On Inference Mechanisms of Fuzzy-Probabilistic Inference Systems

CAO Nhung and HOLČAPEK Michal and VALÁŠEK Radek

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
30. dubna 22, 701 03 Ostrava, Czech Republic
E-mail: {nhung.cao, michal.holcapek, radek.valasek}@osu.cz

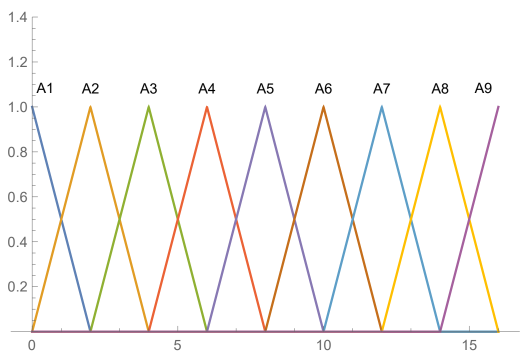
1 Outline of the studied problem

This work concerns the inference mechanism of *fuzzy–probabilistic inference systems* (FPIS), a special class of rule-based models in which antecedents encode fuzzy information while consequents represent conditional probability distributions of the output random variable. Such systems are particularly well suited for modeling uncertainty in time series responses. Specifically, we study a system of m rules taking the form ($k = 1, \ldots, m$) [2]:

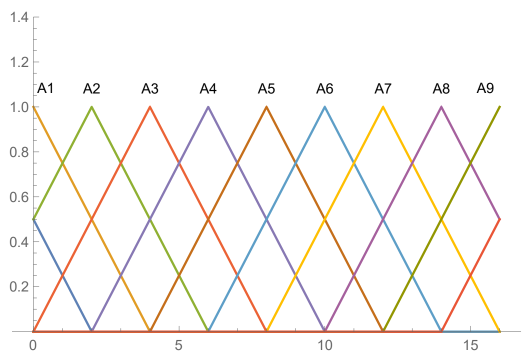
$$R_k := \text{IF } X \text{ is } A_k \text{ THEN } Y \text{ follows the probability distribution given by } Q_k(q), \tag{1}$$

where each consequent $Q_k(q)$, for $q \in [0, 1]$, is an *empirical quantile function* representing the underlying probability measure (distribution) of the output random variable Y conditioned on the fuzzy set A_k defined on a given universe X. We assume familiarity with basic probabilistic notions such as probability measures, cumulative distribution functions, and quantile functions.

In our work, fuzzy sets A_k used in the antecedents form a family Δ that covers the universe X, i.e., for every $x \in X$ there exists an $A_k \in \Delta$ with $A_k(x) > 0$. Recall that the fuzzy sets $A_1, \ldots, A_m \in \Delta$ are defined on the nodes $c_1, \ldots, c_m$, respectively, where each A_k attains its maximum. In practice, we often consider families of fuzzy sets that form a simple or generalized uniform fuzzy partition, see Fig. 1.



(a)



(b)

Figure 1: Examples of uniform fuzzy partitions: (a) simple, (b) generalized.

Acknowledgement The study is supported by the project "Research of Excellence on Digital Technologies and Wellbeing CZ.02.01.01/00/22_008/0004583" which is co-financed by the European Union.

https://www.doi.org/10.15452/978-80-7599-515-5.2026.06

Given a dataset $\{(x_i, y_i)\}_{i=1}^n \subseteq X \times \mathbb{R}$ and a generalized partition Δ, the q-weighted quantile $Q_k(q)$ associated with $A_k \in \Delta$ is obtained as the minimizer of (cf. [1]):

$$\Phi_q(z) = \sum_{i=1}^{n} \rho_q(y_i - z)\, A_k(x_i), \quad \rho_q(u) = \begin{cases} q|u|, & u > 0, \\ (1-q)|u|, & u \leq 0. \end{cases} \tag{2}$$

The inference mechanism that produces an empirical quantile function for an arbitrary input $x \in X$ is defined as a linear combination of the local quantile functions $Q_k(q)$.

Definition 1 *[1, 2] Let $\{(x_i, y_i)\}_{i=1}^n \subseteq X \times \mathbb{R}$ be given, together with a fuzzy partition $\Delta = \{A_1, \ldots, A_m\}$ of X. For each k, let Q_k be the weighted quantile function associated with A_k via the k-th rule in (1). Then, for any $x \in X$, the quantile function $Q_x : [0,1] \to \mathbb{R}$ is defined as*

$$Q_x(q) = \sum_{k=1}^{m} w_k(x) Q_k(q), \tag{3}$$

where $w_k(x) = A_k(x)/\sum_{j=1}^m A_j(x)$, $k = 1, \ldots, m$, denotes the normalized weight.

