Parametric families of continuous belief functions based on generalized Gaussian random fuzzy numbers

Thierry Denœux[a,b]

[a] *Université de technologie de Compiègne, CNRS*
UMR 7253 Heudiasyc, Compiègne, France
[b] *Institut universitaire de France, Paris, France*

Abstract

The theory of epistemic random fuzzy sets is a general theory of uncertainty encompassing both possibility theory and the Dempster-Shafer theory of belief functions as special cases. Within this framework, Gaussian random fuzzy numbers have recently been introduced as a practical model of uncertainty about real variables. However, the limited flexibility of this model does not allow it to represent all kinds of beliefs encountered in applications. In this paper, it is extended in two ways. First, we study one-to-one transformations of random fuzzy sets and show that such transformations commute with combination. This property allows us to define parametric families of easily combinable random fuzzy numbers and vectors on different frames based on the Gaussian model. We then go one step further by studying mixtures of random fuzzy variables, which provide a very flexible model making it possible to construct belief functions on continuous frames with arbitrary complexity. To demonstrate the applicability and practical interest of these models, two applications are studied: the elicitation of expert beliefs about numerical quantities, and generalized Bayesian inference with weak prior information represented by random fuzzy numbers.

Keywords: Belief functions, evidence theory, possibility theory, random fuzzy sets, uncertainty, statistical inference, elicitation.

1. Introduction

The Dempster-Shafer (DS) theory of belief functions [5, 26, 12] and possibility theory [30, 15, 13] 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 associated with vague information such as conveyed by fuzzy sets. In [8, 14], 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), also called "random fuzzy variable", maps each element of a probability space to a fuzzy subset of a set Θ. It is, thus, a model of evidence that can

Email address: `Thierry.Denoeux@utc.fr` (Thierry Denœux)

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 rule, 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 [27], 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 [9], we have proposed Gaussian Random Fuzzy Numbers (GRFNs) as a model for defining belief functions on the real line. A GRFN is a "doubly Gaussian model": it can be seen either as a Gaussian possibility distribution whose mode is a Gaussian random variable, or as a Gaussian probability distribution whose mean is a Gaussian fuzzy set. We have also proposed Gaussian Random Fuzzy Vectors (GRFVs) as a multidimensional extension of GRFNs, which makes it possible to construct belief functions in $\mathbb{R}^p$ for $p \geq 1$. The families of GRFNs and GRFVs are closed under the product-intersection operation, which makes them suitable for evidential reasoning with continuous variables. An application to machine learning was presented in [9, 11].

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, a GRFN is unimodal (the contour function has a unique maximum) and symmetric about the mean μ, i.e., intervals of the form $[\mu - r, \mu]$ and $[\mu, \mu + r]$, for any $r > 0$, have the same degree of belief; these properties may not always reflect an agent's actual beliefs. The GRFV model is also inadequate for representing, e.g., beliefs on probabilities or proportions, for which the domain of interest is the probability simplex. It is thus of interest to define 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.

In this paper[1] the objective stated above is achieved in two ways. We first study *bijective transformations* of RFSs and show that such transformations commute with combination, i.e., applying the transformation before or after the combination yields the same result. This property is exploited to define easily combinable random fuzzy numbers and vectors based on GRFNs and GRFVs. We then go one step further by studying *mixtures* of RFSs, which provide a very flexible model making it possible to construct belief functions on $\mathbb{R}^p$ with virtually unlimited complexity. Finally, we combine the two ideas and propose *mixtures of transformed GRFNs and GRFVs* as a very general model of RFSs (and associated belief functions) easily combinable using the product-intersection rule.

To demonstrate the applicability and practical interest of the flexible models introduced in this paper, two applications will be discussed. The first one concerns the *elicitation*

[1]This paper is an extended version of the short paper [10] to be presented at the 2023 IEEE International Conference on Fuzzy Systems, Songdo Incheon, Korea, August 13-17, 2023.

of expert beliefs about a numerical quantity of interest. Without getting into the delicate methodological problems posed by real elicitation experiments, we will demonstrate how the parameters of a random fuzzy number can be fitted to a small number of expert plausibility statements. The second application is related to *generalized Bayesian inference*. As discussed in [8, 14], the relative likelihood function can be seen as possibility distribution about the parameter of interest. This sample information can be combined with prior knowledge expressed as a RFS, which constitutes a form of "weak prior", ranging continuously from precise knowledge represented by a probability distribution, to complete ignorance. In many cases, the likelihood function (sometimes after transforming the parameter) is well approximated by a Gaussian possibility distribution and can easily be combined with a GRFN or a mixture thereof to compute a posterior RFS.

The rest of this paper is organized as follows. The main definitions related to RFSs and GRFNs are first recalled in Section 2. Transformations of RFSs and mixture models are studied, respectively, in Section 3 and 4. The applications are then discussed in Section 5. Finally, Section 6 concludes the paper.

2. Random fuzzy sets

To make the paper self-contained, the RFS setting and its relation with belief functions will first be briefly reviewed in Section 2.1. The GRFN and GRFV models will then be recalled, respectively, in Sections 2.2 and 2.3.

2.1. General definitions and results

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 (see Figure 1). 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 ${}^\alpha[\widetilde{X}(\omega)] = \{\theta \in \Theta : \widetilde{X}(\omega)(\theta) \geq \alpha\}$ 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 [24], 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*) [4]. The images $\widetilde{X}(\omega)$ of elements of Ω by $\widetilde{X}$ are called the *(fuzzy) focal sets* of $\widetilde{X}$. We define the *support* of $\widetilde{X}$ as the union of the supports of its focal sets, i.e.,

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

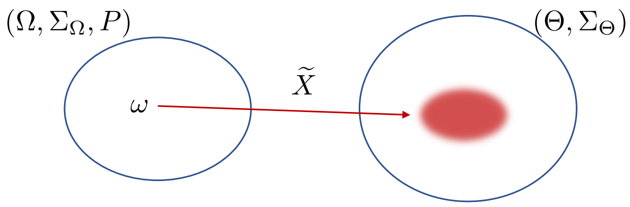


Figure 1: Definition of a random fuzzy set.

If Θ is equal to $\mathbb{R}$ or a real interval, and if the images $\widetilde{X}(\omega)$ are fuzzy numbers (i.e., normal and convex fuzzy subsets of $\mathbb{R}$), $\widetilde{X}$ is said to be a *random fuzzy number* (RFN). For any $\alpha \in [0,1]$, the mapping ${}^{\alpha}\widetilde{X}$ is, then, a random interval.

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 $\widetilde{X}(\omega)$, which represents the *meaning* of the evidence if interpretation ω holds. We do not know which interpretation is true, but we hold beliefs about the true interpretation, expressed by the probability measure P on Ω. Such a RFS represents a piece of evidence with uncertain meaning or, equivalently, 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 [25, 18], or as imprecise information about a "true" random variable [21, 3].

As a toy example illustrating the notion of unreliable and fuzzy evidence, assume that $\boldsymbol{\theta}$ represents John's height in centimeters, with domain $\Theta = [0, 200]$. We receive a testimony telling us that "John is tall", the linguistic term "tall" being represented by a fuzzy subset $\widetilde{T}$ of Θ. Let us denote by $\Omega = \{R, \neg R\}$ the set of interpretations of the evidence, where R stands for "reliable" and $\neg R$ for "not reliable", and assume that we have 80% confidence that this testimony is reliable, i.e., $P(R) = 0.8$ and $P(\neg R) = 0.8$. If the testimony is reliable, John's height is constrained by fuzzy set $\widetilde{T}$; if it is not, we know nothing about John's height; the mapping $\widetilde{X}$ is, thus, defined by $\widetilde{X}(R) = \widetilde{T}$ and $\widetilde{X}(\neg R) = \Theta$.

It is clear that this model can encode different kinds of uncertainty: probabilistic uncertainty represented by P, and imprecision and/or fuzziness represented by mapping $\widetilde{X}$. If each image $\widetilde{X}(\omega)$ is a singleton, only probabilistic uncertainty is present. Pure imprecision or fuzziness corresponds to the case where $\widetilde{X}$ is a constant mapping, i.e., there exists some crisp or fuzzy subset $\widetilde{F}$ of Θ such that for all $\omega \in \Omega$, $\widetilde{X}(\omega) = \widetilde{F}$. If $\widetilde{F}$ is a crisp subset, no fuzziness is present.

Belief and plausibility functions. Just as a random set, a RFS induces a belief function, which 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 $\omega \in \Omega$, the height of $\widetilde{X}(\omega)$, defined as $\mathsf{hgt}(\widetilde{X}(\omega)) = \sup_{\theta \in \Theta} \widetilde{X}(\omega)(\theta)$ is either 0 or 1, i.e., $\widetilde{X}(\omega)$ is either the empty set, or a normal fuzzy set;
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 [31, 4].

Lower and upper expectations of a RFN. Let $\widetilde{X}$ be a RFN, and let F_* and F^* be its lower and upper cumulative distribution functions (cdfs) defined, respectively, as

$$F_*(x) = Bel_{\widetilde{X}}((-\infty, x]) \quad \text{and} \quad F^*(x) = Pl_{\widetilde{X}}((-\infty, x])$$

for all $x \in \mathbb{R}$. We define the *lower and upper expectations* of $\widetilde{X}$ (see [5]) as, respectively, the integrals

$$\mathbb{E}_*(\widetilde{X}) = \int x dF^*(x) \quad \text{and} \quad \mathbb{E}^*(\widetilde{X}) = \int x dF_*(x).$$

It can be shown [4, 14] that these integrals can be computed as the means of the lower and upper expectations of the α-cuts of $\widetilde{X}$, i.e.,

$$\mathbb{E}_*(\widetilde{X}) = \int_0^1 \mathbb{E}_*({}^\alpha\widetilde{X}) d\alpha \quad \text{and} \quad \mathbb{E}^*(\widetilde{X}) = \int_0^1 \mathbb{E}^*({}^\alpha\widetilde{X}) d\alpha.$$

If, for any $\omega \in \Omega$ and any $\alpha \in (0, 1]$, the α-cut ${}^\alpha\widetilde{X}(\omega)$ is an interval with lower bound ${}^\alpha\widetilde{X}^-(\omega)$ and upper bound ${}^\alpha\widetilde{X}^+(\omega)$, the lower and upper expectations of ${}^\alpha\widetilde{X}$ are, respectively, the expectations of these lower and upper bounds. The lower and upper expectations of $\widetilde{X}$ can then be computed by averaging these expectations with respect to α, i.e.,

$$\mathbb{E}_*(\widetilde{X}) = \int_0^1 \mathbb{E}({}^\alpha\widetilde{X}^-) d\alpha \quad \text{and} \quad \mathbb{E}^*(\widetilde{X}) = \int_0^1 \mathbb{E}({}^\alpha\widetilde{X}^+) d\alpha. \tag{3}$$

Combination of RFSs. The combination of independent pieces of evidence by Dempster's rule [26] is a key component of DS theory. In possibility theory, conjunctive combination operators are based on t-norms [16]. In ERFS theory, the product-intersection rule introduced in [8, 14] extends these operators to the general case where evidence is represented by RFSs.

Let us assume that we have a piece of evidence about $\boldsymbol{\theta}$ represented by RFS

$$(\Omega_1, \Sigma_1, P_1, \Theta, \Sigma_\Theta, \widetilde{X}_1),$$

and we receive a second one represented by RFS $(\Omega_2, \Sigma_2, P_2, \Theta, \Sigma_\Theta, \widetilde{X}_2)$. We consider the joint measurable space $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2)$, where $\Sigma_1 \otimes \Sigma_2$ is the tensor product of σ-algebras Σ_1 and Σ_2, and we denote by P the probability measure representing our beliefs about the interpretations of both items of evidence, *before considering the conflict between mappings* $\widetilde{X}_1$ *and* $\widetilde{X}_2$. Naturally, we have $P(A \times \Omega_2) = P_1(A)$ for all $A \in \Sigma_1$ and $P(\Omega_1 \times B) = P_2(B)$ for all $B \in \Sigma_2$. The two pieces of evidence are said to be *independent*[2] iff the following two conditions hold:

1. The meaning of one piece of evidence depends only on its own set of interpretations. That is to say, if we denote by $\widetilde{X}_{1\cdot}(\omega_1, \omega_2)$ the meaning of the first piece of evidence given that the pair of interpretations $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ holds, we have

$$\forall (\omega_2, \omega_2') \in \Omega_2^2, \quad \widetilde{X}_{1\cdot}(\omega_1, \omega_2) = \widetilde{X}_{1\cdot}(\omega_1, \omega_2')$$

and, symmetrically, denoting by $\widetilde{X}_{\cdot 2}(\omega_1, \omega_2)$ the meaning of the second piece of evidence given $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$,

$$\forall (\omega_1, \omega_1') \in \Omega_1^2, \quad \widetilde{X}_{\cdot 2}(\omega_1, \omega_2) = \widetilde{X}_{\cdot 2}(\omega_1', \omega_2).$$

2. For any $A \in \Sigma_1$ and $B \in \Sigma_2$, our degree of belief that the true interpretation of the first evidence belongs to A and the true interpretation of the second evidence belongs to B is $P(A \times B) = P_1(A)P_2(B)$, i.e., P is the product measure $P_1 \times P_2$.

Under these two assumptions, we construct a RFS representing the aggregation of the two pieces of evidence by proceeding in two steps.

Step 1: We define a new mapping from the product space $\Omega_1 \times \Omega_2$ to $[0,1]^\Theta$ that maps each pair of interpretations (ω_1, ω_2) to a fuzzy intersection of $\widetilde{X}_1(\omega_1)$ and $\widetilde{X}_2(\omega_2)$. As argued in [16, 8], the normalized product intersection $\odot$ is the most suitable for combining fuzzy information from independent sources. Furthermore, this operation is associative. We thus define the following mapping,

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

[2] The notion of independence defined here concerns items of evidence, and not RFSs. A formal definition of independence for RFSs could be useful, e.g., in relation with the interpretation of RFSs as models of mechanisms for generating fuzzy data; it is outside the scope of this paper.

with

$$\left(\widetilde{X}_1(\omega_1) \odot \widetilde{X}_2(\omega_2)\right)(\theta) = \begin{cases} \dfrac{\widetilde{X}_1(\omega_1)(\theta) \cdot \widetilde{X}_2(\omega_2)(\theta)}{\mathsf{hgt}(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2))} & \text{if } \mathsf{hgt}\left(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2)\right) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

We assume that each α-cut ${}^\alpha\widetilde{X}_\odot$ is $\Sigma_1 \otimes \Sigma_2$-$\Sigma_\Theta$ strongly measurable.

Step 2: We then need to take into account the conflict between mappings $\widetilde{X}_1$ and $\widetilde{X}_2$. Pair of interpretations (ω_1, ω_2) such that $\mathsf{hgt}(\widetilde{X}_1(\omega_1)\widetilde{X}_2(\omega_2)) = 0$ are obviously inconsistent and they must be eliminated. However, in the fuzzy setting, we also need to consider *partially inconsistent* pairs of interpretation such that $0 < \mathsf{hgt}(\widetilde{X}_1(\omega_1)\widetilde{X}_2(\omega_2)) < 1$. This can be achieved by *soft normalization* proposed in [8, 14], which consists in conditioning the product probability measure $P_1 \times P_2$ by the fuzzy subset $\widetilde{\Theta}^*$ of consistent pairs of interpretations, with membership function

$$\widetilde{\Theta}^*(\omega_1, \omega_2) = \mathsf{hgt}\left(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2)\right). \tag{5}$$

The resulting conditional probability measure $\widetilde{P}_{12} = (P_1 \times P_2)(\cdot \mid \widetilde{\Theta}^*)$ has the following expression, for any $C \in \Sigma_1 \otimes \Sigma_2$:

$$\begin{aligned} \widetilde{P}_{12}(C) &= \frac{(P_1 \times P_2)(C \cap \widetilde{\Theta}^*)}{(P_1 \times P_2)(\widetilde{\Theta}^*)} \\ &= \frac{\int_{\Omega_1}\int_{\Omega_2} C(\omega_1, \omega_2)\mathsf{hgt}\left(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2)\right) dP_2(\omega_2) dP_1(\omega_1)}{\int_{\Omega_1}\int_{\Omega_2} \mathsf{hgt}\left(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2)\right) dP_2(\omega_2) dP_1(\omega_1)}, \end{aligned} \tag{6}$$

where $C(\cdot, \cdot)$ denotes the indicator function of C.

The combined RFS, denoted by $\widetilde{X}_1 \oplus \widetilde{X}_2$ and called the *orthogonal sum* of $\widetilde{X}_1$ and $\widetilde{X}_2$ is, thus, formally defined by the following tuple:

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

The operator $\oplus$ is commutative and associative; it generalizes both Dempster's rule and the normalized product intersection of possibility distributions.

2.2. Gaussian random fuzzy sets

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.

