Combination of dependent and partially reliable Gaussian random fuzzy numbers

Thierry Denœux

Combination of dependent and partially reliable Gaussian random fuzzy numbers

Thierry Denœux[a,b]

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

Abstract

Gaussian random fuzzy numbers are random fuzzy sets generalizing Gaussian random variables and possibility distributions. They define belief functions on the real line that can be conveniently combined by the product-intersection rule under the independence assumption. In this paper, we introduce various extensions of this rule to account for dependence and partial reliability of the pieces of evidence. We first provide formulas for the combination of an arbitrary number of Gaussian random fuzzy numbers whose dependence is described by a correlation matrix, and we introduce a minimum-conflict combination operation. To account for partially reliable evidence, we then introduce two discounting operations called possibilistic and evidential discounting, as well as several combination operators based on different assumptions, each one parameterized by a correlation matrix and a vector of discounting coefficients. We demonstrate the application of these operators to the combination of predictions with different sets of inputs in machine learning, and show that performance can be enhanced by optimizing the parameters of the combination operators.

Keywords: Evidence theory, Dempster-Shafer theory, belief functions, random fuzzy sets, discounting, information fusion, machine learning, regression.

1. Introduction

The Dempster-Shafer (DS) theory of evidence [5][26] is a widely used methodology for information fusion (see, e.g., [3][18][31]). This approach is based on two main concepts: the representation of independent items of evidence by *belief functions*, and their combination by an operator called *Dempster's rule of combination* [26]. Although the mathematical foundations of continuous belief functions have been laid for a long time [27][28], most applications of the theory consider belief functions on finite frames of discernment. The reason for this limitation has been the absence of practical models of belief functions for continuous variables compatible with Dempster's rule of combination. For instance, random interval models such as p-boxes [16] define belief functions on the real line, but they cannot

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

be conveniently used in evidential reasoning because the combination of two p-boxes by Dempster's rule is no longer a p-box.

As an extension of DS theory and possibility theory [33], *epistemic random fuzzy set* (ERFS) theory recently introduced in [7][9][14] makes it possible to overcome this limitation. In this new theoretical framework, uncertain and/or fuzzy pieces of evidence are represented by random fuzzy sets inducing belief functions, and independent items of evidence are combined by the *product-intersection rule* generalizing both Dempster's rule and the normalized product-intersection operator of possibility theory. Within this framework, Gaussian random fuzzy numbers (GRFNs) introduced in [9] are an important practical model making it possible to represent evidence about continuous real variables. A GRFN is characterized by its mean, its variance and its imprecision. Gaussian random variables and Gaussian fuzzy numbers are recovered, respectively, in the special cases of infinite precision and zero variance. As shown in [9], GRFNs define a parametric family of belief functions on the real line, closed under the product-intersection rule. Even more general models based on transformations and mixtures [14] allow us to represent evidence about variables taking values in a subset of the real line, with virtually unlimited flexibility. GRFNs have been used for uncertainty quantification in machine learning in [8] and [12]; applications to statistical inference and belief elicitation are discussed in [14].

As Dempster's rule, the product-intersection rule introduced in [7][9] is based on two assumptions: independence and reliability (or "relevance" [23, 24]), which can be given precise meanings in the random fuzzy set (RFS) framework. Formulas for the combination of reliable and independent GRFNs are given in [9]. However, these assumptions are often too restrictive in applications. For instance, opinions from different experts, or predictions based on correlated features or overlapping datasets (using, e.g., the ENNreg model [8]) often cannot be treated as fully reliable or independent. The combination of dependent and partially reliable GRFNs is, thus, an important problem. In this paper[1], we address this problem and present the following main contributions:

1. We provide formulas for the combination of n GRFNs whose dependence is described by a correlation matrix;
2. We introduce a minimum-conflict conjunctive combination operation allowing one to combine GRFNs with unknown dependence;
3. We introduce two discounting operations for RFSs and new combination rules based on meta-knowledge about the relevance of information sources, allowing us to combine partially reliable GRFNs;
4. We demonstrate the application of these new fusion mechanisms to the combination of predictions based on different sets of inputs in machine learning.

The rest of this paper is organized as follows. Necessary notions about epistemic random fuzzy sets are first recalled in Section 2. Formulas for the combination of n GRFNs with an arbitrary covariance matrix are then derived in Section 3, and discounting operations as well

[1]This paper is an extended version of short paper [10], in which only the combination of dependent GRFN is addressed.

as new rules for combining partially reliable GRFNs are introduced in Section 4. Finally, numerical experiments are reported in Section 5, and Section 6 concludes the paper.

2. Background

General definitions and results about RFSs are first reviewed in Section 2.1. The notion of GRFN as well as those of transformation and mixture of GRFNs are then recalled, respectively, in Sections 2.2 and 2.3.

2.1. Epistemic Random Fuzzy Sets

ERFS theory is based on two main components: the representation of evidence by RFSs (inducing belief and plausibility functions), and a combination mechanism: the product-intersection rule for pooling independent evidence.

Random fuzzy sets. Let $(\Omega, \Sigma_\Omega, P)$ be a probability space, (Θ, Σ_Θ) a measurable space, and $\widetilde{X}$ a mapping from Ω to the set $[0,1]^\Theta$ of fuzzy subsets of Θ (see Figure 1). For any $\alpha \in [0,1]$, let ${}^\alpha\widetilde{X}$ denote the mapping from Ω to 2^Θ that maps each $\omega \in \Omega$ to the (weak) α-cut of $\widetilde{X}(\omega)$,

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

If, for any $\alpha \in [0,1]$, ${}^\alpha\widetilde{X}$ is $\Sigma_\Omega - \Sigma_\Theta$ strongly measurable [21], the tuple $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \widetilde{X})$ is said to be a RFS [4]. We sometimes identify the RFS with mapping $\widetilde{X}$ when there is no ambiguity about the underlying probability and measurable spaces.

Interpretation. In ERFS theory, a RFS represents a piece of evidence, which may be unreliable, vague (fuzzy), or both. The set Ω is seen as a *set of interpretations* of a piece of evidence about a variable X taking values in Θ. If interpretation $\omega \in \Omega$ holds, we only know that X is constrained by the possibility distribution defined by fuzzy set $\widetilde{X}(\omega)$. Standard DS theory only considers the case of unambiguous evidence, in which every image $\widetilde{X}(\omega)$ is crisp; mapping $\widetilde{X}$ is then a *random set*. In contrast, possibility theory only imposes a flexible contraint on X, without considering that this constraint may be itself uncertain. By considering both vagueness and uncertainty, ERFS is, thus, more flexible, allowing for faithful representation of different kinds of evidence.

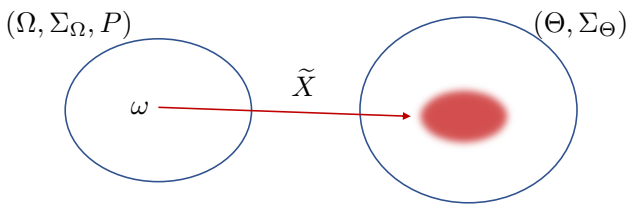


Figure 1: Definition of a random fuzzy set.

Associated belief functions. To any RFS, we can be associate a belief function representing one's beliefs based on the available evidence. For technical reasons, we assume hereafter any RFS $\widetilde{X}$ to verify the following normalization conditions: (1) For all $\omega \in \Omega$, $\widetilde{X}(\omega)$ is either the empty set, or a normal fuzzy set, and (2) the image $\widetilde{X}(\omega)$ is almost surely nonempty, i.e., $P(\{\omega \in \Omega : \widetilde{X}(\omega) = \emptyset\}) = 0$. For any $\omega \in \Omega$, a conditional possibility measure $\Pi_{\widetilde{X}(\omega)}$ and a dual conditional necessity measure $N_{\widetilde{X}(\omega)}$ on Θ can be defined as follows: for any $B \subseteq \Theta$,

$$\Pi_{\widetilde{X}(\omega)}(B) = \sup_{\theta \in B} \widetilde{X}(\omega)(\theta),$$

and

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

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 wrt P:

$$Bel_{\widetilde{X}}(B) = \int_\Omega N_{\widetilde{X}(\omega)}(B) dP(\omega), \tag{1a}$$

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

The corresponding mappings $Bel_{\widetilde{X}} : \Sigma_\Theta \to [0,1]$ and $Pl_{\widetilde{X}} : \Sigma_\Theta \to [0,1]$, are, respectively, belief and plausibility functions [4, 34].

Product-intersection rule. Given two normal fuzzy subsets $\widetilde{F}$ and $\widetilde{G}$ of Θ, their normalized product intersection is defined as

$$(\widetilde{F} \odot \widetilde{G})(\theta) = \begin{cases} \dfrac{\widetilde{F}(\theta)\widetilde{G}(\theta)}{\mathrm{hgt}(\widetilde{F} \cdot \widetilde{G})} & \text{if } \mathrm{hgt}(\widetilde{F} \cdot \widetilde{G}) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

where $\mathrm{hgt}(\widetilde{F} \cdot \widetilde{G}) = \sup_{\theta \in \Theta} \widetilde{F}(\theta)\widetilde{G}(\theta)$ is the height of the product intersection of $\widetilde{F}$ and $\widetilde{G}$. This operation is associative; as shown in [13], it is the only normalized intersection operator having this property. The normalized product intersection can be extended to RFSs as follows. Let $(\Omega_i, \Sigma_i, P_i, \Theta, \Sigma_\Theta, \widetilde{X}_i)$, $i = 1, 2$, be two RFSs representing *independent* pieces of evidence. The product-intersection operation maps these two RFSs to another RFS, called their *orthogonal sum* and defined as

$$(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \widetilde{P}_{12}, \Theta, \Sigma_\Theta, \widetilde{X}_1 \oplus \widetilde{X}_2), \tag{3}$$

where $\widetilde{X}_1 \oplus \widetilde{X}_2$ is the mapping from $\Omega_1 \times \Omega_2$ to $[0,1]^\Theta$ defined as $(\widetilde{X}_1 \oplus \widetilde{X}_2)(\omega_1, \omega_2) = \widetilde{X}_1(\omega_1) \odot \widetilde{X}_1(\omega_2)$, $\Sigma_1 \otimes \Sigma_2$ is the tensor product of Σ_1 and Σ_2, and $\widetilde{P}_{12}$ is the probability

measure on $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2)$ obtained by conditioning the product measure $P_1 \times P_2$ by the fuzzy set of consistent pairs (ω_1, ω_2), defined as $\widetilde{F}(\omega_1, \omega_2) = \mathrm{hgt}\left(\widetilde{X}_1(\omega_1) \cdot \widetilde{X}_2(\omega_2)\right)$, i.e.,

$$\forall A \in \Sigma_1 \otimes \Sigma_2, \widetilde{P}_{12}(A) = \frac{\int_{\Omega_1}\int_{\Omega_2} A(\omega_1,\omega_2)\widetilde{F}(\omega_1,\omega_2)dP_2(\omega_2)dP_1(\omega_1)}{\int_{\Omega_1}\int_{\Omega_2} \widetilde{F}(\omega_1,\omega_2)dP_2(\omega_2)dP_1(\omega_1)}, \tag{4}$$

where $A(\cdot,\cdot)$ denotes the indicator function of A. The *degree of conflict* between the two pieces of evidence is defined as one minus the denominator in the right-hand side of (4). The product intersection of RFSs is commutative and associative. It extends both Dempster's rule for combining random sets, and the normalized product intersection (2) for combining possibility distributions.

As mentioned above, the product-intersection operation can be used to combine independent pieces of evidence. The independence assumption comes into play when considering the product measure $P_1 \times P_2$ for the definition of $\widetilde{P}_{12}$ in (3). It could be relaxed by considering a general probability measure P_{12} in $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2)$, with marginals equal to P_1 and P_2. This idea will be exploited in Section 3 for the special case of GRFNs, whose definition is recalled in the next section.

2.2. Gaussian Random Fuzzy Numbers

Gaussian Fuzzy Numbers (GFNs) play the same role in quantitative possibility theory as Gaussian random variables (GRVs) in probability theory. They are defined as fuzzy subsets of $\mathbb{R}$ with membership function

$$x \mapsto \exp\left(-\frac{h}{2}(x-m)^2\right), \tag{5}$$

where $m \in \mathbb{R}$ is the *mode* and $h \in [0, +\infty]$ is the *precision*. A GFN with mode m and precision h will be denoted by $\mathrm{GFN}(m,h)$. If $h = 0$, parameter m becomes immaterial and we write $\mathrm{GFN}(*,h)$; this GFN is said to be *vacuous*. GFNs are easily combined by the normalized product-intersection operator, as the following property holds:

$$\mathrm{GFN}(m_1,h_1) \odot \mathrm{GFN}(m_2,h_2) = \begin{cases} \mathrm{GFN}\left(\frac{h_1m_1+h_2m_2}{h_1+h_2}, h_1+h_2\right) & \text{if } h_1+h_2 > 0, \\ \mathrm{GFN}\,(*,0) & \text{otherwise.} \end{cases}$$

Let us now consider a GRV M with mean μ and variance σ^2. The random set

$$(\mathbb{R}, \mathcal{B}_\mathbb{R}, P_M, \mathbb{R}, \mathcal{B}_\mathbb{R}, \widetilde{X}),$$

where P_M is the probability distribution of M, $\mathcal{B}_\mathbb{R}$ is the Borel-σ algebra on $\mathbb{R}$, and $\widetilde{X}$ is the mapping $\widetilde{X} : \mathbb{R} \to [0,1]^\mathbb{R}$ such that $\widetilde{X}(m) = \mathrm{GFN}(m,h)$, is called a *Gaussian random fuzzy number* (GRFN) with mean μ, variance σ^2 and precision h. We write, equivalently, $\widetilde{X} \sim \mathrm{GFN}(M,h)$, $M \sim N(\mu,\sigma^2)$, or $\widetilde{X} \sim \widetilde{N}(\mu,\sigma^2,h)$ depending on the context.

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. A GRV or a GFN is recovered when, respectively, $h = +\infty$ or $\sigma^2 = 0$. When $h = 0$, the GRFN is said to be vacuous: it represents total ignorance; the mean and standard deviation are then irrelevant and we write $\widetilde{X} \sim \widetilde{N}(*, *, 0)$. The contour function $pl_{\widetilde{X}} : x \mapsto Pl_{\widetilde{X}}(\{x\})$ as well as the lower and upper cumulative distribution functions (cdfs) $x \mapsto Bel_{\widetilde{X}}((-\infty, x])$ and $x \mapsto Pl_{\widetilde{X}}((-\infty, x])$ are given by the following equations, proved in [9]:

$$pl_{\widetilde{X}}(x) = \frac{1}{\sqrt{1 + h\sigma^2}} \exp\left(-\frac{h(x-\mu)^2}{2(1+h\sigma^2)}\right), \tag{6a}$$

$$Bel_{\widetilde{X}}((-\infty, x]) = \Phi\left(\frac{x-\mu}{\sigma}\right) - pl_{\widetilde{X}}(x)\Phi\left(\frac{x-\mu}{\sigma\sqrt{1+h\sigma^2}}\right), \tag{6b}$$

where Φ is the standard normal cdf, and

$$Pl_{\widetilde{X}}((-\infty, x]) = Bel_{\widetilde{X}}((-\infty, x]) + pl_{\widetilde{X}}(x). \tag{6c}$$

As shown in [9], the family of GRFNs is closed under the product-intersection combination operation $\oplus$. Let $M_1 \sim N(\mu_1, \sigma_1^2)$ and $M_2 \sim N(\mu_2, \sigma_2^2)$ be two independent GRVs, and let $\widetilde{X}_1 = \mathrm{GFN}(M_1, h_1)$ and $\widetilde{X}_2 = \mathrm{GFN}(M_2, h_2)$ be corresponding GRFNs. To combine $\widetilde{X}_1$ and $\widetilde{X}_2$ by the product-intersection rule, we proceed as follows [9]:

1. We condition the joint probability distribution of (M_1, M_2) by the fuzzy subset $\widetilde{F}$ of $\mathbb{R}$ defined by

$$\widetilde{F}(m_1, m_2) = \mathrm{hgt}\left(\mathrm{GFN}(m_1, h_1) \cdot \mathrm{GFN}(m_2, h_2)\right) = \exp\left(-\frac{\overline{h}}{2}(m_1 - m_2)^2\right), \tag{7}$$

   where $\overline{h} = h_1 h_2/(h_1 + h_2)$. This conditional distribution is normal with mean $\widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}_1, \widetilde{\mu}_2)^T$ and covariance matrix $\widetilde{\boldsymbol{\Sigma}}$,

$$\widetilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \widetilde{\sigma}_1^2 & \widetilde{\rho}\widetilde{\sigma}_1\widetilde{\sigma}_2 \\ \widetilde{\rho}\widetilde{\sigma}_1\widetilde{\sigma}_2 & \widetilde{\sigma}_1^2 \end{pmatrix},$$

   where

$$\widetilde{\mu}_i = \frac{\mu_i(1 + \overline{h}\sigma_j^2) + \mu_j\overline{h}\sigma_i^2}{1 + \overline{h}(\sigma_i^2 + \sigma_j^2)}, \quad \widetilde{\sigma}_i^2 = \frac{\sigma_i^2(1 + \overline{h}\sigma_j^2)}{1 + \overline{h}(\sigma_i^2 + \sigma_j^2)} \tag{8a}$$

   for $(i, j) \in \{(1, 2), (2, 1)\}$, and

$$\widetilde{\rho} = \frac{\overline{h}\sigma_1\sigma_2}{\sqrt{(1 + \overline{h}\sigma_1^2)(1 + \overline{h}\sigma_2^2)}}. \tag{8b}$$

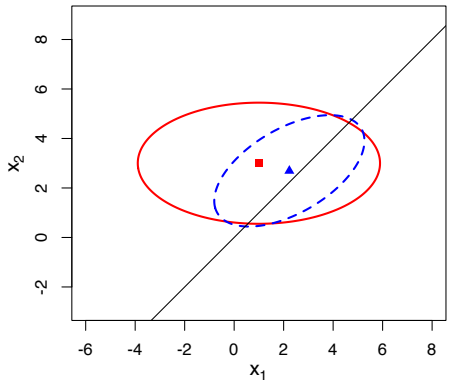


Figure 2: 95% coverage probability ellipses for the unconditional distribution of random vector (M_1, M_2) in Example 1 (solid red curve) and its conditional distribution given $\widetilde{F}$ (dashed blue curve).

