Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 458 (2023) 69–93

www.elsevier.com/locate/fss

Convergence theorems for random elements in convex combination spaces

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 17 February 2022; received in revised form 23 June 2022; accepted 25 June 2022
Available online 30 June 2022

Abstract

A Vitali convergence theorem is proved for subspaces of an abstract convex combination space which admits a complete separable metric. The convergence may be in that metric or, more generally, in a quasimetric satisfying weaker properties. Versions for convergence in probability and in distribution are given. As applications, we show that some dominated convergence theorems in the literature of fuzzy random variables and random compact sets can be recovered or improved, and we derive new convergence theorems in another space of sets and in a space of probability distributions.

Keywords: Convex combination operation; Dominated convergence theorem; Fuzzy random variable; Random set; Vitali convergence theorem

1. Introduction

The usual spaces of fuzzy subsets of $\mathbb{R}^d$, endowed with the ordinary operations of addition and product by scalars which are derived from Zadeh's extension principle, are not linear spaces. There are two main differences with a linear space. First, in general $(-1) \cdot A$ is not the additive inverse of a given fuzzy set A; in fact, most often an additive inverse does not exist. Second, the distributive law $\lambda \cdot A + \mu \cdot A = (\lambda + \mu) \cdot A$ fails. While this includes the first difference (by taking $\lambda = 1, \mu = -1$), it continues to fail even restricting the law to the positive scalars.

That means that many notions, results and techniques which could be immediately applied in linear spaces, are not available. Even the mere notion of subtraction, in the sense of defining $A - B = A + (-1) \cdot B$, is unavailable, which creates a lot of difficulties. For instance, it is not obvious how to define the derivative (or differential) of a fuzzy-valued function, and in fact many definitions have been given through the years. Another example is in statistics with fuzzy data, since many methods for random variables rely on the difference between a sample statistic and the corresponding population value.

* 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.2022.06.019
0165-0114/

Another departure from linear spaces, from a geometric rather than algebraic perspective, is that a set formed by a single point can fail to be convex. This sounds counterintuitive or even confusing, since our intuition has been shaped in linear spaces.

But, even with those obstacles, remarkably some standard metrics in spaces of fuzzy sets preserve some of conventional properties of a norm: if d is such a metric then $d(A, B)$ becomes a convenient surrogate for the non-sensical $\|A - B\|$. In particular, the distance $d(A, I_{\{0\}})$ (where the indicator function $I_{\{0\}}$ of the crisp set $\{0\}$ is the neutral element of the sum) becomes a surrogate for $\|A\|$ which still has all the defining properties of a norm.

A relevant question is whether these spaces of fuzzy sets can be seen as instances of some abstract mathematical structure involving the sum, the product by a scalar, and a metric: a structure weaker than a (normed) linear space but preserving some of the advantages of working with a norm. Maybe the most recent answer is the notion of a quasilinear metric space [24]. In that paper, O'Regan and Lupulescu give notions of derivation and integration which, in particular, apply to spaces of fuzzy sets.

Such generalizations of normed spaces can ultimately be traced back to Rådström's 1952 paper [35] where he showed that some spaces of sets embed into normed spaces by giving a set of sufficient conditions valid in a general abstract space. Among the generalizations inspired by the problem of fitting sets as elements of an abstract space with a list of axioms, we would like to mention metric convex cones (Prolla [30,31]) and near vector lattices (Labuschagne et al. [22]). Labuschagne and Pinchuck applied near vector lattices to fuzzy martingale theory [21]. Other generalizations are commented upon in [24].

In this paper, we are concerned with another such generalization: convex combination spaces (Terán and Molchanov [41]). To motivate the convenience of studying both an abstract structure and convex combination spaces in particular, let us present an interesting example from a quite different setting, which is a convex combination space [40, Lemma 6.2] but not a quasilinear metric space, a metric convex cone or a near vector lattice.

Consider the space of all probability distributions of random variables with finite variance. Distributions can be 'added' using the convolution operation, i.e., $P * Q$ is defined as the distribution of $X + Y$ where X, Y are independent random variables respectively distributed as P and Q. Convolution is of great importance in many applications such as signal processing. They can also be multiplied by scalars by rescaling, namely $\lambda \cdot P$ is the distribution of $\lambda \cdot X$ where X is distributed as P.

