Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 435 (2022) 149–163

FUZZY sets and systems

www.elsevier.com/locate/fss

Some results on convergence and distributions of fuzzy random variables

Miriam Alonso de la Fuente [a], Pedro Terán [b,*]

[a] *Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, E-33071 Oviedo, Spain*

[b] *Escuela Politécnica de Ingeniería, Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, E-33071 Gijón, Spain*

Received 23 November 2020; received in revised form 3 May 2021; accepted 19 June 2021
Available online 25 June 2021

Abstract

Versions of several results from the theory of random variables are proved for fuzzy random variables: the Skorokhod representation theorem, the Vitali convergence theorem, the dominated convergence theorem, the continuous mapping theorem, existence of regular conditional distributions, and a few others.

Keywords: Borel measurability; Convex combination space; Fuzzy random variable; Random set; Skorokhod representation

1. Introduction

Fuzzy random variables aim at providing a sound framework for the simultaneous analysis of random and some non-random uncertainties. Thus, a fuzzy random variable is a random variable whose possible values are fuzzy sets instead of real numbers. There are two broad approaches, although this should be taken to be an orientative simplification rather than a normative distinction as it is sometimes understood, since it does not exhaust all the modeling possibilities with fuzzy random variables.

In the epistemic approach, the fuzzy sets represent descriptions or perceptions of unavailable underlying crisp values. For instance, it is recorded that the temperature was ‘high’ but the numerical value is unknown. This leads to a number of problems which often lend themselves to fuzzification techniques like Zadeh’s extension theorem [21]. As a concrete example, you model temperature as a normal $N(\mu, \sigma)$ random variable and the problem in estimating those parameters is that the data are ‘high’, ‘average’, ‘quite high’, ‘cool’, and so on. Zadeh’s extension allows one to define estimators which, appropriately, will also be fuzzy sets instead of real numbers. Since this approach gravitates towards the properties of the unobservable classical random variable, it is conceptually related to the statistical frameworks of censored data and coarse data [23].

In the ontic approach, the focus is on the postulated mechanism that produces observations in the form of fuzzy sets. Thus, giving the class of possible fuzzy values some amount of mathematical structure is unavoidable as a first

* Corresponding author.
E-mail addresses: alonsofmiriam@uniovi.es, uo233280@uniovi.es (M. Alonso de la Fuente), teranpedro@uniovi.es (P. Terán).

https://doi.org/10.1016/j.fss.2021.06.008
0165-0114/

step to generalizing traditional results to this more general type of data. Researchers have succeeded to do so in ways which are consistent with other parts of probability theory in spaces more general than the Euclidean space. Puri and Ralescu [25] defined fuzzy random variables so that their α-cuts are guaranteed to be random sets. Krätschmer [16] showed that this definition is equivalent to Borel measurability with respect to a certain metric, which brings fuzzy random variables in close contact with probability in metric spaces. Additional bridges with random elements of previously studied spaces include, for instance, càdlàg functions [4], Banach spaces [13,30], among others.

This paper regards fuzzy random variables as random elements of a metric space, specifically the space of d-dimensional (generalized) fuzzy intervals endowed with the d_p-metric. From that point of view, a natural topic is convergence in distribution. Although a number of *ad hoc* views of what a distribution function could be in the fuzzy setting have been presented, and regardless of their application to specific practical problems, their common flaw is that they do not enjoy the properties that make the cumulative distribution function a valuable tool in the theory of random variables [32]. Our view of convergence in distribution, then, disposes of distribution functions entirely. Instead, we just adopt the standard setting of probability in metric spaces which defines convergence in distribution via expectations of bounded continuous functions (weak convergence). In the case of ordinary random variables, the Helly–Bray theorem and its converse ensure that this is equivalent to the definition via convergence of cumulative distribution functions at the continuity points of the limit. But this equivalent definition is valid in an arbitrary metric space and can be directly applied to fuzzy random variables with the d_p-metric.

All results in this paper rely on two basic theorems presented in Section 3:

(a) The space of fuzzy sets with the d_p-metric is Borel measurable in its completion (see Theorem 3.1).
(b) A Skorokhod representation theorem which allows one to obtain an almost surely convergent sequence of fuzzy random variables from a sequence converging in distribution (see Theorem 3.5).

In fact, the proof of (b) uses (a). But since other previous Skorokhod theorems in metric spaces could be used instead, it seems more accurate to say that subsequent results are based on both (a) and (b).

The Skorokhod theorem is an essential tool for studying weak convergence. It allowed for very simple proofs of some theorems, and even to weaken their assumptions in some cases. One such application is presented in Section 4. There, we prove a version of Vitali's convergence theorem showing that convergence in distribution implies convergence of the expectations (both in the d_p-metric) provided the sequence satisfies a uniform integrability condition with respect to the same metric (Theorem 4.6). Note that, since it involves convergence in distribution, this is different from the result also called Vitali's convergence theorem which states that L^1-convergence is equivalent to convergence in probability plus uniform integrability.

Since a sequence dominated by an integrable function is uniformly integrable, a dominated convergence theorem follows easily (Theorem 4.7). This is close to Krätschmer's [20, Theorem 8.2] but the assumption is weakened to convergence in distribution. We also deduce another form of the dominated convergence theorem (Corollary 4.8).

Section 5 collects a number of short applications of results (a) and (b) above. We obtain the continuous mapping theorem for fuzzy random variables and generalize to the d-dimensional case the result in [2] that the distributions of fuzzy random variables are perfect probability measures. From (a) follows also that the space of fuzzy sets we work with is a Lusin space. That implies that known results in Lusin spaces can be applied, in particular the existence of regular conditional distributions, from which rigorous sense can be made of the concept of probabilities of a fuzzy random variable conditional on the value of another one. Another example is that every probability assessment on the elements of a countably generated sub-σ-algebra in the space of fuzzy sets must coincide with the probabilities of some fuzzy random variable. We also apply our recent Choquet theorem for random sets in Lusin metrizable spaces, motivated by the appearance of random sets of fuzzy sets in statistical estimation with fuzzy data.

2. Preliminaries

Let $\mathbb{E}$ be a topological space. We denote by $\mathcal{B}_{\mathbb{E}}$ its *Borel σ-algebra*, i.e., the σ-algebra generated by its open sets. A Borel measurable mapping with values in $\mathbb{E}$ will be generally called a *random element* of $\mathbb{E}$.

Let $\mathcal{F}_c(\mathbb{R}^d)$ be the space of fuzzy subsets of $\mathbb{R}^d$, i.e., functions $U : \mathbb{R}^d \to [0, 1]$ whose α-cuts ($\alpha \in [0, 1]$) are in $\mathcal{K}_c(\mathbb{R}^d)$, the space of all non-empty compact convex subsets of $\mathbb{R}^d$. Recall that the *α-cuts* of a fuzzy set U are

$$U_\alpha = \{x \in \mathbb{R}^d \mid U(x) \geq \alpha\}$$

for each $\alpha \in (0, 1]$, and U_0 denotes the closure of its support.

The *Hausdorff metric* in $\mathcal{K}_c(\mathbb{R}^d)$ is defined by

$$d_H(K, K') = \max\{\sup_{x\in K} \inf_{y\in K'} ||x - y||, \sup_{y\in K'} \inf_{x\in K} ||x - y||\}.$$

The *norm* or *magnitude* of K is

$$\|K\| = d_H(K, \{0\}).$$

For each $p \in [1, \infty)$, the metric d_p in $\mathcal{F}_c(\mathbb{R}^d)$, introduced in [13] and [25], is defined by

$$d_p(U, V) = \left[\int_{[0,1]} (d_H(U_\alpha, V_\alpha))^p \, d\alpha\right]^{1/p}.$$

We define

$$\mathcal{U}_c(\mathbb{R}^d) = \{U : \mathbb{R}^d \to [0, 1] \mid \forall\alpha \in (0, 1]\ U_\alpha \in \mathcal{K}_c(\mathbb{R}^d)\}.$$

We consider the following subset of $\mathcal{U}_c(\mathbb{R}^d)$ in which the definition of d_p still makes sense as a (finite) metric.

$$\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d) = \{U : \mathbb{R}^d \to [0, 1] \mid \forall\alpha \in (0, 1]\ U_\alpha \in \mathcal{K}_c(\mathbb{R}^d),\ d_p(U, I_{\{0\}}) < \infty\}.$$

Therefore

$$\mathcal{F}_c(\mathbb{R}^d) \subseteq \widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d) \subseteq \mathcal{U}_c(\mathbb{R}^d).$$

Given a probability space $(\Omega, \mathcal{A}, P)$, a mapping $X : \Omega \to \mathcal{K}_c(\mathbb{R}^d)$ is called a *random set* (also a *random compact convex set* in the literature) if X is measurable with respect to the Borel σ-algebra $\mathcal{B}_{\mathcal{K}_c(\mathbb{R}^d)}$ generated by the topology of the Hausdorff metric. A random set X is *integrably bounded* if $E[\|X\|] < \infty$, where $\|X\|$ maps each $\omega \in \Omega$ to $\|X(\omega)\|$.