2. Let $\widetilde{X} = \widetilde{X}_1 \oplus \widetilde{X}_2$ be the combined RFS; we have $\widetilde{X} \sim \widetilde{N}(\mu_c, \sigma_c^2, h_1 + h_2)$ with $\mu_c = \boldsymbol{h}^{*T}\widetilde{\boldsymbol{\mu}}$ and $\sigma_c^2 = \boldsymbol{h}^{*T}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{h}^*$, where $\boldsymbol{h}^* = (h_1, h_2)^T/(h_1 + h_2)$ is the vector of normalized precisions.
3. The degree of conflict between $\widetilde{X}_1$ and $\widetilde{X}_2$ is

$$\kappa = 1 - \frac{\widetilde{\sigma}_1\widetilde{\sigma}_2}{\sigma_1\sigma_2}\sqrt{1-\widetilde{\rho}^2}\exp\left\{-\frac{1}{2}\left[\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2}\right] + \frac{1}{2(1-\widetilde{\rho}^2)}\left[\frac{\widetilde{\mu}_1^2}{\widetilde{\sigma}_1^2} + \frac{\widetilde{\mu}_2^2}{\widetilde{\sigma}_2^2} - 2\widetilde{\rho}\frac{\widetilde{\mu}_1\widetilde{\mu}_2}{\widetilde{\sigma}_1\widetilde{\sigma}_2}\right]\right\} \quad (9a)$$

if $\sigma_1, \sigma_2 > 0$, and

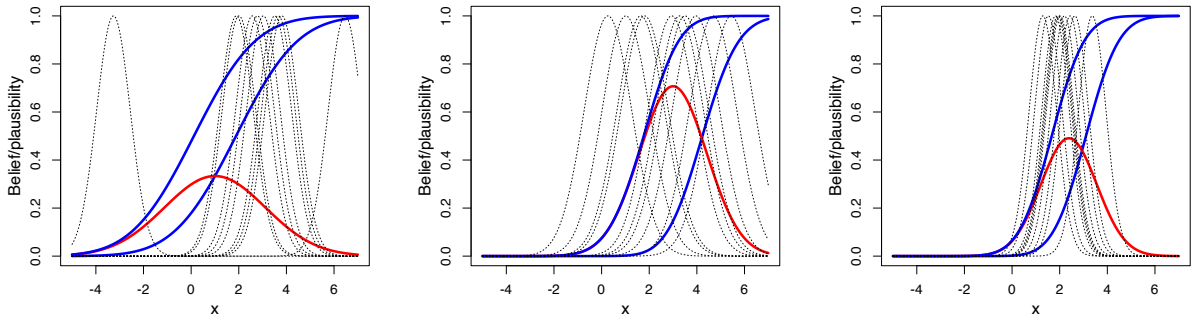
$$\kappa = 1 - \frac{1}{\sqrt{1+\overline{h}\widetilde{\sigma}_1^2}}\exp\left(-\frac{\overline{h}}{2(1+\overline{h}\widetilde{\sigma}_1^2)}(\widetilde{\mu}_1 - \mu_2)^2\right) \quad (9b)$$

if $\sigma_1 \geq 0$ and $\sigma_2 = 0$.

Example 1. *Consider two GRFNs $\widetilde{X}_1 = GFN(M_1, h_1)$ and $\widetilde{X}_2 = GFN(M_2, h_2)$, where $h_1 = 2$, $h_2 = 1$, and (M_1, M_2) has a two-dimensional normal distribution with mean $\boldsymbol{\mu} = (1, 3)^T$ and covariance matrix $\boldsymbol{\Sigma} = diag(4, 1)$. Figure 2 shows ellipses with 95% coverage probability for the unconditional distribution of random vector (M_1, M_2) and its conditional distribution given $\widetilde{F}$. We can see that the mean vector of the conditional distribution is closer to the line $m_1 = m_2$, and that the conditional correlation coefficient $\widetilde{\rho}$ is positive. Figure 3 displays $\widetilde{X}_1$, $\widetilde{X}_2$ and $\widetilde{X}_1 \oplus \widetilde{X}_2$. Each GRFN $\widetilde{X}$ is represented by its contour function $pl_{\widetilde{X}}$ and by its lower and upper cdfs $x \mapsto Bel_{\widetilde{X}}((-\infty, x])$ and $x \mapsto Pl_{\widetilde{X}}((-\infty, x])$.*

2.3. Transformations and Mixtures of Gaussian Random Fuzzy Numbers

As the normal distribution in probability theory, the notion of GRFN is too constrained to represent beliefs about a numerical quantity in all situations. Transformations and mixtures of GRFNs, introduced in [14], allow for the definition of much more flexible models.



(a) (b) (c)

Figure 3: (a) and (b): two GRFN $\widetilde{X}_1$ and $\widetilde{X}_2$; (c): their orthogonal sum $\widetilde{X}_1 \oplus \widetilde{X}_2$. For each GRFN, we plot ten realizations (black dotted curves), the contour function (red curve), as well as the lower and upper cdfs (blue curves).

Transformation of a GRFN. Let $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$ be a GRFN and ψ a one-to-one mapping from $\mathbb{R}$ to $\Lambda \subseteq \mathbb{R}$. We denote by $\widetilde{\psi}$ the extension of ψ to fuzzy subsets, i.e., the mapping $[0,1]^{\mathbb{R}} \to [0,1]^{\Lambda}$ that maps any $\widetilde{F} \in [0,1]^{\mathbb{R}}$ to $\widetilde{\psi}(\widetilde{F}) \in [0,1]^{\Lambda}$ defined as

$$\forall \lambda \in \Lambda, \quad \widetilde{\psi}(\widetilde{F})(\lambda) = \sup_{\lambda = \psi(x)} \widetilde{F}(x) = \widetilde{F}\left(\psi^{-1}(\lambda)\right).$$

As shown in [14], the composed mapping $\widetilde{\psi} \circ \widetilde{X}$ is a RFS, which we call a transformed GRFN (or t-GRFN), and we write $\widetilde{\psi} \circ \widetilde{X} \sim T\widetilde{N}(\mu, \sigma^2, h, \psi^{-1})$. As shown in [14], the belief and plausibility functions associated with $\widetilde{\psi} \circ \widetilde{X}$ can be derived from those associated with $\widetilde{X}$ using the following equalities:

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

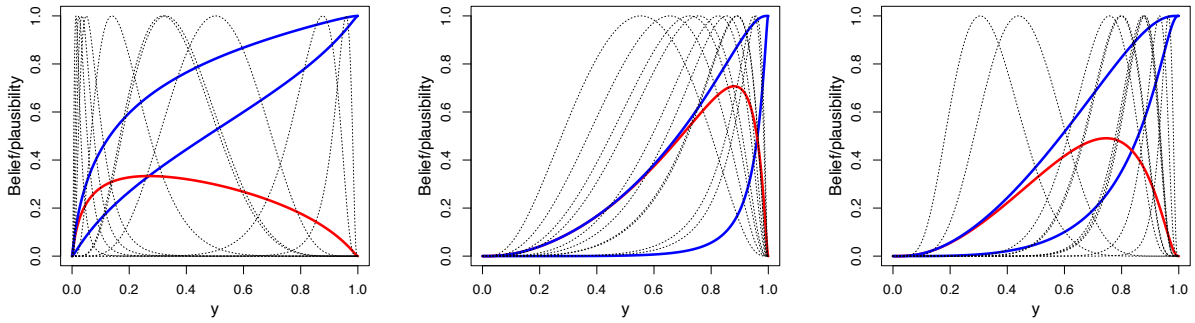
for any measurable subset C of Λ. Two important cases are $\psi = \exp$ and $\psi : x \mapsto (1 + \exp(-x))^{-1}$ corresponding to *lognormal* and *logit-normal* random fuzzy numbers with supports, respectively, $(-\infty, 0]$ and $[0, 1]$.

Finally, the following property was shown in [14]: the orthogonal sum of two transformed RFSs $\widetilde{\psi} \circ \widetilde{X}_1$ and $\widetilde{\psi} \circ \widetilde{X}_2$ is the transformation of the orthogonal sum of $\widetilde{X}_1$ and $\widetilde{X}_2$:

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

Furthermore, the degree of conflict between $\widetilde{\psi} \circ \widetilde{X}_1$ and $\widetilde{\psi} \circ \widetilde{X}_2$ is equal to the degree of conflict between $\widetilde{X}_1$ and $\widetilde{X}_2$. This property allows us to combine, e.g., lognormal or logit-normal RFSs using the formula recalled in Section 2.2 for combining GRFNs.

Example 2. *Figure 4 shows two logit-normal random fuzzy numbers* $\widetilde{Y}_1 \sim T\widetilde{N}(-1, 4, 2, \textit{logit})$, $\widetilde{Y}_2 \sim T\widetilde{N}(2, 1, 1, \textit{logit})$, *and their orthogonal sum* $\widetilde{Y}_1 \oplus \widetilde{Y}_2$.



(a) (b) (c)

Figure 4: (a) and (b): two logit-normal random fuzzy numbers $\widetilde{Y}_1$ and $\widetilde{Y}_2$; (c): their orthogonal sum $\widetilde{Y}_1 \oplus \widetilde{Y}_2$. For each t-GRFN, we plot ten realizations (black dotted curves), the contour function (red curve), as well as the lower and upper cdfs (blue curves).

Mixtures of GRFNs. Another useful notion introduced in [14] is that of (finite) mixture of GRFNs. Let (M, Z) denote a pair of random variables taking values in $\Omega = \mathbb{R} \times \{1, \ldots, K\}$, such that the marginal distribution of Z is defined by $P(Z = k) = \pi_k$, $k = 1, \ldots, K$, 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. We consider the random fuzzy set $\widetilde{X} : \Omega \to [0, 1]^{\mathbb{R}}$ defined as follows,

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

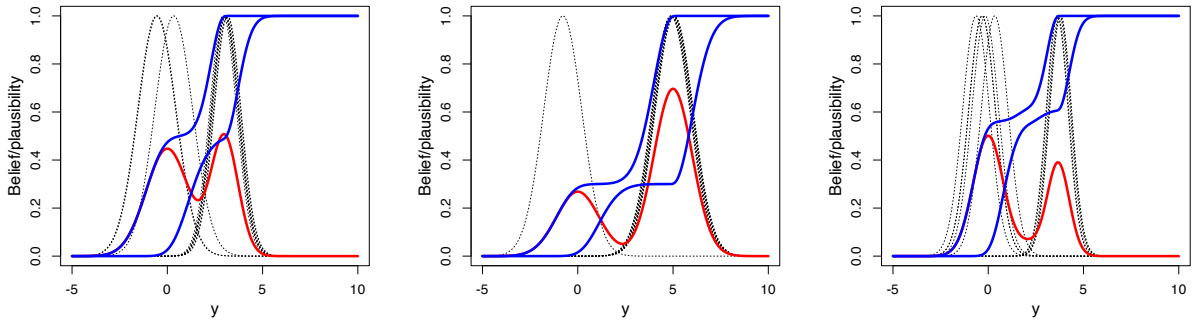
where $Z_k = I(Z = k)$, and $I(\cdot)$ is the indicator function. Conditionally on $Z = k$, $\widetilde{X}$ is a GRFN with mean μ_k, variance σ_k^2 and precision h_k:

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

We say that $\widetilde{X}$ is a *mixture GRFN* (m-GRFN) and we write

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

The following two theorems are proved in [14].



(a) (b) (c)

Figure 5: (a) and (b): two m-GRFNs $\widetilde{X}_1$ and $\widetilde{X}_2$; (c): their orthogonal sum $\widetilde{X}_1 \oplus \widetilde{X}_2$. For each m-GRFN, we plot ten realizations (black dotted curves), the contour function (red curve), as well as the lower and upper cdfs (blue curves).

Theorem 1. *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*

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

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

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

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

where $\kappa_{k\ell}$ denotes the degree of conflict between $\widetilde{X}_{1k}$ and $\widetilde{X}_{2\ell}$ given by (9).

Example 3. *Figure 5 displays two m-GRFNs $\widetilde{X}_1 \sim 0.5\widetilde{N}(0, 0.5^2, 1) + 0.5\widetilde{N}(3, 0.1^2, 2)$ and $\widetilde{X}_2 \sim 0.3\widetilde{N}(0, 0.5^2, 1) + 0.7\widetilde{N}(5, 0.1^2, 1)$, as well as their orthogonal sum $\widetilde{X}_1 \oplus \widetilde{X}_2$.*

3. Combination of dependent GRFNs

The formulas for the product intersection of two GRFNs recalled in Section 2.2 were established under the assumption that the underlying GRVs are independent. In this section, we generalize these formulas to the combination of n *dependent* GRFNs $\text{GFN}(M_i, h_i)$,

$i = 1, \ldots, n$, where random vector $\boldsymbol{M} = (M_1, \ldots, M_n)^T$ has a multidimensional normal distribution with an arbitrary covariance matrix. After preliminaries exposed in Section 3.1, we prove our main result in Section 3.2. The problem of combining GRFNs with unknown dependence is then addressed in Section 3.3.

3.1. Preliminaries

Let us first recall that a Gaussian fuzzy vector (GFV) with mode $\boldsymbol{m} \in \mathbb{R}^n$ and symmetric, positive semidefinite (PSD) precision matrix $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ is a fuzzy subset of $\mathbb{R}^n$ with membership function

$$\boldsymbol{x} \mapsto \exp(-0.5(\boldsymbol{x} - \boldsymbol{m})^T \boldsymbol{H}(\boldsymbol{x} - \boldsymbol{m})).$$

It is denoted as GFV$(\boldsymbol{m}, \boldsymbol{H})$. The results derived in Section 3.2 are based on results about the product of GFNs and GFVs stated in this section.

Proposition 1 below generalizes Propositions 3 in [9] and follows directly from results about the product of univariate normal densities proved in [1].

Proposition 1. *Let GFN(m_i, h_i), $i = 1, \ldots, n$, be n GFNs.*

1. *The height of their product intersection is*

$$\begin{aligned} \widetilde{F}(m_1, \ldots, m_n) &= hgt\left(GFN(m_1, h_1) \cdot \ldots \cdot GFN(m_n, h_n)\right) \\ &= \begin{cases} \exp\left[-\frac{1}{2}\left(\sum_{i=1}^n h_i m_i^2 - \frac{(\sum_{i=1}^n h_i m_i)^2}{\sum_{i=1}^n h_i}\right)\right] & \text{if } \sum_{i=1}^n h_i > 0 \\ 1 & \text{otherwise.} \end{cases} \end{aligned} \tag{10}$$

2. *If $\sum_{i=1}^n h_i > 0$, their normalized product intersection is a GFN with precision $h = \sum_{i=1}^n h_i$ and mode $m = (1/h)\sum_{i=1}^n h_i m_i$. Otherwise, it is the vacuous GFN with precision $h = 0$.*

The following proposition generalizes Proposition 11 in [9], by only assuming one of the two precision matrices to be positive definite (PD).