The space so defined has exactly the same shortcomings and positives discussed above. The neutral element is the degenerate distribution δ_0 (Dirac distribution) at 0. Let P be a normal distribution $\mathcal{N}(0, 1)$. Then $(-1) \cdot P$ is again P, not the additive inverse of P. Moreover, by the reproductive property of the normal distribution,

$$\frac{1}{2} \cdot \mathcal{N}(0, 1) * \frac{1}{2} \cdot \mathcal{N}(0, 1) = \mathcal{N}(0, \frac{1}{\sqrt{2}}),$$

so that the distributive law fails and, further, the singleton $\{P\}$ is not convex. Rather than being counterintuitive, nonconvex points obey the key fact in Statistics that averaging reduces the variance. The deconvolution problem (finding a distribution R such that $Q * R = P$), used in signal denoising, is the equivalent here of defining a subtraction $P - Q$, which is not always possible. Finally, when this space is endowed with the L^2-Wasserstein metric

$$w_2(P, Q) = \inf_{X,Y} \|X - Y\|_2$$

(where the infimum runs over all random variables distributed as P and Q), the definition via the L^2-norm provides $w_2(P, Q)$ with working properties very similar to what we would expect from 'the norm of $P - Q$' if such a thing existed.

This example underlines the interest of finding good abstract approaches which provide unified methods for these very different spaces with similar properties. Convex combination spaces were found when trying to develop an abstract version of the law of large numbers in [38]. A convex combination space which is separable and complete as a metric space allows for a theory of integration against probability measures generalizing Lebesgue and Bochner integration [41]. With that notion of expectation, a law of large numbers holds [41, Theorem 5.1] and further results were proved like a Jensen inequality [40], dominated convergence theorems [40,43] and a Birkhoff ergodic theorem [43]. Further examples of limit theorems in the setting of convex combination spaces can be found in [34,44,33] and references therein.

In this paper we will establish some convergence theorems for integration in convex combination spaces (in probabilistic language, for expectations of random elements). That is, we will give sufficient conditions to ensure that, if a sequence of random elements converges in some sense, also the sequence of their expectations will converge to the expectation of the limit. The central result is an abstract form of the Vitali convergence theorem. This theorem is similar to the dominated convergence theorem, but the condition that the sequence is dominated is replaced by the weaker one that it is uniformly integrable. Following a modern probability approach, subsequent to the Skorokhod representation theorem, the assumption of convergence in probability or almost sure will be weakened to convergence in distribution (weak convergence). Our motivation to begin this study was double.

(i) In [2, Theorem 4.6] we established a Vitali theorem for fuzzy random variables in the d_p-metrics. Although some assumptions are weaker than in the earlier literature, the dominated convergence theorem of Krätschmer [20, Theorem 8.2] does not follow from it, either because his space of fuzzy values is larger (if $p \in [1, \infty)$) or because our result does not apply (if $p = \infty$). It would be interesting to find a common generalization with the best of both theorems.
(ii) In [40, Remark 2], an incorrect way to generalize the dominated convergence theorem for convex combination spaces [40, Theorem 4.2] was postulated by the second author (for a detailed discussion, the reader is referred to Section 7). That brings the validity of [40, Proposition 5.2] into question. It would be interesting to correct this mistake and show, by a different proof, that the claim in the remark is valid (even if the method is not).

Our main result will solve these two questions in the positive, while at the same time being much more general. In fact, in Section 6 we will provide some applications of this general abstract result to specific (subsets of) convex combination spaces, which go beyond these initial motivations. These applications lean towards spaces of fuzzy sets or crisp sets although other applications can be worked out as well.

The structure of the paper is as follows. Section 2 contains the preliminaries. In Section 3 we state the general Vitali convergence theorem in subsets of convex combination spaces, and discuss its assumptions. The theorem is proved in Section 4. Then some variants, for which the proof only requires minor modifications, are presented in Section 5. Moreover, dominated convergence theorems are obtained as corollaries.