In the sequel, unless explicitly stated, $(\Omega, \mathcal{A}, P)$ will be the underlying probability space for all random elements under consideration. Let $X : \Omega \to \mathcal{K}_c(\mathbb{R}^d)$ be an integrably bounded random set. The *Aumann expectation* of X is the compact set

$$E_A[X] = \{E[f] \mid f : \Omega \to \mathbb{R}, f \in L^1(\Omega, \mathcal{A}, P), f \in X \ P\text{-a.s.}\}.$$

A mapping $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R}^d)$ is called a *fuzzy random variable* if, for each $\alpha \in [0, 1]$, the α-cut mapping $\mathcal{X}_\alpha : \Omega \to \mathcal{K}_c(\mathbb{R}^d)$ defined by $\mathcal{X}_\alpha(\omega) = (\mathcal{X}(\omega))_\alpha$ for each $\omega \in \Omega$ is a random set (see [25]). We will denote by σ_L the natural σ-algebra in $\mathcal{F}_c(\mathbb{R}^d)$ with which a mapping is a fuzzy random variable if and only if it is measurable, i.e., the smallest σ-algebra that makes the mappings $U \in \mathcal{F}_c(\mathbb{R}^d) \mapsto U_\alpha \in \mathcal{K}_c(\mathbb{R}^d)$ measurable.

A fuzzy random variable $\mathcal{X}$ is called *integrably bounded* if $E[\|\mathcal{X}_0\|] < \infty$. Then the *expectation* of $\mathcal{X}$ is the unique fuzzy set $\widetilde{E}[\mathcal{X}] \in \mathcal{F}_c(\mathbb{R}^d)$ such that

$$(\widetilde{E}[\mathcal{X}])_\alpha = E_A[\mathcal{X}_\alpha]$$

for each $\alpha \in [0, 1]$ (see [25]). A fuzzy random variable with finite range will be called *simple*.

A sequence of probability measures $\{P_n\}_n$ on σ_L is said to *converge weakly in d_p* to a probability measure P if

$$\int f \, \mathrm{d}P_n \to \int f \, \mathrm{d}P$$

for every $f : \mathcal{F}_c(\mathbb{R}^d) \to \mathbb{R}$ which is d_p-continuous and bounded. A sequence $\{\mathcal{X}_n\}_n$ of fuzzy random variables converges *weakly* or *in distribution in d_p* to a fuzzy random variable X if their distributions P_{X_n} converge weakly to P_X, namely

$$E[f(X_n)] \to E[f(X)]$$

for each bounded d_p-continuous f.

The Lebesgue measure in [0, 1] will be denoted by ℓ. A class F of random variables is called *uniformly integrable* if given $\varepsilon > 0$ there exists $K \geq 0$ such that

$$E[|X| \cdot I_{\{|X|>K\}}] \leq \varepsilon$$

for all $X \in F$.

Let us mention a couple of results which will be used later on.

Proposition 2.1. *If $\mathcal{X}$ is a simple fuzzy random variable with range $\{U_1, \ldots, U_m\}$ then*

$$\widetilde{E}[\mathcal{X}] = p_1 \cdot U_1 + \ldots + p_m \cdot U_m,$$

where $p_i = P(\mathcal{X} = U_i)$.

Proposition 2.2. (Krätschmer [16, Theorem 6.4.(i)]) *Let $p \in [1, \infty)$. A mapping $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R}^d)$ is a fuzzy random variable if and only if it is a random element of the space $(\mathcal{F}_c(\mathbb{R}^d), d_p)$.*

3. Measurability and Skorokhod representation

In this section, we present the two main tools to be used in the remainder of the paper. First we prove that $\mathcal{F}_c(\mathbb{R}^d)$ is an element of the Borel σ-algebra of $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$. In [2], that problem was studied in $\mathbb{R}$ and answered positively by a laborious constructive process. While the constructive nature of the proof is interesting in itself, it does not seem to admit a generalization to $\mathbb{R}^d$. We will present here a shorter proof which relies on earlier results of Krätschmer [19]. The measurability result will then be used to obtain a Skorokhod representation theorem for fuzzy random variables with the d_p metric.

Theorem 3.1. *The set $\mathcal{F}_c(\mathbb{R}^d)$ is measurable in $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$.*

Proof. Let Σ denote the minimal σ-algebra generated by the mappings $L_\alpha : U \in \mathcal{U}_c(\mathbb{R}^d) \mapsto U_\alpha \in \mathcal{K}_c(\mathbb{R}^d)$ for $\alpha \in (0, 1]$. (Note that this is not the same as σ_L since Σ is a σ-algebra in $\mathcal{U}_c(\mathbb{R}^d)$, not $\mathcal{F}_c(\mathbb{R}^d)$.)

By [19, Theorem 1.(1) and Lemma 3.(3)], $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d) \in \Sigma$. Moreover, by [19, Theorem 1.(2)], $\mathcal{B}_{\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)}$ is the trace σ-algebra of Σ in $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$, i.e., it is the σ-algebra generated by the inclusion mapping $i : \widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d) \to \mathcal{U}_c(\mathbb{R}^d)$:

$$\mathcal{B}_{\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)} = \{i^{-1}(A) \mid A \in \Sigma\}. \tag{1}$$

Denoting by $\varphi : \mathcal{K}_c(\mathbb{R}^d) \to \mathbb{R}$ the continuous mapping $K \in \mathcal{K}_c(\mathbb{R}^d) \mapsto \|K\|$, we have

$$\begin{aligned}\mathcal{F}_c(\mathbb{R}^d) &= \bigcup_{k\in\mathbb{N}} \{\widetilde{U} \in \mathcal{U}_c(\mathbb{R}^d) \mid \|U_0\| \leq k\} \\ &= \bigcup_{k\in\mathbb{N}} \bigcap_{n\in\mathbb{N}} \{\widetilde{U} \in \mathcal{U}_c(\mathbb{R}^d) \mid \|U_{1/n}\| \leq k\} \\ &= \bigcup_{k\in\mathbb{N}} \bigcap_{n\in\mathbb{N}} (\varphi \circ L_{1/n})^{-1}((-\infty, k]) \in \Sigma.\end{aligned}$$

Accordingly, by (1),

$$\mathcal{F}_c(\mathbb{R}^d) = i^{-1}(\mathcal{F}_c(\mathbb{R}^d)) \in \mathcal{B}_{\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)}. \quad \square$$

The main tool to establish our Skorokhod theorem is Skorokhod's original version [28].

Theorem 3.2. (Skorokhod) *Let $(\mathbb{E}, d)$ be a complete separable metric space. Let P_n and P be probability measures on $\mathcal{B}_{\mathbb{E}}$, such that $P_n \to P$ weakly. Then there exist random elements $X_n, X : ([0, 1], \mathcal{B}_{[0,1]}, \ell) \to (\mathbb{E}, \mathcal{B}_{\mathbb{E}})$ for which*

(a) *The distributions of X_n and X are P_n and P, respectively.*

(b) $X_n(t) \to X(t)$ *for each* $t \in [0,1]$.

We denote by i_p the inclusion embedding of $\mathcal{F}_c(\mathbb{R}^d)$ into $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$.