Proposition 2. *Let GFV$(\boldsymbol{m}_1, \boldsymbol{H}_1)$ and GFV$(\boldsymbol{m}_2, \boldsymbol{H}_2)$ be two GRVs. Assuming that $\boldsymbol{H}_1$ is PD and $\boldsymbol{H}_2$ is PSD,*

1. *The height of their product intersection is*

$$\exp\left(-\frac{1}{2}(\boldsymbol{m}_1 - \boldsymbol{m}_2)^T \boldsymbol{H}_2(\boldsymbol{I}_n + \boldsymbol{H}_1^{-1}\boldsymbol{H}_2)^{-1}(\boldsymbol{m}_1 - \boldsymbol{m}_2)\right), \tag{11}$$

   where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix;
2. *Their normalized product intersection is a GFV with precision matrix $\boldsymbol{H} = \boldsymbol{H}_1 + \boldsymbol{H}_2$ and mode $\boldsymbol{m} = \boldsymbol{H}^{-1}(\boldsymbol{H}_1\boldsymbol{m}_1 + \boldsymbol{H}_2\boldsymbol{m}_2)$.*

Proof. See Appendix A. □

3.2. Conjunctive combination of GRFNs with arbitrary correlation matrix

Let us consider n GRFNs $\widetilde{X}_1, \ldots, \widetilde{X}_n$ such that $\widetilde{X}_i \sim \widetilde{N}(\mu_i, \sigma_i^2, h_i)$ for $i = 1, \ldots, n$. Let M_i denote the random mode of $\widetilde{X}_i$, and assume that random vector $\boldsymbol{M} = (M_1, \ldots, M_n)^T$ has a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Matrix $\boldsymbol{\Sigma}$ can be written as $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma})\boldsymbol{R}\,\text{diag}(\boldsymbol{\sigma})$, where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ is the vector of standard deviations of the components of $\boldsymbol{M}$, $\text{diag}(\boldsymbol{\sigma})$ is the diagonal matrix with i-th diagonal elements equal to σ_i, and $\boldsymbol{R}$ is the correlation matrix of random vector $\boldsymbol{M}$.

We define the conjunctive combination of $\widetilde{X}_1, \ldots, \widetilde{X}_n$ with correlation matrix $\boldsymbol{R}$ as the RFS $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \widetilde{P}_{\boldsymbol{M}}, \mathbb{R}, \mathcal{B}_{\mathbb{R}}, \widetilde{X})$, where $\widetilde{X}$ is the mapping from $\mathbb{R}^n$ to $[0,1]^{\mathbb{R}^p}$ such that $(m_1, \ldots, m_n) \mapsto \widetilde{X}_1(m_1) \odot \ldots \odot \widetilde{X}_n(m_n)$, and $\widetilde{P}_{\boldsymbol{M}}$ is the multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ conditioned by the fuzzy set

$$\widetilde{F}(m_1, \ldots, m_n) = \text{hgt}\left(\widetilde{X}_1(m_1) \cdot \ldots \cdot \widetilde{X}_n(m_n)\right)$$

of consistent tuples $(m_1, \ldots, m_n)$ given by (10). We denote this RFS as $\mathcal{C}_{\boldsymbol{R}}(\widetilde{X}_1, \ldots, \widetilde{X}_n)$, and we refer to this operation as the *$\boldsymbol{R}$-conjunctive combination*. Obviously, we have $\mathcal{C}_{\boldsymbol{R}}(\widetilde{X}_1, \ldots, \widetilde{X}_n) = \widetilde{X}_1 \oplus \ldots \oplus \widetilde{X}_n$ if $\boldsymbol{R} = \boldsymbol{I}_n$, i.e., if the n GRFNs are independent.

To derive the expression of $\mathcal{C}_{\boldsymbol{R}}(\widetilde{X}_1, \ldots, \widetilde{X}_n)$, we start with the following lemma.

Lemma 1. *Let $\boldsymbol{M} = (M_1, \ldots, M_n)^T$ be a random vector having a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Its conditional distribution given the fuzzy subset $\widetilde{F}$ of $\mathbb{R}^n$ expressed by* (10) *is multivariate normal with mean*

$$\widetilde{\boldsymbol{\mu}} = (\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{\mu} \tag{12a}$$

and covariance matrix

$$\widetilde{\boldsymbol{\Sigma}} = (\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{\Sigma}, \tag{12b}$$

in which $\boldsymbol{A}$ is the symmetric and PSD matrix

$$\boldsymbol{A} = \begin{cases} \boldsymbol{0}_{n,n} & \text{if } h_1 = \ldots = h_n = 0, \\ \text{diag}(\boldsymbol{h}) - \frac{\boldsymbol{h}\boldsymbol{h}^T}{\boldsymbol{1}^T\boldsymbol{h}} & \text{otherwise,} \end{cases} \tag{12c}$$

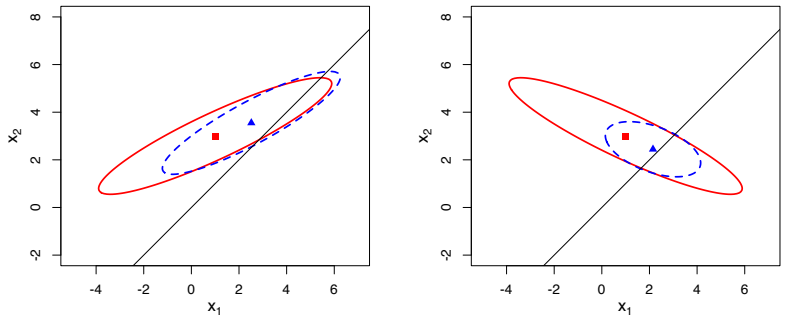
where $\boldsymbol{0}_{n,n}$ denotes the matrix of size $n \times n$ whose entries are zero, $\boldsymbol{h} = (h_1, \ldots, h_n)^T$ and $\boldsymbol{1} = (1, \ldots, 1)^T$.

Proof. See Appendix B. □

We can now state the main result in this section.

Theorem 3. *Let $\widetilde{X}_1, \ldots, \widetilde{X}_n$ be n GRFNs such that $\widetilde{X}_i \sim \widetilde{N}(\mu_i, \sigma_i^2, h_i)$, $i = 1, \ldots, n$. Let $M_i \sim N(\mu_i, \sigma_i^2)$ be the random mode of $\widetilde{X}_i$, and assume that random vector $\boldsymbol{M} = (M_1, \ldots, M_n)^T$ has a multivariate normal distribution with correlation matrix $\boldsymbol{R}$ and covariance matrix $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma})\boldsymbol{R}\,\text{diag}(\boldsymbol{\sigma})$ with $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$. We have $\mathcal{C}_{\boldsymbol{R}}(\widetilde{X}_1, \ldots, \widetilde{X}_n) \sim \widetilde{N}\left(\mu_c, \sigma_c^2, \sum_{i=1}^n h_i\right)$, with*

$$\mu_c = \boldsymbol{h}^{*T}(\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{\mu} \quad \text{and} \quad \sigma_c^2 = \boldsymbol{h}^{*T}(\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{\Sigma}\boldsymbol{h}^*, \tag{13}$$



(a) (b)

Figure 6: 95% coverage probability ellipses for the unconditional distribution of random vector (M_1, M_2) in Example 4 (solid red curve) and its conditional distribution given $\widetilde{F}$ (dashed blue curve), for $\rho = 0.9$ (a) and $\rho = -0.9$ (b).

where $\boldsymbol{A}$ *is given by* (12c) *and* $\boldsymbol{h}^* = (h_1, \ldots, h_n)^T / \sum_{i=1}^n h_i$ *is the vector of normalized precisions. Furthermore, the degree of conflict is*

$$\kappa_{\boldsymbol{R}}(\widetilde{X}_1, \ldots, \widetilde{X}_n) = 1 - |\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\mu}^T \boldsymbol{A}[\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A}]^{-1}\boldsymbol{\mu}\right). \tag{14}$$

Proof. See Appendix C. □

We can remark that the equations derived in [9] and recalled in (8) and (9) for the product intersection of two independent GRFNs and their degree of conflict can be recovered, respectively, from (12) and (14) when $n = 2$ and $\boldsymbol{\Sigma}$ is diagonal.

Example 4. *Let us consider again the two GRFNs of Example 1, and let us denote by* ρ *the correlation coefficient between* M_1 *and* M_2*. Figures 6a and 6b show ellipses with 95% coverage probability for the unconditional distribution of random vector* (M_1, M_2) *and its conditional distribution given* $\widetilde{F}$ *for, respectively,* $\rho = 0.9$ *and* $\rho = -0.9$*. The combined GRFNs assuming* $\rho = -1$*,* $\rho = 0$ *and* $\rho = 1$ *are shown in Figure 7. It is clear that the assumed correlation coefficient strongly influences the result of the combination. Figures 8a and 8b show, respectively, the mean and standard deviation of the combined GRFN as functions of* ρ*. The standard deviation appears to be particularly sensitive to the value of* ρ*. Figure 8c shows the degree of conflict as a function of* ρ*. It reaches a minimum value of 0.629 for* $\rho = 0.625$*.*

Case of complete positive dependence. Complete positive dependence corresponds to the case where all the correlation coefficients are equal to 1, i.e., $\boldsymbol{R} = \boldsymbol{J}_n$, where $\boldsymbol{J}_n$ is the square matrix of size n whose entries are all equal to 1. Operator $\mathcal{C}_{\boldsymbol{J}_n}$ is not idempotent, because the normalized product intersection $\odot$ is not. However, the degree of conflict κ

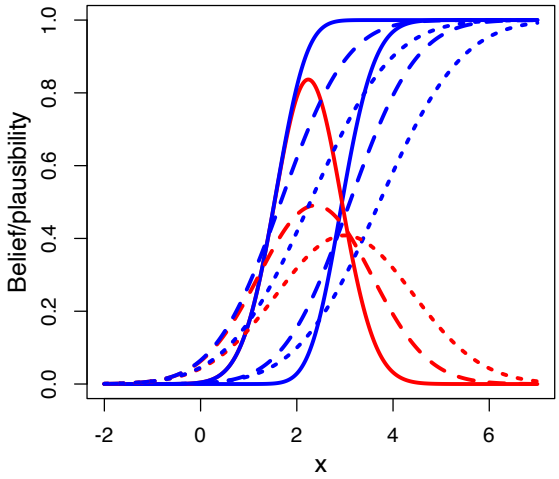
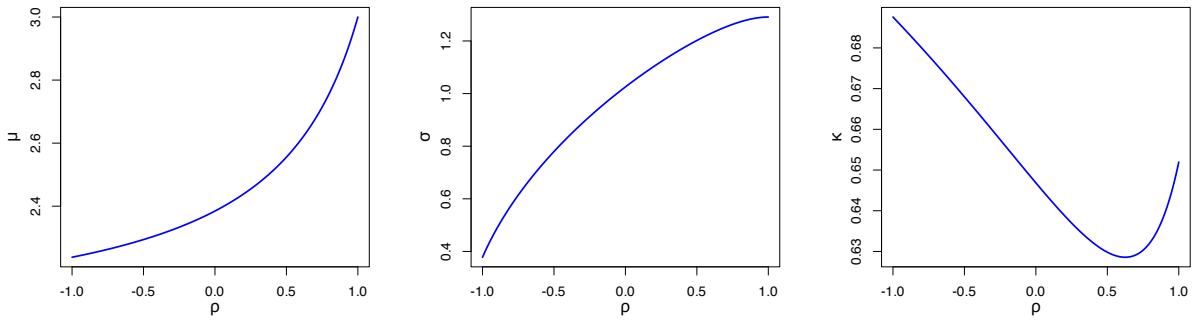


Figure 7: Combination of the two GRFNs of Example 1 assuming $\rho = -1$ (solid lines), $\rho = 0$ (dashed lines) and $\rho = 1$ (dotted lines). Each GRFN is represented by its contour function (red curve) and by its lower and upper cdfs (blue curves).



(a) (b) (c)

Figure 8: Mean (a), standard deviation (b) and degree of conflict (c) as functions of ρ for the combination of the two GRFNs of Example 4.

when combining a GRFN $\widetilde{X}$ n times is null. These results are expressed in the following proposition.

Proposition 3. *For any GRFN* $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$,

$$\mathcal{C}_{\boldsymbol{J}_n}(\underbrace{\widetilde{X}, \ldots, \widetilde{X}}_{n \text{ times}}) \sim \widetilde{N}(\mu, \sigma^2, nh), \quad \text{and} \quad \kappa_{\boldsymbol{J}_n}(\underbrace{\widetilde{X}, \ldots, \widetilde{X}}_{n \text{ times}}) = 0.$$

Proof. See Section Appendix E. □

When applied to identical GRFNs, the conjunctive combination operator with complete positive dependence thus gives the smallest degree of conflict. However, this is not true in general when combining different GRFNs. The idea of combining GRFNs while minimizing the conflict is explored in Section 3.3.

Combination of transformed GRFNs. Just as the product-intersection rule, the $\boldsymbol{R}$-conjunctive combination operation can be extended to transformed GRFNs in a simple way. Let n t-GRFNs $\widetilde{Y}_1, \ldots, \widetilde{Y}_n$ such that $\widetilde{Y}_i = \widetilde{\psi} \circ \widetilde{X}_i$ with $\widetilde{X}_i \sim \widetilde{N}(\mu_1, \sigma_i^2, h_i)$. We define their $\boldsymbol{R}$-conjunctive combination as

$$\mathcal{C}_{\boldsymbol{R}}(\widetilde{Y}_1, \ldots, \widetilde{Y}_n) = \widetilde{\psi} \circ \left(\mathcal{C}_{\boldsymbol{R}}(\widetilde{X}_1, \ldots, \widetilde{X}_n)\right).$$

Example 5. *Consider again the two logit-normal RFNs* $\widetilde{Y}_1$ *and* $\widetilde{Y}_2$ *of Example 2. Figure 9 shows their conjunctive combination with* $\rho = 0$ *(product-intersection rule),* $\rho = 1$ *(complete positive dependence) and* $\rho = 1$ *(complete negative dependence).*

Combination of mixtures of GRFNs. The $\boldsymbol{R}$-conjunctive combination operation can also be extended to mixtures of GRFNs. For ease of exposition, consider two mGRFNs $\widetilde{X}_1 \sim \sum_{k=1}^{K} \pi_{1k} \widetilde{N}(\mu_{1k}, \sigma_{1k}^2, h_{1k})$ and $\widetilde{X}_2 \sim \sum_{l=1}^{L} \pi_{2l} \widetilde{N}(\mu_{2l}, \sigma_2^2, h_{2l})$. Let use denote by $M_{11}, \ldots, M_{1K}$ and $M_{21}, \ldots, M_{2L}$ the random modes of, respectively, the components of $\widetilde{X}_1$ and $\widetilde{X}_2$. In the most general case, we need to specify the correlation coefficient ρ_{kl} between M_{1k} and M_{2l} for each $(k, l) \in \{1, \ldots, K\} \times \{1, \ldots, L\}$. In real applications, such detailed information is unlikely to be available. A useful working hypothesis is to assume that these correlation coefficients all have a common value ρ. In the general case of n mGRFNs we thus have a single correlation matrix $\boldsymbol{R}$. This model is illustrated by the following example.

Example 6. *Let us consider again the two mGRFNs of Example 3 in Section 2.3. Figure 10 shows the contour functions and lower/upper cdfs of the combined mGRFNs with* $\rho \in \{-1, 0, 1\}$.

3.3. Minimum-conflict combination

In real applications, it is often the case that the sources cannot be assumed to be independent, but the correlation matrix $\boldsymbol{R}$ is unknown. One could then consider the set of all possible combinations for all possible values of $\boldsymbol{R}$. However, it is not clear how such a set of GRFNs could be manipulated and exploited for further combination or decision-making.

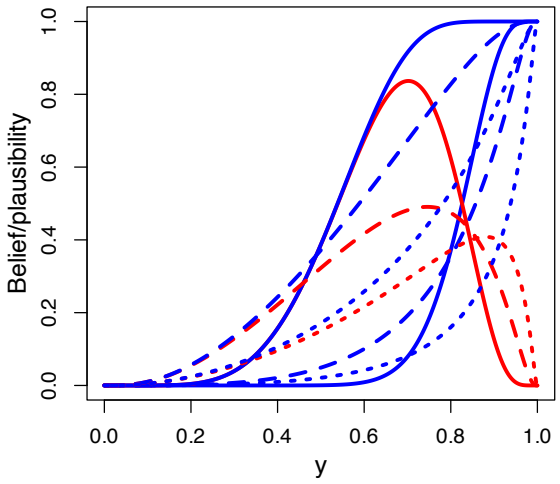


Figure 9: Combination of the two logit-normal RFNs of Example 5 assuming assuming $\rho = -1$ (solid lines), $\rho = 0$ (dashed lines) and $\rho = 1$ (dotted lines). Each GRFN is represented by its contour function (red curve) and by its lower and upper cdfs (blue curves).