2 Inference mechanisms: theoretical foundation

Fuzzy rule-based systems express the relationship between two variables in a rough manner, while the inference mechanism provides a means to refine this rough relationship into a complete relationship, which can be used in practice. In [7], we experimentally analyzed and compared the standard weighted average of quantile functions as the inference mechanism for FPIS and other alternatives on synthetic and real datasets. This work left open a deeper theoretical analysis, particularly regarding the motivation and justification of the inference mechanisms used, including the original weighted average - all of which were introduced in a largely ad hoc manner. This knowledge gap motivates us to a deeper investigation of the foundation of the inference mechanism for FPIS, in which one part of the IF–THEN rules follows a fuzzy framework, while the other adopts a probabilistic one. For this purpose, we use the Wasserstein metric space $(\mathcal{P}_p(\mathbb{R}), W_p)$ of the order $p > 0$ on $\mathbb{R}$ (see, [4]), where

$$\mathcal{P}_p(\mathbb{R}) = \left\{ \mu \text{ is a probability measure on } (\mathbb{R}, \mathcal{B}(\mathbb{R})) : \int_{\mathbb{R}} |x|^p \, d\mu(x) < \infty \right\}$$

and W_p is the p-Wasserstein distance given for any probability distribution μ and ν as follows

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} |x - y|^p \, d\pi(x,y) \right)^{1/p},$$

where $\Pi(\mu, \nu)$ denotes the set of all couplings of μ and ν, i.e. the set of all probability measures on $\mathbb{R} \times \mathbb{R}$ with marginals μ and ν. Note that a probability measure μ on $\mathbb{R}$ is defined on the Borel algebra on $\mathbb{R}$ and can be equivalently expressed by the respective probability mass function f_μ, cumulative distribution function F_μ or quantile function Q_μ. For example, we have $F(x) = \mu((-\infty, x))$ and $Q_\mu(q) = F^{-1}(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\}$. The p-Wasserstein distance for probability measures on $\mathbb{R}$ can be simply expressed in terms of quantile functions as follows

$$W_p(\mu, \nu) = \left(\int_0^1 |Q_\mu(q) - Q_\nu(q)|^p \, dq \right)^{1/p} = \|Q_\mu - Q_\nu\|_{L_p(0,1)},$$

In what follows, we show that the inference mechanism for FPIS, originally defined ad hoc as the weighted average of quantile functions, can be rigorously formulated using the 2-Wasserstein distance. Specifically, it corresponds to the Wasserstein barycenter, also known as the Fréchet mean. This observation allows us to introduce a broader class of inference mechanisms based on p-Wasserstein distances. For convenience, we assume a generalized fuzzy partition $\Delta = \{A_1, \ldots, A_m\}$ of X, and define $w_k(x) = A_k(x)/\sum_{j=1}^m A_j(x)$ for any $k = 1, \ldots, m$ and $x \in X$.

Inference mechanism as the Wasserstein barycenter in $\mathcal{P}_2(\mathbb{R})$

Let $\mu_1, \ldots, \mu_m \in \mathcal{P}_2(\mathbb{R})$ with weights $w_i \geq 0$ such that $\sum_{i=1}^m w_i = 1$. The *Wasserstein barycenter* (or *Fréchet mean*) of $\mu_1, \ldots, \mu_m$ is a probability measure $\gamma \in \mathcal{P}_2(\mathbb{R})$ given as

$$\gamma = \arg \min_{\eta \in \mathcal{W}_2(\mathbb{R})} \sum_{i=1}^m w_i W_2^2(\eta, \mu_i).$$

It is well known that the Wasserstein barycenter of probability measures on $\mathbb{R}$ always exists and admits a closed-form expression in terms of quantile functions:

$$Q_\gamma(q) = \sum_{i=1}^m w_i Q_{\mu_i}(q), \tag{4}$$

where Q_{μ_i} denotes the quantile function of μ_i. Thus, the inference rule (3) coincides with the Wasserstein barycenter of empirical probability measures, which motivates the following equivalent formulation of the inference mechanism.

Definition 2 (Inference mechanism as the Fréchet mean) *Let $\{(x_i, y_i)\}_{i=1}^n \subseteq X \times \mathbb{R}$ be given data, and let Δ be a generalized fuzzy partition. We define $f^\star : X \to \mathcal{W}_2(\mathbb{R})$ as*

$$f^\star(x) = \arg \min_{\gamma \in \mathcal{P}_2(\mathbb{R})} \sum_{k=1}^m w_k(x)\, W_2^2(\gamma, \mu_k), \tag{5}$$

where $(\mu_k)_{k=1}^m$ are the probability measures with the empirical quantile functions $(Q_k)_{k=1}^m$ associated with the fuzzy sets $(A_k)_{k=1}^m$.

By the uniqueness of the Wasserstein barycenter in $\mathcal{P}_2(\mathbb{R})$, $f^\star(x)$ is well-defined, and its quantile function is given by (4). Assume that X is a metric space denoted as (X, d). The following proposition shows a local convergence property of f^* at nodes c_k of fuzzy sets $A_k \in \Delta$.