Lemma 3.3. *For every* $p \in [1, \infty)$*, the mapping* i_p *is measurable.*

Proof. It follows from Theorem 3.1. □

Lemma 3.4. *Let* P_n, P *be probability distributions in* $(\mathcal{F}_c(\mathbb{R}^d), \sigma_L)$*. If* $P_n \to P$ *weakly in* d_p*, then* $P_n \circ i_p^{-1} \to P \circ i_p^{-1}$ *weakly in* $(\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d), d_p)$*.*

Proof. By the portmanteau lemma [14, Theorem 13.16, p. 254], weak convergence is equivalent to the property that $\liminf_n P_n(G) \geq P(G)$ for every open set G. Since, for every d_p-open subset $G \subseteq \widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$ its preimage $i_p^{-1}(G) = G \cap \mathcal{F}_c(\mathbb{R}^d)$ is d_p-open in $\mathcal{F}_c(\mathbb{R}^d)$ and the metric d_p in $\mathcal{F}_c(\mathbb{R}^d)$ generates the σ-algebra σ_L (by Proposition 2.2),

$$\liminf_{n\to\infty} P_n \circ i_p^{-1}(G) = \liminf_{n\to\infty} P_n(i_p^{-1}(G)) \geq P(i_p^{-1}(G)) = P \circ i_p^{-1}(G),$$

proving the weak convergence $P_n \circ i_p^{-1} \to P \circ i_p^{-1}$. □

We are ready now to prove a variant of the Skorokhod representation theorem for the d_p-metrics. Given a weakly convergent sequence, Skorokhod's theorem provides an almost surely convergent sequence which preserves the distributions of the variables involved.

It may be observed that the Skorokhod representation theorem in Dudley's version [7] for separable metric spaces would suffice to obtain the applications presented later on in this paper. The difference lies in the simplification of the probability space on which the new variables are defined. Specifically, if X_n, X are fuzzy random variables with distributions P_{X_n}, P_X, Dudley's variables are defined on the product of countably many copies of $\mathcal{F}_c(\mathbb{R}^d) \times [0,1]$ with the product measure $(P_X \otimes \ell) \otimes (P_{X_1} \otimes \ell) \otimes (P_{X_2} \otimes \ell) \otimes \ldots$. More recent versions of the theorem [36] involve even more complex probability spaces. In contrast, our variables are defined on $[0,1]$ with the uniform distribution.

Theorem 3.5. *Let* $p \in [1, \infty)$*. Let* P_n, P *be probability measures on* σ_L*, such that* $P_n \to P$ *weakly, and set* $\mathbb{P} = \ell$*. Then there exist fuzzy random variables* $\mathcal{X}_n, \mathcal{X} : ([0,1], \mathcal{B}_{[0,1]}, \mathbb{P}) \to (\mathcal{F}_c(\mathbb{R}^d), d_p)$*, such that*

(a) *The distributions of* $\mathcal{X}_n$ *and* $\mathcal{X}$ *are* P_n *and* P*, respectively.*
(b) $\mathcal{X}_n(t) \to \mathcal{X}(t)$ *in* d_p *for every* $t \in [0,1]$.

Proof. By Lemma 3.4, $P_n \circ i_p^{-1} \to P \circ i_p^{-1}$ weakly. By [18, Corollary 3.3], the space $(\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d), d_p)$ is separable and a completion of $(\mathcal{F}_c(\mathbb{R}^d), d_p)$. By Theorem 3.2, there exist $\mathcal{Y}_n, \mathcal{Y} : [0,1] \to (\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d), \mathcal{B}_{(\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d), d_p)})$ such that

- The distributions of $\mathcal{Y}_n$ and $\mathcal{Y}$ are $P_n \circ i_p^{-1}$ and $P \circ i_p^{-1}$, respectively.
- $\mathcal{Y}_n(t) \to \mathcal{Y}(t)$ for every $t \in [0,1]$.

Note that, by construction,

$$\mathbb{P}(\mathcal{Y} \in \mathcal{F}_c(\mathbb{R}^d)) = P(i_p^{-1}(\mathcal{F}_c(\mathbb{R}^d)) = P(\mathcal{F}_c(\mathbb{R}^d)) = 1$$

and analogously, $\mathbb{P}(\mathcal{Y}_n \in \mathcal{F}_c(\mathbb{R}^d)) = 1$ for each n.

Let N be a measurable null set containing the null set

$$\{\mathcal{Y} \notin \mathcal{F}_c(\mathbb{R}^d)\} \cup \bigcup_{n\in\mathbb{N}} \{\mathcal{Y}_n \notin \mathcal{F}_c(\mathbb{R}^d)\}$$

and set

$$\mathcal{X}_n(t) = \begin{cases} \mathcal{Y}_n(t) & \text{if } t \notin N \\ I_{\{0\}} & \text{if } t \in N \end{cases}$$

and

$$\mathcal{X}(t) = \begin{cases} \mathcal{Y}(t) & \text{if } t \notin N \\ I_{\{0\}} & \text{if } t \in N. \end{cases}$$

Next, we have to show that $\mathcal{X}_n$ and $\mathcal{X}$ are fuzzy random variables. For each $B \in \mathcal{B}_{(\mathcal{F}_c(\mathbb{R}^d),d_p)}$,

$$\begin{aligned} &\mathcal{X}^{-1}(B) = (N \cap \mathcal{X}^{-1}(B)) \cup (N^c \cap \mathcal{X}^{-1}(B)) \\ &= \begin{cases} N \cup (N^c \cap \mathcal{Y}^{-1}(B)) \in \mathcal{B}_{[0,1]} & \text{if } I_{\{0\}} \in B \\ \emptyset \cup (N^c \cap \mathcal{Y}^{-1}(B)) \in \mathcal{B}_{[0,1]} & \text{if } I_{\{0\}} \notin B. \end{cases} \end{aligned}$$

Analogously,

$$\mathcal{X}_n^{-1}(B) \in \mathcal{B}_{[0,1]}.$$

Since $\mathcal{X}_n$ and $\mathcal{X}$ are Borel measurable in d_p, by Proposition 2.2 both $\mathcal{X}_n$ and $\mathcal{X}$ are fuzzy random variables.

For the pointwise convergence, let $t \in [0, 1]$. Obviously, if $t \in N$, we have $\mathcal{X}_n(t) \to \mathcal{X}(t)$. If $t \notin N$, then $\mathcal{X}_n(t) = \mathcal{Y}_n(t)$, with $\mathcal{Y}_n(t) \to \mathcal{Y}(t) = \mathcal{X}(t)$. Therefore $\mathcal{X}_n(t) \to \mathcal{X}(t)$.

Finally, there remains to check $\mathbb{P}(\mathcal{X} \in A) = P(A)$ for $A \in \mathcal{B}_{(\mathcal{F}_c(\mathbb{R}^d),d_p)}$. Indeed,

$$\begin{aligned} &\mathbb{P}(\mathcal{X} \in A) = \mathbb{P}(\{t \in [0, 1] \mid \mathcal{X}(t) \in A\}) \\ &= \mathbb{P}(\{t \in N^c \mid \mathcal{X}(t) \in A\}) = \mathbb{P}(\{t \in N^c \mid \mathcal{Y}(t) \in A\}) \\ &= \mathbb{P}(\{t \in [0, 1] \mid \mathcal{Y}(t) \in A\}) = \mathbb{P}(\mathcal{Y} \in A) = (P \circ i_p^{-1})(A) = P(A), \end{aligned}$$

since, by [24, Theorem 1.9], $\{\mathcal{Y} \in A\}$ is measurable. □

To the best of our knowledge, there is only one earlier adaptation of the Skorokhod representation theorem to fuzzy random variables [31, Proposition 10]. In that case, convergence in is the metric d_∞ given by

$$d_\infty(U, V) = \sup_{\alpha \in [0,1]} d_H(U_\alpha, V_\alpha),$$

(see [25] for more details) which means both the required weak convergence and the obtained almost sure convergence are stronger. The new variables are defined on [0, 1] as ours are too.

That result, however, is valid only for fuzzy random variables which are Borel measurable with respect to the stronger metric d_∞, a rather strong requirement as very simple fuzzy random variables fail to satisfy it. In fact, since measurable mappings from [0, 1] to a metric space must take on values, almost surely, on a closed separable subset [15, Theorem 1.2], the characterization of d_∞-separable subsets in [31, Proposition 10] yields the following: for any d_∞-Borel fuzzy random variable X there exists a subset of Ω with null complement for which

$$\alpha \in [0, 1] \mapsto X_\alpha(\omega) \in \mathcal{K}_c(\mathbb{R}^d)$$

is continuous for all α out of an at most countable subset independent of ω. Thus the applicability of Theorem 3.5 is much wider.

4. Vitali convergence theorem

In this section, we generalize Vitali's convergence theorem to fuzzy random variables. Taking advantage of Theorem 3.5, the traditional assumption of almost sure convergence (or convergence in probability) will be relaxed to convergence in distribution. Two variants of the dominated convergence theorem will be obtained as applications.

