. Springer Verlag, 2020, vol. 1, ch. 3, pp. 69–117.

[7] T. Denœux, "Belief functions induced by random fuzzy sets: A general framework for representing uncertain and fuzzy evidence," *Fuzzy Sets and Systems*, vol. 424, pp. 63–91, 2021.

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

[9] G. Shafer, "Allocations of probability," *Annals of Probability*, vol. 7, no. 5, pp. 827–839, 1979.

[10] T. Denœux, "Quantifying Prediction Uncertainty in Regression using Random Fuzzy Sets: the ENNreg model," *IEEE Transactions on Fuzzy Systems*, 2023, doi:10.1109/TFUZZ.2023.3268200.

[11] H. T. Nguyen, "On random sets and belief functions," *Journal of Mathematical Analysis and Applications*, vol. 65, pp. 531–542, 1978.

[12] I. Couso and L. Sánchez, "Upper and lower probabilities induced by a fuzzy random variable," *Fuzzy Sets and Systems*, vol. 165, no. 1, pp. 1–23, 2011.

[13] M. L. Puri and D. A. Ralescu, "Fuzzy random variables," *Journal of Mathematical Analysis and Applications*, vol. 114, no. 2, pp. 409–422, 1986.

[14] M. A. Gil, M. López-Díaz, and D. A. Ralescu, "Overview on the development of fuzzy random variables," *Fuzzy Sets and Systems*, vol. 157, no. 19, pp. 2546–2557, 2006.

[15] R. Kruse and K. D. Meyer, *Statistics with vague data*. Dordrecht: Kluwer, 1987.

[16] I. Couso and L. Sánchez, "Higher order models for fuzzy random variables," *Fuzzy Sets and Systems*, vol. 159, no. 3, pp. 237–258, 2008.

[17] L. A. Zadeh, "Fuzzy sets and information granularity," in *Advances in Fuzzy Sets Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds. Amsterdam: North-Holland, 1979, pp. 3–18.

[18] D. Dubois, H. Prade, and R. Yager, "Merging fuzzy information," in *Fuzzy sets in approximate reasoning and information systems*, J. C. Bezdek, D. Dubois, and H. Prade, Eds. Boston: Kluwer Academic Publishers, 1999, pp. 335–401.

[19] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning –I," *Information Sciences*, vol. 8, pp. 199–249, 1975.

[20] J. Aitchison and S. M. Shen, "Logistic-normal distributions: Some properties and uses," *Biometrika*, vol. 67, no. 2, pp. 261–272, 1980.

[21] T. Denœux, *evreg: Evidential Regression*, 2023, R package version 1.0.1. [Online]. Available: https://CRAN.R-project.org/package=evreg

[22] F. Cribari-Neto and A. Zeileis, "Beta regression in R," *Journal of Statistical Software*, vol. 34, no. 2, pp. 1–24, 2010.