Gaussian fuzzy number. 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 $\mathsf{GFN}(m, h)$. GFNs are easily combined by the normalized product intersection operator (4), as the following property holds: $\mathsf{GFN}(m_1, h_1) \odot \mathsf{GFN}(m_2, h_2) = \mathsf{GFN}(m_{12}, h_1+h_2)$, with $m_{12} = (h_1 m_1 + h_2 m_2)/(h_1 + h_2)$. Furthermore, the height of $\mathsf{GFN}(m_1, h_1) \cdot \mathsf{GFN}(m_2, h_2)$ is

$$\mathsf{hgt}\left[\mathsf{GFN}(m_1, h_1) \cdot \mathsf{GFN}(m_2, h_2)\right] = \exp\left(-\frac{h_1 h_2 (m_1 - m_2)^2}{2(h_1 + h_2)}\right). \tag{7}$$

Gaussian random fuzzy number. 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

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

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 $\mathsf{GFN}(\mu, h)$, i.e., $\widetilde{N}(\mu, 0, h) = \mathsf{GFN}(\mu, h)$. 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.*

Belief and plausibility. 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 [14]. In particular, the contour function of $\widetilde{X}$ is given by

$$pl_{\widetilde{X}}(x) = \frac{1}{\sqrt{1 + h\sigma^2}} \exp\left(-\frac{h(x-\mu)^2}{2(1 + h\sigma^2)}\right). \tag{8}$$

The lower and upper cdfs have the following expressions:

$$Bel_{\widetilde{X}}((-\infty, x]) = \Phi\left(\frac{x-\mu}{\sigma}\right) - pl_{\widetilde{X}}(x)\Phi\left(\frac{x-\mu}{\sigma\sqrt{h\sigma^2+1}}\right), \tag{9a}$$

where Φ is the standard normal cdf, and

$$Pl_{\widetilde{X}}((-\infty, x]) = Bel_{\widetilde{X}}((-\infty, x]) + pl_{\widetilde{X}}(x). \tag{9b}$$

Lower and upper expectations. Let $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$ be a GRFN with $h > 0$. As shown in [14], its lower and upper expectations are, respectively,

$$\mathbb{E}_*(\widetilde{X}) = \mu - \sqrt{\frac{\pi}{2h}} \quad \text{and} \quad \mathbb{E}^*(\widetilde{X}) = \mu + \sqrt{\frac{\pi}{2h}}. \tag{10}$$

Combination of GRFNs. Most importantly, as shown in [14], the family of GRFNs is closed under the product-intersection rule: 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

$$\widetilde{\mu}_{12} = \frac{h_1\widetilde{\mu}_1 + h_2\widetilde{\mu}_2}{h_1 + h_2}, \quad \widetilde{\sigma}_{12}^2 = \frac{h_1^2\widetilde{\sigma}_1^2 + h_2^2\widetilde{\sigma}_2^2 + 2\rho h_1 h_2 \widetilde{\sigma}_1\widetilde{\sigma}_2}{(h_1 + h_2)^2}, \tag{11a}$$

where

$$\widetilde{\mu}_1 = \frac{\mu_1(1 + \overline{h}\sigma_2^2) + \mu_2\overline{h}\sigma_1^2}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}, \tag{11b}$$

$$\widetilde{\mu}_2 = \frac{\mu_2(1 + \overline{h}\sigma_1^2) + \mu_1\overline{h}\sigma_2^2}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}, \tag{11c}$$

$$\widetilde{\sigma}_1^2 = \frac{\sigma_1^2(1 + \overline{h}\sigma_2^2)}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}, \quad \widetilde{\sigma}_2^2 = \frac{\sigma_2^2(1 + \overline{h}\sigma_1^2)}{1 + \overline{h}(\sigma_1^2 + \sigma_2^2)}, \tag{11d}$$

$$\rho = \frac{\overline{h}\sigma_1\sigma_2}{\sqrt{(1 + \overline{h}\sigma_1^2)(1 + \overline{h}\sigma_2^2)}}, \tag{11e}$$

and $\overline{h} = h_1h_2/(h_1 + h_2)$. The degree of conflict between $\widetilde{X}_1$ and $\widetilde{X}_2$ is given by the following proposition.

Proposition 1. *The degree of conflict between $\widetilde{X}_1 \sim \widetilde{N}(\mu_1, \sigma_1^2, h_1)$ and $\widetilde{X}_2 \sim \widetilde{N}(\mu_2, \sigma_2^2, h_2)$ is*

$$\kappa = \begin{cases} 1 - \frac{\widetilde{\sigma}_1\widetilde{\sigma}_2}{\sigma_1\sigma_2}\sqrt{1-\rho^2}\exp\left\{-\frac{1}{2}\left[\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2}\right] + \frac{1}{2(1-\rho^2)}\left[\frac{\widetilde{\mu}_1^2}{\widetilde{\sigma}_1^2} + \frac{\widetilde{\mu}_2^2}{\widetilde{\sigma}_2^2} - 2\rho\frac{\widetilde{\mu}_1\widetilde{\mu}_2}{\widetilde{\sigma}_1\widetilde{\sigma}_2}\right]\right\} & \textit{if } \sigma_1, \sigma_2 > 0 \\ 1 - \frac{1}{\sqrt{1+\overline{h}\widetilde{\sigma}_1^2}}\exp\left(-\frac{\overline{h}}{2(1+\overline{h}\widetilde{\sigma}_1^2)}(\widetilde{\mu}_1 - \mu_2)^2\right) & \textit{if } \sigma_1 \geq 0, \sigma_2 = 0, \end{cases}$$

where $\widetilde{\mu}_1$, $\widetilde{\mu}_2$, $\widetilde{\sigma}_1^2$, $\widetilde{\sigma}_2^2$, ρ are given by (11).

Proof. The formula for the case $\sigma_1 > 0, \sigma_2 > 0$ is proved in [14]. The case $\sigma_2 = 0$ can be treated by replacing $\widetilde{\mu}_1$, $\widetilde{\mu}_2$, $\widetilde{\sigma}_1^2$, $\widetilde{\sigma}_2^2$, by their expressions given by (11), and letting σ_2 tend to zero. □

2.3. Gaussian random fuzzy vectors

Gaussian random fuzzy vectors (GRFVs) are multidimensional extensions of GRFNs: they are defined as Gaussian fuzzy vectors (GFVs), whose modes are random vectors with a multidimensional Gaussian distribution. We start by giving the definition of GFVs, after which we recall that of GRFVs as well as some properties.

Gaussian fuzzy vectors. A p-dimensional GFV with mode $\boldsymbol{m} \in \mathbb{R}^p$ and $p \times p$ symmetric and positive semidefinite precision matrix $\boldsymbol{H}$ is defined as the normalized 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 $\mathsf{GFV}(\boldsymbol{m}, \boldsymbol{H})$. The normalized product of two GFV's is still a GFV; more precisely, the following equality holds: $\mathsf{GFV}(\boldsymbol{m}_1, \boldsymbol{H}_1) \odot \mathsf{GFV}(\boldsymbol{m}_2, \boldsymbol{H}_2) = \mathsf{GFV}(\boldsymbol{m}_{12}, \boldsymbol{H}_{12})$, with $\boldsymbol{m}_{12} = (\boldsymbol{H}_1 + \boldsymbol{H}_2)^{-1}(\boldsymbol{H}_1\boldsymbol{m}_1 + \boldsymbol{H}_2\boldsymbol{m}_2)$ and $\boldsymbol{H}_{12} = \boldsymbol{H}_1 + \boldsymbol{H}_2$.

Gaussian random fuzzy vectors. Let $(\Omega, \Sigma_\Omega, P)$ be a probability space, $\boldsymbol{M} : \Omega \to \mathbb{R}^p$ a p-dimensional Gaussian random vector with mean $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{H}$ a $p \times p$ symmetric and positive semidefinite real matrix. The random fuzzy set $\widetilde{X} : \Omega \to [0,1]^{\mathbb{R}^p}$ defined as $\widetilde{X}(\omega) = \mathsf{GFV}(\boldsymbol{M}(\omega), \boldsymbol{H})$ is called a *Gaussian random fuzzy vector (GRFV)*, which we denote as $\widetilde{X} \sim \widetilde{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{H})$.

As shown in [14], the contour function of a GRFV is given by the following equation, which generalizes (8):

$$pl_{\widetilde{X}}(\boldsymbol{x}) = \frac{1}{|\boldsymbol{I}_p + \boldsymbol{\Sigma}\boldsymbol{H}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{H}^{-1} + \boldsymbol{\Sigma})^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \tag{12}$$

where $\boldsymbol{I}_p$ is the p-dimensional identity matrix.

The orthogonal sum of two GRFVs is still a GRFV. Formulas for the mean, variance matrix and precision matrix of the combined GRFV were derived in [14]. For completeness, they are recalled in Appendix A.

3. Transformations of Gaussian random fuzzy variables

As mentioned in Section 1, the GRFN and GRFV models are very convenient for uncertain reasoning with real variables due to their closure property with respect to the $\oplus$ operator, but they also have several limitations. In particular, the support of a GRFN is the whole real line, making it unsuitable for representing 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 3.1, is that the image of the orthogonal sum of two RFSs by a bijective mapping is the orthogonal sum of the images. Some useful transformations of GRFNs are studied in Section 3.2, and a particular transformation of GRFVs is considered in Section 3.3.

3.1. Transformation of a random fuzzy set

Let $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \widetilde{X})$ be a RFS, and $\psi : \Theta \to \Lambda$ a one-to-one mapping from Θ to some set Λ. Zadeh's extension principle [29] allows us to extend mapping ψ to fuzzy subsets of Θ; specifically, we can define a mapping $\widetilde{\psi} : [0,1]^\Theta \to [0,1]^\Lambda$ such that

$$\forall \widetilde{F} \in [0,1]^\Theta, \quad \widetilde{\psi}(\widetilde{F})(\lambda) = \sup_{\lambda=\psi(\theta)} \widetilde{F}(\theta) = \widetilde{F}(\psi^{-1}(\lambda)). \tag{13}$$

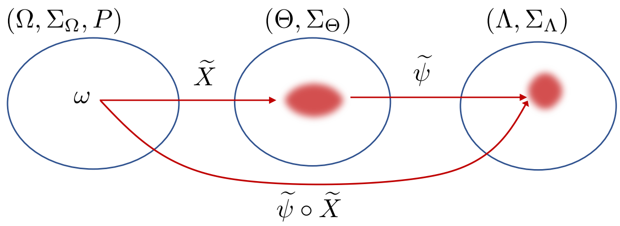


Figure 2: Transformation of a random fuzzy set $\widetilde{X}$ by a one-to-one mapping $\psi : \Theta \to \Lambda$.

We note that mapping $\widetilde{\psi}$ is also one-to-one, and its inverse is the extension of ψ^{-1}: for any $\widetilde{F} \in [0,1]^\Theta$,

$$
\begin{aligned}
(\widetilde{\psi^{-1}} \circ \widetilde{\psi})(\widetilde{F})(\theta) &= \widetilde{\psi^{-1}}(\widetilde{\psi}(\widetilde{F}))(\theta) \\
&= \sup_{\theta=\psi^{-1}(\lambda)} \widetilde{\psi}(\widetilde{F})(\lambda) \\
&= \sup_{\theta=\psi^{-1}(\lambda)} \widetilde{F}(\psi^{-1}(\lambda)) = \widetilde{F}(\theta).
\end{aligned}
$$

Similarly, $(\widetilde{\psi} \circ \widetilde{\psi^{-1}})(\widetilde{F}) = \widetilde{F}$, so we can write $\widetilde{\psi^{-1}} = \widetilde{\psi}^{-1}$.

We now consider the composed mapping $\widetilde{\psi} \circ \widetilde{X}$ from Ω to $[0,1]^\Lambda$, such that $(\widetilde{\psi} \circ \widetilde{X})(\omega) = \widetilde{\psi}[\widetilde{X}(\omega)]$ (see Figure 2). To show that it is a RFS, we start by the following lemma.

Lemma 1. *The set Σ_Λ containing the images of all elements of Σ_Θ by ψ,*

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

is a σ-algebra on Λ.

Proof. Since Σ_Θ is a σ-algebra, it contains the empty set $\emptyset$; consequently, $\psi(\emptyset) = \emptyset \in \Sigma_\Lambda$. Now, for any $A \in \Sigma_\Lambda$, $B = \psi^{-1}(A) \in \Sigma_\Theta$; hence, $\psi(\Theta \setminus B) = \Lambda \setminus \psi(B) \in \Sigma_\Lambda$. Finally, let (A_i), $i \in I$ be a collection of elements of Σ_Λ, and $B_i = \psi^{-1}(A_i)$, $i \in I$ their inverse images. We have

$$\bigcup_{i\in I} A_i = \bigcup_{i\in I} \psi(B_i) = \psi\left(\bigcup_{i\in I} B_i\right) \in \Sigma_\Lambda.$$

$\square$

Proposition 2. *Let $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \widetilde{X})$ be a RFS, $\psi : \Theta \to \Lambda$ a one-to-one mapping from Θ to Λ, and $\Sigma_\Lambda = \{\psi(B) : B \in \Sigma_\Theta\}$. The tuple $(\Omega, \Sigma_\Omega, P, \Lambda, \Sigma_\Lambda, \widetilde{\psi} \circ \widetilde{X})$ is a RFS.*

Proof. We need to prove that, for any $\alpha \in [0,1]$ the mapping ${}^\alpha(\widetilde{\psi} \circ \widetilde{X})$ is $\Sigma_\Omega - \Sigma_\Lambda$ strongly measurable [24], i.e., for any $A \in \Sigma_\Lambda$,

$$A^* = \{\omega \in \Omega : {}^\alpha(\widetilde{\psi} \circ \widetilde{X})(\omega) \cap A \neq \emptyset\} \in \Sigma_\Omega.$$

Now,

$$\begin{aligned}
{}^{\alpha}(\widetilde{\psi} \circ \widetilde{X})(\omega) &= \{\lambda \in \Lambda : (\widetilde{\psi} \circ \widetilde{X})(\omega)(\lambda) \geq \alpha\} \\
&= \{\lambda \in \Lambda : \widetilde{X}(\omega)[\psi^{-1}(\lambda)] \geq \alpha\} \\
&= \{\lambda \in \Lambda : \psi^{-1}(\lambda) \in {}^{\alpha}\widetilde{X}(\omega)\} \\
&= \psi[{}^{\alpha}\widetilde{X}(\omega)].
\end{aligned}$$

Consequently,

$$\begin{aligned}
A^* &= \{\omega \in \Omega : \psi[{}^{\alpha}\widetilde{X}(\omega)] \cap A \neq \emptyset\} \\
&= \{\omega \in \Omega : \psi\left[{}^{\alpha}\widetilde{X}(\omega) \cap \psi^{-1}(A)\right] \neq \emptyset\} \\
&= \{\omega \in \Omega : {}^{\alpha}\widetilde{X}(\omega) \cap \psi^{-1}(A) \neq \emptyset\}.
\end{aligned}$$

As ${}^{\alpha}\widetilde{X}$ is $\Sigma_\Omega - \Sigma_\Theta$ strongly measurable and $\psi^{-1}(A) \in \Sigma_\Theta$, it follows that $A^* \in \Sigma_\Omega$. □

Example 1. *Let* $\Omega = \{\omega_1, \omega_2\}$*,* P *the probability measure on* Ω *defined by* $P(\{\omega_1\}) = 0.7$*,* $P(\{\omega_2\}) = 0.3$*, and* $\widetilde{X} : \Omega \to [0,1]^{\mathbb{R}}$ *a random fuzzy number with two fuzzy focal sets defined by* $\widetilde{X}(\omega_1) = \widetilde{F}_1 \sim \mathsf{GFN}(1,1)$ *and* $\widetilde{X}(\omega_2) = \widetilde{F}_2 \sim \mathsf{GFN}(2, 0.5)$ *(see Figure 3a). Let* ψ *the bijection from* $\Theta = \mathbb{R}$ *to* $\Lambda = \mathbb{R}_+$ *defined by* $\psi : \theta \mapsto \exp(\theta)$*. From* (13)*, the images of* $\widetilde{F}_1$ *and* $\widetilde{F}_2$ *by* $\widetilde{\psi}$ *are, respectively, the following fuzzy subsets of* $\mathbb{R}_+$*:*

$$\widetilde{\psi}(\widetilde{F}_1)(\lambda) = \widetilde{F}_1(\log \lambda) = \exp\left(-\frac{1}{2}(\log \lambda - 1)^2\right)$$

and

$$\widetilde{\psi}(\widetilde{F}_2)(\lambda) = \widetilde{F}_2(\log \lambda) = \exp\left(-\frac{1}{4}(\log \lambda - 2)^2\right).$$

They are plotted in Figure 3b. Let $\widetilde{Y} = \widetilde{\psi} \circ \widetilde{X}$ *the transformation of* $\widetilde{X}$ *by* $\widetilde{\psi}$*; it is defined by* $\widetilde{Y}(\omega_1) = \widetilde{\psi}(\widetilde{F}_1)$ *and* $\widetilde{Y}(\omega_2) = \widetilde{\psi}(\widetilde{F}_2)$*.*

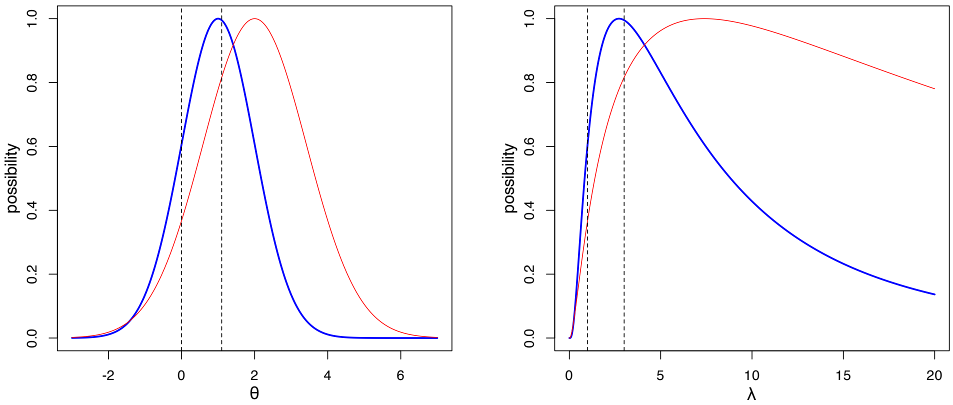
Belief and plausibility. Interestingly, the belief and plausibility functions induced by the transformed RFS $\widetilde{\psi} \circ \widetilde{X}$ have a simple expression in terms of corresponding functions induced by $\widetilde{X}$, as expressed by the following theorem.