We will be using a number of notions and results about convex combination spaces [34,35]. Since they are needed only for this section, they are collected here rather than in Section 2.

Definition 4.1. Let $(\mathbb{E}, d)$ be a metric space with a *convex combination operation* $[\cdot, \cdot]$ which, for any $n \geq 2$ numbers $\lambda_1, \ldots \lambda_n > 0$ satisfying $\sum_{i=1}^n \lambda_i = 1$, and all $u_1, \ldots, u_n \in \mathbb{E}$, produces an element $[\lambda_i, u_i]_{i=1}^n$ of $\mathbb{E}$. We will say that $\mathbb{E}$ is a *convex combination space* if the following axioms are satisfied:

1. (Commutativity) $[\lambda_i, u_i]_{i=1}^n = [\lambda_{\sigma(i)}, u_{\sigma(i)}]_{i=1}^n$ for every permutation σ of $\{1, \dots, n\}$;
2. (Associativity) $[\lambda_i, u_i]_{i=1}^{n+2} = [\lambda_1, u_1; \dots, \lambda_n, u_n; \lambda_{n+1} + \lambda_{n+2}, [\frac{\lambda_{n+j}}{\lambda_{n+1}+\lambda_{n+2}}; u_{n+j}]_{j=1}^2]$;
3. (Continuity) If $u, v \in \mathbb{E}$ and $\lambda^{(k)} \to \lambda \in (0, 1)$, then
$$[\lambda^{(k)}, u; 1 - \lambda^{(k)}, v] \to [\lambda, u; 1 - \lambda, v];$$
4. (Negative curvature) For all $u_1, u_2, v_1, v_2 \in \mathbb{E}$ and $\lambda \in (0, 1)$
$$d([\lambda, u_1; 1 - \lambda, u_2], [\lambda, u_1; 1 - \lambda, u_2]) \le \lambda d(u_1, v_1) + (1 - \lambda) d(u_2, v_2);$$
5. (Convexification) For each $u \in \mathbb{E}$, there exists $\lim_{n\to\infty}[n^{-1}, u]_{i=1}^n$, which will be denoted by $\mathbf{K}u$.

In [34, Theorems 2 and 4] it is shown that $(\mathcal{K}_c(\mathbb{R}^d), d_H)$ and $(\mathcal{F}_c(\mathbb{R}^d), d_p)$, for every $p \in [1, \infty]$, are convex combination spaces.

Definition 4.2. Let $(\mathbb{E}, d)$ be a convex combination space and $u_0 \in \mathbb{E}$. A random element $X : \Omega \to (\mathbb{E}, d)$ is called *integrable* if $d(u_0, X)$ is an integrable random variable.

Notice that the definition does not depend on the chosen point, since for any $u, v \in \mathbb{E}$ we have

$$E[d(u, X)] \le E[d(u, v) + d(v, X)] = E[d(u, v)] + E[d(v, X)]$$
$$= d(u, v) + E[d(v, X)].$$

The expectation in a convex combination space is defined through approximation by simple random elements [35].

Definition 4.3. Let $(\mathbb{E}, d)$ be a complete and separable convex combination space and let X be a random element. If X is simple, i.e., has the form $X = \sum_{j=1}^r I_{\Omega_j} u_j$, its *expectation* is $E[X] = [P(\Omega_j), \mathbf{K}u_j]_{j=1}^r$. If X is integrable then there exist sequences $\{X_k\}_k$ of simple functions converging almost surely to X and with $E[d(X_k, X)] \to 0$, and for any such sequence the d-limit of $E[X_k]$ exists and is the same element $E[X] \in \mathbb{E}$, which is called the *expectation* of X.

In [35] we can also find the following properties.

Theorem 4.1. *Let $(\mathbb{E}, d)$ be a complete separable convex combination space.*

1. *There exists a sequence of measurable simple functions $\{\phi_k\}_k$ with $\phi_k : \mathbb{E} \to \mathbb{E}$ such that $d(\phi_k(X), X) \searrow 0$ a.s. and $E[d(\phi_k(X), X)] \to 0$.*
2. *Let $X, Y : \Omega \to \mathbb{E}$ be integrable random elements. Then*
$$d(E[X], E[Y]) \le E[d(X, Y)].$$
3. *Let $X : \Omega \to \mathbb{E}$ an integrable random element and let $\{X_n\}_n$ be a sequence of pairwise independent random elements distributed as X. Then*
$$[n^{-1}, X_i]_{i=1}^n \to E[X] \text{ almost surely.}$$

For the following result, the reader is referred to [35, p. 887] or [34, Example 1].

Lemma 4.2. *Let $X : \Omega \to (\mathcal{K}_c(\mathbb{R}^d), d_H)$ be an integrably bounded random set. Then X is integrable and $E[X] = E_A[X]$.*

The following lemma corresponds to Theorem 5 in [34].

Lemma 4.3. *Let $\mathcal{X} : \Omega \to (\mathcal{F}_c(\mathbb{R}^d), d_p)$ be an integrably bounded fuzzy random variable. Then $\mathcal{X}$ is integrable and its expectation in the sense of convex combination spaces is $E[X] = \widetilde{E}[X]$.*

It will be convenient to state explicitly, for future reference, the following consequence of Theorem 4.1.(2) and Lemma 4.3.

Lemma 4.4. *Let $\mathcal{X}, \mathcal{Y} : \Omega \to \mathcal{F}_c(\mathbb{R}^d)$ be integrably bounded fuzzy random variables. Then*

$$d_p(\widetilde{E}[\mathcal{X}], \widetilde{E}[\mathcal{Y}]) \leq E[d_p(\mathcal{X}, \mathcal{Y})].$$

Vitali's convergence theorem for real random variables, in the form we wish to extend, is as follows.

Lemma 4.5 *(Vitali's convergence theorem). Let X_n and X be random variables such that $\{X_n\}_n$ is uniformly integrable. If $X_n \to X$ almost surely, then $E[X_n] \to E[X]$.*

In order to obtain an adaptation of this theorem for fuzzy random variables, we will use the following concept.

Definition 4.4. A class F of fuzzy random variables will be called *uniformly integrable in d_p* if given $\varepsilon > 0$ there exists $K \geq 0$ such that

$$E[d_p(\mathcal{X}, I_{\{0\}}) \cdot I_{\{d_p(\mathcal{X}, I_{\{0\}}) > K\}}] \leq \varepsilon$$

for all $\mathcal{X} \in F$.

The version we will prove weakens the assumption from almost sure convergence to weak convergence.

Theorem 4.6. *Let $\mathcal{X}_n$ and $\mathcal{X}$ be integrably bounded fuzzy random variables such that $\{\mathcal{X}_n\}_n$ is uniformly integrable in d_p. If $\mathcal{X}_n \to \mathcal{X}$ weakly in d_p, then $\widetilde{E}[\mathcal{X}_n] \to \widetilde{E}[\mathcal{X}]$ in d_p.*

Proof. By Theorem 3.5, there exist fuzzy random variables $\mathcal{Y}_n, \mathcal{Y} : ([0,1], \mathcal{B}_{[0,1]}, \mathbb{P}) \to \mathcal{F}_c(\mathbb{R}^d)$ such that $P_{\mathcal{Y}_n} = P_{\mathcal{X}_n}$, $P_{\mathcal{Y}} = P_{\mathcal{X}}$ and $\mathcal{Y}_n(t) \to \mathcal{Y}(t)$ in d_p for every $t \in [0,1]$. Then, by Lemma 4.4,

$$d_p(\widetilde{E}[\mathcal{Y}_n], \widetilde{E}[\mathcal{Y}]) \leq E[d_p(\mathcal{Y}_n, \mathcal{Y})].$$

Furthermore, $\{d_p(\mathcal{Y}_n, \mathcal{Y})\}_n$ is a sequence of measurable functions which converges pointwise to the null function. Let us show that $\{d_p(\mathcal{Y}_n, \mathcal{Y})\}_n$ is uniformly integrable. By the triangle inequality,

$$d_p(\mathcal{Y}_n, \mathcal{Y}) \leq d_p(\mathcal{Y}_n, I_{\{0\}}) + d_p(I_{\{0\}}, \mathcal{Y}).$$

Since $\mathcal{Y}_n$ has the same distribution as $\mathcal{X}_n$, for each $n \in \mathbb{N}$ and $K \geq 0$ it follows that $d_p(\mathcal{Y}_n, I_{\{0\}}) \cdot I_{\{d_p(\mathcal{Y}_n, I_{\{0\}}) > K\}}$ has the same distribution as $d_p(\mathcal{X}_n, I_{\{0\}}) \cdot I_{\{d_p(\mathcal{X}_n, I_{\{0\}}) > K\}}$. Hence

$$E\left[d_p(\mathcal{Y}_n, I_{\{0\}}) \cdot I_{\{d_p(\mathcal{Y}_n, I_{\{0\}}) > K\}}\right] = E\left[d_p(\mathcal{X}_n, I_{\{0\}}) \cdot I_{\{d_p(\mathcal{X}_n, I_{\{0\}}) > K\}}\right].$$

Then $\{d_p(\mathcal{Y}_n, I_{\{0\}})\}_n$ is a uniformly integrable sequence.