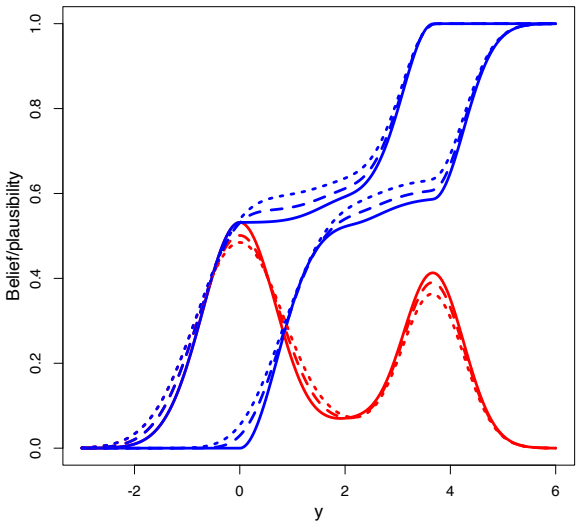


Figure 10: Combination of the two mGRFNs of Example 3 assuming $\rho = -1$ (solid lines), $\rho = 0$ (dashed lines) and $\rho = 1$ (dotted lines). Each mGRFN is represented by its contour function (red curve) and by its lower and upper cdfs (blue curves).

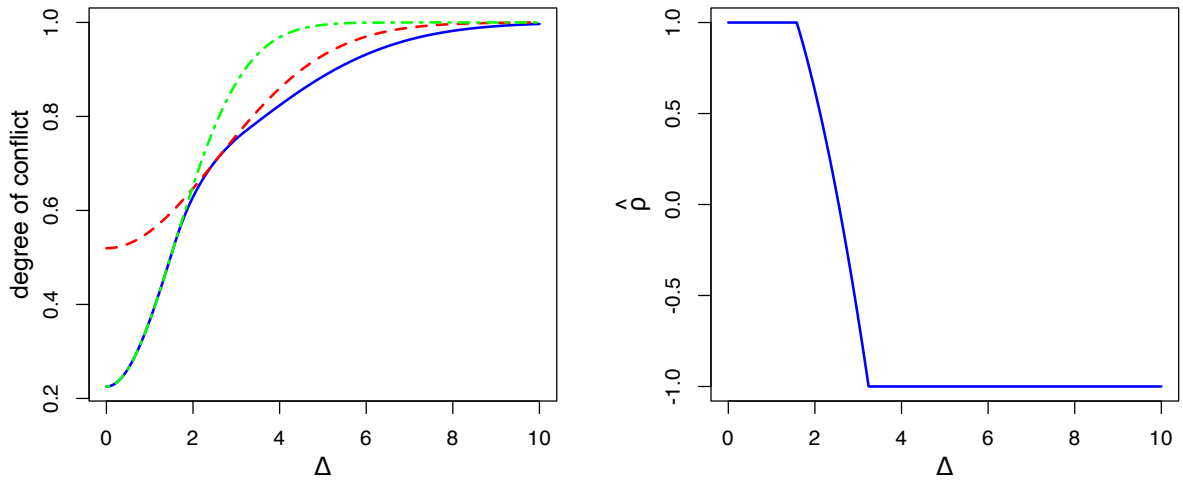
A more tractable approach is to assume a priori the sources to be consistent: among all correlation matrices, those yielding the *smallest degree of conflict* are thus regarded as the most plausible. A similar strategy was proposed in [2] for combining dependent belief functions on a finite frame of discernment. To justify this strategy, we can remark that the idea of minimizing the conflict or inconsistency when combining pieces of information is natural in many situations. For instance, in natural language understanding, disambiguation consists in choosing, among different possible meanings of a word or a sentence, one making it maximally consistent with the context. As another example, as discussed in [6], maximum likelihood estimation in statistics can be seen as finding the parameter value that minimizes the conflict between the statistical model and the observations. In the following we consider, successively, the case of $n = 2$ sources, and the case where $n > 2$.

Case $n = 2$. In the case of two sources, there is only one parameter to determine: the correlation coefficient $\rho \in [-1, 1]$ between the modes of the two GRFNs to be combined. We then have a univariate optimization problem, which is easy to solve. This is illustrated by the following example.

Example 7. *Let us consider two GRFNs* $\widetilde{X}_1 \sim \widetilde{N}(\mu_1, 4, 2)$ *and* $\widetilde{X}_2 \sim \widetilde{N}(\mu_2, 1, 1)$ *and let* $\Delta = |\mu_1 - \mu_2|$ *denote the distance between their means. Figure 11a shows the minimum degree of conflict as a function of* $\Delta \in [0, 10]$*, as well as the degrees conflict for the product-intersection rule (corresponding to* $\rho = 0$*) and the complete positive dependence rule (corresponding to* $\rho = 1$*). As expected, the minimum degree of conflict increases with* Δ *and is always smaller than the degree of conflict of the product-intersection rule. For* $\Delta > 1.6$*, the complete positive dependence rule no longer minimizes the conflict, and it even yields a larger conflict than the product-intersection rule. Figure 11b displays the correlation coefficient* $\widehat{\rho}$ *corresponding to the minimum conflict as a function of* Δ*. Consistently with Proposition 3, the minimum-conflict combination is achieved for* $\widehat{\rho} = 1$ *when* Δ *is small (low conflict), but it is achieved for* $\widehat{\rho} = -1$ *(complete negative dependence) when* Δ *is large (high conflict), and* $\widehat{\rho}$ *takes values between* -1 *and* 1 *for intermediate distances.*

Case $n > 2$. When the number of sources is greater than 2, finding the minimum-conflict combination of GRFNs becomes more delicate. The correlation matrix must be parameterized in such a way that the problem can be solved using an unconstrained optimization algorithm. Parameterizations of a covariance matrix are reviewed in [25]. For correlation matrices, the most widely used approach is the spherical parameterization [19]. It is based on the Cholesky decomposition $\boldsymbol{R} = \boldsymbol{C}\boldsymbol{C}^T$, where $\boldsymbol{C} = (c_{i,j})$ is an $n \times n$ lower triangular matrix with positive diagonal elements. Matrix $\boldsymbol{C}$ can be parameterized by $n(n-1)/2$ angles $\omega_{i,j} \in [0, \pi]$ with $i \in \{2, \ldots, n\}$, $j \in \{1, \ldots, i-1\}$ as follows:

$$c_{i,j} = \begin{cases} \cos(\omega_{i,j}) & \text{if } j = 1 \\ \cos(\omega_{i,j}) \prod_{k=1}^{j-1} \sin(\omega_{i,k}) & \text{if } 2 \leq j \leq i-1 \\ \prod_{k=1}^{j-1} \sin(\omega_{i,k}) & \text{if } j = i. \end{cases}$$



(a) (b)

Figure 11: (a): Minimum degree of conflict (solid blue curve), degree of conflict for $\rho = 0$ (dashed red curve) and $\rho = 1$ (green dashed-dotted curve) vs. $\Delta = |\mu_1 - \mu_2|$. (b): Minimum-conflict correlation coefficient $\widehat{\rho}$ as a function of Δ.

To satisfy the constraints $\omega_{i,j} \in [0, \pi]$, we introduce new parameters $\theta_{i,j} \in \mathbb{R}$ and write

$$\omega_{i,j} = \frac{\pi}{1 + \exp(-\theta_{i,j})}.$$

Using this parameterization, there is a one-to-one mapping between correlation matrices $\boldsymbol{R}$ and vectors $\boldsymbol{\theta} \in \mathbb{R}^{n(n-1)/2}$, and any unconstrained optimization procedure, such as the BFGS algorithm, can be used to minimize the conflict. We can remark that the number of parameters grows as the square of the number n of sources, making the optimization problem increasingly difficult when n increases. We denote by $\mathcal{C}_{\boldsymbol{R}^*}(\widetilde{X}_1, \ldots, \widetilde{X}_n)$ the minimum-conflict combination of $\widetilde{X}_1, \ldots, \widetilde{X}_n$, with

$$\boldsymbol{R}^* = \arg\min_{\boldsymbol{R}} \kappa_{\boldsymbol{R}}(\widetilde{X}_1, \ldots, \widetilde{X}_n).$$

Example 8. *We consider three GRFNs* $\widetilde{X}_1 \sim \widetilde{N}(\mu_1, 1, 1)$, $\widetilde{X}_2 \sim \widetilde{N}(0, 1, 1)$ *and* $\widetilde{X}_3 \sim \widetilde{N}(\mu_3, 1, 1)$ *with* $(\mu_1, \mu_3) \in \{-2, -1.5, -1, 0\} \times \{0, 1, 1.5, 2\}$. *Table 1 shows the degrees of conflict* $\kappa_{\boldsymbol{R}^*}$ *and* $\kappa_{\boldsymbol{I}_3}$ *obtained by, respectively, the minimum-conflict and product-intersection rules for different values of* μ_1 *and* μ_3. *As expected, the conflict increases with the distance* $\mu_1 - \mu_3$. *The conflict reduction achieved by the minimum-conflict rule is larger when the distance is smaller. The minimum-conflict correlation coefficients* ρ^*_{12}, ρ^*_{13} *and* ρ^*_{23} *are reported in Table 2. We can see that the complete positive dependence rule* $\mathcal{C}_{\boldsymbol{J}_3}$ *minimizes the conflict*

Table 1: Degrees of conflict $(\kappa_{\boldsymbol{R}^*}, \kappa_{\boldsymbol{I}_3})$ obtained by, respectively, the minimum-conflict and product-intersection rules for different values of μ_1 and μ_3 in Example 8.

$\mu_1 \diagdown \mu_3$	0	1	1.5	2.0
0	(0.00,0.50)	(0.28,0.58)	(0.50,0.66)	(0.63,0.74)
-1	(0.28,0.58)	(0.58,0.70)	(0.68,0.77)	(0.76,0.84)
-1.5	(0.50,0.66)	(0.68,0.77)	(0.76,0.84)	(0.82,0.89)
-2	(0.63,0.74)	(0.76,0.84)	(0.82,0.89)	(0.87,0.93)

Table 2: Minimum-conflict correlation coefficients $(\rho^*_{12}, \rho^*_{13}, \rho^*_{23})$ for different values of μ_1 and μ_3 in Example 8.

$\mu_1 \diagdown \mu_3$	0	1	1.5	2.0
0	(1,1,1)	(1,1,1)	(1,0.62,0.63)	(1,-0.25,-0.25)
-1	(1,1,1)	(0.79,0.26,0.79)	(0.79,-0.26,0.38)	(0.85,-0.66,-0.17)
-1.5	(0.63,0.62,1)	(0.38,-0.26,0.79)	(0.48,-0.54,0.48)	(0.58,-0.75,0.10)
-2	(-0.25,-0.25,1)	(-0.17,-0.66,0.85)	(0.10,-0.75,0.58)	(0.28,-0.84,0.28)

when $(\mu_1, \mu_3) \in \{(0,0), (0,1), (-1,0)\}$. *We have* $\rho^*_{12} = 1$ *when* $\mu_1 = 0$ *and, symmetrically* $\rho^*_{23} = 1$ *when* $\mu_3 = 0$.

4. Combination of partially reliable GRFNs

The product-intersection rule recalled in Section 2.2 is based on two main assumptions: the independence and reliability of the combined evidence. The former issue has been dealt with in Section 3; the latter is addressed in this section. The notion of reliability, or relevance of a source of information has been well studied in classical DS theory and in the finite setting [29, 20, 23, 24]. In the ERFS framework, it can be given the following precise meaning. Let $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \widetilde{X})$ be a RFS modeling a piece of evidence about a variable X taking values in Θ. If interpretation $\omega \in \Omega$ holds, the source tells us that the value of X is constrained by possibility distribution $\widetilde{X}(\omega)$. If we consider the source as fully reliable, we accept $\widetilde{X}(\omega)$ as our possibility distribution for X when ω holds. If we consider the source as fully unreliable, we regard the information it gives us as irrelevant and we are left in a state of total ignorance, which can be represented by the vacuous possibility distribution assigning a maximum possibility degree to all $\theta \in \Theta$. In intermediate situations where the source is believed to be partially reliable, $\widetilde{X}$ must be transformed into a weaker, less informative RFS. Such an operation is referred to as *discounting*. In Section 4.1 below, we briefly review notions of discounting in DS and possibility theories. We then extend these notions to random fuzzy sets in Section 4.2. Finally, in Section 4.3, we propose mechanisms for combining partially reliable GRFNs based on different assumptions about the reliability

of the sources.

4.1. Discounting in DS and possibility theories

Discounting plays an important role in DS theory, where it is often used to "weaken" evidence before combination [26]. Similarly, a notion of discounting taking into account the credibility of evidence was also introduced in possibility theory [32]. These two notions are reviewed below.

Evidential discounting. The notion of discounting plays an important role in DS theory. It was introduced by Shafer [26] for belief functions on finite frames. Let $m : 2^\Theta \to [0,1]$ be a mass function on a finite set Θ, and let $m_?$ denote the vacuous mass such that $m_?(\Theta) = 1$. For any $\alpha \in [0,1]$, the discounted mass function with discount rate $1-\alpha$ is mass function $m' = D_e(\alpha, m)$ such that

$$m' = \alpha m + (1-\alpha) m_?. \tag{15}$$

Obviously, $D_e(1,m) = m$ and $D_e(0,m) = m_?$. Let Pl and Pl' denote the pausibility functions corresponding, respectively, to m and m'. For any $A \subseteq \Theta$,

$$Pl'(A) = \alpha Pl(A) + (1-\alpha) Pl_?(A) \tag{16a}$$

$$= \begin{cases} \alpha Pl(A) + 1 - \alpha & \text{if } A \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases} \tag{16b}$$

Hence, for all $A \subseteq \Theta$, $Pl'(A) \geq Pl(A)$ or, equivalently, $Bel'(A) \leq Bel(A)$: m' is thus less committed, or less informative than m. In [29], Smets proposed a model of source reliability by introducing a variable $Z \in \{0,1\}$ and a probability distribution p_Z such that $p_Z(1) = \alpha$. If $Z = 1$, the source is reliable and we accept m as a representation of our beliefs on Θ; otherwise, the source is not reliable and our beliefs on Θ are represented by the vacuous mass function $m_?$. Smets then showed that the discounting operation can be deduced from fundamental operations of DS theory. In this model, α has a clear interpretation as the probability that the source is reliable.

Possibilistic discounting. A notion of discounting was also introduced in approximate reasoning and possibility theory by Yager [32]. Let us assume that a piece of evidence on a variable X taking values in Θ can be expressed as "X is $\widetilde{A}$", where $\widetilde{A}$ is a fuzzy subset of Θ, and let α denote the "credibility" of the sources measured on the scale $[0,1]$. Upon consideration of the source's credibility, Yager [32] proposes to transform fuzzy set $\widetilde{A}$ into $\widetilde{A}' = D_p(\alpha, \widetilde{A})$ such that

$$\forall \theta \in \Theta, \quad \widetilde{A}'(\theta) = \delta(\alpha, \widetilde{A}(\theta)),$$

where δ is a *discounting function*, defined as a mapping from $[0,1]^2$ to $[0,1]$ verifying the following conditions:

1. $\forall a \in [0,1]$, $\delta(0,a) = 1$;
2. $\forall a \in [0,1]$, $\delta(1,a) = a$;

3. $\forall a \in [0,1]$, $\forall (\alpha_1, \alpha_2) \in [0,1]^2$, $\alpha_1 > \alpha_2 \Rightarrow \delta(\alpha_2, a) \geq \delta(\alpha_1, a)$;
4. $\forall (a,b) \in [0,1]^2$, $\forall \alpha \in [0,1]$, $a > b \Rightarrow \delta(\alpha, a) \geq \delta(\alpha, b)$.

Examples of discounting functions are $\delta(\alpha, a) = S(1-\alpha, a)$, where S is a t-conorm, and $\delta(\alpha, a) = a^\alpha$. It is clear that, if $\widetilde{A}' = D_p(\alpha, \widetilde{A})$, we have $\widetilde{A}'(\theta) \geq \widetilde{A}(\theta)$ for all $\theta \in \Theta$. Consequently, for all $A \subseteq \Theta$, $\Pi_{\widetilde{A}'}(A) \geq \Pi_{\widetilde{A}}(A)$, i.e., the possibilistic discounting operation transforms possibility measure $\Pi_{\widetilde{A}}$ into a less informative one.

4.2. Extension to random fuzzy sets

Both evidential and possibilistic discounting operations recalled in Section 4.1 can be extended to ERFSs.