Theorem 1. *Let* $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \widetilde{X})$ *be a RFS,* $\psi : \Theta \to \Lambda$ *a one-to-one mapping from* Θ *to* Λ*,* $\Sigma_\Lambda = \{\psi(B) : B \in \Sigma_\Theta\}$*, and* $\widetilde{\psi} \circ \widetilde{X}$ *the RFS resulting from the transformation of* $\widetilde{X}$ *by* ψ*. For any* $C \in \Sigma_\Lambda$*,*

$$Bel_{\widetilde{\psi} \circ \widetilde{X}}(C) = Bel_{\widetilde{X}}(\psi^{-1}(C)), \tag{14a}$$

and

$$Pl_{\widetilde{\psi} \circ \widetilde{X}}(C) = Pl_{\widetilde{X}}(\psi^{-1}(C)). \tag{14b}$$



(a) (b)

Figure 3: Focal fuzzy sets of $\widetilde{X}$ (left) and $\widetilde{Y} = \widetilde{\psi} \circ \widetilde{X}$ (right) in Example 1. The vertical broken lines indicate the intervals $C = [1, 3]$ (right) and $\log(C) = [0, \log(3)]$ (left).

Proof. 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,

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

The result follows directly using the definition of belief and plausibility function in (2). $\square$

Example 2. *Continuing Example 1, let $C = [1, 3]$. We can compute $Pl_{\widetilde{Y}}(C)$ and $Bel_{\widetilde{Y}}(C)$ in two ways (see Figure 3):*

1. *Using the definitions* (2)*:*

$$\begin{aligned}
Pl_{\widetilde{Y}}(C) &= 0.7 \sup_{\lambda \in [1,3]} \widetilde{\psi}(\widetilde{F}_1)(\lambda) + 0.3 \sup_{\lambda \in [1,3]} \widetilde{\psi}(\widetilde{F}_2)(\lambda) \\
&= 0.7 \times 1 + 0.3 \times \exp\left(-\frac{1}{4}(\log(3) - 2)^2\right) \approx 0.94 \\
Bel_{\widetilde{Y}}(C) &= 1 - \left(0.7 \sup_{\lambda \notin [1,3]} \widetilde{\psi}(\widetilde{F}_1)(\lambda) + 0.3 \sup_{\lambda \notin [1,3]} \widetilde{\psi}(\widetilde{F}_2)(\lambda)\right) \\
&= 1 - \left(0.7 \times \exp\left(-\frac{1}{2}(\log(3) - 1)^2\right) + 0.3 \times 1\right) \approx 0.0034.
\end{aligned}$$

2. *Using Theorem 1:*

$$
\begin{aligned}
Pl_{\widetilde{Y}}(C) &= Pl_{\widetilde{X}}(\psi^{-1}(C)) \\
&= 0.7 \sup_{\theta\in[0,\log(3)]} \widetilde{F}_1(\theta) + 0.3 \sup_{\theta\in[0,\log(3)]} \widetilde{F}_2(\theta) \\
&= 0.7 \times 1 + 0.3 \times \exp\left(-\frac{1}{4}(\log 3 - 2)^2\right) \approx 0.94 \\
Bel_{\widetilde{Y}}(C) &= Bel_{\widetilde{X}}(\psi^{-1}(C)) \\
&= 1 - \left(0.7 \sup_{\theta\notin[0,\log(3)]} \widetilde{F}_1(\theta) + 0.3 \sup_{\theta\notin[0,\log(3)]} \widetilde{F}_2(\theta)\right) \\
&= 1 - \left(0.7 \times \exp\left(-\frac{1}{2}(\log 3 - 1)^2\right) + 0.3 \times 1\right) \approx 0.0034.
\end{aligned}
$$

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 2. *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

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

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

$$
\begin{aligned}
\widetilde{\psi}(\widetilde{F} \odot \widetilde{G})(\lambda) &= (\widetilde{F} \odot \widetilde{G})[\psi^{-1}(\lambda)] && (15a) \\
&= \frac{\widetilde{F}[\psi^{-1}(\lambda)]\widetilde{G}[\psi^{-1}(\lambda)]}{\sup_{\lambda'} \widetilde{F}[\psi^{-1}(\lambda')]\widetilde{G}[\psi^{-1}(\lambda')]} && (15b) \\
&= \frac{\widetilde{\psi}(\widetilde{F})(\lambda)\widetilde{\psi}(\widetilde{G})(\lambda)}{\sup_{\lambda'} \widetilde{\psi}(\widetilde{F})(\lambda')\widetilde{\psi}(\widetilde{G})(\lambda')} && (15c) \\
&= (\widetilde{\psi}(\widetilde{F}) \odot \widetilde{\psi}(\widetilde{G}))(\lambda). && (15d)
\end{aligned}
$$

Now, the degree of conflict between $\widetilde{\psi}(\widetilde{F})$ and $\widetilde{\psi}(\widetilde{G})$ is the denominator on the right-hand side of (15b). It is equal to

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

□

We can now state the main result of this section.

Theorem 2. *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 2.1, 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

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

Now, from Lemma 2,

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

Hence, the mappings from Ω to $[0,1]^\Lambda$ associated to $\widetilde{\psi} \circ (\widetilde{X}_1 \oplus \widetilde{X}_2)$ and $(\widetilde{\psi} \circ \widetilde{X}_1) \oplus (\widetilde{\psi} \circ \widetilde{X}_2)$ are identical. Furthermore, from Lemma 2,

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

The fuzzy conditioning event in $[0,1]^{\Omega_1 \times \Omega_2}$ associated to $\widetilde{X}_1 \oplus \widetilde{X}_2$ and $(\widetilde{\psi} \circ \widetilde{X}_1) \oplus (\widetilde{\psi} \circ \widetilde{X}_2)$ are, thus, also identical, which completes the proof. □

Example 3. *Continuing Examples 1 and 2, let us consider another RFS defined by* $\Omega' = \{\omega'_1, \omega'_2\}$, $P'(\{\omega'_1\}) = 0.6$, $P'(\{\omega'_2\}) = 0.4$, $\widetilde{X}'(\omega'_1) = \widetilde{F}'_1 \sim \mathsf{GFN}(0, 1)$, $\widetilde{X}'(\omega'_2) = \widetilde{F}'_2 \sim \mathsf{GFN}(1.5, 1)$. *Let* $\widetilde{Y} = \widetilde{\psi} \circ \widetilde{X}$ *and* $\widetilde{Y}' = \widetilde{\psi} \circ \widetilde{X}'$ *and assume that we want to compute their orthogonal sum* $\widetilde{Y} \oplus \widetilde{Y}'$. *From Theorem 2, we can first combine* $\widetilde{X}$ *and* $\widetilde{X}'$, *and then compose the result with* $\widetilde{\psi}$. *Using* (7), *we obtain*

$$\mathsf{hgt}(\widetilde{F}_1 \cdot \widetilde{F}'_1) = 0.78, \quad \mathsf{hgt}(\widetilde{F}_1 \cdot \widetilde{F}'_2) = 0.94$$

$$\mathsf{hgt}(\widetilde{F}_2 \cdot \widetilde{F}'_1) = 0.51, \quad \mathsf{hgt}(\widetilde{F}_2 \cdot \widetilde{F}'_2) = 0.96$$

Consequently, from (6), *the joint probability measure on* $\Omega \times \Omega'$ *conditioned on fuzzy subset* $\widetilde{\Theta}$ *defined by* (5) *is*

$$P''(\{(\omega_1, \omega'_1)\}) \propto 0.7 \times 0.6 \times 0.78$$

$$P''(\{(\omega_1, \omega'_2)\}) \propto 0.7 \times 0.4 \times 0.94$$

$$P''(\{(\omega_2, \omega'_1)\}) \propto 0.3 \times 0.6 \times 0.51$$

$$P''(\{(\omega_2, \omega'_2)\}) \propto 0.3 \times 0.4 \times 0.96.$$

After normalization, we get

$$P''(\{(\omega_1,\omega_1')\}) \approx 0.41, \quad P''(\{(\omega_1,\omega_2')\}) \approx 0.33$$

$$P''(\{(\omega_2,\omega_1')\}) \approx 0.12, \quad P''(\{(\omega_2,\omega_2')\}) \approx 0.14.$$

Now, the combined mapping is $\widetilde{X} \oplus \widetilde{X}'$ is

$$\begin{aligned}
(\widetilde{X} \oplus \widetilde{X}')(\omega_1,\omega_1') &= \widetilde{F}_1 \odot \widetilde{F}_1' \sim \textsf{\textit{GFN}}(0.5, 2)\\
(\widetilde{X} \oplus \widetilde{X}')(\omega_1,\omega_2') &= \widetilde{F}_1 \odot \widetilde{F}_2' \sim \textsf{\textit{GFN}}(1.25, 2)\\
(\widetilde{X} \oplus \widetilde{X}')(\omega_2,\omega_1') &= \widetilde{F}_2 \odot \widetilde{F}_1' \sim \textsf{\textit{GFN}}(2/3, 1.5)\\
(\widetilde{X} \oplus \widetilde{X}')(\omega_1,\omega_1') &= \widetilde{F}_2 \odot \widetilde{F}_2' \sim \textsf{\textit{GFN}}(5/3, 1.5).
\end{aligned}$$

Finally, we have

$$(\widetilde{Y} \oplus \widetilde{Y}')(\omega_1,\omega_1') = \widetilde{\psi}(\widetilde{F}_1 \odot \widetilde{F}_1'), \quad (\widetilde{Y} \oplus \widetilde{Y}')(\omega_1,\omega_2') = \widetilde{\psi}(\widetilde{F}_1 \odot \widetilde{F}_2')$$

$$(\widetilde{Y} \oplus \widetilde{Y}')(\omega_2,\omega_1') = \widetilde{\psi}(\widetilde{F}_2 \odot \widetilde{F}_1'), \quad (\widetilde{Y} \oplus \widetilde{Y}')(\omega_2,\omega_2') = \widetilde{\psi}(\widetilde{F}_2 \odot \widetilde{F}_2').$$

3.2. Transformed Gaussian Random Fuzzy Numbers

Applying the idea developed in Section 3.1 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 $\widetilde{\psi} \circ \widetilde{X}$ be the result of the transformation of $\widetilde{X}$ by ψ. We will say that $\widetilde{\psi} \circ \widetilde{X}$ is a *transformed GRFN* (or t-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

$$\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \psi^{-1}) \Leftrightarrow \widetilde{\psi}^{-1} \circ \widetilde{Y} \sim \widetilde{N}(\mu, \sigma^2, h). \tag{16}$$

From Theorem 2, given two t-GRFNs $\widetilde{Y}_i \sim T\widetilde{N}(\mu_i, \sigma_i^2, h_i, \psi^{-1})$, $i = 1, 2$, we have $Y_1 \oplus Y_2 \sim T\widetilde{N}(\widetilde{\mu}_{12}, \widetilde{\sigma}_{21}^2, h_1 + h_2, \psi^{-1})$, where $\widetilde{\mu}_{12}$ and $\widetilde{\sigma}_{21}^2$ are given by (11).

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

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 *lognormal random fuzzy number (RFN)* $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \log)$. From (16), $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \log)$ if and only if $\widetilde{\log}(\widetilde{Y}) \sim \widetilde{N}(\mu, \sigma^2, h)$. A lognormal random variable is recovered when $h = +\infty$. From (8) and (14b), 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 (9) and (14) as

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

and $Pl_{\widetilde{Y}}((-\infty, y]) = Bel_{\widetilde{Y}}((-\infty, y]) + pl_{\widetilde{Y}}(y)$.

Expressions for the lower and upper expectations of a lognormal RFN are given in the following proposition.

Proposition 3. *The lower and upper expectation of a lognormal RFN $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \log)$ are given, respectively, by*

$$\mathbb{E}_*(\widetilde{Y}) = \sqrt{2\pi}\exp\left(\mu + \frac{\sigma^2}{2} + \frac{1}{2h}\right)\left[\phi\left(\frac{1}{\sqrt{h}}\right) - \frac{1}{\sqrt{h}}\left(1 - \Phi\left(\frac{1}{\sqrt{h}}\right)\right)\right] \tag{17a}$$

and

$$\mathbb{E}^*(\widetilde{Y}) = \sqrt{2\pi}\exp\left(\mu + \frac{\sigma^2}{2} + \frac{1}{2h}\right)\left[\phi\left(\frac{1}{\sqrt{h}}\right) + \frac{1}{\sqrt{h}}\Phi\left(\frac{1}{\sqrt{h}}\right)\right], \tag{17b}$$

where ϕ is the standard normal probability density function.

Proof. See Appendix B □

We note that the expectation of a lognormal random variable is recovered in the limit when h tends to infinity, as

$$\lim_{h\to+\infty}\mathbb{E}_*(\widetilde{Y}) = \lim_{h\to+\infty}\mathbb{E}^*(\widetilde{Y}) = \exp\left(\mu + \frac{\sigma^2}{2}\right).$$

Example 4. *Figure 4 displays two lognormal RFNs*

$$\widetilde{Y}_1 \sim T\widetilde{N}(1, 1, 5, \log) \quad \textit{and} \quad \widetilde{Y}_2 \sim T\widetilde{N}(2, 0.1, 2, \log),$$

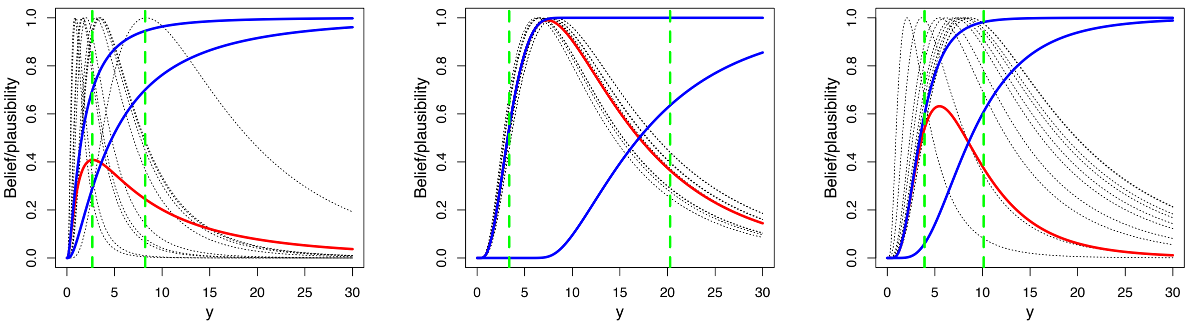
as well as their orthogonal sum $\widetilde{Y}_1 \oplus \widetilde{Y}_2$. For each RFN, we plot ten realizations, the contour function, the lower and upper expectations, as well as the lower and upper cdfs.

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,

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

A RFN $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma, h, \mathsf{logit})$ will said to be *logit-normal.* A logit-normal random variable [1] is recovered when $h = +\infty$. From (8) and (14b), the contour function of $\widetilde{Y}$ is

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



(a) (b) (c)

Figure 4: (a) and (b): From left to right, two lognormal 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 lognormal 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), the lower and upper expectations (vertical broken lines), as well as the lower and upper cdfs (blue curves).

The lower and upper cdfs of $\widetilde{Y}$ can be computed from (9) and (14) in a similar manner. The expectation of the logit-normal probability does not have any analytical expression. Consequently, this is also true for the lower and upper expectations of a logit-normal RFS. A multidimensional extension of logit-normal RFNs will be studied in Section 3.3.

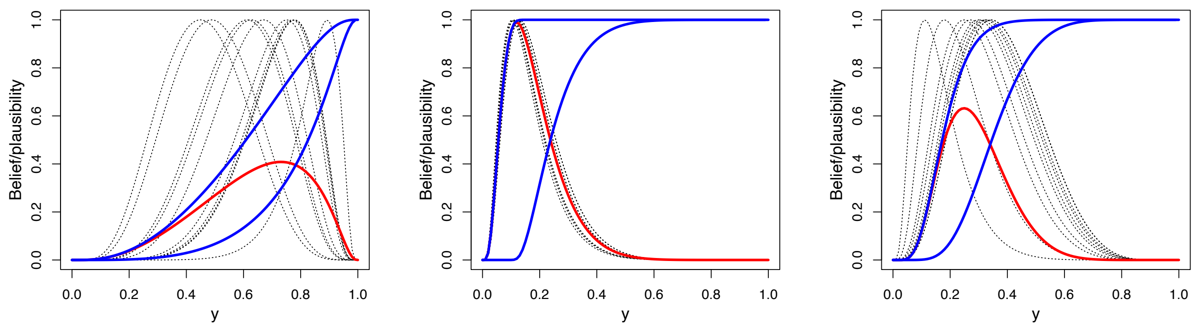
Example 5. *Figure 5 shows representations of logit-normal RFNs* $\widetilde{Y}_1 \sim T\widetilde{N}(1,1,5,\mathit{logit})$ *and* $\widetilde{Y} \sim T\widetilde{N}(-2,0.1,2,\mathit{logit})$*, as well as their orthogonal sum* $Y_1 \oplus Y_2$*.*

Parameterized families of transformations. Whereas the fixed transformations considered above allow us to define RFNs with various supports, it may be useful to consider more general transformations belonging to parameterized families. We then obtain a parametrized family of RFNs with parameters μ, σ^2, h, and the parameters of the transformation. Such flexible families may be useful, for instance, in a belief elicitation context where we attempt to fit an expert's belief statements with a t-GRFN, as will be seen in Section 5.1. The idea of transforming the normal distribution to obtained parameterized families of distributions with varying skewness and kurtosis can be traced back, at least, to Ref. [20]. Johnson [20] actually considers the inverse problem of finding a transformation ψ^{-1} of a random variable Y such that $X = \psi^{-1}(Y)$ has, approximately, a standard normal distribution. In addition to the logarithmic and logit transformations, he considers a "system S_U" based on the following transformation:

$$X = \psi^{-1}(Y) = \gamma + \delta \sinh^{-1} \frac{Y - \xi}{\lambda}, \tag{18}$$

where γ, δ, ξ and λ are four parameters, $\delta > 0$, $\lambda > 0$, and $X \sim N(0,1)$. Inverting (18), we get

$$Y = \psi(X) = \xi + \lambda \sinh \frac{X - \gamma}{\delta}. \tag{19}$$



(a) (b) (c)

Figure 5: (a) and (b): From left to right, two logit-normal random fuzzy numbers $\widetilde{Y} \sim T\widetilde{N}(1, 1, 5, \mathsf{logit})$ and $\widetilde{Y} \sim T\widetilde{N}(-2, 0.1, 2, \mathsf{logit})$; (c): combined lognormal random fuzzy number $\widetilde{Y}_1 \oplus \widetilde{Y}_2$. For each RFN, we plot ten realizations (black dotted curves), the contour function (red curve), as well as lower and upper cdfs (blue curves).