In Section 6, the general Vitali type theorems will be applied to several specific spaces. Applications to spaces of fuzzy sets include the following.

- The space of d-dimensional (generalized) fuzzy numbers with compact support, with the metrics d_p and d_∞.
- The space of d-dimensional (generalized) fuzzy numbers with L^p-type support function, with the metrics d_p.

Applications to spaces of sets are the following.

- The space of compact subsets of a convex combination space with the Hausdorff metric.
- The space of compact convex subsets of $\mathbb{R}^d$ with the Bartels–Pallaschke metric introduced by Diamond et al. [11].

Moreover, another application is presented to the above-mentioned space of probability distributions.

To obtain these results, some propositions with independent interest will be proved. For instance, it follows from known facts that d_∞, regarded as a bivariate function, is measurable with respect to d_p. We will need to establish the stronger result that d_∞ is lower semicontinuous with respect to d_p. Similarly, we will prove that the Bartels–Pallaschke metric between compact convex sets is lower semicontinuous with respect to the Hausdorff metric. This will imply that the Bartels–Pallaschke distance between two random compact convex sets is a random variable, even when the random sets are *not* measurable with respect to that metric.

Finally, a comparison to the extant literature will be made in Section 7.

2. Preliminaries

Let $(\mathbb{E}, \tau)$ be a topological space. We will denote by $\mathcal{B}_\mathbb{E}$ the Borel σ-algebra in $\mathbb{E}$, i.e. the smallest σ-algebra containing all open sets of $\mathbb{E}$. We will denote by ℓ the Lebesgue measure on $\mathcal{B}_{[0,1]}$ and by $\mathcal{L}_{[0,1]}$ the Lebesgue σ-algebra in $[0, 1]$ (the Lebesgue measure on $\mathcal{L}_{[0,1]}$ will eventually be used as a probability measure and denoted $\mathbb{P}$). A general probability space will be denoted by $(\Omega, \mathcal{A}, P)$.

In [41], Terán and Molchanov established a list of conditions for a metric space provided with a convex combination operation to be a convex combination space.

Definition 2.1. 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 any $v_1, \ldots, v_n \in \mathbb{E}$ this operation produces an element of $\mathbb{E}$, denoted $[\lambda_i, v_i]_{i=1}^n$ or $[\lambda_1, v_1; \cdots; \lambda_n, v_n]$. We will say that $\mathbb{E}$ is a *convex combination space* if the following axioms are satisfied:

(CC1) (Commutativity) For every permutation σ of $\{1, \ldots, n\}$,

$$[\lambda_i, v_i]_{i=1}^n = [\lambda_{\sigma(i)}, v_{\sigma(i)}]_{i=1}^n;$$

(CC2) (Associativity) $[\lambda_i, v_i]_{i=1}^{n+2} = [\lambda_1, v_1; \ldots, \lambda_n, v_n; \lambda_{n+1} + \lambda_{n+2}, [\frac{\lambda_{n+j}}{\lambda_{n+1}+\lambda_{n+2}}; v_{n+j}]_{j=1}^2];$

(CC3) (Continuity) If $u, v \in \mathbb{E}$ and $\lambda^{(k)} \to \lambda \in (0,1)$, then

$$[\lambda^{(k)}, u; 1-\lambda^{(k)}, v] \to [\lambda, u; 1-\lambda, v];$$

(CC4) (Negative curvature) For all $u_1, u_2, v_1, v_2 \in \mathbb{E}$ and $\lambda \in (0,1)$,

$$d([\lambda, u_1; 1-\lambda, u_2], [\lambda, v_1; 1-\lambda, v_2]) \leq \lambda d(u_1, v_1) + (1-\lambda) d(u_2, v_2);$$

(CC5) (Convexification) For each $v \in \mathbb{E}$, there exists $\lim_{n\to\infty}[n^{-1}, v]_{i=1}^n$, which will be denoted by $\mathbf{K}(v)$.

As seen, the metric d considered in the definition of a convex combination space must satisfy some strong conditions, so it would be interesting to know if some convergence results for metric spaces can be proved for convex combination spaces with respect to a distance function related to d. We can find different definitions of a quasimetric in literature, sometimes under the name of near-metric or with more restrictive conditions, as in [9].