Evidential discounting. Let $(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \widetilde{X})$ be an ERFS representing some piece of evidence. Let beliefs about reliability of the source of this evidence be represented by a binary random variable $Z \in \{0,1\}$ (such that $Z = 1$ means that the source is reliable) with probability distribution P_Z such that $P_Z(1) = \alpha$. A discounted RFS can be defined as

$$(\Omega \times \{0,1\}, \Sigma_\Omega \otimes 2^{\{0,1\}}, P \times P_Z, \Theta, \Sigma_\Theta, \widetilde{X}'),$$

where $\widetilde{X}'$ is the mapping from $\Omega \times \{0,1\}$ to $[0,1]^\Theta$ defined as

$$\begin{aligned} \widetilde{X}'(\omega, 1) &= \widetilde{X}(\omega) \\ \widetilde{X}'(\omega, 0) &= \Theta. \end{aligned}$$

We call this operation *evidential discounting* with discount rate $1-\alpha$ and we write $\widetilde{X}' = D_e(\alpha, \widetilde{X})$. We have, for all measurable subset $A \subseteq \Theta$,

$$\begin{aligned} Pl_{\widetilde{X}'}(A) &= \alpha \int \Pi_{\widetilde{X}(\omega)}(A) dP(\omega) + (1-\alpha)\Pi_\Theta(A) \\ &= \begin{cases} \alpha Pl_{\widetilde{X}}(A) + 1 - \alpha & \text{if } A \neq \emptyset \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

which parallels (16).

Possibilistic discounting. Alternatively, we may extend the notion of possibilistic discounting by considering the RFS

$$(\Omega, \Sigma_\Omega, P, \Theta, \Sigma_\Theta, \widetilde{X}''),$$

where $\widetilde{X}''$ is the mapping from Ω to $[0,1]^\Theta$ such that $\omega \mapsto \widetilde{X}''(\omega) = D_p(\alpha, \widetilde{X}(\omega))$, where D_p is a possibilitic discounting operation based on some discounting function δ. We call this operation *possibilistic discounting* with discount rate $1-\alpha$ and we write $\widetilde{X}'' = D_p(\alpha, \widetilde{X})$. As, for all $\omega \in \Omega$ and all $\theta \in \Theta$, $\widetilde{X}''(\omega)(\theta) \geq \widetilde{X}(\omega)(\theta)$, we have, for all $A \subseteq \Theta$, $\Pi_{\widetilde{X}''}(A) \geq \Pi_{\widetilde{X}}(A)$ and, for all measurable $A \subseteq \Theta$,

$$Pl_{\widetilde{X}''}(A) = \int \Pi_{\widetilde{X}''}(A) dP(\omega) \geq Pl_{\widetilde{X}}(A).$$

Application to GRFNs. Let $\widetilde{X} \sim \widetilde{N}(\mu, \sigma^2, h)$ be a GRFN. The result of the evidential discounting operation with discount rate $1-\alpha$ is a mixture of $\widetilde{X}$ with proportion α and the vacuous GRFN with proportion $1-\alpha$,

$$\widetilde{X}' = D_e(\alpha, \widetilde{X}) \sim \alpha \widetilde{N}(\mu, \sigma^2, h) + (1-\alpha)\widetilde{N}(*, *, 0).$$

Let us now consider possibilistic discounting with discounting function $\delta(a) = a^\alpha$. Let $\widetilde{X}' = D_p(\alpha, \widetilde{X})$. We have

$$\widetilde{X}'(\omega)(x) = [\widetilde{X}'(\omega)(x)]^\alpha = \left[\exp\left(-\frac{h}{2}(x - M(\omega))^2\right)\right]^\alpha = \exp\left(-\frac{\alpha h}{2}(x - M(\omega))^2\right).$$

Consequently, $\widetilde{X}'' = D_p(\alpha, \widetilde{X}) \sim \widetilde{N}(\mu, \sigma^2, \alpha h)$.

In both cases, the belief function induced by the discounted RFS tends to the vacuous belief function when $\alpha \to 0$, i.e., for any event $A \subset \mathbb{R}$,

$$\lim_{\alpha \to 0} Bel_{D_e(\alpha, \widetilde{X})}(A) = 0 \quad \text{and} \quad \lim_{\alpha \to 0} Bel_{D_p(\alpha, \widetilde{X})}(A) = 0$$

and, for any nonempty event A,

$$\lim_{\alpha \to 0} Pl_{D_e(\alpha, \widetilde{X})}(A) = 1 \quad \text{and} \quad \lim_{\alpha \to 0} Pl_{D_p(\alpha, \widetilde{X})}(A) = 1.$$

In practice, the discount rate α can sometimes be elicited from experts, but it will more often be determined as the solution of an optimization problem, as will be illustrated in Section 5.

Example 9. *Let us consider the GRFN $\widetilde{X} \sim \widetilde{N}(0, 1, 2)$. Figure 12 shows the lower/upper cdfs and contour functions of $D_p(\alpha, \widetilde{X})$ and $D_e(\alpha, \widetilde{X})$ for $\alpha \in \{1, 0.8, 0.5, 0.2\}$. We can see that, for a given α, evidential discounting is more "drastic" than possibilistic discounting, as the plausibility of any nonempty event A after evidential discounting with rate $1-\alpha$ is at least equal to $1-\alpha$.*

4.3. Combination of discounted GRFNs

In classical DS theory, a classical strategy for fusing evidence from unreliable sources is to combine discounted mass functions by Dempster's rule (see, e.g., [15]). Discounting has, in particular, the effect of reducing the conflict. A similar approach can be adopted for combining unreliable GRFNs. We discuss this approach below considering, successively, possibilistic and evidential discounting.

Conjunctive combination after possibilistic discounting. As possibilistic discounting yields GRFNs, its use with the conjunctive combination operation introduced in Section 3 is immediate. Applying this operation to n GRFNs $\widetilde{X}_1, \ldots, \widetilde{X}_n$ after possibilistic discounting, we get a new GRFN denoted as

$$\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}(\widetilde{X}_1, \ldots, \widetilde{X}_n) = \mathcal{C}_{\boldsymbol{R}}\left(D_p(\alpha_1, \widetilde{X}_1), \ldots, D_p(\alpha_n, \widetilde{X}_n)\right), \tag{17}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$. We can observe that the new $\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ depends on two vector parameters: the correlation matrix $\boldsymbol{R}$ and the vector of coefficients $\boldsymbol{\alpha}$.

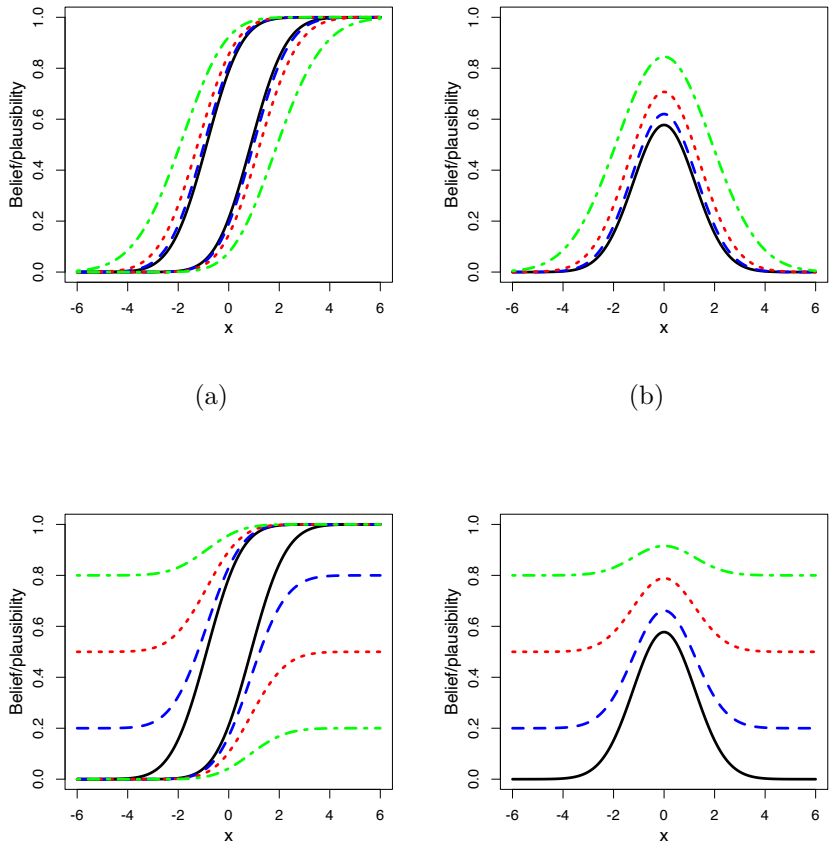


Figure 12: Lower/upper cdfs and contour functions for discounted GRFNs (Example 9), for $\alpha = 1$ (solid black lines), $\alpha = 0.8$ (dashed blue lines), $\alpha = 0.5$ (dotted red lines), $\alpha = 0.2$ (dashed-dotted green lines). (a): lower/upper cdfs, possibilistic discounting; (b) contour functions, possibilistic discounting; (c): lower/upper cdfs, evidential discounting; (d) contour functions, evidential discounting.

Conjunctive combination after evidential discounting. As before, let us assume that we have n GRFNs $\widetilde{X}_1, \ldots, \widetilde{X}_n$ with $\widetilde{X}_i \sim \widetilde{N}(\mu_1, \sigma_i^2, h_i)$. Let $Z_i \sim \mathcal{B}(\alpha_i)$ be a Bernoulli random variable indicating the reliability of $\widetilde{X}_i$. After evidential discounting, we obtain n mGRFNs $\widetilde{X}'_1, \ldots, \widetilde{X}'_n$ with

$$\widetilde{X}'_i = D_e(\alpha_i, \widetilde{X}_i) \sim \alpha_i \widetilde{N}(\mu_i, \sigma_i^2, h_i) + (1 - \alpha_i)\widetilde{N}(*, *, 0).$$

Each mGRFN $\widetilde{X}'_i$ can, alternatively, be written as

$$\widetilde{X}'_i = GFN(M_i, h_i Z_i),$$

where M_i is a GRV with mean μ_i and variance σ_i^2, and it is assumed that M_i and Z_i are independent. To combine these n mGRFNs, we need to make assumptions about the joint distribution of random vectors $\boldsymbol{M} = (M_1, \ldots, M_n)^T$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$. As in Section 3, we assume that $\boldsymbol{M}$ has a multivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ and covariance matrix $\boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{\sigma})\boldsymbol{R}\operatorname{diag}(\boldsymbol{\sigma})$, where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ and $\boldsymbol{R}$ is a correlation matrix. We further assume that $\boldsymbol{Z}$ has a probability mass function $p_{\boldsymbol{Z}}$ with marginals $p_{Z_i}(1) = \alpha_i$, and that $\boldsymbol{M}$ and $\boldsymbol{Z}$ are independent. Under these assumptions, the following theorem gives the expression of the combined mGRFNs.

Theorem 4. *Under the above assumptions, the conjunctive combination of $\widetilde{X}_1, \ldots, \widetilde{X}_n$ after evidential discounting is*

$$\widetilde{X} = \mathcal{C}^{(e)}_{\boldsymbol{R}, p_{\boldsymbol{Z}}}(\widetilde{X}_1, \ldots, \widetilde{X}_n) \sim \sum_{\boldsymbol{z} \in \{0,1\}^n} \pi_{\boldsymbol{z}} \widetilde{N}(\mu_{\boldsymbol{z}}, \sigma_{\boldsymbol{z}}, h_{\boldsymbol{z}}), \tag{18a}$$

with $h_{\boldsymbol{z}} = \sum_{i=1}^n z_i h_i$,

$$\mu_{\boldsymbol{z}} = \boldsymbol{h}_{\boldsymbol{z}}^{*T}(\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A}_{\boldsymbol{z}})^{-1}\boldsymbol{\mu}, \quad \sigma_{\boldsymbol{z}}^2 = \boldsymbol{h}_{\boldsymbol{z}}^{*T}(\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A}_{\boldsymbol{z}})^{-1}\boldsymbol{\Sigma}\boldsymbol{h}_{\boldsymbol{z}}^*, \tag{18b}$$

$$\pi_{\boldsymbol{z}} = \frac{(1 - \kappa_{\boldsymbol{z}})p_{\boldsymbol{Z}}(\boldsymbol{z})}{\sum_{\boldsymbol{z}'}(1 - \kappa_{\boldsymbol{z}'})p_{\boldsymbol{Z}}(\boldsymbol{z}')}, \tag{18c}$$

where $\boldsymbol{h}_{\boldsymbol{z}}^* = \boldsymbol{h}_{\boldsymbol{z}}/h_{\boldsymbol{z}}$, $\boldsymbol{h}_{\boldsymbol{z}} = (h_1 z_1, \ldots, h_n z_n)^T$,

$$\boldsymbol{A}_{\boldsymbol{z}} = \begin{cases} \boldsymbol{0}_{n,n} & \text{if } z_1 = \ldots = z_n = 0, \\ \operatorname{diag}(\boldsymbol{h}_{\boldsymbol{z}}) - \boldsymbol{h}_{\boldsymbol{z}}\boldsymbol{h}_{\boldsymbol{z}}^T/h_{\boldsymbol{z}} & \text{otherwise,} \end{cases} \tag{18d}$$

and

$$\kappa_{\boldsymbol{z}} = 1 - |\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\mu}^T \boldsymbol{A}_{\boldsymbol{z}}[\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A}_{\boldsymbol{z}}]^{-1}\boldsymbol{\mu}\right). \tag{18e}$$

Proof. See Appendix D. □

This new conjunctive operation is, thus, indexed by a correlation matrix $\boldsymbol{R}$ and a probability mass function $p_{\boldsymbol{Z}}$ in $\{0,1\}^n$. It essentially computes a mixture of GRFNs obtained by conjunctively combining the GRFNs $\widetilde{X}_i$ such that $z_i = 1$, for all combinations of reliable sources encoded in $\boldsymbol{z}$, with proportions given by (18c). For small n, the probability mass function $p_{\boldsymbol{Z}}$ can be learnt from data (cf. Section 5) or specified by the user. For large n, some additional assumptions can be useful to reduce the number of parameters. Hereafter, we describe three models with n degrees of freedom, based on different simplifying assumptions:

Model 1: RVs Z_i are assumed to be independent, so that

$$p_{\boldsymbol{Z}}(\boldsymbol{z}) = \prod_{i=1}^{n} P(Z_i = z_i) = \prod_{i=1}^{n} \alpha_i^{z_i}(1-\alpha_i)^{1-z_i}.$$

This assumption is often plausible, and it reduces the number of parameters from $2^n - 1$ to n. The corresponding operator can be denoted by $\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$. It has the same number of parameters as $\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ in (17). This model includes as a special cases the conjunctive combination of the n GRFNs (when $\alpha_1 = \ldots = \alpha_n = 1$), as well as the conjunctive combination of any subset of the n GRFNs (when $\alpha_i = 1$ for $i \in I \subset \{1, \ldots, n\}$ and $\alpha_i = 0$ for $i \notin I$).

Model 2: we assume that either there is exactly one reliable source, or all sources are reliable, i.e., the number of reliable sources is either 1 or n. Let β_i be the probability that source i is the only reliable source, and let β_{n+1} be the probability that all sources are reliable. These numbers must satisfy the constraint $\sum_{i=1}^{n+1} \beta_i = 1$, and we have $\alpha_i = \beta_i + \beta_{n+1}$. In that case, the combined mGRFN is a mixture of the $\widetilde{X}_i$'s and their conjunctive combination,

$$\widetilde{X} = \mathcal{C}^{(e)}_{\boldsymbol{R},p_{\boldsymbol{Z}}}(\widetilde{X}_1, \ldots, \widetilde{X}_n) \sim \sum_{i=1}^{n} \pi_i \widetilde{N}(\mu_i, \sigma_i, h_i) + \pi_{n+1} \widetilde{N}\left(\mu_c, \sigma_c, \sum_{i=1}^{n} h_i\right),$$

with μ_c and σ_c given by (13), and

$$\pi_i = \frac{\beta_i}{\sum_{j=1}^{n} \beta_j + (1-\kappa_{\boldsymbol{R}})\beta_{n+1}}, \quad \pi_{n+1} = \frac{(1-\kappa_{\boldsymbol{R}})\beta_{n+1}}{\sum_{j=1}^{n} \beta_j + (1-\kappa_{\boldsymbol{R}})\beta_{n+1}},$$

where $\kappa_{\boldsymbol{R}}$ is the degree of conflict of the conjunctive combination given by (14). This operator is interesting, because it includes mixtures of the n GRFNs as well as the conjunctive combination as special cases. We can observe that it behaves similarly to a mixture combination of the n GRFNs when their degree of conflict $\kappa_{\boldsymbol{R}}$ is close to 1. This operator will be denoted by $\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$, with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{n+1})$.