To be consistent with our previous notations, we can rewrite (19) as

$$Y = \psi_{\xi,\lambda}(X) = \xi + \lambda \sinh X, \tag{20}$$

where $X \sim N(\mu, \sigma^2)$, and the four parameters are now: μ, σ^2 defining the distribution of X on the one hand, and ξ and λ defining the transformation on the other hand. This transformation makes it possible to define a four-parameter family of probability distributions on the whole real line, with varying skewness and kurtosis. The same transformation applied to GRFNs defines a parametric family of RFNs with different shapes. We note that Theorem 2 allows us to combine two t-GRFNs $\widetilde{Y}_1 \sim T\widetilde{N}(\mu_1, \sigma_1^2, h_1, \psi_{\xi,\lambda}^{-1})$ and $\widetilde{Y}_2 \sim T\widetilde{N}(\mu_2, \sigma_2^2, h_2, \psi_{\xi,\lambda}^{-1})$ with different means, variances and precisions, but the same transformation $\psi_{\xi,\lambda}$.

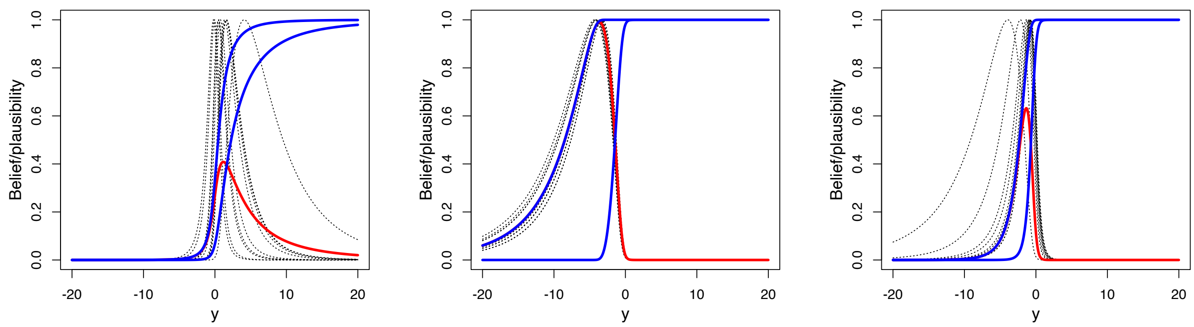
Example 6. *Figure 6 shows representations of RFNs $\widetilde{Y}_1 \sim T\widetilde{N}(1, 1, 5, \psi_{\xi,\lambda}^{-1})$ and $\widetilde{Y} \sim T\widetilde{N}(-2, 0.1, 2, \psi_{\xi,\lambda}^{-1})$ with $\xi = 0$ and $\lambda = 1$, as well as their orthogonal sum $Y_1 \oplus Y_2$. We can see that the contour functions of $\widetilde{Y}_1$ and $\widetilde{Y}_2$ are, respectively, right-skewed and left-skewed, while the contour function of $\widetilde{Y}_1 \oplus \widetilde{Y}_2$ is left-skewed. In general, the positive or negative skewness depends, for this transformation, on the sign of μ.*

An even more general approach to create parametric families of probability distributions was more recently proposed in [2]. Given three random variables X, T and R, Aljarrah et al. [2] define a new random variable

$$Y = \psi_{RT}(X) = (F_R^{-1} \circ F_T)(X),$$

where F_R and F_T are the cdfs of R and T. The cdf of Y is, thus,

$$F_Y(y) = P(Y \le y) = P(F_T(X) \le F_R(y)) = P(X \le F_T^{-1}(F_R(y))) = (F_X \circ F_T^{-1} \circ F_R)(y).$$



(a) (b) (c)

Figure 6: (a) and (b): From left to right, two t-GRFNs $\widetilde{Y} \sim T\widetilde{N}(1, 1, 5, \psi^{-1}_{\xi,\lambda})$ and $\widetilde{Y} \sim T\widetilde{N}(-2, 0.1, 2, \psi^{-1}_{\xi,\lambda})$, with $\xi = 0$ and $\lambda = 1$; (c): combined lognormal random fuzzy number $\widetilde{Y}_1 \oplus \widetilde{Y}_2$. For each RFN, we plot ten realizations (black dotted curves), the contour function (red curve), as well as lower and upper cdfs (blue curves).

Taking $X \sim N(\mu, \sigma^2)$ it is possible, using this approach, to define infinitely many parametric families of probability distributions by choosing different parametric families for T and R. We note that the support of Y is included in the support of R. Also, choosing $F_T = F_X$ gives us $Y = R$, while choosing $F_T = F_R$ yields $Y = X$. By extension, we can define parametric families of RFNs $\widetilde{Y} \sim T\widetilde{N}(\mu, \sigma^2, h, \psi^{-1}_{RT})$ based on parametric families for R and T.

3.3. Transformations of Gaussian Random Vectors

The general approach introduced in Section 3.1 can also be applied to GRFVs. Of special interest is the multivariate extension of the notion of logit-normal RFNs, which is obtained from a multidimensional normal distribution $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in $\mathbb{R}^{p-1}$ ($p \geq 2$) and the following *softmax* transformation from $\mathbb{R}^{p-1}$ to the simplex $\mathcal{S}_p$ of p-dimensional probability vectors:

$$\psi_S(\boldsymbol{x}) = \left[\frac{\exp(x_1)}{1 + \sum_{j=1}^{p-1}\exp(x_j)}, \ldots, \frac{\exp(x_{p-1})}{1 + \sum_{j=1}^{p-1}\exp(x_j)}, \frac{1}{1 + \sum_{j=1}^{p-1}\exp(x_j)}\right]^T, \tag{21}$$

with inverse

$$\psi_S^{-1}(\boldsymbol{y}) = \left[\log\left(\frac{y_1}{y_p}\right), \ldots, \log\left(\frac{y_{p-1}}{y_p}\right)\right]^T. \tag{22}$$

This transformation is used in [1] to define the multidimensional logistic-normal probability distribution. Here, given a GRVN $\widetilde{X} \sim \widetilde{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{H})$, we can define the random fuzzy vector $\widetilde{Y} = \widetilde{\psi}_S \circ \widetilde{X}$, where $\widetilde{\psi}_S$ is the extension of the softmax transformation (21). We will say that $\widetilde{Y}$ is a *logistic-normal RFV*, and we will write $\widetilde{Y} \sim T\widetilde{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{H}, \psi_S^{-1})$. The domain of $\widetilde{Y}$ is the simplex $\mathcal{S}_p$. Such a RFV can be used to represent beliefs about a vector of probabilities or proportions.

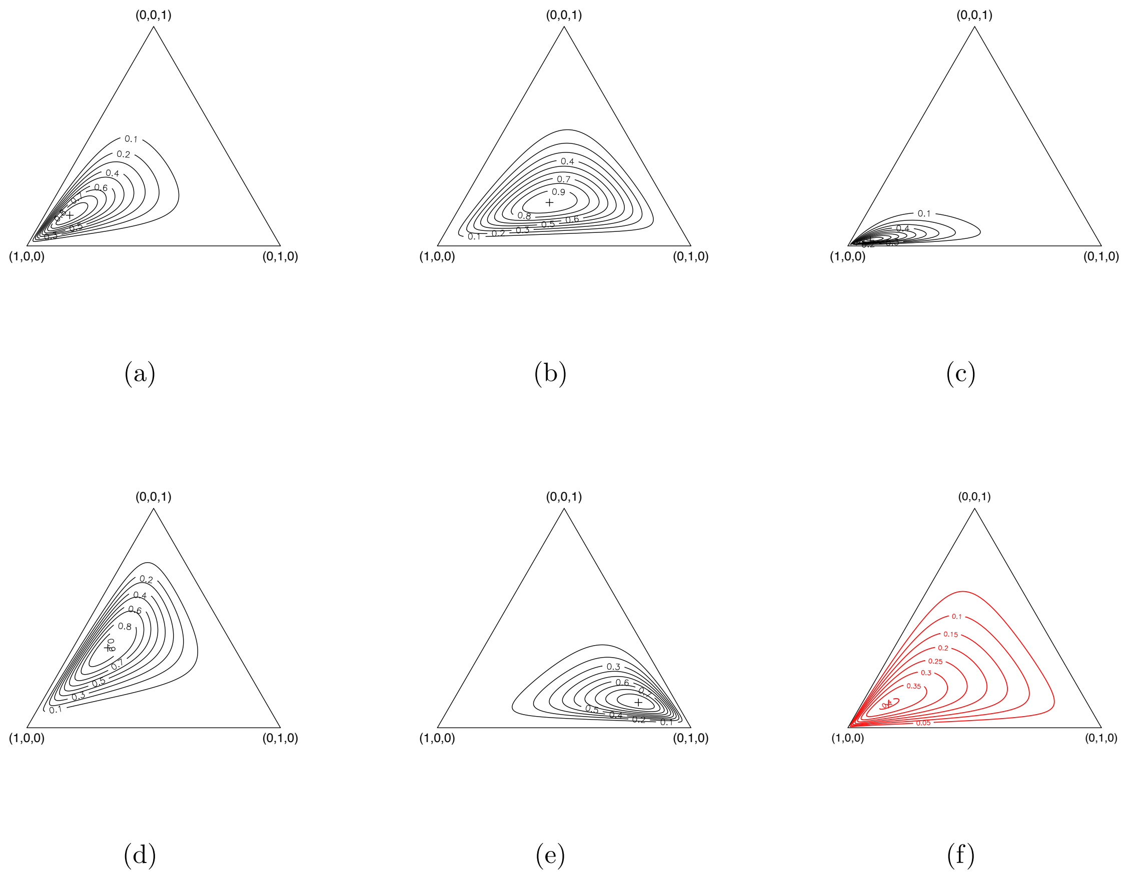


Figure 7: (a)-(e): Five focal fuzzy sets $\widetilde{Y}(\omega)$ for a logistic-normal RFV $\widetilde{Y} \sim \widetilde{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{H}, \psi_S^{-1})$ (see Example 7); (f): corresponding contour function.

Example 7. *Figures 7a-7e show five focal fuzzy sets $\widetilde{Y}(\omega)$ in barycentric coordinates for a logistic RFV $\widetilde{Y} \sim \widetilde{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{H}, \psi_S^{-1})$ with $p = 3$, $\boldsymbol{\mu} = (2, 0)^T$,*

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.5 & 0.2 \\ 0.2 & 1 \end{pmatrix} \quad \textit{and} \quad \boldsymbol{H} = \begin{pmatrix} 1 & -0.2 \\ -0.2 & 1.5 \end{pmatrix}.$$

The corresponding contour function is displayed in Figure 7f.

4. Mixtures of Gaussian random fuzzy variables

In probability and statistics, finite mixtures of probability distributions and, in particular, finite mixtures of Gaussians are commonly used to obtain distributions with arbitrarily complex shapes [23]. In this section, we extend this approach to define mixtures of (transformed) GRFNs and GRFVs. Mixtures of GRFNs and GRFVs will first be defined in

Section 4.1. Their properties will then be studied in Section 4.2, and their combination will be addressed in Section 4.3. A summarization procedure allowing us to approximate a mixture with a large number of components by a simpler one will be described in Section 4.4. Finally, mixtures of transformed GRFNs and GRFVs will be introduced in Section 4.5.

4.1. Definitions

Mixture of GRFNs. We consider a pair of random variables (M, Z) from a probability space $(\Omega, \Sigma_\Omega, P)$ to $\mathbb{R} \times \{1, \dots, K\}$, such that the marginal distribution of Z is defined by $P(Z = k) = \pi_k$, $k = 1, \dots, K$, and the conditional distribution of M given $Z = k$ is univariate normal:

$$M \mid (Z = k) \sim N(\mu_k, \sigma_k^2).$$

The marginal distribution of M is, thus, a mixture of K normal distributions. Now, consider the random fuzzy set $\widetilde{X} : \Omega \to [0, 1]^{\mathbb{R}}$ defined as follows,

$$\widetilde{X}(\omega) = \mathsf{GFN}\left(M(\omega), \prod_{k=1}^{K} h_k^{Z_k(\omega)}\right),$$

where $Z_k(\omega) = I(Z(\omega) = k)$, and $I(\cdot)$ is the indicator function. Conditionally on $Z = k$, $\widetilde{X}$ is a GRFN with mean μ_k, variance σ_k^2 and precision h_k:

$$\widetilde{X} \mid (Z = k) \sim \widetilde{N}(\mu_k, \sigma_k^2, h_k).$$

We denote this conditional GRFN by $\widetilde{X}_k$. We say that $\widetilde{X}$ is a *mixture GRFN* (m-GRFN) and we write

$$\widetilde{X} \sim \sum_{k=1}^{K} \pi_k \widetilde{N}(\mu_k, \sigma_k^2, h_k).$$

Mixture of GRFVs. Similarly, we can define a mixture of GRFVs by a mapping $\widetilde{X} : \Omega \to [0, 1]^{\mathbb{R}^p}$ with $p \geq 2$ such that

$$\widetilde{X}(\omega) = \mathsf{GFV}(\boldsymbol{M}(\omega), \prod_{k=1}^{K} \boldsymbol{H}_k^{Z_k(\omega)}),$$

where, as usual, $\boldsymbol{H}_k^0$ is the identity matrix, $\boldsymbol{M}$ is a random vector having a mixture of multivariate normal distributions,

$$\boldsymbol{M} \sim \sum_{k=1}^{K} \pi_k N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$

and $\boldsymbol{H}_k$, $k = 1, \dots, K$ are positive definite precision matrices. We will use the following notation:

$$\widetilde{X} \sim \sum_{k=1}^{K} \pi_k \widetilde{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_k).$$

4.2. Properties

Belief and plausibility functions. The belief and plausibility functions associated with an m-GRFN are the weighted sums of, respectively, the belief and plausibility associated with the components of the mixture. This property is expressed by the following theorem.

Theorem 3. *Let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algera on $\mathbb{R}$, and $A \in \mathcal{B}(\mathbb{R})$ be a measurable subset of $\mathbb{R}$. The degrees of belief and plausibility of A induced by an m-GRFN $\widetilde{X} \sim \sum_{k=1}^{K} \pi_k \widetilde{N}(\mu_k, \sigma_k^2, h_k)$ are*

$$Bel_{\widetilde{X}}(A) = \sum_{k=1}^{K} \pi_k Bel_{\widetilde{X}_k}(A) \tag{23a}$$

$$Pl_{\widetilde{X}}(A) = \sum_{k=1}^{K} \pi_k Pl_{\widetilde{X}_k}(A), \tag{23b}$$

with $\widetilde{X}_k \sim \widetilde{N}(\mu_k, \sigma_k^2, h_k)$.

Proof. Let us start with (23b). By definition, $Pl_{\widetilde{X}}(A)$ is defined as the following expectation,

$$\begin{aligned}
Pl_{\widetilde{X}}(A) &= \mathbb{E}_{M,Z}\left[\sup_{u\in A} \varphi(u, M, \prod_{k=1}^{K} h_k^{Z_k})\right] \\
&= \mathbb{E}_Z \mathbb{E}_{M|Z}\left[\sup_{u\in A} \varphi(u, M, \prod_{k=1}^{K} h_k^{Z_k})\right] \\
&= \sum_{k=1}^{K} \pi_k \mathbb{E}_{M|Z}\left[\sup_{u\in A} \varphi(u, M, h_k) \mid Z = k\right] \\
&= \sum_{k=1}^{K} \pi_k Pl_{\widetilde{X}_k}(A).
\end{aligned}$$

Eq. (23a) can be proved in the same way, since $Bel_{\widetilde{X}}(A)$ is also defined as an expectation. □

Using Theorem 3 and the closed-form expressions given in [14], we can compute the degrees of belief and plausibility of any real interval $[x, y]$. In particular, setting $x = y$ gives the following corollary:

Corollary 1. *The contour function of m-GRFN $\widetilde{X} \sim \sum_{k=1}^{K} \pi_k \widetilde{N}(\mu_k, \sigma_k^2, h_k)$ is*

$$pl_{\widetilde{X}}(x) = \sum_{k=1}^{K} \frac{\pi_k}{\sqrt{1 + h_k \sigma_k^2}} \exp\left(-\frac{h_k(x - \mu_k)^2}{2(1 + h_k \sigma_k^2)}\right). \tag{24}$$

Proof. Immediate from Theorem 3 and Eq. (8). □

Example 8. *Figure 8 shows ten focal sets of an m-GRFN $\widetilde{X} \sim 0.4\widetilde{N}(-2, 1, 5) + 0.6\widetilde{N}(2, 0.1^2, 1)$, its contour function as well as its lower and upper cfds.*

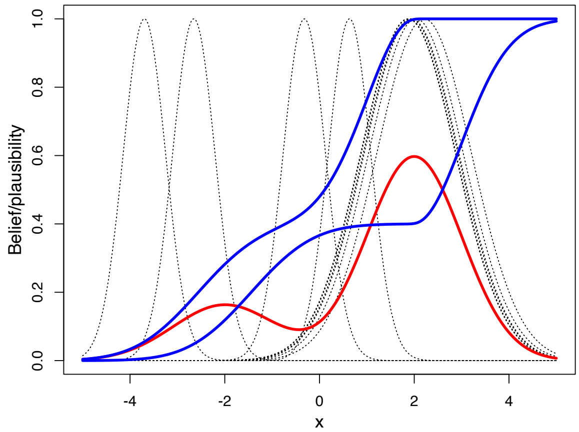


Figure 8: Ten realizations (black dotted curves), contour function (red curve), and lower/upper cdfs (blue curves) of an m-GRFN $\widetilde{X} \sim 0.4\widetilde{N}(-2, 1, 5) + 0.6\widetilde{N}(2, 0.1^2, 1)$.