By an analogous reasoning,

$$E[d_p(I_{\{0\}}, \mathcal{Y})] = E[d_p(I_{\{0\}}, \mathcal{X})] = E\left[\left[\int_{[0,1]} (d_H(\{0\}, \mathcal{X}_\alpha))^p \, d\ell(\alpha)\right]^{1/p}\right]$$

$$\leq E\left[\left[\int_{[0,1]} (d_H(\{0\}, \mathcal{X}_0))^p \, d\ell(\alpha)\right]^{1/p}\right] = E[d_H(\{0\}, \mathcal{X}_0)] = E[\|\mathcal{X}_0\|] < \infty,$$

since $\mathcal{X}$ is integrably bounded, so $d_p(I_{\{0\}}, \mathcal{Y})$ is an integrable random variable.

The space $L^1([0,1], \mathcal{B}_{[0,1]}, \mathbb{P})$ is a Banach space, hence the mapping

$$T : f \in L^1([0,1], \mathcal{B}_{[0,1]}, \mathbb{P}) \to f + d_p(I_{\{0\}}, Y) \in L^1([0,1], \mathcal{B}_{[0,1]}, \mathbb{P})$$

is well defined and continuous.

By [3, Theorem 13.6, p. 140], $\{d_p(Y_n, I_{\{0\}})\}_n$ is uniformly integrable if and only if it is relatively weakly compact (i.e., its closure is compact in the weak topology of $L^1([0,1], \mathcal{B}_{[0,1]}, \mathbb{P})$). Since the continuous image of a relatively compact set is relatively compact [10, Theorem 6.8, p. 254], it follows that the set

$$T(\{d_p(\mathcal{Y}_n, I_{\{0\}})\}_n) = \{d_p(\mathcal{Y}_n, I_{\{0\}}) + d_p(I_{\{0\}}, \mathcal{Y})\}_n$$

is relatively compact. Applying again [3, Theorem 13.6], the sequence $\{d_p(\mathcal{Y}_n, I_{\{0\}}) + d_p(I_{\{0\}}, \mathcal{Y})\}_n$ is uniformly integrable, i.e., for any fixed $\varepsilon > 0$ there exists $K > 0$ such that

$$E\left[(d_p(\mathcal{Y}_n, I_{\{0\}}) + d_p(I_{\{0\}}, \mathcal{Y})) \cdot I_{\{(d_p(\mathcal{Y}_n, I_{\{0\}}) + d_p(I_{\{0\}}, \mathcal{Y})) > K\}}\right] < \varepsilon.$$

Then

$$E\left[d_p(\mathcal{Y}_n, \mathcal{Y}) \cdot I_{\{d_p(\mathcal{Y}_n, \mathcal{Y}) > K\}}\right] \leq E\left[(d_p(Y_n, I_{\{0\}}) + d_p(I_{\{0\}}, \mathcal{Y})) \cdot I_{\{d_p(\mathcal{Y}_n, \mathcal{Y}) > K\}}\right]$$
$$\leq E\left[(d_p(\mathcal{Y}_n, I_{\{0\}}) + d_p(I_{\{0\}}, \mathcal{Y})) \cdot I_{\{(d_p(\mathcal{Y}_n, I_{\{0\}}) + d_p(I_{\{0\}}, \mathcal{Y})) > K\}}\right] < \varepsilon.$$

Thus $\{d_p(\mathcal{Y}_n, \mathcal{Y})\}_n$ is uniformly integrable. By Lemma 4.5, $E[d_p(\mathcal{Y}_n, \mathcal{Y})] \to 0$. As we have seen that

$$d_p(\widetilde{E}[\mathcal{Y}_n], \widetilde{E}[\mathcal{Y}]) \leq E[d_p(\mathcal{Y}_n, \mathcal{Y})],$$

it follows that

$$d_p(\widetilde{E}[\mathcal{Y}_n], \widetilde{E}[\mathcal{Y}]) \to 0.$$

To finish the proof, we have to check that the expectations of the fuzzy random variables provided by Skorokhod's theorem are the same as those of the original variables. Assume first that $\mathcal{X}$ and $\mathcal{Y}$ are simple fuzzy random variables, i.e.,

$$\mathcal{X}(\Omega) = \{\widetilde{U}_1, ..., \widetilde{U}_m\} = \mathcal{Y}([0,1]).$$

By Proposition 2.1 we have

$$\widetilde{E}[\mathcal{X}] = P(\{\omega \in \Omega \mid \mathcal{X}(\omega) = \widetilde{U}_1\}) \cdot \widetilde{U}_1 + ... + P(\{\omega \in \Omega \mid \mathcal{X}(\omega) = \widetilde{U}_m\}) \cdot \widetilde{U}_m$$
$$= P(\{t \in [0,1] \mid \mathcal{Y}(t) = \widetilde{U}_1\}) \cdot \widetilde{U}_1 + ... + P(\{t \in [0,1] \mid \mathcal{Y}(t) = \widetilde{U}_m\}) \cdot \widetilde{U}_m = \widetilde{E}[\mathcal{Y}].$$

For the general case, by Theorem 4.1 there exists a sequence $\{\phi_k\}_k$ of measurable simple functions such that $\phi_k(\mathcal{X}) \to \mathcal{X}$ and $\phi_k(\mathcal{Y}) \to \mathcal{Y}$.

$$d_p(\widetilde{E}[\mathcal{X}], \widetilde{E}[\mathcal{Y}]) \leq d_p(\widetilde{E}[\mathcal{X}], \widetilde{E}[\phi_k(\mathcal{X})]) + d_p(\widetilde{E}[\phi_k(\mathcal{X})], \widetilde{E}[\mathcal{Y}])$$
$$\leq d_p(\widetilde{E}[\mathcal{X}], \widetilde{E}[\phi_k(\mathcal{X})]) + d_p(\widetilde{E}[\phi_k(\mathcal{X})], \widetilde{E}[\phi_k(\mathcal{Y})]) + d_p(\widetilde{E}[\phi_k(\mathcal{Y})], \widetilde{E}[\mathcal{Y}]).$$

On the one hand, $d_p(\widetilde{E}[\mathcal{X}], \widetilde{E}[\phi_k(\mathcal{X})]) \to 0$, by Theorem 4.1.(1,2). Analogously, $d_p(\widetilde{E}[\mathcal{Y}], \widetilde{E}[\phi_k(\mathcal{Y})]) \to 0$.

On the other hand, since $\phi_k(\mathcal{X})$ and $\phi_k(\mathcal{Y})$ are simple fuzzy random variables with the same distribution, we have $d_p(\widetilde{E}[\phi_k(\mathcal{X})], \widetilde{E}[\phi_k(\mathcal{Y})]) = 0$.

Then $d_p(\widetilde{E}[\mathcal{X}], \widetilde{E}[\mathcal{Y}]) = 0$, namely $\widetilde{E}[\mathcal{X}] = \widetilde{E}[\mathcal{Y}]$. For the same reason, also $\widetilde{E}[\mathcal{X}_n] = \widetilde{E}[\mathcal{Y}_n]$. In conclusion,

$$d_p(\widetilde{E}[\mathcal{X}_n], \widetilde{E}[\mathcal{X}]) \to 0. \quad \square$$

Remark 4.1. The proof above uses some results from [35] which are under the assumption that the convex combination space is complete as a metric space, specifically the approximation scheme by simple functions. That is not satisfied by $(\mathcal{F}_c(\mathbb{R}^d), d_p)$, whence it might appear that the proof is doubtful. However, completeness is used in the construction of the expectation: writing $E[X]$ in the sense of convex combination spaces already involves considering the completion of $(\mathcal{F}_c(\mathbb{R}^d), d_p)$ (see [34]).