Model 3: we assume that either there are exactly $n-1$ reliable sources, or all sources are reliable, i.e., the number of reliable sources is either $n-1$ or n. Let β_i be the probability all sources except source i are reliable, and let β_{n+1} be the probability that

all sources are reliable. These numbers must satisfy the constraint $\sum_{i=1}^{n+1} \beta_i = 1$, and we have $\alpha_i = 1 - \beta_i$, $i = 1, \ldots, n$. Denoting by $\boldsymbol{z}_{(-i)}$ the vector with all components equal to 1, except component i equal to 0, the combined mGRFN can be written as

$$\widetilde{X} = \mathcal{C}^{(e)}_{\boldsymbol{R},p_{\boldsymbol{Z}}}(\widetilde{X}_1, \ldots, \widetilde{X}_n) \sim \sum_{i=1}^{n} \pi_i \widetilde{N}(\mu_{\boldsymbol{z}_{(-i)}}, \sigma_{\boldsymbol{z}_{(-i)}}, h_{\boldsymbol{z}_{(-i)}}) + \pi_{n+1} \widetilde{N}\left(\mu_c, \sigma_c, \sum_{i=1}^{n} h_i\right),$$

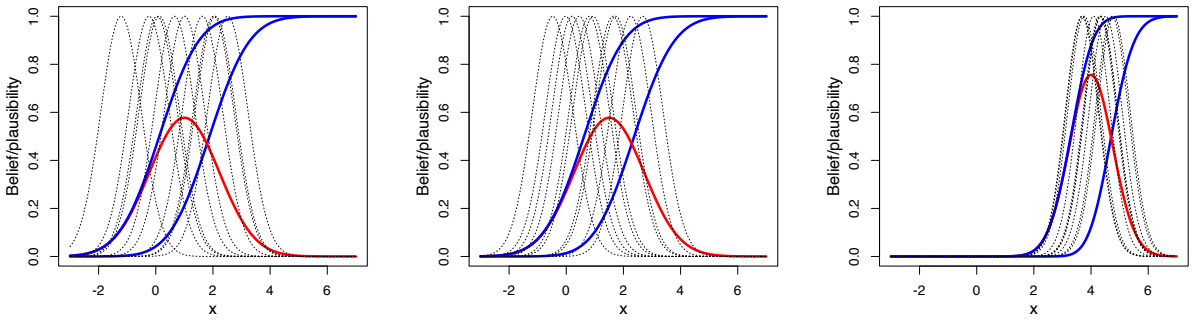
where $h_{\boldsymbol{z}_{(-i)}} = \sum_{j \neq i} h_j$, $\mu_{\boldsymbol{z}_{(-i)}}$ and $\sigma_{\boldsymbol{z}_{(-i)}}$ are defined by (18b), μ_c and σ_c given by (13), and

$$\pi_i = \frac{(1 - \kappa_{\boldsymbol{R}}^{-i})\beta_i}{\sum_{j=1}^{n}(1 - \kappa_{\boldsymbol{R}}^{-j})\beta_j + (1 - \kappa_{\boldsymbol{R}})\beta_{n+1}}, \quad \pi_{n+1} = \frac{(1 - \kappa_{\boldsymbol{R}})\beta_{n+1}}{\sum_{j=1}^{n}(1 - \kappa_{\boldsymbol{R}}^{-j})\beta_j + (1 - \kappa_{\boldsymbol{R}})\beta_{n+1}},$$

where $\kappa_{\boldsymbol{R}}$ is the degree of conflict of $\widetilde{X}_1, \ldots, \widetilde{X}_n$ given by (14), and $\kappa_{\boldsymbol{R}}^{-i}$ is the degree of conflict of the $n-1$ GRFNs $\widetilde{X}_j$, $j \neq i$. This fusion operator will be denoted by $\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$.

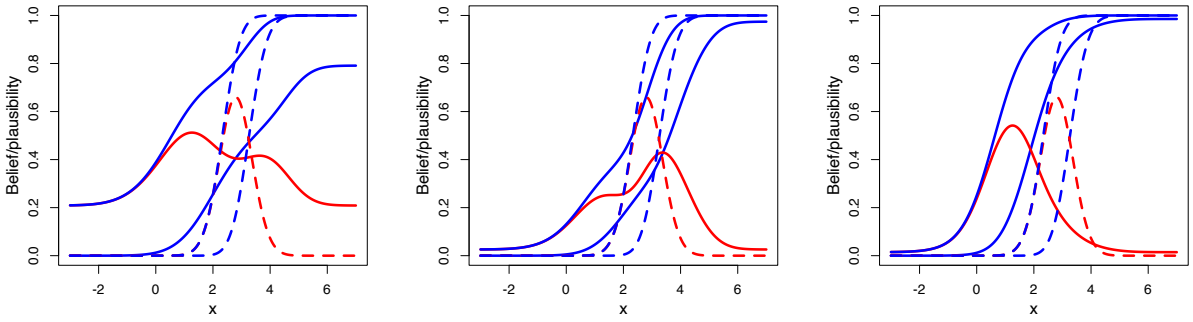
Other rules could be defined based on more complex assumptions. For instance, we could assume that q out of n sources are reliable, for some $q \in \{2, \ldots, n-1\}$. However, the three simple operations $\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$, $\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$ and $\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$ defined above, together with $\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ given by (17), have the advantage of depending on a small number of parameters and can be expected to be sufficient for most applications. In practice, different models can be compared empirically, as will be shown in Section 5.

Example 10. *Figure 13 shows three GRFNs* $\widetilde{X}_1 \sim \widetilde{N}(1, 1, 2)$, $\widetilde{X}_2 \sim \widetilde{N}(1.5, 1, 2)$ *and* $\widetilde{X}_3 \sim \widetilde{N}(4, 0.5^2, 3)$. *We assume these GRFNs to be independent, i.e., the correlation matrix is equal to the identity matrix* $\boldsymbol{I}$. *Combinations of these three GRFNs under Models 1, 2 and 3 with different parameter values are illustrated in Figures 14 to 16. Figure 14 shows the combined GRFN* $\mathcal{C}^{(e)}_{\boldsymbol{I},\boldsymbol{\alpha}}$ *using Model 1 (independence) with (a):* $\boldsymbol{\alpha} = (0.5, 0.5, 0.5)$*; (b):* $\boldsymbol{\alpha} = (0.5, 0.9, 0.9)$*, and (c):* $\boldsymbol{\alpha} = (0.9, 0.9, 0.5)$. *Figures 15 and 16 display, respectively, the combined GRFNs using Models 2 and 3 with (a):* $\boldsymbol{\beta} = (1/3, 1/3, 1/3, 0)$*; (b):* $\boldsymbol{\beta} = (0.1/3, 0.1/3, 0.1/3, 0.9)$*, and (c):* $\boldsymbol{\beta} = (0.01/3, 0.01/3, 0.01/3, 0.99)$.



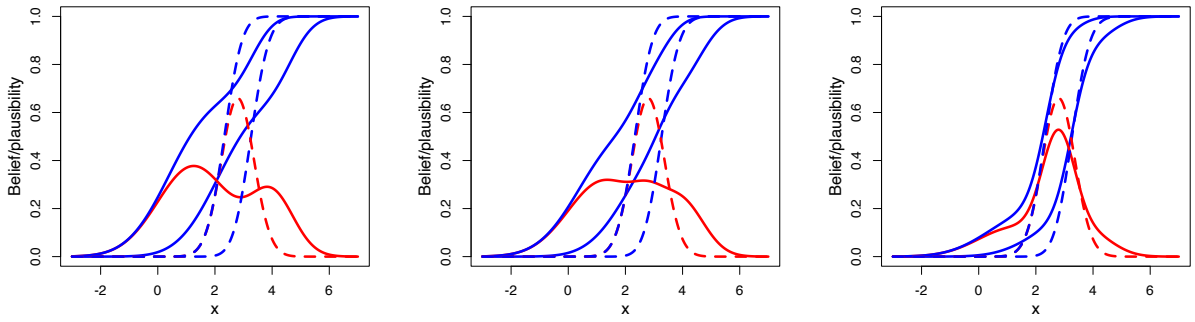
(a) (b) (c)

Figure 13: The three GRFNs of Example 10. Each GRFN is represented by its contour function (red curve), its lower and upper cdfs (blue curves) and 10 realizations (dotted curves).



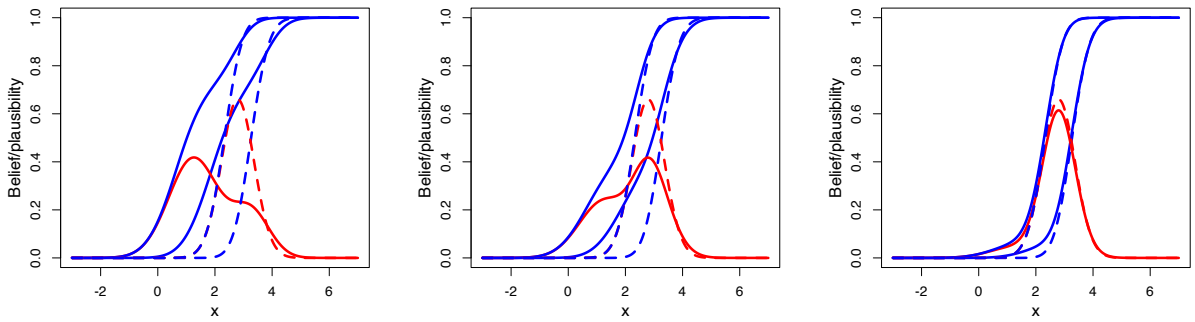
(a) (b) (c)

Figure 14: Combined GRFN $\mathcal{C}^{(e)}_{\boldsymbol{I},\boldsymbol{\alpha}}$ using Model 1 (independence) with (a): $\boldsymbol{\alpha} = (0.5, 0.5, 0.5)$; (b): $\boldsymbol{\alpha} = (0.5, 0.9, 0.9)$; (c): $\boldsymbol{\alpha} = (0.9, 0.9, 0.5)$. Each GRFN is represented by its contour function (red curve) and by its lower and upper cdfs (blue curves). The conjunctive combination is represented by dashed curves.



(a) (b) (c)

Figure 15: Combined GRFN $\mathcal{M}_{\boldsymbol{I},\boldsymbol{\beta}}$ using Model 2 with (a): $\boldsymbol{\beta} = (1/3, 1/3, 1/3, 0)$; (b): $\boldsymbol{\beta} = (0.1/3, 0.1/3, 0.1/3, 0.9)$; (c): $\boldsymbol{\beta} = (0.01/3, 0.01/3, 0.01/3, 0.99)$. Each GRFN is represented by its contour function (red curve) and by its lower and upper cdfs (blue curves). The conjunctive combination is represented by dashed curves.



(a) (b) (c)

Figure 16: Combined GRFN $\overline{\mathcal{M}}_{\boldsymbol{I},\boldsymbol{\beta}}$ using Model 3 with (a): $\boldsymbol{\beta} = (1/3, 1/3, 1/3, 0)$; (b): $\boldsymbol{\beta} = (0.1/3, 0.1/3, 0.1/3, 0.9)$; (c): $\boldsymbol{\beta} = (0.01/3, 0.01/3, 0.01/3, 0.99)$. Each GRFN is represented by its contour function (red curve) and by its lower and upper cdfs (blue curves). The conjunctive combination is represented by dashed curves.

5. Numerical experiments

To demonstrate the usefulness of the fusion methods introduced in this paper, we consider their application to the combination of predictions of a quantitative variable in machine learning. The ENNreg model introduced in [8, 11] is a neural network model for regression that quantifies prediction uncertainty using GRFNs. The mean of the output GRFN indicates the most plausible value of the response for the given input, while its standard deviation and precision correspond, respectively, to aleatory and epistemic uncertainty.

In the experiment described in this section, we consider the combination of several models trained with different, but overlapping sets of predictors. In the real world, model combination can be useful in different situations: for instance, different models may be trained using different datasets or with different sets of inputs to account for possible unavailability of some predictors (due to sensor failure or other reason). We can remark that model combination in evidential machine learning is an important topic that, in full generality, goes beyond the scope of this paper. Here, our objective is to illustrate the practical use of the combination operators studied in this paper, and to show that they can be optimized to enhance prediction performance. Experimental settings will first be described in Section 5.1; results will then be reported and discussed in Section 5.2.

5.1. Experimental settings

We considered four regression datasets: the Boston dataset (506 observations, 13 inputs) from the R package `MASS` [30], two datasets from the UCI Machine Learning Repository[2]: Concrete Compressive Strength (1030 observations, 8 inputs) and Energy efficiency (769 observations, 8 inputs), and the kin8nm dataset downloaded from the OpenML web site[3] (8192 obervations, 8 inputs). For each dataset, we proceeded as follows:

1. The data were split randomly into training, validation and test sets containing, respectively, 60%, 20% and 20% of the observations;
2. Three overlapping subsets of $\lfloor p/2 \rfloor + 1$ inputs were randomly selected out of the p inputs;
3. An ENNreg model with $K = 30$ prototypes was trained with each of the three input subsets, yielding three models (hyperparameters ξ and ρ were tuned by five-fold cross-validation);
4. The parameters of the following five combination operators were fitted using the validation set:
   (a) $\mathcal{C}_{\boldsymbol{R}}$ (conjunctive combination with correlation matrix $\boldsymbol{R}$);
   (b) $\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ (conjunctive combination with correlation matrix $\boldsymbol{R}$ and possibilistic discounting with coefficients $\boldsymbol{\alpha}$);
   (c) $\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ (conjunctive combination with correlation matrix $\boldsymbol{R}$ and model-1 evidential discounting with coefficients $\boldsymbol{\alpha}$);

[2] `https://archive.ics.uci.edu/ml/`.
[3] `https://www.openml.org`

(d) $\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$ (conjunctive combination with correlation matrix $\boldsymbol{R}$ and model-2 evidential discounting with coefficients $\boldsymbol{\beta}$);
(e) $\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$ (conjunctive combination with correlation matrix $\boldsymbol{R}$ and model-3 evidential discounting with coefficients $\boldsymbol{\beta}$).

5. The quality of the combined predictions using each of the optimized combination operators as well as the product-intersection $\mathcal{C}_{\boldsymbol{I}}$ was assessed on the test set.

The whole process was repeated 30 times (with different random data partitions and input subsets). The ENNreg model was trained by minimizing a generalized log-likelihood (GLL) loss, as explained in [8]. Similarly, the combination operators were fitted in Step 4 above by minimizing the following loss function,

$$\mathcal{L}_\epsilon(y, \widetilde{Y}) = -\frac{1}{2}\left[\ln Bel_{\widetilde{Y}}([y-\epsilon, y+\epsilon]) + \ln Pl_{\widetilde{Y}}([y-\epsilon, y+\epsilon])\right], \tag{19}$$

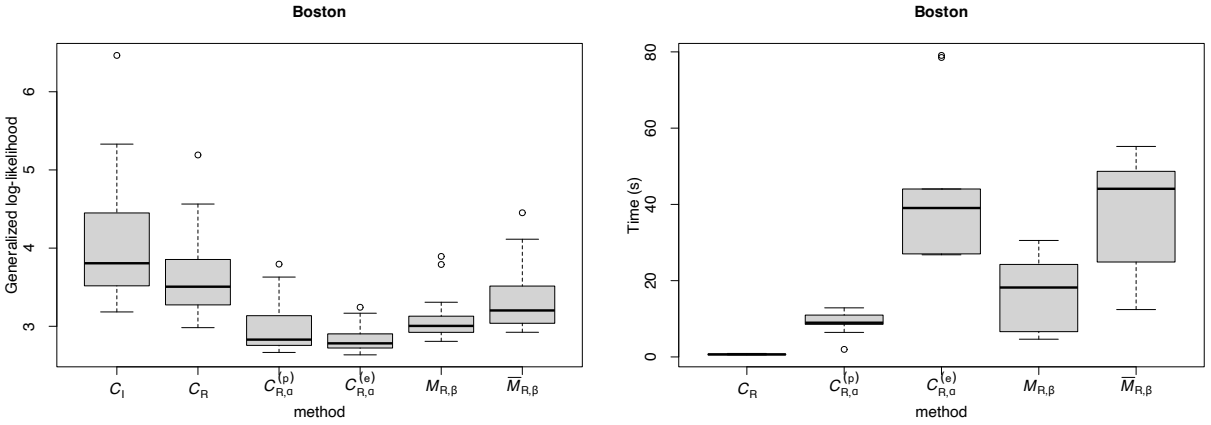
where y is the true value of the response, $\widetilde{Y}$ the prediction (a random fuzzy number), and ϵ a hyperparameter that was fixed to 0.01 times the standard deviation of the response Y. The rationale for criterion (19) is that a good prediction should assign a high degree of belief to a small interval $[y-\epsilon, y+\epsilon]$ centered on the true value, and a low degree of belief to the complement of this interval, i.e., a high plausibility degree to $[y-\epsilon, y+\epsilon]$. The same criterion was used to assess the quality of the combined predictions in Step 5.