Similar results can be obtained in the same way for mixtures of GRFVs. In particular, the contour function of a mixture of GRFVs is the weighted sum of the contour functions of its components, which is expressed by the following proposition.

Proposition 4. *The contour function of m-GRFV* $\widetilde{X} \sim \sum_{k=1}^{K} \pi_k \widetilde{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_k)$ *is*

$$pl_{\widetilde{X}}(x) = \sum_{k=1}^{K} \frac{\pi_k}{|\boldsymbol{I}_p + \boldsymbol{\Sigma}_k \boldsymbol{H}_k|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_k)^T (\boldsymbol{H}_k^{-1} + \boldsymbol{\Sigma}_k)^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k)\right). \tag{25}$$

Lower and upper expectations of an m-GRFN. The lower and upper expectations of an m-GRFN can easily be computed from those of its components, as stated in the following proposition.

Proposition 5. *The lower and upper expectations of m-GRFN* $\widetilde{X} \sim \sum_{k=1}^{K} \pi_k \widetilde{N}(\mu_k, \sigma_k^2, h_k)$ *are given by*

$$\mathbb{E}_*(\widetilde{X}) = \sum_{k=1}^{K} \pi_k \mu_k - \sum_{k=1}^{K} \sqrt{\frac{\pi}{2h_k}}, \quad \textit{and}$$

$$\mathbb{E}^*(\widetilde{X}) = \sum_{k=1}^{K} \pi_k \mu_k + \sum_{k=1}^{K} \sqrt{\frac{\pi}{2h_k}}.$$

Proof. The α-cut of $\widetilde{X}(\omega)$ is the closed interval

$${}^{\alpha}\widetilde{X}(\omega) = \left[M(\omega) - \sqrt{\frac{-2\ln\alpha}{\prod_{k=1}^{K} h_k^{Z_k(\omega)}}}, M(\omega) + \sqrt{\frac{-2\ln\alpha}{\prod_{k=1}^{K} h_k^{Z_k(\omega)}}} \right].$$

From (3), the lower expectation of $\widetilde{X}$ is

$$\begin{aligned}
\mathbb{E}_*(\widetilde{X}) &= \int_0^1 \mathbb{E}_{M,Z}\left[M - \sqrt{\frac{-2\ln\alpha}{\prod_{k=1}^K h_k^{Z_k}}}\right] d\alpha \\
&= \int_0^1 \mathbb{E}_Z \mathbb{E}_{M|Z}\left[M - \sqrt{\frac{-2\ln\alpha}{\prod_{k=1}^K h_k^{Z_k}}}\right] d\alpha \\
&= \int_0^1 \sum_{k=1}^K \pi_k \mathbb{E}_{M|Z}\left[M - \sqrt{\frac{-2\ln\alpha}{h_k}} | Z = k\right] d\alpha \\
&= \sum_{k=1}^K \pi_k \int_0^1 \mathbb{E}_{M|Z}\left[M - \sqrt{\frac{-2\ln\alpha}{h_k}} | Z = k\right] d\alpha \\
&= \sum_{k=1}^K \pi_k \left(\mu_k - \sqrt{\frac{\pi}{2h_k}}\right),
\end{aligned}$$

where the last equality is derived from (10). The upper expectation of $\widetilde{X}$ can be computed in the same way. □

4.3. Combination

Combination of m-GRFNs. Let us now consider two independent m-GRFNs

$$\widetilde{X}_1 \sim \sum_{k=1}^K \pi_{1k}\widetilde{N}(\mu_{1k}, \sigma_{1k}^2, h_{1k}) \quad \text{and} \quad \widetilde{X}_2 \sim \sum_{\ell=1}^L \pi_{2\ell}\widetilde{N}(\mu_{2\ell}, \sigma_{2\ell}^2, h_{2\ell}).$$

The following theorem states that their orthogonal sum $\widetilde{X}_1 \oplus \widetilde{X}_2$ is an m-GRFN.

Theorem 4. *Given two independent m-GRFNs* $\widetilde{X}_1 \sim \sum_{k=1}^K \pi_{1k}\widetilde{N}(\mu_{1k}, \sigma_{1k}^2, h_{1k})$ *and* $\widetilde{X}_2 \sim \sum_{\ell=1}^L \pi_{2\ell}\widetilde{N}(\mu_{2\ell}, \sigma_{2\ell}^2, h_{2\ell})$*,*

1. *The orthogonal sum of* $\widetilde{X}_1$ *and* $\widetilde{X}_2$ *is an m-GRFN,*

$$\widetilde{X}_1 \oplus \widetilde{X}_2 \sim \sum_{k=1}^K \sum_{\ell=1}^L \widetilde{\pi}_{k\ell}\left[\widetilde{N}(\mu_{1k}, \sigma_{1k}^2, h_{1k}) \oplus \widetilde{N}(\mu_{2\ell}, \sigma_{2\ell}^2, h_{2\ell})\right],$$

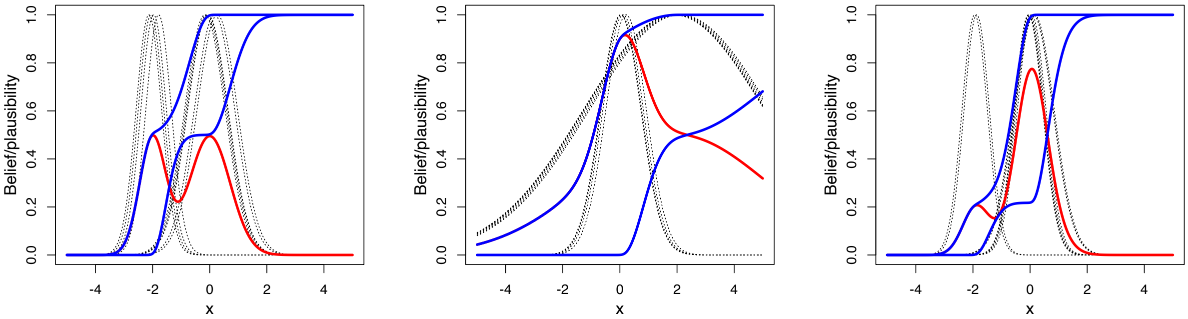
with

$$\widetilde{\pi}_{k\ell} = \frac{(1-\kappa_{k\ell})\pi_{1k}\pi_{2\ell}}{\sum_{k'\ell'}(1-\kappa_{k'\ell'})\pi_{1k'}\pi_{2\ell'}},$$

$\kappa_{k\ell}$ *denoting the degree of conflict between* $\widetilde{X}_{1k}$ *and* $\widetilde{X}_{2\ell}$ *given by Proposition 1.*

2. *The degree of conflict between* $\widetilde{X}_1$ *and* $\widetilde{X}_2$ *is*

$$\kappa = \sum_{k=1}^K \sum_{\ell=1}^L \kappa_{k\ell}\pi_{1k}\pi_{2\ell}.$$



(a) (b) (c)

Figure 9: (a) and (b): From left to right, two m-GRFNs $\widetilde{X}_1 \sim 0.5\widetilde{N}(-2, 0.1^2, 5) + 0.5\widetilde{N}(0, 0.1^2, 2)$ and $\widetilde{X}_2 \sim 0.5\widetilde{N}(0.1, 0.1^2, 2) + 0.5\widetilde{N}(2, 0.1^2, 0.1)$; (c): combined mGRFN $\widetilde{X}_1 \oplus \widetilde{X}_2$. For each RFN, we plot ten realizations (black dotted curves), the contour function (red curve), as well as the lower and upper cdfs (blue curves).

Proof. See Section Appendix C. $\square$

Example 9. *Figure 9 displays two m-GRFNs* $\widetilde{X}_1 \sim 0.5\widetilde{N}(-2, 0.1^2, 5) + 0.5\widetilde{N}(0, 0.1^2, 2)$ *and* $\widetilde{X}_2 \sim 0.5\widetilde{N}(0.1, 0.1^2, 2) + 0.5\widetilde{N}(2, 0.1^2, 0.1)$*, as well as their orthogonal sum*

$$\widetilde{X}_1 \oplus \widetilde{X}_2 \sim 0.020\widetilde{N}(-1.38, 0.0767^2, 7) + 0.426\widetilde{N}(0.05, 0.0707^2, 4) + \\ 0.197\widetilde{N}(-1.92, 0.0980^2, 5.1) + 0.357\widetilde{N}(0.0970, 0.0953^2, 2.1).$$

We note that $\widetilde{X}_1 \oplus \widetilde{X}_2$ *has four components but the first one, resulting from the combination of* $\widetilde{X}_{11}$ *and* $\widetilde{X}_{21}$*, has a small proportion because of high conflict. When combining a large number of m-GRFNs, the number of components grows exponentially. However, the combined m-GRFN can be approximated by a simpler one, using a technique that will be introduced in Section 4.4.*

Combination of m-GRFVs. Theorem 4 can easily be generalized to GRFVs, using the results in presented in [14, Proposition 13]. Basically, when combining two GRFVs, each component of the first GRFV is combined with each component of the second GRFV, and the proportions are adjusted based on degrees of conflict. This result is formally stated in the next theorem.

Theorem 5. *Let* $\widetilde{X}_1 \sim \sum_{k=1}^{K} \pi_{1k}\widetilde{N}(\boldsymbol{\mu}_{1k}, \boldsymbol{\Sigma}_{1k}^2, \boldsymbol{H}_{1k})$*, and* $\widetilde{X}_2 \sim \sum_{\ell=1}^{L} \pi_{2\ell}\widetilde{N}(\boldsymbol{\mu}_{2\ell}, \boldsymbol{\Sigma}_{2\ell}^2, \boldsymbol{H}_{2\ell})$ *be two p-dimensional m-GRFVs and assume that the precision matrices* $\boldsymbol{H}_{1k}$ *and* $\boldsymbol{H}_{2\ell}$ *are all positive definite. Then:*

1. *The orthogonal sum of $\widetilde{X}_1$ and $\widetilde{X}_2$ is an m-GRFV,*

$$\widetilde{X}_1 \oplus \widetilde{X}_2 \sim \sum_{k=1}^{K}\sum_{\ell=1}^{L} \widetilde{\pi}_{k\ell} \left[\widetilde{N}(\boldsymbol{\mu}_{1k}, \boldsymbol{\Sigma}^2_{1k}, \boldsymbol{H}_{1k}) \oplus \widetilde{N}(\boldsymbol{\mu}_{2\ell}, \boldsymbol{\Sigma}^2_{2\ell}, \boldsymbol{H}_{2\ell})\right],$$

   with

$$\widetilde{\pi}_{k\ell} = \frac{(1-\kappa_{k\ell})\pi_{1k}\pi_{2\ell}}{\sum_{k'\ell'}(1-\kappa_{k'\ell'})\pi_{1k'}\pi_{2\ell'}},$$

   where $\kappa_{k\ell}$ is the degree of conflict between $\widetilde{X}_{1k}$ and $\widetilde{X}_{2\ell}$ given by (A.1)*.*
2. *The degree of conflict between $\widetilde{X}_1$ and $\widetilde{X}_2$ is*

$$\kappa = \sum_{k=1}^{K}\sum_{\ell=1}^{L} \kappa_{k\ell}\pi_{1k}\pi_{2\ell}.$$

Proof. The theorem can be proved by following a similar line of reasoning as in the proofs of Lemma 3 and Theorem 4, using the results in Lemma 2 and Proposition 13 in [14]. □

4.4. Summarization of an m-GRFN

As already mentioned, the number of components grows exponentially when combining m-GRFNs: if all of N m-GRFNs have the same number K of components, the combined m-GRFN has, in general, K^N components. However, components of a combined m-GRFN resulting from the combination of highly conflicting GRFNs have a very small proportion. When combining many m-GRFNs, we can expect to obtain a large number of components, many of which will have a proportion close to zero. A similar problem occurs when combining a large number of mass functions, in which case we often observe a proliferation of focal sets with very small masses. Several strategies have been proposed for controlling the number of focal sets by transferring some masses to supersets, resulting in an outer approximation of the original mass function [22, 6, 19]. The simplest approach is to transfer all masses less than some threshold to the union of the corresponding focal sets [22]. Here, we do not have a notion of "union" for GRFNs that would result in a GRFN, but we can transfer the small proportions to a vacuous GRFN $\widetilde{X}_0 \sim \widetilde{N}(\mu, \sigma^2, 0)$ with arbitrary μ and σ^2. This procedure, formally described in Algorithm 1, yields a conservative approximation of the original m-GRFN. It is illustrated by the following example.

Example 10. *Consider the following three m-GRFNs, each one with three components:*

$$\widetilde{X}_1 \sim \frac{1}{3}\widetilde{N}(-2, (0.1)^2, 2) + \frac{1}{3}\widetilde{N}(0, (0.1)^2, 2) + \frac{1}{3}\widetilde{N}(2, (0.1)^2, 2),$$

$$\widetilde{X}_2 \sim \frac{1}{3}\widetilde{N}(-2.1, (0.5)^2, 2) + \frac{1}{3}\widetilde{N}(0.1, (0.5)^2, 2) + \frac{1}{3}\widetilde{N}(1/9, (0.5)^2, 2),$$

$$\widetilde{X}_3 \sim \frac{1}{3}\widetilde{N}(-1.9, (0.5)^2, 5) + \frac{1}{3}\widetilde{N}(-0.1, (0.5)^2, 5) + \frac{1}{3}\widetilde{N}(2.1, (0.5)^2, 5).$$

Algorithm 1 Summarization of a m-GRFN.

Input: m-GRFN $\widetilde{X} \sim \sum_{k=1}^{K} \pi_k \widetilde{N}(\mu_k, \sigma_k^2, h_k)$, threshold $\epsilon < 1$
  Reorder the components of $\widetilde{X}$ such that $\pi_{(1)} \leq \pi_{(2)} \leq \ldots \leq \pi_{(K)}$
  if $\pi_{(1)} \geq \epsilon$ **then**
    $\widetilde{X}' \leftarrow \widetilde{X}$
  else
    Find the largest k_0 such that $\sum_{k=1}^{k_0} \pi_{(k)} \leq \epsilon$
    $\pi_0 \leftarrow \sum_{k=1}^{k_0} \pi_{(k)}$
    $\widetilde{X}' \leftarrow \pi_0 \widetilde{N}(0, 1, 0) + \sum_{k=k_0+1}^{K} \pi_{(k)} \widetilde{N}(\mu_{(k)}, \sigma_{(k)}^2, h_{(k)})$
  end if
Output: Approximate m-GRFN $\widetilde{X}'$

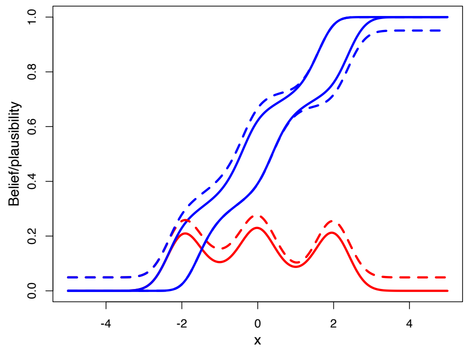


Figure 10: Contour functions (red curves), and lower/upper cdfs (blue curves) of m-GRFN $\widetilde{X}$ with 27 components in Example 10 (solid lines) and its approximation $\widetilde{X}'$ with only 13 components (broken lines).

The combined m-GRFN $\widetilde{X} = \widetilde{X}_1 \oplus \widetilde{X}_2 \oplus \widetilde{X}_3$ *has 27 components. Summarizing* $\widetilde{X}$ *with a threshold* $\epsilon = 0.05$ *yields an n-GRFN with 13 components. The contour function as well as the lower and upper cdfs of* $\widetilde{X}$ *and* $\widetilde{X}'$ *are shown in Figure 10. As we can see,* $\widetilde{X}'$ *is a good conservative approximation of* $\widetilde{X}$*, while having less than half as many components.*

4.5. Mixtures of transformed GRFNs

Definition. The ideas developed in Section 3 and in the current section can naturally be combined to define mixtures of transformed fuzzy numbers. Let us consider a one-to-one mapping ψ from $\mathbb{R}$ to $\Lambda \subseteq \mathbb{R}$ and, as in Section 4.1, an m-GRFN $\widetilde{X} \sim \sum_{k=1}^{K} \pi_k \widetilde{N}(\mu_k, \sigma_k^2, h_k)$, originating from a pair of random variables (M, Z) from a probability space $(\Omega, \Sigma_\Omega, P)$ to $\mathbb{R} \times \{1, \ldots, K\}$, such that $P(Z = k) = \pi_k$, $k = 1, \ldots, K$, and $M \mid (Z = k) \sim N(\mu_k, \sigma_k^2)$, $k = 1, \ldots, K$. From Proposition 2, the composed mapping

$$\widetilde{Y}(\omega) = (\widetilde{\psi} \circ \widetilde{X})(\omega) = \widetilde{\psi}\left[\mathsf{GFN}(M(\omega), \prod_{k=1}^{K} h_k^{Z_k(\omega)})\right],$$

where $Z_k(\omega) = I(Z(\omega) = k)$, is a RFS. Conditionally on $Z = k$, $\widetilde{Y}$ is a t-GRFN with mean μ_k, variance σ_k^2 and precision h_k:

$$\widetilde{Y} \mid (Z = k) \sim T\widetilde{N}(\mu_k, \sigma_k^2, h_k, \psi^{-1}).$$

We say that $\widetilde{Y}$ is a mixture of transformed GRFNs (mt-GRFN) and we write

$$\widetilde{Y} \sim \sum_{k=1}^{K} \pi_k T\widetilde{N}(\mu_k, \sigma_k^2, h_k, \psi^{-1}).$$

In a similar way, we can define a mixture of transformed GRFVs (mt-GRFV) as the transformation of an m-GRFV $\widetilde{X} \sim \sum_{k=1}^{K} \pi_k \widetilde{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_k)$ by a one-to-one mapping ψ from $\mathbb{R}^p$ to $\Lambda \subseteq \mathbb{R}^p$ and write

$$\widetilde{Y} = \widetilde{\psi} \circ \widetilde{X} \sim \sum_{k=1}^{K} \pi_k T\widetilde{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_k, \psi^{-1}).$$

Properties. The properties of mt-GRFNs can be derived directly from those of m-GRFNs and t-GRFNs. Some of them are described in the following propositions. Similar properties hold for mt-GRFVs.