Once the expectation has been constructed as an element of the completion of $(\mathcal{F}_c(\mathbb{R}^d), d_p)$ which is ensured to lie within $\mathcal{F}_c(\mathbb{R}^d)$ (thanks to Lemma 4.3 and the assumption of integrable boundedness), clearly the invoked properties continue to hold as properties of the completion which involve only elements of $\mathcal{F}_c(\mathbb{R}^d)$.

From the Vitali theorem we can deduce a variant of the dominated convergence theorem for fuzzy random variables.

Theorem 4.7. *Let $\mathcal{X}_n$ and $\mathcal{X}$ be integrably bounded fuzzy random variables. If $\mathcal{X}_n \to \mathcal{X}$ weakly in d_p and there exists $g \in L^1(\Omega, \mathcal{A}, P)$ such that $d_p(\mathcal{X}_n, I_{\{0\}}) \le g$ for all $n \in \mathbb{N}$, then $\widetilde{E}[\mathcal{X}_n] \to \widetilde{E}[\mathcal{X}]$ in d_p.*

Proof. The sequence of random variables $\{d_p(\mathcal{X}_n, I_{\{0\}})\}_n$ is dominated by a function in $L^1(\Omega, \mathcal{A}, P)$. Every dominated sequence is uniformly integrable, see e.g. [5, Remark 3.13.(b), p. 72]. Then, for all $\varepsilon > 0$, there exists $K \ge 0$ such that

$$E\left[d_p(\mathcal{X}_n, I_{\{0\}}) \cdot I_{\{d_p(\mathcal{X}_n, I_{\{0\}}) > K\}}\right] \le \varepsilon,$$

that is, $\{\mathcal{X}_n\}_n$ is uniformly integrable in d_p. It suffices now to apply Theorem 4.6. □

A number of versions of the dominated convergence theorem for fuzzy random variables have been proved so far. The most comprehensive one seems to be Krätschmer's [20, Theorem 8.2], which includes and/or improves upon earlier versions like [25, Theorem 4.3], [13, Theorem 4.1], [19, Theorem 3], [17, Theorem 3.6]. Krätschmer's result considers the metrics ρ_p and d_∞, of which ρ_p is uniformly equivalent to d_p and d_∞ is strictly stronger.

Let us comment the case of the L^p-type metrics first, since it is the one directly related to our setting. By the uniform equivalence, Krätschmer's integrability conditions are equivalent to those in Theorem 4.7. The conclusion is the same in both theorems. He considers fuzzy random variables with values in $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$, which is more general than $\mathcal{F}_c(\mathbb{R}^d)$. On the other hand, he assumes almost sure convergence while we do with weak convergence.

In the case of a dominated convergence theorem for the d_∞-metrics, convergence in both the assumption and the conclusion is stronger than in the d_p-metrics, so no version implies the other. However, the weak convergence assumed in Theorem 4.7 is still a weaker type of requirement than the almost sure convergence in [20, Theorem 8.2] and the convergence in probability in [9, Theorem 3.4].

Remark 4.2. It seems worth pointing out that our dominated convergence theorem does not seem to follow from successive application of the dominated convergence theorem with almost sure convergence (e.g., [20, Theorem 8.2]) and Skorokhod's theorem (to obtain an almost sure convergence from a weak one). The reason is that Skorokhod's theorem does not ensure that the sequence of almost surely converging fuzzy random variables it provides will still be dominated.

To close this section, we present a variant which does not require the calculation of $d_p(\mathcal{X}_n, I_{\{0\}})$.

Corollary 4.8. *Let $\mathcal{X}_n$ and $\mathcal{X}$ be integrably bounded fuzzy random variables. If $\mathcal{X}_n \to \mathcal{X}$ weakly in d_p and there exists an integrably bounded fuzzy random variable $\mathcal{Y}$ such that $\mathcal{X}_n \subseteq \mathcal{Y}$ for all $n \in \mathbb{N}$, then $\widetilde{E}[\mathcal{X}_n] \to \widetilde{E}[\mathcal{X}]$ in d_p.*

Proof. We have $d_p(\mathcal{X}_n, I_{\{0\}}) \le \|(\mathcal{X}_n)_0\|$, as shown in the proof of Theorem 4.6. Since $\mathcal{X}_n \subseteq \mathcal{Y}$ for all $n \in \mathbb{N}$, we have $\|(\mathcal{X}_n)_0\| \le \|\mathcal{Y}_0\|$ and since $\mathcal{Y}$ is integrably bounded, $E[\|\mathcal{Y}_0\|] < \infty$.

Finally, $\|\mathcal{Y}_0\|$ is an $L^1(\Omega, \mathcal{A}, P)$ function dominating the sequence $\{d_p(\mathcal{X}_n, I_{\{0\}})\}_n$, so Theorem 4.7 applies. □

5. Miscellaneous applications

This section provides a number of further consequences of the results in Section 3.

5.1. Continuous mapping theorem

Theorem 3.5 allows one to obtain the continuous mapping theorem for fuzzy random variables with the assumption of weak convergence.

Theorem 5.1. *Let $\mathcal{X}_n$ and $\mathcal{X}$ be fuzzy random variables such that $\mathcal{X}_n \to \mathcal{X}$ weakly in d_p. If $f : \mathcal{F}_c(\mathbb{R}^d) \to \mathcal{F}_c(\mathbb{R}^d)$ is a $P_{\mathcal{X}}$-almost surely continuous function, then $f(\mathcal{X}_n) \to f(\mathcal{X})$ weakly in d_p.*

Proof. By Theorem 3.5, there exist fuzzy random variables $\mathcal{Y}_n, \mathcal{Y} : ([0,1], \mathcal{B}_{[0,1]}, \mathbb{P}) \to \mathcal{F}_c(\mathbb{R}^d)$ with $\mathbb{P} = \ell$ such that $P_{\mathcal{Y}_n} = P_{\mathcal{X}_n}$, $P_{\mathcal{Y}} = P_{\mathcal{X}}$ and $\mathcal{Y}_n(t) \to \mathcal{Y}(t)$ in d_p for every $t \in [0,1]$. Since f is a continuous function $P_{\mathcal{Y}}$-a.s., there exists $\mathcal{M} \subseteq \mathcal{F}_c(\mathbb{R}^d)$ such that $P_{\mathcal{Y}}(\mathcal{M}) = 1$, and if $\mathcal{Y} \in \mathcal{M}$ then $f(\mathcal{Y}_n) \to f(\mathcal{Y})$. Therefore

$$\{t \in [0,1] \mid f(\mathcal{Y}_n(t)) \nrightarrow f(\mathcal{Y}(t))\} \subseteq \{\mathcal{Y} \notin \mathcal{M}\},$$

which is a null set since $\mathbb{P}(\mathcal{Y} \notin \mathcal{M}) = 1 - P_{\mathcal{Y}}(\mathcal{M}) = 0$. Then $f(\mathcal{Y}_n) \to f(\mathcal{Y})$ $\mathbb{P}$-a.s. Combining Lemma 3.2 and Lemma 3.7 in [12], almost sure convergence implies weak convergence, hence $f(\mathcal{Y}_n) \to f(\mathcal{Y})$ weakly. Finally, from the fact that $\mathcal{X}$ and $\mathcal{Y}$ and also $\mathcal{X}_n$ and $\mathcal{Y}_n$ have the same distribution, it follows that $f(\mathcal{X}_n) \to f(\mathcal{X})$ weakly. □

In the case of random variables, the codomain of f is $\mathbb{R}$ and so preservation of weak convergence by continuous mappings follows easily from the definition of weak convergence (since the composition of continuous functions is continuous). The difficulty in establishing the continuous mapping theorem in the real case is rather to prove it from convergence in distribution defined as pointwise convergence of the cumulative distribution functions in the continuity points of the limit.

5.2. *Perfect distributions*

The notion of a perfect probability measure was introduced by Gnedenko and Kolmogorov [11] in order to avoid some counter-intuitive behaviours of probability measures in arbitrary sample spaces which began to be detected in the late 1940s.