5.2. Results

Boxplots of test GLL values for the four datasets are shown in Figures 17a, 18a, 19a and 20a. We can see that optimizing the correlation coefficient allows for a performance improvement over the product-intersection combination, and that both possibilistic and evidential discounting bring further improvements. The best performing combination operators were conjunctive combination with possibilistic discounting $\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ and conjunctive combination with evidential discounting and independent reliability assumption $\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$, with the latter outperforming the former on the Boston and Energy datasets.

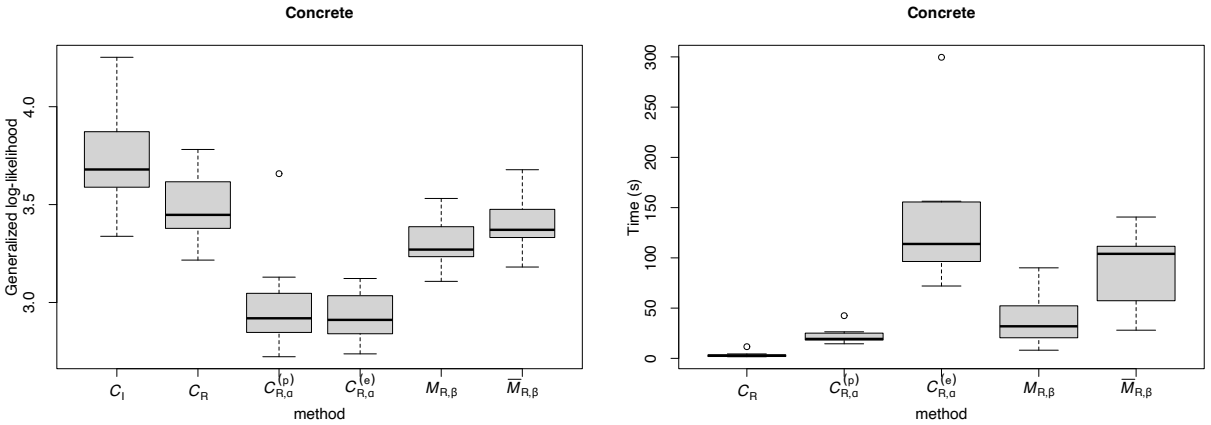
P-values of pairwise paired t-tests with Bonferroni correction for the comparison of GLL values are shown in Tables 3 to 6. We can see that most observed differences are highly significant. In particular, $\mathcal{C}_{\boldsymbol{R}}$ significantly outperforms $\mathcal{C}_{\boldsymbol{I}}$ on the four datasets, and the additional performance gain obtained by possibilistic or evidential discounting is also highly significant. Combination operator $\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ significantly outperforms $\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ on the Boston (p-value=0.031) and Energy (p-value=0.00339) datasets, while the two methods yield similar results on the Concrete and kin8nm datasets. Operators $\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$ and $\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$ perform significantly worse than $\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ and $\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ for the four datasets. Whether these operators can be useful for prediction combination or other information fusion applications remains an open question.

Finally, boxplots of computing times are displayed in Figures 17b, 18b, 19b and 20b. The reported times are the times to learn the parameters of the fusion operators using the validation set. We can see that learning the correlation matrix alone is quite fast, while



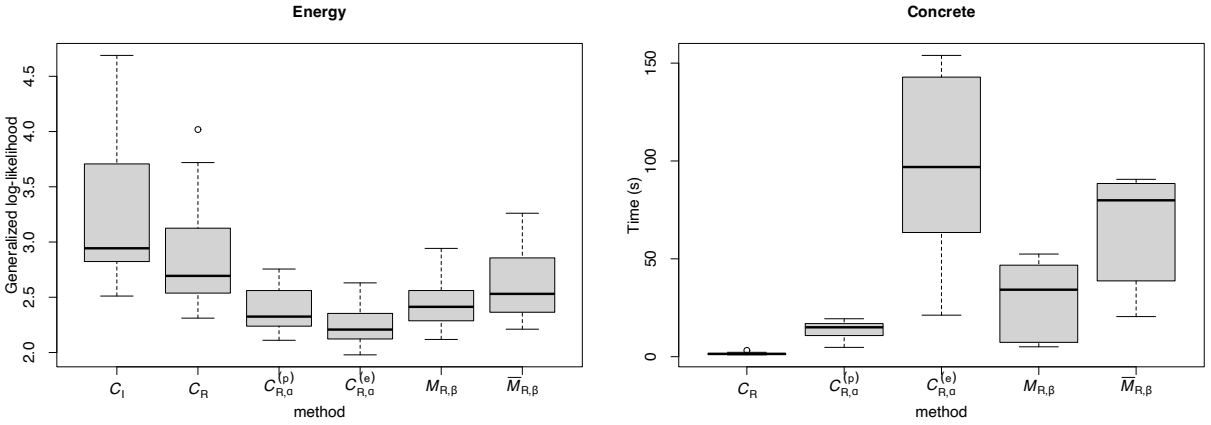
(a) (b)

Figure 17: Generalized log-likelihood (a) and learning time (b) for the Boston dataset with different fusion methods.



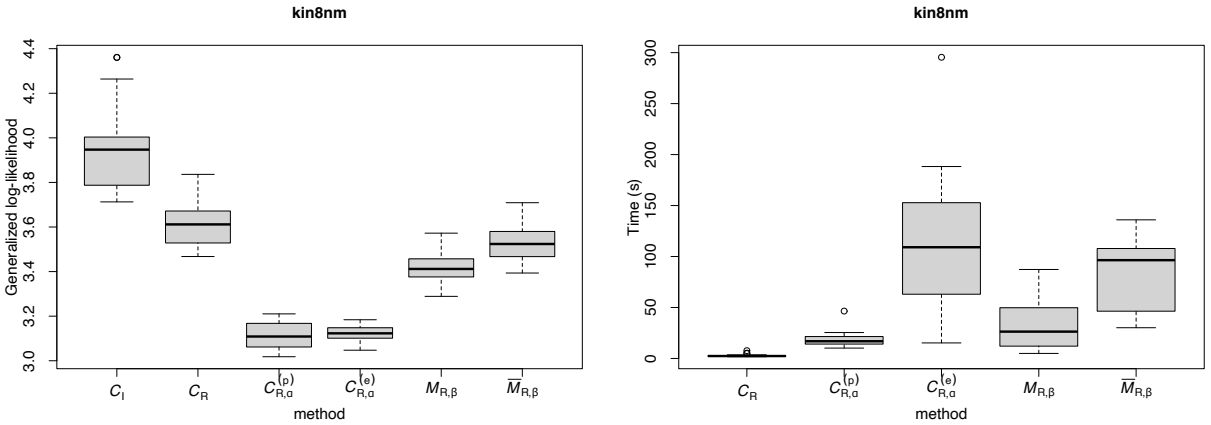
(a) (b)

Figure 18: Generalized log-likelihood (a) and learning time (b) for the Concrete dataset with different fusion methods.



(a) (b)

Figure 19: Generalized log-likelihood (a) and learning time (b) for the Energy dataset with different fusion methods.



(a) (b)

Figure 20: Generalized log-likelihood (a) and learning time (b) for the kin8nm dataset with different fusion methods.

learning discounting coefficients takes significantly more time. We can also observe that operators computing a mixture of GRFNs ($\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$, $\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$ and $\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$) are more computationally costly than operator $\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ based on possibilistic discounting. Since $\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$ performs almost as well as $\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$, this may be an argument in favor of the former when computational resources are limited.

Table 3: P-values with Bonferroni correction of pairwise paired t-tests between the GLL values obtained with different fusion methods for the Boston dataset.

	$\mathcal{C}_{\boldsymbol{I}}$	$\mathcal{C}_{\boldsymbol{R}}$	$\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	$\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$	$\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$
$\mathcal{C}_{\boldsymbol{R}}$	1.5×10^{-7}	-	-	-	-
$\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	1.1×10^{-9}	6.9×10^{-10}	-	-	-
$\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$	3.4×10^{-9}	2.2×10^{-8}	5.2×10^{-8}	-	-
$\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$	4.7×10^{-9}	1.2×10^{-7}	3.3×10^{-9}	2.9×10^{-6}	-
$\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	1.1×10^{-11}	1.2×10^{-13}	0.031	0.129	3.6×10^{-13}

Table 4: P-values with Bonferroni correction of pairwise paired t-tests between the GLL values obtained with different fusion methods for the Concrete dataset.

	$\mathcal{C}_{\boldsymbol{I}}$	$\mathcal{C}_{\boldsymbol{R}}$	$\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	$\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$	$\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$
$\mathcal{C}_{\boldsymbol{R}}$	3.3×10^{-14}	-	-	-	-
$\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	-	-	-
$\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$	$< 2 \times 10^{-16}$	1.4×10^{-13}	$< 2 \times 10^{-16}$	-	-
$\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$	8.5×10^{-16}	3.6×10^{-8}	$< 2 \times 10^{-16}$	1.4×10^{-13}	-
$\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	1	8.9×10^{-14}	$< 2 \times 10^{-16}$

Table 5: P-values with Bonferroni correction of pairwise paired t-tests between the GLL values obtained with different fusion methods for the Energy dataset.

	$\mathcal{C}_{\boldsymbol{I}}$	$\mathcal{C}_{\boldsymbol{R}}$	$\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	$\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$	$\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$
$\mathcal{C}_{\boldsymbol{R}}$	1.4×10^{-10}	-	-	-	-
$\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	4.1×10^{-10}	7.2×10^{-8}	-	-	-
$\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$	1.2×10^{-8}	2.5×10^{-5}	3.6×10^{-12}	-	-
$\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$	9.3×10^{-10}	4.1×10^{-5}	3.9×10^{-9}	0.00094	-
$\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	5.3×10^{-10}	2.6×10^{-7}	0.00339	0.03339	3.6×10^{-7}

Table 6: P-values with Bonferroni correction of pairwise paired t-tests between the GLL values obtained with different fusion methods for the kin8nm dataset.

	$\mathcal{C}_{\boldsymbol{I}}$	$\mathcal{C}_{\boldsymbol{R}}$	$\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	$\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$	$\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$
$\mathcal{C}_{\boldsymbol{R}}$	2.4×10^{-15}	-	-	-	-
$\mathcal{C}^{(e)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	-	-	-
$\mathcal{M}_{\boldsymbol{R},\boldsymbol{\beta}}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	-	-
$\overline{\mathcal{M}}_{\boldsymbol{R},\boldsymbol{\beta}}$	$< 2 \times 10^{-16}$	2.6×10^{-15}	$< 2 \times 10^{-16}$	3.6×10^{-13}	-
$\mathcal{C}^{(p)}_{\boldsymbol{R},\boldsymbol{\alpha}}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$

6. Conclusions

The theory of epistemic random fuzzy sets generalizes DS and possibility theories by considering random fuzzy sets as a model of uncertain and imprecise information. In this approach, the basic mechanism for combining information is the product-intersection rule, which extends and plays the same role as Dempster's rule in DS theory. Among other benefits, this new approach makes it possible to define parametric families of random fuzzy numbers (GRFNs and extensions), allowing one to represent and reason with evidence about continuous variables.

As Dempster's rule, the product-intersection rule is, however, too restrictive to apply to all situations encountered in practice. Specifically, it is based on strong assumptions about the independence and reliability of the pieces of evidence being combined. In this paper, we have shown how these assumptions can be relaxed when combining GRFNs and related random fuzzy sets. First, we have extended the product-intersection rule by providing formulas for the combination of any number of GRFNs whose dependence is represented by an arbitrary correlation matrix (instead of the identity matrix assumed in the product-intersection rule). To accommodate frequent situations in which the dependence structure of the evidence is not precisely known, we have proposed a minimum-conflict combination operation in which the correlation matrix is determined to minimize the degree of conflict between pieces of evidence.

To address the issue of partial reliability of the evidence, we have introduced two discounting operations for random fuzzy sets: possibilistic discounting "weakens" a RFS by making the fuzzy focal sets more imprecise, whereas evidential discounting mixes the RFS with a vacuous one as done by the discounting operation in classical DS theory. When applied to a GRFN, the former operation yields another GRFN with a smaller precision parameter, while the latter yield a mixture of two GRFNs, one of which is vacuous. We have then studied different operations based on the conjunctive combination (with a given correlation matrix) of discounted evidence. When used within a combination mechanism, evidential discounting appears to be richer as it allows for the definition of different combination operations based on different assumptions about the reliability of items of evidence. These operations all amount to computing a mixture of conjunctive combinations of subsets of GRFNs.

To illustrate the application of the combination operations introduced in this paper, we

have considered the combination of predictions made by the ENNreg model, a regression neural network model introduced in [8], which quantifies prediction uncertainty using a GRFN. We have considered the particular situation in which different models are trained with partially overlapping subsets of features. We have shown that optimizing the parameters of different combination operators using a validation set allows for improved performance on test data, as compared to the baseline product-intersection operation. For this application, the best fusion strategy was found to be the combination of evidentially discounted GRFNs with the reliability independence assumption.

The methods introduced in this paper considerably extend the toolbox of techniques for combining evidence about real variables. One of the promising applications of these techniques is evidential machine learning. For instance, parameterized combination operations could be embedded in evidential regression models such as ENNreg [8]; alternatively, combination schemes for ensembles of models built in different ways could be devised. The combination of knowledge elicited from multiple experts is another research avenue that will be explored in future work.

References

[1] P. A. Bromiley. Products and convolutions of Gaussian probability density functions. Technical Report 2003-003, TINA, 2014. `https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.583.3007&rep=rep1&type=pdf`.

[2] M. E. G. V. Cattaneo. Combining belief functions issued from dependent sources. In J. M. Bernard, T. Seidenfeld, and M. Zaffalon, editors, *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications (ISIPTA'03)*, pages 133–147, Lugano, Switzerland, 2003. Carleton Scientific.

[3] A. Chen, X. Tang, B. Cheng, and J. He. Multi-source monitoring information fusion method for dam health diagnosis based on Wasserstein distance. *Information Sciences*, 632:378–389, 2023.

[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. Maximum likelihood estimation from uncertain data in the belief function framework. *IEEE Transactions on Knowledge and Data Engineering*, 25(1):119–130, 2013.

[7] 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.

[8] T. Denœux. Quantifying prediction uncertainty in regression using random fuzzy sets: the ENNreg model. *IEEE Transactions on Fuzzy Systems*, 31:3690–3699, 2023.

[9] 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.

[10] T. Denœux. Combination of dependent Gaussian random fuzzy numbers. In *Proceedings of the Eighth International Conference on Belief Functions (BELIEF 2024)*, Belfast, UK, September 2024. Springer International Publishing. `https://hal.science/hal-04621423`.

[11] T. Denœux. *evreg: Evidential Regression*, 2024. R package version 1.0.4, `https://CRAN.R-project.org/package=evreg`.

[12] T. Denœux. Uncertainty quantification in regression neural networks using likelihood-based belief functions. In *Proceedings of the Eighth International Conference on Belief Functions (BELIEF 2024)*, Belfast, UK, September 2024. Springer International Publishing. `https://hal.science/hal-04621414`.

[13] T. Denœux and V. Kreinovich. Algebraic product is the only "and-like"-operation for which normalized intersection is associative: A proof. In *Fifth International Conference on Artificial Intelligence and Computational Intelligence (AICI 2024)*, Hanoi, Vietnam, January 2024. `https://hal.science/hal-04436177`.

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

[15] Z. Elouedi, K. Mellouli, and P. Smets. Assessing sensor reliability for multisensor data fusion within the Transferable Belief Model. *IEEE Transactions on Systems, Man and Cybernetics B*, 34(1):782–787, 2004.

[16] S. Ferson, V. Kreinovitch, L. Ginzburg, D. S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002-4015, Sandia National Laboratories, Albuquerque, NM, 2003.

[17] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, New-York, 2nd edition, 2013.

[18] Z. Ji, J. Tian, H. Chen, and S. Liu. A new method for weighted fusion of evidence based on the unified trust distribution mechanism and the reward-punishment mechanism. *Information Sciences*, 629:798–815, 2023.

[19] R. Lucchetti and L. Pedini. The spherical parametrisation for correlation matrices and its computational advantages. *Computational Economics*, 2023. `https://doi.org/10.1007/s10614-023-10467-3`.

[20] D. Mercier, B. Quost, and T. Denœux. Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*, 9(2):246–258, 2008.

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

[22] K. B. Petersen and M. S. Pedersen. The matrix cookbook, November 2012. `http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html`.

[23] F. Pichon, T. Denœux, and D. Dubois. Relevance and truthfulness in information correction and fusion. *International Journal of Approximate Reasoning*, 53(2):159–175, 2012.

[24] F. Pichon, D. Dubois, and T. Denœux. Quality of information sources in information fusion. In É. Bossé and G. L. Rogova, editors, *Information Quality in Information Fusion and Decision Making*, pages 31–49. Springer International Publishing, Cham, 2019.

[25] J. C. Pinheiro and D. M. Bates. Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6(3):289–296, 1996.