Proposition 6. *Let $\widetilde{Y} \sim \sum_{k=1}^{K} \pi_k T\widetilde{N}(\mu_k, \sigma_k^2, h_k, \psi^{-1})$ be an mt-GRFN, where ψ is a one-to-one mapping from $\mathbb{R}$ to $\Lambda \subseteq \mathbb{R}$. For any event $A \in \Sigma_\Lambda$,*

$$Bel(A) = \sum_{k=1}^{K} \pi_k Bel_{\widetilde{X}_k}(\psi^{-1}(A)), \quad \text{and} \quad Pl(A) = \sum_{k=1}^{K} \pi_k Pl_{\widetilde{X}_k}(\psi^{-1}(A)),$$

where $\widetilde{X}_k \sim \widetilde{N}(\mu_k, \sigma_k^2, h_k)$. In particular the contour function of $\widetilde{Y}$ is

$$pl_{\widetilde{Y}}(y) = \sum_{k=1}^{K} \frac{\pi_k}{\sqrt{1 + h_k \sigma_k^2}} \exp\left(-\frac{h_k(\psi^{-1}(y) - \mu_k)^2}{2(1 + h_k \sigma_k^2)}\right).$$

Proof. Immediate from Theorem 1, Theorem 3 and Corollary 1. □

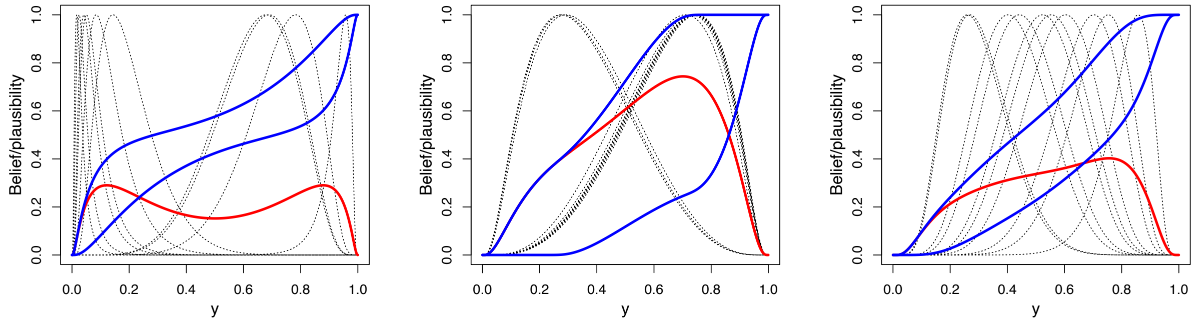
Proposition 7. *Let $\widetilde{Y}_1 \sim \sum_{k=1}^{K} \pi_{1k} T\widetilde{N}(\mu_{1k}, \sigma_{1k}^2, h_{1k}, \psi^{-1})$ and $\widetilde{Y}_2 \sim \sum_{k=1}^{L} \pi_{2\ell} T\widetilde{N}(\mu_{2\ell}, \sigma_{2\ell}^2, h_{2\ell}, \psi^{-1})$ be two mt-GRFNs, where ψ is a one-to-one mapping from $\mathbb{R}$ to $\Lambda \subseteq \mathbb{R}$. Their orthogonal sum is the mt-GRFN*

$$\widetilde{Y}_1 \oplus \widetilde{Y}_2 \sim \sum_{k=1}^{K} \sum_{k=1}^{L} \widetilde{\pi}_{k\ell} T\widetilde{N}(\widetilde{\mu}_{k\ell}, \widetilde{\sigma}_{k\ell}^2, h_{1k} + h_{2\ell}, \psi^{-1}),$$

where $\widetilde{\mu}_{k\ell}$ and $\widetilde{\sigma}_{k\ell}^2$ are the mean and variance of the orthogonal sum $\widetilde{N}(\mu_{1k}, \sigma_{1k}^2, h_{1k}) \oplus \widetilde{N}(\mu_{2\ell}, \sigma_{2\ell}^2, h_{2\ell})$ given by (11), and

$$\widetilde{\pi}_{k\ell} = \frac{(1 - \kappa_{k\ell})\pi_{1k}\pi_{2\ell}}{\sum_{k'\ell'}(1 - \kappa_{k'\ell'})\pi_{1k'}\pi_{2\ell'}},$$

where $\kappa_{k\ell}$ is the degree of conflict between $\widetilde{N}(\mu_{1k}, \sigma_{1k}^2, h_{1k})$ and $\widetilde{N}(\mu_{2\ell}, \sigma_{2\ell}^2, h_{2\ell})$ given by Proposition 1.



(a) (b) (c)

Figure 11: (a) and (b): From left to right, two mixtures of logit-normal RFNs $\widetilde{Y}_1 \sim 0.5T\widetilde{N}(2, 1, 2, \mathsf{logit}) + 0.5T\widetilde{N}(-2, 1, 2, \mathsf{logit})$ and $\widetilde{Y}_2 \sim 0.3T\widetilde{N}(-1, 0.1^2, 1, \mathsf{logit}) + 0.7T\widetilde{N}(1, 0.1^2, 1, \mathsf{logit})$; (c): orthogonal sum $\widetilde{Y}_1 \oplus \widetilde{Y}_2$. For each RFN, we plot ten realizations (black dotted curves), the contour function (red curve), as well as the lower and upper cdfs (blue curves).

Proof. Immediate from Theorems 2 and 4. $\square$

Example 11. *Figure 11 displays two mixtures of logit-normal RFNs* $\widetilde{Y}_1 \sim 0.5T\widetilde{N}(2, 1, 2, \mathit{logit}) + 0.5T\widetilde{N}(-2, 1, 2, \mathit{logit})$ *and* $\widetilde{Y}_2 \sim 0.3T\widetilde{N}(-1, 0.1^2, 1, \mathit{logit}) + 0.7T\widetilde{N}(1, 0.1^2, 1, \mathit{logit})$*, as well as their orthogonal sum* $\widetilde{Y}_1 \oplus \widetilde{Y}_2$*.*

5. Applications

In this section, we will discuss two applications of the models introduced in this paper. Belief elicitation will first be addressed in Section 5.1. The use of m-GRFNs to represent weak prior knowledge in generalized Bayesian inference will then be discussed in Section 5.2.

5.1. Elicitation

According to Garthwaite et al. [17], "elicitation is the process of formulating a person's knowledge and beliefs about one or more uncertain quantities into a (joint) probability distribution for those quantities". There is, however, no reason to limit oneself to probability distributions for representing a person's beliefs. In this section, we briefly discuss the application of the RFN models introduced in this paper to represent expert beliefs about a numerical quantity. It is clear that an in-depth treatment of this topic would require to delve into complex methodological issues arising when interviewing human experts. Here, we will only be concerned with the use of already obtained plausibility assessments to fit the parameters of a RFN.

For single-expert probabilistic elicitation, the two main approaches are the *fixed interval method*, in which the expert is asked to give his subjective probabilities for some fixed intervals, and the *variable interval method* in which the expert is invited to provide points corresponding to specified percentiles of his subjective probability distribution. Assuming the latter approach is used, a sequence of questions for eliciting an expert's beliefs about a numerical quantity X could be the following:

1. What is the most plausible value m_0 of X?
2. Given two numbers $0 < \alpha < \beta < 1$ (such as, e.g., $\alpha = 0.1$ and $\beta = 0.5$), give values x_α and x_β such that $Pl(X \leq x_\alpha) = \alpha$ and $Pl(X \leq x_\beta) = \beta$.
3. Give values x'_α and x'_β such that $Pl(X > x'_\alpha) = \alpha$ and $Pl(X > x'_\beta) = \beta$.

This procedure yields a maximum-plausibility value m_0 and the plausibility degrees of four intervals $(-\infty, x_\alpha]$, $(-\infty, x_\beta]$, $[x'_\alpha, +\infty)$ and $[x'_\beta, +\infty)$. Whatever the details of the elicitation procedure, we can assume that we have obtained m_0 and the plausibilities $pl_1, \ldots, pl_n$ of n real intervals $I_1, \ldots, I_n$.

Several parametric families of RFN proposed in Sections 3.3 and 4 could be fitted to such data. As the number n of intervals will typically be small, simpler models should be preferred. Mixture models may not be the most suitable because they depend on many parameters (at least seven for the simplest two-component case) and they can yield multimodal contour functions. Let us, thus, consider a parameterized family of t-GRFNs $\widetilde{X}_\theta \sim T\widetilde{N}(\mu, \sigma^2, h, \psi_\eta^{-1})$, where η is a vector of parameters for the transformation function ψ, and $\theta = (\mu, \sigma^2, h, \eta)$ is the vector of all parameters. We can then identify θ by minimizing the following mean squared error function:

$$\mathsf{MSE}(\theta) = \sum_{i=1}^{n} (Pl_{\widetilde{X}_\theta}(I_i) - pl_i)^2. \tag{26}$$

subject to the constraint $\psi_\eta(\mu) = m_0$. The following example illustrate this approach with the family of t-GRFNs defined by transformation $\psi_\eta = \psi_{\xi,\lambda}$ given by (20).

Example 12. *Assume that an expert gives us $m_0 = 1$ as the most plausible value of X, and the following plausibility assessments:*

$$Pl(X \leq -7) = 0.1, \quad Pl(X \leq -1) = 0.5, \quad Pl(X > 2) = 0.5, \quad Pl(X > 5) = 0.1.$$

The constraint $\psi_{\xi,\lambda}(\mu) = m_0$ gives us $\mu = \sinh^{-1}\left(\frac{m_0 - \xi}{\lambda}\right)$. Substituting μ by its expression as a function of ξ and λ and minimizing (26) w.r.t. σ^2, h, ξ and λ yields the following estimates:

$$\widehat{\sigma}^2 = 0.73, \quad \widehat{h} = 10.90, \quad \widehat{\xi} = 3.37, \quad \widehat{\lambda} = 2.71,$$

and $\widehat{\mu} = -0.79$. Figure 12 shows of the lower and upper cdf of the fitted t-GRFN, as well as its contour function. By comparison, we also show the corresponding functions for a GRFN fitted on the same data. Obviously, a GRFN is a poor fit given the asymmetry of the plausibility assessments.

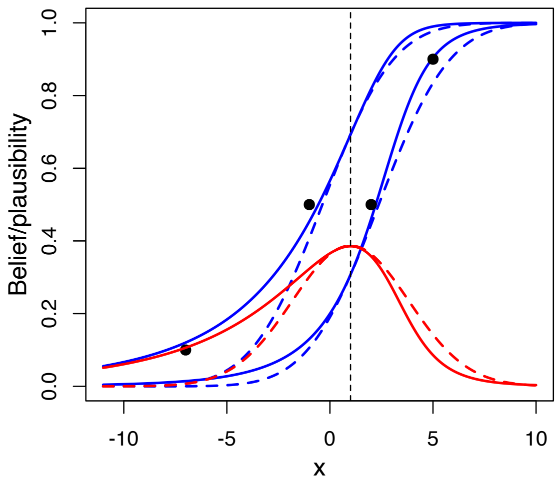


Figure 12: Contour functions (red curves), and lower/upper cdfs (blue curves) of the fitted t-GRFN $\widetilde{X}_{\widehat{\theta}} \sim T\widetilde{N}(\widehat{\mu}, \widehat{\sigma}^2, \widehat{h}, \psi^{-1}_{\widehat{\xi},\widehat{\lambda}})$ of Example 12 (solid lines) and a fitted GRFN (broken lines). The data points are shown as black dots, and the vertical broken lines marks the most plausible value m_0.

5.2. Generalized Bayesian inference

Let us consider a statistical model in which observed data X are drawn randomly from a probability distribution P_θ, where $\theta \in \Theta$ is some unknown parameter. Having observed a realization x of X, we define the likelihood function as the mapping

$$\begin{aligned} L(\cdot; x) : \Theta &\to [0, +\infty) \\ \theta &\mapsto L(\theta; x) = f(x; \theta), \end{aligned}$$

where $f(x;\theta)$ denotes the probability mass or density function of X. Assuming $\sup_{\theta\in\Theta} L(\theta;x) < +\infty$, the relative likelihood function can be defined as

$$\begin{aligned} \widetilde{L}_x : \Theta &\to [0, +\infty) \\ \theta &\mapsto \widetilde{L}_x(\theta) = \frac{L(\theta; x)}{\sup_{\theta'\in\Theta} L(\theta'; x)}. \end{aligned}$$

In [8], we showed that $\widetilde{L}_x$ can be interpreted as the membership function of a fuzzy subset of Θ (the fuzzy subset of "likely" values of θ after observing x). Equivalently, $\widetilde{L}_x$ can be seen as a constant RFS. As shown in [8], $\widetilde{L}_x$ is, in some sense, the least committed RFS verifying the following two requirements:

1. Compatibility with Bayesian inference: let P_0 be a prior probability measure on Θ; combining it with $\widetilde{L}_x$ using the product-intersection rule yields the Bayesian posterior distribution $P(\cdot|x)$:
$$P_0 \oplus \widetilde{L}_x = P(\cdot|x). \tag{27}$$

2. Combination of independent observations: if x and y are realizations of two independent observations X and Y, then the RFS induced by (x, y) is the orthogonal sum of the RFS induced by x, and that induced by y:

$$\widetilde{L}_x \oplus \widetilde{L}_y = \widetilde{L}_{x,y}.$$

From the perspective of a general theory of epistemic random fuzzy sets, there is no reason to limit oneself to probability distributions for representing prior information. Indeed, prior knowledge is often vague and the relevance of representing it by precise probabilities is questionable. Equation (27) can, thus, be generalized as

$$\widetilde{\theta}_0 \oplus \widetilde{L}_x = \widetilde{\theta}_x, \tag{28}$$

where $\widetilde{\theta}_0$ is a RFS representing *weak prior information*, and $\widetilde{\theta}_x$ is posterior RFS resulting from the combination of prior information with observations. Considering the case where $\Theta \subseteq \mathbb{R}$, (28) lends itself to easy computation if $\widetilde{L}_x$ is a GFN: in this case, modeling prior information as a GRFN $\widetilde{\theta}_0 \sim \widetilde{N}(\mu_0, \sigma_0^2, h_0)$ or, more generally, an m-GRFN $\widetilde{\theta}_0 \sim \sum_{k=1}^K \pi_{0k} \widetilde{N}(\mu_{0k}, \sigma_{0k}^2, h_{0k})$ will result, respectively, in a posterior GRFN $\widetilde{\theta}_x \sim \widetilde{N}(\mu_x, \sigma_x^2, h_x)$ or a posterior m-GRFN $\widetilde{\theta}_x \sim \sum_{k=1}^K \pi_{xk} \widetilde{N}(\mu_{xk}, \sigma_{xk}^2, h_{xk})$.

Except in a few simple cases, the relative likelihood function is not exactly a GFN, but this model can be used as an approximation [7]. Indeed, a Taylor series expansion of $\log \widetilde{L}_x(\theta)$ about the maximum likelihood estimate (MLE) $\widehat{\theta}$ up to the second order gives us [28, p. 33]:

$$\log \widetilde{L}_x(\theta) = \log \widetilde{L}_x(\widehat{\theta}) + (\theta - \widehat{\theta}) \left. \frac{\partial \log \widetilde{L}_x(\theta)}{\partial \theta} \right|_{\theta=\widehat{\theta}} + \frac{1}{2}(\theta - \widehat{\theta})^2 \left. \frac{\partial^2 \log \widetilde{L}_x(\theta)}{\partial \theta^2} \right|_{\theta=\widehat{\theta}} + \cdots$$

The first term on the right-hand is equal to zero by definition, and the second term is zero in the usual case where $\widehat{\theta}$ is a stationary point of the likelihood function. Neglecting the higher-order terms, we get the approximation

$$\widetilde{L}_x(\theta) \approx \exp\left[-\frac{1}{2} I(\widehat{\theta}; x)(\theta - \widehat{\theta})^2\right], \tag{29}$$

i.e., $\widetilde{L}_x \approx \mathsf{GFN}(\widehat{\theta}, I(\widehat{\theta}; x))$, where the precision $I(\widehat{\theta}; x)$ is the observed Fisher information

$$I(\widehat{\theta}; x) = -\left. \frac{\partial^2 \log \widetilde{L}_x(\theta)}{\partial \theta^2} \right|_{\theta=\widehat{\theta}} = -\left. \frac{\partial^2 \log L(\theta; x)}{\partial \theta^2} \right|_{\theta=\widehat{\theta}}.$$

In the multidimensional case where $\Theta \subseteq \mathbb{R}^p$, the same line of reasoning as above yields $\widetilde{L}_x \approx \mathsf{GFV}(\widehat{\theta}, I(\widehat{\theta}; x))$, where $I(\widehat{\theta}; x)$ is the observed information matrix.