Definition 5.1. A probability measure P in a measurable space $(\Omega, \mathcal{A})$ is called *perfect* if for every $A \subseteq \mathbb{R}$ and every random variable $X : \Omega \to \mathbb{R}$ such that $\{X \in A\} \in \mathcal{A}$, there exist $A_1, A_2 \in \mathcal{B}_{\mathbb{R}}$ such that

- $A_1 \subseteq A \subseteq A_2$.
- $P(X \in A_2 \setminus A_1) = 0$.

Hence a probability measure is perfect when all the probabilistic information about its random variables is determined by the probabilities of the Borel sets. For a more detailed (but still introductory) discussion, the reader is referred to [2, Section 4]. The following result ensures that fuzzy random variables always have perfect distributions. Its proof is analogous to that of the $\mathcal{F}_c(\mathbb{R})$ case in [2, Proposition 4.7] and therefore omitted. It is based on the measurability of $\mathcal{F}_c(\mathbb{R}^d)$ in $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$ (Theorem 3.1), the fact that $(\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d), d_p)$ is a complete separable metric space, and the fact that σ_L is the Borel σ-algebra in $\mathcal{F}_c(\mathbb{R}^d)$ generated by d_p (Proposition 2.2).

Theorem 5.2. *Let $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R}^d)$ be a fuzzy random variable. Then its distribution $P_{\mathcal{X}}$ is perfect.*

5.3. *$\mathcal{F}_c(\mathbb{R}^d)$ is a Lusin space*

Definition 5.2. A topological space is called *Polish* if its topology is generated by some complete separable metric.

Definition 5.3. A *Lusin* space is the image of a Polish space under a continuous bijective function.

Clearly, every Polish space is Lusin. The following result can be found in [29].

Lemma 5.3. *A topological subspace of a Lusin space is Lusin if and only if it is Borel measurable.*

The metric space $(\mathcal{F}_c(\mathbb{R}^d), d_p)$ is not complete, but it is separable [6, Theorem 3]; as a consequence of Theorem 3.1, it is Lusin.

Proposition 5.4. *The space $(\mathcal{F}_c(\mathbb{R}^d), d_p)$ is a Lusin space for every $p \in [1, \infty)$.*

Proof. As mentioned in the proof of Theorem 3.5, $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$ is the completion of a separable metric space, therefore it is Polish, in particular it is a Lusin space. Moreover, by Theorem 3.1, $\mathcal{F}_c(\mathbb{R}^d)$ is a measurable subspace of $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$. Finally, from Lemma 5.3 we deduce that $\mathcal{F}_c(\mathbb{R}^d)$ is Lusin. □

This will be applied in the next three subsections.

5.4. Regular conditional distributions

In this subsection, we discuss an approach to defining probabilities of events concerning a fuzzy random variable, conditional on the value of another fuzzy random variable.

Definition 5.4. A *standard measurable space* is a measurable space isomorphic to a Borel subset of a Polish space.

Definition 5.5. Let (Y, T, μ) be a probability space and (X, S) be a measurable space. A *Markov kernel* is a function $\nu : Y \times S \to [0, 1]$ such that $\nu(y, \cdot)$ is a probability measure on (X, S) for each $y \in Y$ and $\nu(\cdot, F)$ is a T-measurable function for each $F \in S$.

Definition 5.6. Let $(X \times Y, S \otimes T, \lambda)$ be a probability space. A *regular conditional probability* is a Markov kernel satisfying

$$\lambda(F \times E) = \int_E \nu(y, F)\lambda_y(dy) \tag{2}$$

for all $E \in T$, $F \in S$, where λ_y is the Y-marginal of λ.

Theorem 5.5. *Any space $(\mathcal{F}_c(\mathbb{R}^d) \times Y, \sigma_L \otimes T, \lambda)$ has a regular conditional probability.*

Proof. By Theorem 3.1, $\mathcal{F}_c(\mathbb{R}^d)$ is a measurable subset of a Polish space, hence standard. In particular, it is pre-standard in the sense of [8]. The result follows by an application of [8, Theorem 5]. □

In particular, this provides a rigorous way to define conditional distributions for fuzzy random variables. Let $\mathcal{X}$ and $\mathcal{Y}$ be fuzzy random variables with values in $\mathcal{F}_c(\mathbb{R}^d)$ and $\mathcal{F}_c(\mathbb{R}^{d'})$, respectively. Taking λ to be the joint distribution $P_{(\mathcal{X},\mathcal{Y})}$ and ν provided by Theorem 5.5 we define

$$P(\mathcal{X} \in F \mid \mathcal{Y} = U) = \nu(U, F)$$

for all $U \in \mathcal{F}_c(\mathbb{R}^d)$ and $F \in \sigma_L$. Thus (2) becomes

$$P(\mathcal{X} \in F, \mathcal{Y} \in E) = \int_E P(\mathcal{X} \in F \mid \mathcal{Y} = V)\mathrm{d}P_{\mathcal{Y}}(V).$$

From the definition of a Markov kernel, the conditional distributions $P_{\mathcal{X}|\mathcal{Y}=U}$ so defined are actual probability measures. Of course, if $U \in \mathcal{F}_c(\mathbb{R}^d)$ is such that $P(\mathcal{Y} = U) > 0$ then

$$P(\mathcal{X} \in F, \mathcal{Y} = U) = \int_{\{U\}} P(\mathcal{X} \in F \mid \mathcal{Y} = V)\mathrm{d}P_{\mathcal{Y}}(V) = P(\mathcal{X} \in F \mid \mathcal{Y} = U) \cdot P_{\mathcal{Y}}(\{U\})$$

whence

$$P(\mathcal{X} \in F \mid \mathcal{Y} = U) = \frac{P(\mathcal{X} \in F, \mathcal{Y} = U)}{P_{\mathcal{Y}}(\{U\})} = \frac{P(\mathcal{X} \in F, \mathcal{Y} = U)}{P(\mathcal{Y} = U)}$$

consistently with the ordinary definition $P(A \mid B) = P(A \cap B)/P(B)$ for $P(B) > 0$.

One can define a conditional expectation $E[\mathcal{X} \mid \mathcal{Y} = U]$ to be the expectation of the identity mapping $id : (\mathcal{F}_c(\mathbb{R}^d), \sigma_L, P_{\mathcal{X}|\mathcal{Y}=U}) \to \mathcal{F}_c(\mathbb{R}^d)$, provided the latter is integrably bounded. Other notions (variance, median, and so on) are adapted similarly.

Let X, Y be random vectors of $\mathbb{R}^d$ and $\mathbb{R}^{d'}$, respectively. We can define conditional probabilities and distributions of the types $X|\mathcal{Y} = U$ and $\mathcal{X}|Y = y$ with the formulas

$$P(X \in A|\mathcal{Y} = U) = P(I_{\{X\}} \in \{I_{\{x\}} \mid x \in A\} \mid \mathcal{Y} = U)$$

and

$$P(\mathcal{X} \in F|Y = y) = P(\mathcal{X} \in F|I_{\{Y\}} = I_{\{y\}}).$$

López-Díaz and Gil [22] considered regular conditional distributions in the context of fuzzy random variables, but their problem is very different. They consider one fuzzy random variable defined on a product space which is assumed to be endowed with a regular conditional distribution. Here we consider a joint distribution of two fuzzy random variables and establish the existence of a regular conditional distribution (as an application of Faden's general result).

5.5. An extension theorem

Every probability measure on a sub-σ-algebra of σ_L extends to σ_L, whence there exist some fuzzy random variable whose induced distribution agrees on every event with the original probability, provided the sub-σ-algebra is countably generated.

Definition 5.7. A *Suslin* space is the image of a Polish space under a continuous surjective function.

Theorem 5.6. *Let $\mathcal{C} \subseteq \sigma_L$ be a countably generated sub-σ-algebra, and let $\mu : \mathcal{C} \to [0, 1]$ be a probability measure. Then there exists a fuzzy random variable $\mathcal{X}$ on a probability space $(\Omega, \mathcal{A}, P)$ such that $P(\mathcal{X} \in A) = \mu(A)$ for every $A \in \mathcal{C}$.*

Proof. By Proposition 5.4, $\mathcal{F}_c(\mathbb{R}^d)$ is a Lusin space and obviously Suslin. By [1, Corollary 5.3] (note that Suslin spaces are called analytic there), μ has an extension $\widehat{\mu} : \sigma_L \to [0, 1]$. One fuzzy random variable $\mathcal{X}$ such that $P_{\mathcal{X}} = \widehat{\mu}$ is the identity mapping in $\mathcal{F}_c(\mathbb{R}^d)$. □

Therefore a partial assignment of probabilities in a (countably generated) sub-σ-algebra always agrees with the probabilities assigned by some fuzzy random variable.