Proposition 3 *Let (X, d) be a metric space, and let $\Delta = \{A_1, \ldots, A_m\}$ be a generalized fuzzy partition such that each $A_k \in \Delta$, $k = 1, \ldots, m$, is K-Lipschitz continuous. Assume that the quantile function $Q_k \in L_2(0, 1)$ associated with $A_k \in \Delta$ satisfies $\|Q_k\|_{L_2(0,1)} \leq B$ for any $k = 1, \ldots, m$. Moreover, assume that $S(x) = \sum_{i=1}^m A_k(x) \geq S_0 > 0$ for every $x \in X$. Then for any c_k and $x \in X$, we have*

$$W_2\big(f^\star(x), f^\star(c_k)\big) \leq \frac{2mBK}{S_0}\, d(x, c_k).$$

Inference mechanism as the Wasserstein barycenter in $\mathcal{P}_p(\mathbb{R})$

The observation in the previous paragraph suggests a natural generalization of the inference mechanism in which the case $p = 2$ is extended to an arbitrary p. Particularly, the function $f^\star_p : X \to \mathcal{W}_p(\mathbb{R})$ is defined as the Wasserstein barycenter of $\mu_1, \dots, \mu_m$ in $\mathcal{P}_p(\mathbb{R})$ for which

$$f^\star_p(x) \in \arg \min_{\gamma \in \mathcal{P}_p(\mathbb{R})} \sum_{k=1}^{m} w_k(x)\, W_p^p(\gamma, \mu_k). \tag{6}$$

Unlike the case $p = 2$, the Wasserstein barycenter is generally not unique for arbitrary p. To define $f^*_p(x)$ properly, one must select a specific solution. For example, for $p = 1$ (the Wasserstein barycenter coincides with the Fréchet median [5]), a minimizing probability measure $\gamma_x = f^\star_1(x)$ has a quantile function

$$Q^*_{\gamma_x}(q) \in \arg \min_{z \in \mathbb{R}} \sum_{k=1}^{m} w_k(x)|z - Q_k(q)|,$$

where the set of minimizers for each q may form a closed interval. To guarantee the uniqueness of f^*_1, we select the left boundary (a canonical choice) of the closed interval of minimizers.

Definition 4 (Inference mechanism as the Fréchet median) *Let $\{(x_i, y_i)\}_{i=1}^n \subseteq X \times \mathbb{R}$ be given data, and let Δ be a generalized fuzzy partition. We define the function $Q^* : X \times [0, 1] \to \mathbb{R}$ as follows*

$$Q^*(x, q) = \min\Big\{\arg \min_{z \in \mathbb{R}} \sum_{k=1}^{m} w_k(x)\ |z - Q_k(q)|\Big\},$$

where $(Q_k)_{k=1}^m$ denote the empirical quantile functions associated with the fuzzy sets $(A_k)_{k=1}^m$. The function $f^*_1 : X \to \mathcal{W}_1(\mathbb{R})$ is defined by $f^*_1(x) = \gamma_x$ which quantile function is $Q^*(x, -)$.

The general definition of the inference mechanism for FPIS via the Wasserstein barycenter in $\mathcal{P}_p(\mathbb{R})$ provides a theoretical framework for further investigation of its properties, such as the local approximation property formulated in Lemma 3. Moreover, applying these theoretical results to practical implementations of the inference mechanism is an important direction for future research.

References

[1] Guerra, M.L., Sorini, L. and Stefanini, L.: *Quantile and expectile smoothing based on L1-norm and L2-norm fuzzy transforms.* International Journal of Approximate Reasoning **107** (2019) 17–43.

[2] Madrid, N.: *Significance measures for rules in probabilistic-fuzzy inference systems based on fuzzy transforms.* Fuzzy Sets and Systems **467** (2023) Article 108575.

[3] Brizzi, C., Friesecke, G. and Ried, T.: *p-Wasserstein barycenters.* Nonlinear Analysis **185** (2019) 113687.

[4] Panaretos, V.M. and Zemel, Y.: *An invitation to statistics in Wasserstein space.* Springer Nature, Cham (2020).

[5] Koenker, R.: *The median is the message: Toward the Fréchet median.* Journal de la Société Française de Statistique **147(2)** (2006) 61–64.

[6] Novák, V., Perfilieva, I., Dvořák, A.: *Insight into fuzzy modeling.* John Wiley & Sons (2016).

[7] Cao, N., Holčapek, M., Valášek, R., Madrid, N., Neděla, D.: *An Investigation of Alternative Methods for the Inference of Probabilistic-Fuzzy Systems.* In: International Conference on Modeling Decisions for Artificial Intelligence. Cham: Springer Nature Switzerland (2025) 104–116.