As noted in [28], the quality of the normality assumption can sometimes be improved for small samples by applying some transformation to the parameter. Let $\theta = \psi(\delta)$ for some one-to-one differentiable mapping ψ and alternative parameter δ, and assume that the relative likelihood as a function of δ, $\widetilde{L}_x^{(\delta)}(\delta)$, is approximately Gaussian, $\widetilde{L}_x^{(\delta)}(\delta) \approx \mathsf{GFN}(\widehat{\delta}, I(\widehat{\delta}, x))$. Then, the likelihood as a function of θ is $\widetilde{L}_x^{(\theta)}(\theta) = \widetilde{\psi}(\widetilde{L}_x^{(\delta)})(\theta) = \widetilde{L}_x^{(\delta)}(\psi^{-1}(\theta))$, and it can be approximated by

$$\widetilde{L}_x^{(\delta)}(\psi^{-1}(\theta)) \approx \exp\left[-\frac{1}{2}I(\widehat{\delta};x)\left(\psi^{-1}(\theta) - \psi^{-1}(\widehat{\theta})\right)^2\right] \tag{30a}$$

$$\approx \exp\left[-\frac{1}{2}I(\widehat{\theta};x)[\psi'(\widehat{\delta})]^2\left(\psi^{-1}(\theta) - \psi^{-1}(\widehat{\theta})\right)^2\right], \tag{30b}$$

where we have used the equality $I(\widehat{\delta};x) = I(\widehat{\theta};x)[\psi'(\widehat{\delta})]^2$ (see [28, p. 35]), and ψ' denotes the first derivative of ψ. The fuzzy set $\widetilde{L}_x^{(\theta)}$ can be combined with a t-GRFN prior $\widetilde{\theta}_0 \sim \widetilde{TN}(\mu_0, \sigma_0^2, h_0, \psi^{-1})$ or, more generally, an mt-GRFN $\widetilde{\theta}_0 \sim \sum_{k=1}^K \pi_{0k}\widetilde{TN}(\mu_{0k}, \sigma_{0k}^2, h_{0k}, \psi^{-1})$ to obtain, respectively, a posterior t-GRFN $\widetilde{\theta}_x \sim \widetilde{TN}(\mu_x, \sigma_x^2, h_x, \psi^{-1})$, or a posterior mt-GRFN

$$\widetilde{\theta}_x \sim \sum_{k=1}^K \pi_{xk}\widetilde{TN}(\mu_{xk}, \sigma_{xk}^2, h_{xk}, \psi^{-1}).$$

Example 13. *Consider an iid sample $x = (x_1, \ldots, x_n)$ from a Poisson distribution with mean θ. Let $t = \sum_{i=1}^n x_i$. The likelihood function is*

$$L(\theta;x) = \prod_{i=1}^n \frac{\theta^{x_i}\exp(-\theta)}{x_i!} = \frac{\theta^t \exp(-n\theta)}{\prod x_i!},$$

and the MLE of θ is $\widehat{\theta} = t/n$, which gives us the following expression for the relative likelihood:

$$\widetilde{L}_x(\theta) = \left(\frac{\theta}{\widehat{\theta}}\right)^t \exp\left[n(\widehat{\theta} - \theta)\right].$$

It can easily be shown that $I(\widehat{\theta};x) = n/\widehat{\theta}$, so that (29) *yields*

$$\widetilde{L}_x(\theta) \approx \exp\left[-\frac{n}{2\widehat{\theta}}(\theta - \widehat{\theta})^2\right] = \exp\left[-\frac{n}{2}\left(\frac{\theta}{\widehat{\theta}} - 1\right)^2\right]. \tag{31}$$

Alternatively, let $\theta = \exp(\delta)$. Approximation (30) *gives us*

$$\widetilde{L}_x(\theta) \approx \exp\left[-\frac{n}{2\widehat{\theta}}\exp(2\widehat{\theta})(\log\theta - \log\widehat{\theta})^2\right]. \tag{32}$$

Figure 13 displays the exact relative likelihood $\widetilde{L}_x(\theta)$ for $n = 10$ and $t = 30$, as well as its approximations (31) *and* (32)*. As we can see, the lognormal approximation* (32) *is more accurate, and it takes into account the positivity of θ.*

Figure 14a shows two priors: a lognormal Bayesian prior $\widetilde{\theta}_0 \sim \widetilde{TN}(0.5, 0.1, +\infty, \log)$, represented by its cdf (blue dotted line) and a weaker lognormal t-GRFN prior

$$\widetilde{\theta}_0' \sim \widetilde{TN}(0.5, 0.1, 10, \log),$$

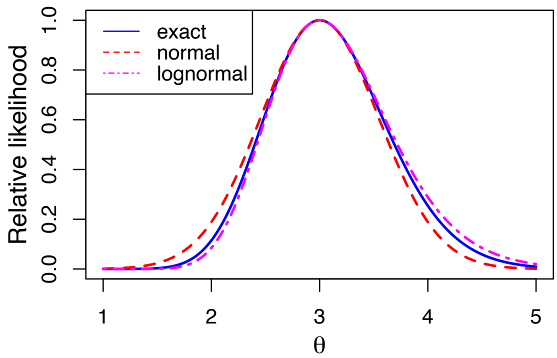


Figure 13: Relative likelihood $\widetilde{L}_x(\theta)$ and its normal and lognormal approximations for an iid sample for the Poisson distribution with mean θ, for $n = 10$ observations and $t = 30$.

represented by its contour function (red solid line), its upper and lower cdfs (blues solid lines) and 10 realizations (black dotted lines). We also show the lognormal approximation to the relative likelihood as a red broken line. Figure 14b shows the corresponding lognormal t-GRFN posteriors $\widetilde{\theta}_x = \widetilde{\theta}_0 \oplus \widetilde{L}_x$ *(equal to the posterior probability distribution) and* $\widetilde{\theta}'_x = \widetilde{\theta}'_0 \oplus \widetilde{L}_x$*, as well as the approximated relative likelihood. As expected, the posterior lognormal t-GRFN corresponding to the weaker prior is more imprecise, and closer to the relative likelihood. It is clear from Figure 14b that our approach is not a robust Bayes method: although the Bayes prior cdf is comprised between the lower and upper weak prior cdfs, this is not the case for the posteriors. In the extreme situation where the prior is vacuous,* $\widetilde{\theta}^v_0 \sim T\widetilde{N}(0, 1, 0, \log)$*, the posterior is the relative likelihood, as* $\widetilde{\theta}^v_0 \oplus \widetilde{L}_x = \widetilde{L}_x$*. This is in contrast with the robust Bayes approach, which yields a (useless) vacuous posterior with the same vacuous prior information.*

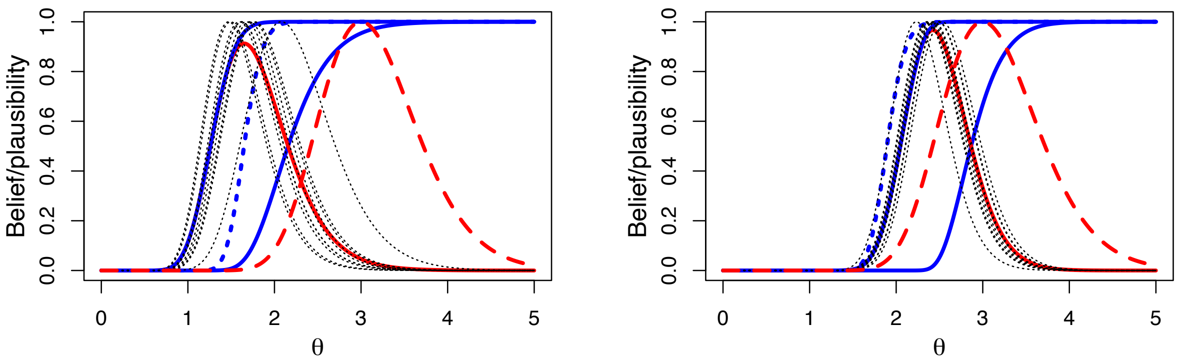
To illustrate the possibility of taking into account more complex prior information, we show a bimodal mt-GRFN prior

$$\widetilde{\theta}_0 \sim 0.5\, T\widetilde{N}(0.5, 0.1, 30, \log) + 0.5\, T\widetilde{N}(1.5, 0.01, 20, \log)$$

in Figure 15a, and the corresponding posterior mt-GRFN

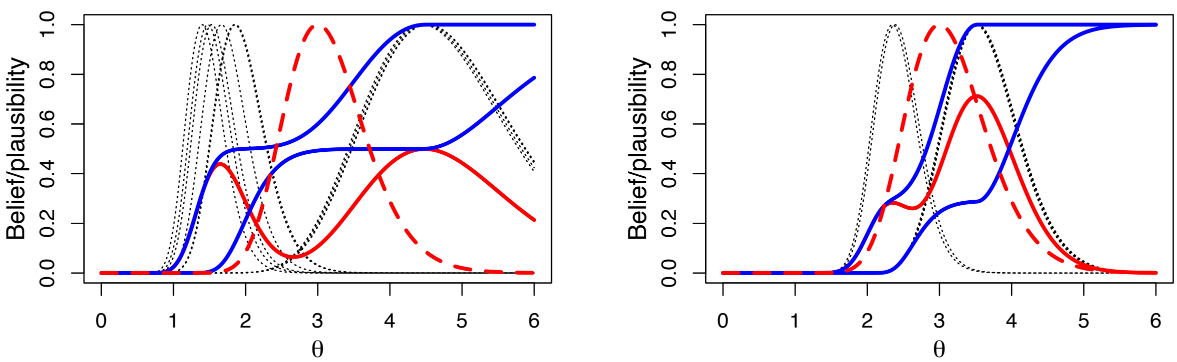
$$\widetilde{\theta}_x = \widetilde{\theta}_0 \oplus \widetilde{L}_x = 0.29\, T\widetilde{N}(0.84, 0.047, 60, \log) + 0.71\, T\widetilde{N}(1.26, 0.0040, 50, \log).$$

in Figure 15b. In practice, the prior RFS could be elicited by a method similar to that proposed in Section 5.1.



(a) (b)

Figure 14: (a) Lognormal Bayesian prior $\widetilde{\theta}_0 \sim T\widetilde{N}(0.5, 0.1, +\infty, \log)$ (blue dotted line), lognormal t-GRFN prior $\widetilde{\theta}_0' \sim T\widetilde{N}(0.5, 0.1, 10, \log)$ (contour function: red solid line, upper and lower cdfs: blues solid lines, and 10 realizations: black dotted lines), and lognormal approximation to the relative likelihood $\widetilde{L}_x$ (red broken line); (b): Bayesian lognormal posterior distribution $\widetilde{\theta}_x = \widetilde{\theta}_0 \oplus \widetilde{L}_x$ (blue dotted line), lognormal t-GRFN posterior $\widetilde{\theta}_x' = \widetilde{\theta}_0' \oplus \widetilde{L}_x$ (contour function: red solid line, upper and lower cdfs: blues solid lines, and 10 realizations: black dotted lines), and lognormal approximation to the relative likelihood (red broken line).



(a) (b)

Figure 15: (a) Lognormal mt-GRFN prior $\widetilde{\theta}_0 \sim 0.5T\widetilde{N}(0.5, 0.1, 30, \log) + 0.5T\widetilde{N}(1.5, 0.01, 20, \log)$ (contour function: red solid line, upper and lower cdfs: blues solid lines, and 10 realizations: black dotted lines), and lognormal approximation to the relative likelihood (red broken line); (b): mt-GRFN posterior $\widetilde{\theta}_x' = \widetilde{\theta}_0' \oplus \widetilde{L}_x$ (contour function: red solid line, upper and lower cdfs: blues solid lines, and 10 realizations: black dotted lines), and lognormal approximation to the relative likelihood (red broken line).

6. Conclusions

Until recently, the application of evidential reasoning to problems involving continuous variables has been limited due to the lack of practical models of belief functions in $\mathbb{R}^p$, $p \geq 1$ allowing for easy computation and combination. In [14], we have proposed a solution to this problem by considering "epistemic random fuzzy sets" as the basic construct, from which belief functions can be derived. In this framework, random fuzzy sets represent items of evidence and can be combined by a generalized product-intersection rule extending both the normalized product of possibility theory, and Dempster's rule of Dempster-Shafer theory. In this framework, we have introduced in [14] simple models of random fuzzy sets based on the normal distribution. The proposed GRFN model and its multidimensional generalization define parameterized families of belief functions in $\mathbb{R}^p$ that can be easily used in calculations and combined using simple mathematical formulas. However, GRFNs, indexed by three parameters (mean, standard deviation and precision) are not flexible enough to represent the wide variety of belief functions needed in applications.

In this paper, we have presented two extensions of the GRFN model. The first one consists in transforming a RFS $\widetilde{X}$ defined in $\mathbb{R}^p$ by a bijective mapping ψ from $\mathbb{R}^p$ to $\Lambda \subseteq \mathbb{R}^p$. Such a transformation allows us to define, e.g., belief functions on real intervals $[a, b]$ with $-\infty \leq a < b \leq +\infty$, or on the probability simplex $\mathcal{S}_p$. The second extension consists in considering mixtures of (transformed) GRFNs, which make it possible to define belief functions of arbitrary "shape" and complexity. The rich families of generalized GRFNs introduced in this paper are closed under the product-intersection rule and can be used for a variety of evidential reasoning tasks involving real variables.

We have discussed two important applications: the elicitation of expert beliefs about numerical quantities, and generalized Bayesian inference with weak priors defined as generalized GRFNs. Elicitation has been widely studied in the probabilistic context, much less in other settings such as possibility theory or belief functions. The new models introduced in this paper could be tested as representations of expert beliefs in real experiments. As for statistical inference, the epistemic random fuzzy set perspective provides a simple and consistent model in which data and prior knowledge are treated symmetrically as pieces of evidence represented by random fuzzy sets. The parametric families of random fuzzy numbers introduced in this paper can be used with a variety of statistical models without resorting to Monte Carlo simulation. It would be interesting to compare this approach with alternative methods such as, e.g., robust Bayesian analysis, both conceptually and practically.

As another important application of the models introduced in this paper, we can mention the quantification of prediction uncertainty in machine learning, for regression problems characterized by asymmetric or heavy-tailed noise distributions, or in which target variables are subject to some constraints. The EVREG model, an evidential neural network model proposed in [9, 11] for "classical" regression problems, models prediction uncertainty using GRFNs. It could be adapted, for instance, to compositional regression tasks (in which target variables are proportions) by transforming its outputs to logistic normal random fuzzy vectors. This and other research directions will be investigated in future publications.

References

[1] J. Aitchison and S. M. Shen. Logistic-normal distributions: Some properties and uses. *Biometrika*, 67(2):261–272, 1980.

[2] M. A. Aljarrah, C. Lee, and F. Famoye. On generating T-X family of distributions using quantile functions. *Journal of Statistical Distributions and Applications*, 1(1):2, 2014.

[3] I. Couso and L. Sánchez. Higher order models for fuzzy random variables. *Fuzzy Sets and Systems*, 159(3):237–258, 2008.

[4] I. Couso and L. Sánchez. Upper and lower probabilities induced by a fuzzy random variable. *Fuzzy Sets and Systems*, 165(1):1–23, 2011.

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

[6] T. Denœux. Inner and outer approximation of belief structures using a hierarchical clustering approach. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(4):437–460, 2001.

[7] T. Denœux. Likelihood-based belief function: justification and some extensions to low-quality data. *International Journal of Approximate Reasoning*, 55(7):1535–1547, 2014.

[8] 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.

[9] T. Denœux. An evidential neural network model for regression based on random fuzzy numbers. In S. Le Hégarat-Mascle, I. Bloch, and E. Aldea, editors, *Belief Functions: Theory and Applications*, pages 57–66, Cham, 2022. Springer International Publishing.

[10] T. Denœux. Belief functions on the real line defined by transformed Gaussian random fuzzy numbers. In *2023 IEEE International Conference on Fuzzy Systems*, Songdo Incheon, Korea, August 2023. IEEE.

[11] 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.

[12] T. Denœux, D. Dubois, and H. Prade. Representations of uncertainty in artificial intelligence: Beyond probability and possibility. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of Artificial Intelligence Research*, volume 1, chapter 4, pages 119–150. Springer Verlag, 2020.

[13] T. Denœux, D. Dubois, and H. Prade. Representations of uncertainty in artificial intelligence: Probability and possibility. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of Artificial Intelligence Research*, volume 1, chapter 3, pages 69–117. Springer Verlag, 2020.

[14] 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.

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

[16] D. Dubois, H. Prade, and R. Yager. Merging fuzzy information. In J. C. Bezdek, D. Dubois, and H. Prade, editors, *Fuzzy sets in approximate reasoning and information systems*, pages 335–401. Kluwer Academic Publishers, Boston, 1999.

[17] P. H. Garthwaite, J. B. Kadane, and A. O'Hagan. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470):680–700, 2005.

[18] M. A. Gil, M. López-Díaz, and D. A. Ralescu. Overview on the development of fuzzy random variables. *Fuzzy Sets and Systems*, 157(19):2546–2557, 2006.

[19] R. Haenni and N. Lehmann. Resource bounded and anytime approximation of belief function computations. *International Journal of Approximate Reasoning*, 31(1):103–154, 2002.

[20] N. L. Johnson. Systems of frequency curves generated by methods of translation. *Biometrika*, 36(1/2):149–176, 1949.

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

[22] J. D. Lowrance, T. D. Garvey, and T. M. Strat. A framework for evidential-reasoning systems. In T. K. et al., editor, *Proceedings of AAAI'86*, volume 2, pages 896–903, Philadelphia, August 1986. AAAI.

[23] G. J. McLachlan and D. Peel. *Finite Mixture Models.* Wiley, New York, 2000.