5.6. Choquet theorem for random closed sets in $\mathcal{F}_c(\mathbb{R}^d)$

The Choquet theorem is a central result about distributions of random closed sets, analogous to the theorem that identifies the properties a function must have in order to be the cumulative distribution function of some random variable.

In some situations there appear random sets formed by points in a space of fuzzy sets. A motivating example comes from statistical estimation with fuzzy data. Indeed, whenever estimators are obtained by optimizing some objective function, the estimator may or may not be unique (for instance, in the case of random variables, the mean is unique while the median, in general, is not). If uniqueness is not achieved, the set of optimizers is a random set (since it depends on the sample) each point of which is a fuzzy set. The reader is referred to the papers [26,27] for such examples of estimators (M-estimators) in the setting of fuzzy data.

Definition 5.8. Let $\mathbb{E}$ be a topological space. A *capacity* in $\mathbb{E}$ is a set function $c : \mathcal{L} \to [0, 1]$ on a lattice of sets $\mathcal{L} \subseteq \mathcal{P}(\mathbb{E})$ such that $\emptyset, \mathbb{E} \in \mathcal{L}$, $c(\emptyset) = 0$ and $c(\mathbb{E}) = 1$ hold, and moreover $c(A) \le c(B)$ whenever $A \subseteq B$.

A *random closed set* in a metric space $\mathbb{E}$ is a mapping, from a probability space to the class of all non-empty closed subsets of $\mathbb{E}$, for which the events $\{X \cap G \neq \emptyset\}$ are measurable whenever G is open.

Theorem 5.7. *Let $p \in [1, \infty)$. The identities*

$$P(X \cap G \neq \emptyset) = T(G),$$

for every d_p-open set G in $\mathcal{F}_c(\mathbb{R}^d)$, establish a bijection between the distributions P_X of random closed sets in $(\mathcal{F}_c(\mathbb{R}^d), d_p)$ and capacities T in the space of d_p-open sets in $\mathcal{F}_c(\mathbb{R}^d)$ such that $c(A_n) \to c(A)$ whenever $A_n \nearrow A$ and

$$c(\bigcap_{i=1}^{n} A_i) \leq \sum_{I\subseteq\{1,\ldots,n\}, I\neq\emptyset} (-1)^{|I|+1} c(\bigcup_{i\in I} A_i).$$

Proof. Since $(\mathcal{F}_c(\mathbb{R}^d), d_p)$ is a metrizable Lusin space for each $p \in [1, \infty)$ by Proposition 5.4, we can apply [33, Theorem 7]. □

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Research in this paper was partially funded by grants from Spain (MINECO-16-MTM2015-63971-P, PID2019-104486GB-I00), the Principality of Asturias (FC-GRUPIN-IDI/2018/000132), and the University of Oviedo (PAPI-20-PF-21). Their contribution is gratefully acknowledged.

References

[1] J.M. Aldaz, H. Render, Borel measure extensions of measures defined on sub-σ-algebras, Adv. Math. 150 (2000) 233–263.
[2] M. Alonso de la Fuente, P. Terán, Harmonizing two approaches to fuzzy random variables, Fuzzy Optim. Decis. Mak. 19 (2020) 177–189.
[3] N.L. Carothers, A Short Course on Banach Space Theory, Cambridge Univ. Press, Cambridge, 2005.
[4] A. Colubi, J.S. Domínguez-Menchero, M. López-Díaz, R. Körner, A method to derive strong laws of large numbers for random upper semi-continuous functions, Stat. Probab. Lett. 53 (2001) 269–275.
[5] E. Çınlar, Probability and Stochastics, Springer, New York, 2011.
[6] P. Diamond, P. Kloeden, Metric spaces of fuzzy sets, Fuzzy Sets Syst. 35 (1990) 241–249.
[7] R.M. Dudley, Distances of probability measures and random variables, Ann. Math. Stat. 39 (1968) 1563–1572.
[8] A.M. Faden, The existence of regular conditional probabilities: necessary and sufficient conditions, Ann. Probab. 13 (1985) 288–298.
[9] Y. Feng, Convergence theorems for fuzzy random variables and fuzzy martingales, Fuzzy Sets Syst. 103 (1999) 435–441.
[10] S.G. Georgiev, K. Zennir, Functional Analysis with Applications, De Gruyter, Berlin, 2019.
[11] B.V. Gnedenko, A.N. Kolmogorov, Limit Distributions for Sums of Independent Random Variables, Addison-Wesley, Reading, 1954.
[12] O. Kallenberg, Foundations of Modern Probability, Springer, New York, 1997.
[13] E.P. Klement, M.L. Puri, D.A. Ralescu, Limit theorems for fuzzy random variables, Proc. R. Soc. Ser. A 407 (1986) 171–182.
[14] A. Klenke, Probability Theory, second edition, Springer, London, 2014.
[15] G. Koumoullis, K. Prikry, Perfect measurable spaces, Ann. Pure Appl. Log. 30 (1986) 219–248.
[16] V. Krätschmer, A unified approach to fuzzy random variables, Fuzzy Sets Syst. 123 (2001) 1–9.
[17] V. Krätschmer, Limit theorems for fuzzy-random variables, Fuzzy Sets Syst. 126 (2002) 253–263.
[18] V. Krätschmer, Some complete metrics on spaces of fuzzy subsets, Fuzzy Sets Syst. 130 (2002) 357–365.
[19] V. Krätschmer, Probability theory in fuzzy sample spaces, Metrika 60 (2004) 167–189.
[20] V. Krätschmer, Integrals of random fuzzy sets, Test 15 (2006) 433–469.
[21] R. Kruse, K.D. Meyer, Statistics with Vague Data, Reidel, Dordrecht, 1987.
[22] M. López-Díaz, M.A. Gil, Reversing the order of integration in iterated expectations of fuzzy random variables, and statistical applications, J. Stat. Plan. Inference 74 (1998) 11–29.
[23] H.T. Nguyen, B. Wu, Fundamentals of Statistics with Fuzzy Data, Springer, Berlin, 2006.
[24] K.R. Parthasarathy, Probability Measures on Metric Spaces, Academic Press, New York, 1967.
[25] M.L. Puri, D.A. Ralescu, Fuzzy random variables, J. Math. Anal. Appl. 114 (1986) 409–422.
[26] B. Sinova, S. Van Aelst, Advantages of M-estimators of location for fuzzy numbers based on Tukey's biweight loss function, Int. J. Approx. Reason. 93 (2018) 219–237.
[27] B. Sinova, S. Van Aelst, P. Terán, M-estimators and trimmed means: from Hilbert-valued to fuzzy set-valued data, Adv. Data Anal. Classif. 15 (2021) 267–288.
[28] A.V. Skorokhod, Limit theorems for stochastic processes, Theory Probab. Appl. 1 (1953) 261–290.
[29] L. Schwartz, Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures, Oxford University Press, Oxford, 1973.
[30] P. Terán, An embedding theorem for convex fuzzy sets, Fuzzy Sets Syst. 152 (2005) 191–208.
[31] P. Terán, On Borel measurability and large deviations for fuzzy random variables, Fuzzy Sets Syst. 157 (2006) 2558–2568.

[32] P. Terán, Characterizing the distributions of fuzzy random variables, in: Talk at the Fifth Int. Conf. ERCIM Working Group on Computing and Statistics (Oviedo, Spain), December 1st, 2012., 2012, p. 22. Abstract: Book of Abstracts CFE–ERCIM.

[33] P. Terán, Choquet theorem for random sets in Polish spaces and beyond, in: S. Destercke, T. Denoeux, M.Á. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), Uncertainty Modelling in Data Science, Springer, Cham, 2019, pp. 208–215.

[34] P. Terán, I. Molchanov, A general law of large numbers, with applications, in: J. Lawry, E. Miranda, A. Bugarin, S. Li, M.Á. Gil (Eds.), Soft Methods for Integrated Uncertainity Modelling, Springer, Berlin, 2006, pp. 153–160.

[35] P. Terán, I. Molchanov, The law of large numbers in a metric space with a convex combination operation, J. Theor. Probab. 9 (2006) 875–898.

[36] A.W. Van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, Nueva York, 1996.