[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] G. Shafer. Dempster's rule of combination. *International Journal of Approximate Reasoning*, 79:26–40, 2016.

[29] P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.

[30] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002.

[31] X. Xu, H. Guo, Z. Zhang, S. Yu, L. Chang, F. Steyskal, and G. Brunauer. A cloud model-based interval-valued evidence fusion method and its application in fault diagnosis. *Information Sciences*, 658:119995, 2024.

[32] R. R. Yager. On considerations of credibility of evidence. *International Journal of Approximate Reasoning*, 7(1):45–72, 1992.

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

[34] 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. Proof of Proposition 2

From [22, page 42], we have (with our notations), for $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ PD,

$$\exp\left((\boldsymbol{x}-\boldsymbol{m}_1)^T\boldsymbol{H}_1(\boldsymbol{x}-\boldsymbol{m}_1)\right)\cdot\exp\left((\boldsymbol{x}-\boldsymbol{m}_2)^T\boldsymbol{H}_2(\boldsymbol{x}-\boldsymbol{m}_2)\right) = C\ \exp\left((\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{H}(\boldsymbol{x}-\boldsymbol{m})\right)$$

with $\boldsymbol{H}=\boldsymbol{H}_1+\boldsymbol{H}_2$, $\boldsymbol{m}=\boldsymbol{H}^{-1}(\boldsymbol{H}_1\boldsymbol{m}_1+\boldsymbol{H}_2\boldsymbol{m}_2)$, and

$$C=\exp\left(-\frac{1}{2}(\boldsymbol{m}_1-\boldsymbol{m}_2)^T(\boldsymbol{H}_1^{-1}+\boldsymbol{H}_2^{-1})^{-1}(\boldsymbol{m}_1-\boldsymbol{m}_2)\right).$$

Consequently, the following equation holds:

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

and the height of the product intersection between $\mathrm{GFV}(\boldsymbol{m}_1,\boldsymbol{H}_1)$ and $\mathrm{GFV}(\boldsymbol{m}_1,\boldsymbol{H}_2)$ is

$$\mathrm{hgt}\left(\mathrm{GFV}(\boldsymbol{m}_1,\boldsymbol{H}_1)\cdot\mathrm{GFV}(\boldsymbol{m}_1,\boldsymbol{H}_2)\right)=\exp\left(-\frac{1}{2}(\boldsymbol{m}_1-\boldsymbol{m}_2)^T(\boldsymbol{H}_1^{-1}+\boldsymbol{H}_2^{-1})^{-1}(\boldsymbol{m}_1-\boldsymbol{m}_2)\right),$$

which is the result given in [9]. We can remark that

$$(\boldsymbol{H}_1^{-1}+\boldsymbol{H}_2^{-1})^{-1}=\left[(\boldsymbol{H}_1^{-1}\boldsymbol{H}_2+\boldsymbol{I}_n)\boldsymbol{H}_2^{-1}\right]^{-1}=\boldsymbol{H}_2(\boldsymbol{I}_n+\boldsymbol{H}_1^{-1}\boldsymbol{H}_2)^{-1},$$

which gives us the expression in (11). However, this result was established assuming $\boldsymbol{H}_2$ to be PD. We now need to show it is still true when $\boldsymbol{H}_2$ is only PSD. Assume that $\boldsymbol{H}_2$ is PSD. We can observe [17, page 432] that $\boldsymbol{H}_2$ is the limit when $k\to\infty$ of PD matrices $\boldsymbol{H}_2^{(k)}=\boldsymbol{H}_2+k^{-1}\boldsymbol{I}_n$. From the continuity of the product, maximum and matrix inverse operations, we have

$$\begin{aligned}\mathrm{hgt}\left(\mathrm{GFV}(\boldsymbol{m}_1,\boldsymbol{H}_1)\cdot\mathrm{GFV}(\boldsymbol{m}_1,\boldsymbol{H}_2)\right)&=\lim_{k\to\infty}\mathrm{hgt}\left(\mathrm{GFV}(\boldsymbol{m}_1,\boldsymbol{H}_1)\cdot\mathrm{GFV}(\boldsymbol{m}_1,\boldsymbol{H}_2^{(k)})\right)\\&=\lim_{k\to\infty}\boldsymbol{H}_2^{(k)}(\boldsymbol{I}_n+\boldsymbol{H}_1^{-1}\boldsymbol{H}_2^{(k)})^{-1}\\&=\boldsymbol{H}_2(\boldsymbol{I}_n+\boldsymbol{H}_1^{-1}\boldsymbol{H}_2)^{-1}.\end{aligned}$$

Appendix B. Proof of Lemma 1

We first remark that (10) can be written in vector form as

$$\widetilde{F}(\boldsymbol{m})=\exp\left[-\frac{1}{2}\boldsymbol{m}^T\boldsymbol{A}\boldsymbol{m}\right],$$

with $\boldsymbol{m}=(m_1,\ldots,m_n)^T$. For all $\boldsymbol{m}\in\mathbb{R}^n$, $\widetilde{F}(\boldsymbol{m})\leq 1$ or, equivalently, $\boldsymbol{m}^T\boldsymbol{A}\boldsymbol{m}\geq 0$; consequently, $\boldsymbol{A}$ is PSD. We consider two cases.

- Case 1: $\boldsymbol{\Sigma}$ is PD. Random vector $\boldsymbol{M}$ then has a density, and its conditional density given fuzzy event $\widetilde{F}$ is

$$f(\boldsymbol{m}|\widetilde{F}) = \frac{f(\boldsymbol{m})\widetilde{F}(\boldsymbol{m})}{\int f(\boldsymbol{m})\widetilde{F}(\boldsymbol{m})d\boldsymbol{m}}. \tag{B.1}$$

From Proposition 2, $f(\boldsymbol{m}|\widetilde{F}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{m}-\widetilde{\boldsymbol{\mu}})^T\widetilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{m}-\widetilde{\boldsymbol{\mu}})\right)$ with

$$\widetilde{\boldsymbol{\mu}} = \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{0}) = \widetilde{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \quad \text{and} \quad \widetilde{\boldsymbol{\Sigma}}^{-1} = \boldsymbol{\Sigma}^{-1} + \boldsymbol{A}.$$

Consequently, the conditional distribution of $\boldsymbol{M}$ given $\widetilde{F}$ is normal with mean $\widetilde{\boldsymbol{\mu}}$ and PD covariance matrix $\widetilde{\boldsymbol{\Sigma}}$. Writing $\widetilde{\boldsymbol{\Sigma}}^{-1} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A})$, we get

$$\widetilde{\boldsymbol{\Sigma}} = (\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{\Sigma} \quad \text{and} \quad \widetilde{\boldsymbol{\mu}} = (\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{\mu}. \tag{B.2}$$

- Case 2: The distribution of $\boldsymbol{M}$ is degenerate. We can remark that the expressions in (B.2) still make sense, as $\boldsymbol{I}_n + \boldsymbol{\Sigma}\boldsymbol{A}$ is nonsingular. Indeed, as $\boldsymbol{A}$ is symmetric and PSD, it has a symmetric and PSD square root $\boldsymbol{A}^{1/2}$. Matrix $\boldsymbol{A}^{1/2}\boldsymbol{\Sigma}\boldsymbol{A}^{1/2}$ is symmetric and, for any $\boldsymbol{m} \in \mathbb{R}^n$,

$$\boldsymbol{m}^T\boldsymbol{A}^{1/2}\boldsymbol{\Sigma}\boldsymbol{A}^{1/2}\boldsymbol{m} = (\boldsymbol{A}^{1/2}\boldsymbol{m})^T\boldsymbol{\Sigma}(\boldsymbol{A}^{1/2}\boldsymbol{m}) \geq 0,$$

hence it is PSD. Now, $\boldsymbol{A}\boldsymbol{\Sigma} = \boldsymbol{A}^{1/2}(\boldsymbol{A}^{1/2}\boldsymbol{\Sigma})$ and $(\boldsymbol{A}^{1/2}\boldsymbol{\Sigma})\boldsymbol{A}^{1/2}$ have the same eigenvalues (as for any square matrices P and Q, PQ and QP have the same eigenvalues). Therefore, the eigenvalues of $\boldsymbol{A}\boldsymbol{\Sigma}$ are all nonnegative. Let us now consider an eigenvector $\boldsymbol{u}$ of $\boldsymbol{I} + \boldsymbol{\Sigma}\boldsymbol{A}$ and the corresponding eigenvalue λ. We have

$$(\boldsymbol{I} + \boldsymbol{\Sigma}\boldsymbol{A})\boldsymbol{v} = \boldsymbol{v} + \boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \Rightarrow \boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{v} = (\lambda - 1)\boldsymbol{v}.$$

Hence $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{A}\boldsymbol{\Sigma}$ and $\lambda - 1 \geq 0$, i.e., $\lambda \geq 1$. As the eigenvalues of $\boldsymbol{I} + \boldsymbol{\Sigma}\boldsymbol{A}$ are all positive, $\boldsymbol{I} + \boldsymbol{\Sigma}\boldsymbol{A}$ is nonsingular. As, when singular, $\boldsymbol{\Sigma}$ is the limit of a sequence of nonsingular covariance matrices, (B.2) remains true by continuity.

Appendix C. Proof of Theorem 3

From Proposition 1, the product intersection of $\widetilde{X}_1(M_1), \ldots, \widetilde{X}_n(M_n)$ is the GRFN with precision $\sum_{i=1}^n h_i$ and random mode

$$M_c = \frac{\sum_{i=1}^n h_i M_i}{\sum_{i=1}^n h_i} = \boldsymbol{h}^{*T}\boldsymbol{M}.$$

From Lemma 1, $\boldsymbol{M}$ has, conditionally on $\widetilde{F}$, a multivariate normal distribution with mean $\widetilde{\boldsymbol{\mu}}$ and covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ given, respectively, by (12a) and (12b). Hence, $M_c \sim N(\mu_c, \sigma_c^2)$ with $\mu_c = \boldsymbol{h}^{*T}\widetilde{\boldsymbol{\mu}}$ and $\sigma_c^2 = \boldsymbol{h}^{*T}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{h}^*$.

Now, assuming random vector $\boldsymbol{M}$ to have a density, the degree of conflict between the n GRFNs is defined as one minus the denominator on the right-hand side of (B.1). The term on the left-hand side is

$$A = (2\pi)^{n/2}|\widetilde{\boldsymbol{\Sigma}}|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{m}-\widetilde{\boldsymbol{\mu}})^T\widetilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{m}-\widetilde{\boldsymbol{\mu}})\right).$$

The numerator on the right-hand side of (B.1) is

$$B = (2\pi)^{n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{m}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}-\boldsymbol{\mu})\right)\exp\left(-\frac{1}{2}(\boldsymbol{m}^T\boldsymbol{A}\boldsymbol{m}\right).$$

From Proposition 2,

$$B = (2\pi)^{n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{A}[\boldsymbol{I}+\boldsymbol{\Sigma}\boldsymbol{A}]^{-1}\boldsymbol{\mu}\right)\exp\left(-\frac{1}{2}(\boldsymbol{m}-\widetilde{\boldsymbol{\mu}})^T\widetilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{m}-\widetilde{\boldsymbol{\mu}})\right).$$

Consequently, noticing that $|\widetilde{\boldsymbol{\Sigma}}|/|\boldsymbol{\Sigma}| = |\widetilde{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| = |\boldsymbol{I}+\boldsymbol{\Sigma}\boldsymbol{A}|^{-1}$, we get

$$\begin{aligned}\kappa_{\boldsymbol{R}}(\widetilde{X}_1,\ldots,\widetilde{X}_n) &= 1-\frac{B}{A}\\ &= 1-\left(\frac{|\widetilde{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}|}\right)^{1/2}\exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{A}[\boldsymbol{I}+\boldsymbol{\Sigma}\boldsymbol{A}]^{-1}\boldsymbol{\mu}\right)\\ &= 1-|\boldsymbol{I}+\boldsymbol{\Sigma}\boldsymbol{A}|^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{A}[\boldsymbol{I}+\boldsymbol{\Sigma}\boldsymbol{A}]^{-1}\boldsymbol{\mu}\right).\end{aligned}$$

As before, this expression remains valid when $\boldsymbol{\Sigma}$ is singular, by continuity.

Appendix D. Proof of Theorem 4

The proof parallels that of Theorem 4 in [14]. The conjunctive combination of $\widetilde{X}'_1,\ldots,\widetilde{X}'_n$ corresponds to the mapping

$$(\boldsymbol{m},\boldsymbol{z}) \mapsto GFN(m_1,h_1z_1)\odot\ldots\odot GFN(m_n,h_nz_n),$$

with $\boldsymbol{m}=(m_1,\ldots,m_n)^T$. The fuzzy subset $\widetilde{F}$ of consistent interpretations is

$$\widetilde{F}(\boldsymbol{m},\boldsymbol{z}) = \exp\left(-\frac{1}{2}\boldsymbol{m}^T\boldsymbol{A}_{\boldsymbol{z}}\boldsymbol{m}\right),$$

where matrix $\boldsymbol{A}_{\boldsymbol{z}}$ is given by (18d). The conditional probability distribution of $(\boldsymbol{M},\boldsymbol{Z})$ given $\widetilde{F}$ can be described as follows:

- The conditional distribution of $\boldsymbol{M}$ given $\boldsymbol{Z}=\boldsymbol{z}$ and $\widetilde{F}$ is given by Lemma 1: it is a multivariate normal distribution with mean

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{z}} = (\boldsymbol{I}_n+\boldsymbol{\Sigma}\boldsymbol{A}_{\boldsymbol{z}})^{-1}\boldsymbol{\mu}$$

and covariance matrix

$$\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} = (\boldsymbol{I}_n+\boldsymbol{\Sigma}\boldsymbol{A}_{\boldsymbol{z}})^{-1}\boldsymbol{\Sigma}.$$

- From Bayes' theorem and Theorem 3, the conditional probability distribution of $\boldsymbol{Z}$ given $\widetilde{F}$ is

$$\pi_{\boldsymbol{z}} = P(\boldsymbol{Z} = \boldsymbol{z} \mid \widetilde{F}) = \frac{P(\widetilde{F} \mid \boldsymbol{Z} = \boldsymbol{z}) p_{\boldsymbol{Z}}(\boldsymbol{z})}{P(\widetilde{F})} = \frac{(1 - \kappa_{\boldsymbol{z}}) p_{\boldsymbol{Z}}(\boldsymbol{z})}{\sum_{\boldsymbol{z}'} (1 - \kappa_{\boldsymbol{z}'}) p_{\boldsymbol{Z}}(\boldsymbol{z}')},$$

  where $\kappa_{\boldsymbol{z}}$ is given by (18e).

From Theorem 3, given $\boldsymbol{Z} = \boldsymbol{z}$ and $\widetilde{F}$, the mean of $\boldsymbol{M}$ is $\mu_z = \boldsymbol{h}_{\boldsymbol{z}}^{*T} \widetilde{\boldsymbol{\mu}}_{\boldsymbol{z}}$, and its variance is $\sigma_{\boldsymbol{z}}^2 = \boldsymbol{h}_{\boldsymbol{z}}^{*T} \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} \boldsymbol{h}_{\boldsymbol{z}}^{*}$. The result follows.

Appendix E. Proof of Proposition 3

Let $\mathbf{1}_n$ be the vector of length n whose components are all equal to 1. We assume $\boldsymbol{h} = h\mathbf{1}_n$, $\boldsymbol{\mu} = \mu \mathbf{1}_n$ and $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{J}_n$. We then have

$$\boldsymbol{A} = h\boldsymbol{I}_n - \frac{h}{n}\boldsymbol{J}_n$$

and

$$\boldsymbol{\Sigma A} = \sigma^2 \boldsymbol{J}_n \left(h\boldsymbol{I}_n - \frac{h}{n}\boldsymbol{J}_n \right) = \sigma^2 h \boldsymbol{J}_n - \sigma^2 \frac{h}{n} n \boldsymbol{J}_n = \mathbf{0}_{n,n}.$$

Consequently, from (13),

$$\mu_c = \frac{\mu}{n} \mathbf{1}_n^T \mathbf{1}_n = \mu \quad \text{and} \quad \sigma_c^2 = \frac{\sigma^2}{n^2} \mathbf{1}_n^T \boldsymbol{J}_n \mathbf{1}_n = \frac{\sigma^2}{n^2} n^2 = \sigma^2.$$

Finally, $\boldsymbol{\mu}^T \boldsymbol{A} \boldsymbol{\mu} = \mu^2 \mathbf{1}^T (h\boldsymbol{I}_n - \frac{h}{n}\boldsymbol{J}_n)\mathbf{1} = \mu^2 (hn - \frac{h}{n} n^2) = 0$, hence $\kappa_{\boldsymbol{J}_n}(\widetilde{X}, \ldots, \widetilde{X}) = 0$.