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

[25] M. L. Puri and D. A. Ralescu. Fuzzy random variables. *Journal of Mathematical Analysis and Applications*, 114(2):409–422, 1986.
[26] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, Princeton, N.J., 1976.
[27] G. Shafer. Allocations of probability. *Annals of Probability*, 7(5):827–839, 1979.
[28] D. A. Sprott. *Statistical Inference in Science.* Springer-Verlag, Berlin, 2000.
[29] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning –I. *Information Sciences*, 8:199–249, 1975.
[30] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.
[31] L. A. Zadeh. Fuzzy sets and information granularity. In M. M. Gupta, R. K. Ragade, and R. R. Yager, editors, *Advances in Fuzzy Sets Theory and Applications*, pages 3–18. North-Holland, Amsterdam, 1979.

Appendix A. Combination of Gaussian Random Fuzzy vectors

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

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

with

$$\boldsymbol{H}_{12} = \boldsymbol{H}_1 + \boldsymbol{H}_2, \quad \widetilde{\boldsymbol{\mu}}_{12} = \boldsymbol{A}\widetilde{\boldsymbol{\mu}}, \quad \text{and} \quad \widetilde{\boldsymbol{\Sigma}}_{12} = \boldsymbol{A}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{A}^T,$$

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

$$\boldsymbol{A} = \boldsymbol{H}_{12}^{-1} \begin{pmatrix} \boldsymbol{H}_1 & \boldsymbol{H}_2 \end{pmatrix},$$

$$\widetilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Sigma}_1^{-1} + \overline{\boldsymbol{H}} & -\overline{\boldsymbol{H}} \\ -\overline{\boldsymbol{H}} & \boldsymbol{\Sigma}_2^{-1} + \overline{\boldsymbol{H}} \end{pmatrix}^{-1},$$

$$\widetilde{\boldsymbol{\mu}} = \begin{pmatrix} \overline{\boldsymbol{H}}^{-1}\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{I}_p & -\boldsymbol{I}_p \\ -\boldsymbol{I}_p & \overline{\boldsymbol{H}}^{-1}\boldsymbol{\Sigma}_2^{-1} + \boldsymbol{I}_p \end{pmatrix}^{-1} \begin{pmatrix} \overline{\boldsymbol{H}}^{-1}\boldsymbol{\Sigma}_1^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \overline{\boldsymbol{H}}^{-1}\boldsymbol{\Sigma}_2^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix},$$

and

$$\overline{\boldsymbol{H}} = (\boldsymbol{H}_1^{-1} + \boldsymbol{H}_2^{-1})^{-1}.$$

Furthermore, the degree of conflict between $\widetilde{X}_1$ and $\widetilde{X}_2$ is

$$\kappa = 1 - \sqrt{\frac{|\widetilde{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}_1||\boldsymbol{\Sigma}_2|}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 - \widetilde{\boldsymbol{\mu}}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\mu}}\right]\right\}. \tag{A.1}$$

Appendix B. Proof of Proposition 3

The membership function of $\widetilde{Y}(\omega) \sim T\widetilde{N}(\mu, \sigma^2, h, \log)$ is

$$\widetilde{Y}(\omega)(y) = \exp\left(-\frac{h}{2}(\log y - M(\omega))^2\right),$$

with $M \sim N(\mu, \sigma^2)$. The α-cut of $\widetilde{Y}(\omega)$ is

$$ {}^{\alpha}\widetilde{Y}(\omega) = \left[\exp\left(M(\omega) - \sqrt{\frac{-2\log\alpha}{h}}\right), \exp\left(M(\omega) + \sqrt{\frac{-2\log\alpha}{h}}\right)\right]. \tag{B.1} $$

Now, $M - \sqrt{\frac{-2\log\alpha}{h}} \sim N\left(\mu - \sqrt{\frac{-2\log\alpha}{h}}, \sigma^2\right)$, hence the lower bound of (B.1) has a lognormal distribution $LN\left(\mu - \sqrt{\frac{-2\log\alpha}{h}}, \sigma^2\right)$ and an expectation

$$ \mathbb{E}_*({}^{\alpha}\widetilde{Y}) = \exp\left(\mu - \sqrt{\frac{-2\log\alpha}{h}} + \frac{\sigma^2}{2}\right). \tag{B.2} $$

Similarly,

$$ \mathbb{E}^*({}^{\alpha}\widetilde{Y}) = \exp\left(\mu + \sqrt{\frac{-2\log\alpha}{h}} + \frac{\sigma^2}{2}\right). \tag{B.3} $$

The lower and upper expectations of $\widetilde{Y}$ are, respectively, the integrals of (B.2) and (B.3) from $\alpha = 0$ to $\alpha = 1$. Let us start with

$$ \mathbb{E}_*(\widetilde{Y}) = \int_0^1 \exp\left(\mu - \sqrt{\frac{-2\log\alpha}{h}} + \frac{\sigma^2}{2}\right) d\alpha. $$

By the change of variable $\beta = \sqrt{\frac{-2\log\alpha}{h}}$, we get

$$ \mathbb{E}_*(\widetilde{Y}) = h\exp\left(\mu + \frac{\sigma^2}{2}\right)\int_0^{+\infty} \beta\exp\left(-\frac{h}{2}\beta^2 - \beta\right) d\beta. $$

Completing the square gives us

$$ \mathbb{E}_*(\widetilde{Y}) = h\exp\left(\mu + \frac{\sigma^2}{2} + \frac{1}{2h}\right)\underbrace{\int_0^{+\infty} \beta\exp\left[-\frac{h}{2}\left(\beta + \frac{1}{h}\right)^2\right] d\beta}_{I}. \tag{B.4} $$

Integral I is related to the mean of the left-truncated normal distribution on $[0, +\infty)$, with mean $-1/h$, standard deviation $1/h$, and density

$$ f(y) = \frac{\sqrt{h}}{\sqrt{2\pi}\left(1 - \Phi\left(\frac{1}{\sqrt{h}}\right)\right)}\exp\left[-\frac{h}{2}\left(\beta + \frac{1}{h}\right)^2\right]. $$

The mean of this truncated normal distribution is

$$ -\frac{1}{h} + \frac{1}{\sqrt{h}}\frac{\phi(1/\sqrt{h})}{1 - \Phi(1/\sqrt{h})}, $$

where ϕ is the standard normal pdf. Consequently, we have

$$I = \frac{-\frac{1}{h} + \frac{1}{\sqrt{h}} \frac{\phi(1/\sqrt{h})}{1-\Phi(1/\sqrt{h})}}{\frac{\sqrt{h}}{\sqrt{2\pi}\left(1-\Phi\left(\frac{1}{\sqrt{h}}\right)\right)}} = \sqrt{2\pi}\left[-h^{-3/2}\left(1-\Phi\left(\frac{1}{\sqrt{h}}\right)\right) + \frac{1}{h}\phi\left(\frac{1}{\sqrt{h}}\right)\right]. \tag{B.5}$$

From (B.4) and (B.5), we get (17a). Similar calculations starting from the upper bound of (B.1) yield (17b).

Appendix C. Proof of Theorem 4

We start by the following lemma, which generalizes Lemma 1 in [14].

Lemma 3. *Let (M_1, Z_1) and (M_2, Z_2) be two independent two-dimensional random vectors such that*

$$P(Z_1 = k) = \pi_{1k}, \quad k = 1, \ldots, K,$$
$$M_1 \mid (Z_1 = k) \sim N(\mu_{1k}, \sigma^2_{1k}), \quad k = 1, \ldots, K$$

and

$$P(Z_2 = \ell) = \pi_{2\ell}, \quad \ell = 1, \ldots, L,$$
$$M_2 \mid (Z_2 = \ell) \sim N(\mu_{2\ell}, \sigma^2_{2\ell}), \quad \ell = 1, \ldots, L.$$

Let $\widetilde{F}$ be the fuzzy subset of $\mathbb{R}^2 \times \{1, \ldots, K\} \times \{1, \ldots, L\}$ with membership function

$$\begin{aligned}\widetilde{F}(m_1, m_2, y_1, y_2) &= \mathsf{hgt}\left(\mathsf{GFN}(m_1, h_1(y_1)) \cdot \mathsf{GFN}(m_2, h_2(y_2))\right) \\ &= \exp\left(-\frac{h_1(y_1)h_2(y_2)(m_1 - m_2)^2}{2(h_1(y_1) + h_2(y_2))}\right),\end{aligned}$$

where $h_1(y_1) = \prod_{k=1}^K h_{1k}^{y_{1k}}$ and $h_2(y_2) = \prod_{\ell=1}^L h_{2k}^{y_{2k}}$ and, as before, $y_{1k} = I(y_1 = k)$ and $y_{2k} = I(y_2 = \ell)$.

The conditional probability distribution of (M_1, M_2, Z_1, Z_2) given $\widetilde{F}$ can be described as follows:

- *The conditional probability distribution of (M_1, M_2) given $\widetilde{F}$ and $(Z_1, Z_2) = (k, \ell)$ is two-dimensional Gaussian with mean vector $\widetilde{\boldsymbol{\mu}}_{k\ell} = (\widetilde{\mu}_{1k\ell}, \widetilde{\mu}_{2k\ell})^T$ and covariance matrix*

$$\widetilde{\boldsymbol{\Sigma}}_{kl} = \begin{pmatrix} \widetilde{\sigma}^2_{1k\ell} & \rho_{k\ell}\widetilde{\sigma}_{1k\ell}\widetilde{\sigma}_{2kl} \\ \rho_{k\ell}\widetilde{\sigma}_{1k\ell}\widetilde{\sigma}_{2kl} & \widetilde{\sigma}^2_{2k\ell} \end{pmatrix},$$

with

$$\widetilde{\mu}_{1k\ell} = \frac{\mu_{1k}(1 + \overline{h}\sigma^2_{2\ell}) + \mu_{2\ell}\overline{h}_{k\ell}\sigma^2_{1k}}{1 + \overline{h}_{k\ell}(\sigma^2_{1k} + \sigma^2_{2\ell})}, \quad \widetilde{\mu}_{2k\ell} = \frac{\mu_{2\ell}(1 + \overline{h}_{k\ell}\sigma^2_{1k}) + \mu_{1k}\overline{h}_{k\ell}\sigma^2_{2}}{1 + \overline{h}_{k\ell}(\sigma^2_{1k} + \sigma^2_{2\ell})} \tag{C.1a}$$

$$\widetilde{\sigma}^2_{1k\ell} = \frac{\sigma^2_{1k}(1 + \overline{h}_{kl}\sigma^2_{2\ell})}{1 + \overline{h}_{kl}(\sigma^2_{1k} + \sigma^2_{2\ell})}, \quad \widetilde{\sigma}^2_{2\ell} = \frac{\sigma^2_{2\ell}(1 + \overline{h}_{k\ell}\sigma^2_{1k})}{1 + \overline{h}_{k\ell}(\sigma^2_{1k} + \sigma^2_{2\ell})} \tag{C.1b}$$

$$\rho_{k\ell} = \frac{\overline{h}_{k\ell}\sigma_{1k}\sigma_{2\ell}}{\sqrt{(1 + \overline{h}_{k\ell}\sigma^2_{1k})(1 + \overline{h}_{k\ell}\sigma^2_{2\ell})}}, \tag{C.1c}$$

where

$$\overline{h}_{k\ell} = \frac{h_{1k}h_{2\ell}}{h_{1k}+h_{2\ell}}. \tag{C.1d}$$

- *The conditional probability distribution of* (Z_1, Z_2) *given* $\widetilde{F}$ *is*

$$P(Z_1 = k, Z_2 = \ell \mid \widetilde{F}) = \widetilde{\pi}_{k\ell} = \frac{(1-\kappa_{k\ell})\pi_{1k}\pi_{2\ell}}{\sum_{k'\ell'}(1-\kappa_{k'\ell'})\pi_{1k'}\pi_{2\ell'}}, \tag{C.2}$$

where $\kappa_{k\ell}$ *is the degree of conflict between two independent GRFNs* $\widetilde{X}_{1k} \sim \widetilde{N}(\mu_{1k}, \sigma_{1k}^2, h_{1k})$ *and* $\widetilde{X}_{2\ell} \sim \widetilde{N}(\mu_{2\ell}, \sigma_{2\ell}^2, h_{2\ell})$ *given by Proposition 1.*

Proof. Given (Z_1, Z_2), M_1 and M_2 have normal distributions. The conditional probability distribution of (M_1, M_2) given $\widetilde{F}$ and $(Z_1, Z_2) = (k, \ell)$ results directly from Lemma 1 of [14]. Now, from Bayes' theorem,

$$\begin{aligned} P(Z_1 = k, Z_2 = \ell \mid \widetilde{F}) &= \frac{P(\widetilde{F} \mid Z_1 = k, Z_2 = \ell)P(Z_1 = k, Z_2 = \ell)}{P(\widetilde{F})} \\ &= \frac{P(\widetilde{F} \mid Z_1 = k, Z_2 = \ell)\pi_{1k}\pi_{2\ell}}{\sum_{k',\ell'} P(\widetilde{F} \mid Z_1 = k', Z_2 = \ell')\pi_{1k'}\pi_{2\ell'}}. \end{aligned}$$

From Lemma 1 in [14], $P(\widetilde{F} \mid Z_1 = k, Z_2 = \ell) = 1 - \kappa_{k\ell}$, where $\kappa_{k\ell}$ is given by Proposition 1, which completes the proof. □

We can now give the proof of Theorem 4.

Proof of Theorem 4. Let (M_1, Z_1) and (M_2, Z_2) be pairs of random variables from $(\Omega_1, \Sigma_1, P_1)$ and $(\Omega_2, \Sigma_2, P_2)$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ corresponding, respectively, to m-GRFNs $\widetilde{X}_1$ and $\widetilde{X}_2$. The orthogonal sum of $\widetilde{X}_1$ and $\widetilde{X}_2$ is the RFS $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \widetilde{P}_{12}, \mathbb{R}, \mathcal{B}(\mathbb{R}), \widetilde{X}_\odot)$, where $\widetilde{X}_\odot$ is the mapping

$$\widetilde{X}_\odot : (\omega_1, \omega_2) \to \mathsf{GFN}(M_{12}(\omega_1, \omega_2), h_1(\omega) + h_2(\omega)),$$

with

$$M_{12}(\omega_1, \omega_2) = \frac{h_1(\omega)M_1(\omega_1) + h_2(\omega)M_2(\omega_2)}{h_1(\omega) + h_2(\omega)},$$

and $\widetilde{P}_{12}$ is the probability measure on $\Omega_1 \times \Omega_2$ obtained by conditioning $P_1 \times P_2$ on the fuzzy set $\widetilde{\Theta}^*(\omega_1, \omega_2) = \mathsf{hgt}\,(\mathsf{GFN}(M_1(\omega_1), h_1(\omega)), \mathsf{GFN}(M_2(\omega_2), h_2(\omega)))$. The pushforward measure of $\widetilde{P}_{12}$ by the random vector (M_1, M_2, Z_1, Z_2) is the conditional distribution given $\widetilde{F}$ described in Lemma 3, with parameters $(\widetilde{\mu}_{1k\ell}, \widetilde{\mu}_{2k\ell}, \widetilde{\sigma}_{1k\ell}, \widetilde{\sigma}_{2k\ell}, \rho_{k\ell}, \widetilde{\pi}_{j\ell})$, $k = 1, \ldots, K$, $\ell = 1, \ldots, L$. The conditional expectation of M_{12} given $(Z_1, Z_2) = (k, \ell)$ and $\widetilde{F}$ is

$$\begin{aligned} \mathbb{E}(M_{12}|Z_1 = k, Z_2 = \ell, \widetilde{F}) &= \frac{h_{1k}\mathbb{E}(M_1|Z_1 = k, Z_2 = \ell, \widetilde{F}) + h_{2\ell}\mathbb{E}(M_2|Z_1 = k, Z_2 = \ell, \widetilde{F})}{h_{1k} + h_{2\ell}} \\ &= \frac{h_{1k}\widetilde{\mu}_{1k} + h_{2\ell}\widetilde{\mu}_{2\ell}}{h_{1k} + h_{2\ell}}, \end{aligned}$$

and its conditional variance is

$$\mathrm{Var}(M_{12}|Z_1 = k, Z_2 = \ell, \widetilde{F}) = \frac{1}{(h_{1k} + h_{2\ell})^2}(h_1^2 \mathrm{Var}(M_1|Z_1 = k, Z_2 = \ell, \widetilde{F}) + \\ h_2^2 \mathrm{Var}(M_2|Z_1 = k, Z_2 = \ell, \widetilde{F}) + 2h_{1k}h_{2\ell}\mathrm{Cov}(M_1, M_2|Z_1 = k, Z_2 = \ell, \widetilde{F})),$$

which gives

$$\mathrm{Var}(M_{12}|Z_1 = k, Z_2 = \ell, \widetilde{F}) = \frac{h_{1k}^2 \widetilde{\sigma}_{1k}^2 + h_2^2 \widetilde{\sigma}_2^{2\ell} + 2\rho_{k\ell} h_{1k} h_{2\ell} \widetilde{\sigma}_{1k} \widetilde{\sigma}_{2\ell}}{(h_{1k} + h_{2\ell})^2}.$$

Finally, $P(Z_1 = k, Z_2 = \ell|\widetilde{F}) = \widetilde{\pi}_{k\ell}$, which completes the proof of the first part of the theorem. The second part is obtained directly by noticing that

$$\begin{aligned}\kappa = 1 - P(\widetilde{F}) &= 1 - \sum_{k=1}^{K}\sum_{\ell=1}^{L} P(\widetilde{F} \mid Z_1 = k, Z_2 = \ell) P(Z_1 = k, Z_2 = \ell) \\ &= 1 - \sum_{k=1}^{K}\sum_{\ell=1}^{L}(1 - \kappa_{k\ell})\pi_{1k}\pi_{2\ell} = \sum_{k=1}^{K}\sum_{\ell=1}^{L}\kappa_{k\ell}\pi_{1k}\pi_{2\ell}.\end{aligned}$$