Definition 2.2. A mapping $\rho : \mathbb{E} \times \mathbb{E} \to [0, \infty)$ is a *quasimetric* if

1. $\rho(v, v) = 0$ for each $v \in \mathbb{E}$,
2. (Relaxed triangle inequality) There exists an $R \in \mathbb{R}$ such that for every $v_1, v_2, v_3 \in \mathbb{E}$

$$\rho(v_1, v_2) \leq R \cdot (\rho(v_1, v_3) + \rho(v_3, v_2)).$$

If, further, ρ is such that

$$\rho(v, u) \leq C_0 \cdot \rho(u, v)$$

for some constant C_0 valid for all $u, v \in \mathbb{E}$, it will be called a *quasisymmetric quasimetric*.

A quasimetric ρ defines naturally a topology $\tau(\rho)$ formed by all sets $A \subseteq \mathbb{E}$ such that every point $v \in A$ satisfies

$$\{u \in \mathbb{E} \mid \rho(u, v) < \varepsilon\} \subseteq A$$

for some $\varepsilon > 0$.

Definition 2.3. A Borel measurable function $X : (\Omega, \mathcal{A}, P) \to (\mathbb{E}, \tau)$ will be called a *random element* of $\mathbb{E}$.

We will denote by P_X the induced distribution of X and by $L^1(\Omega, \mathcal{A}, P)$ the space of integrable random variables. Notice that ℓ is a probability measure, hence we will denote by ℓ_X the induced distribution of X considering the Lebesgue measure on $\mathcal{B}_{[0,1]}$.

Definition 2.4. Let $(\mathbb{E}, d)$ be a metric space. Let $v_0 \in \mathbb{E}$ be an arbitrary point. A random element $X : (\Omega, \mathcal{A}, P) \to \mathbb{E}$ is called *integrable* if $d(v_0, X)$ is an integrable random variable.

Notice that $d(v_0, X)$ is necessarily a random variable, since $\{d(v_0, X) < t\}$ is the event that X is in the open ball with center v_0 and radius t, for any $t > 0$. The definition of integrability does not depend on the chosen point, since for any $v_1, v_2 \in \mathbb{E}$

$$\begin{aligned}&E[\rho(v_2, X)] \le E[R \cdot (\rho(v_2, v_1) + \rho(v_1, X))]\\&= R \cdot (\rho(v_2, v_1) + E[\rho(v_1, X)]).\end{aligned}$$

The expectation in a convex combination space is defined through the expectation of simple random elements (see Section 4 in [41] for more details).

Definition 2.5. 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 $X = \sum_{j=1}^{r} I_{\Omega_j} v_j$, its *expectation* is $E[X] = [P(\Omega_j), \mathbf{K}(v_j)]_{j=1}^{r}$. 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.

Definition 2.6. Let $\mathbb{E}$ be a topological space. A function $\varphi : \mathbb{E} \to \mathbb{R}$ is *lower* (respectively, *upper*) *semicontinuous* if $\liminf_{v\to v_0} \varphi(v) \ge \varphi(v_0)$ (respectively, $\limsup_{v\to v_0} \varphi(v) \le \varphi(v_0)$) for every $v_0 \in \mathbb{E}$.

A function is lower semicontinuous if and only if its lower level sets $\{v \in \mathbb{E} : f(v) \le a\}$ are closed, for all $a \in \mathbb{R}$.

Definition 2.7. Let $\mathbb{E}$ be a convex combination space and let C be a subset of $\mathbb{E}$. A function $f : C \to \mathbb{R}$ is *convex* in C if

$$f([\lambda_i, v_i]_{i=1}^{n}) \le \sum_{i=1}^{n} \lambda_i f(v_i)$$

whenever $x_i \in C$. It is *midpoint convex* in C if

$$f([1/2, v_1; 1/2, v_2]) \le (\varphi(v_1) + \varphi(v_2))/2$$

for all $v_1, v_2 \in \mathbb{E}$. The function f will be called convex, or midpoint convex, if it is so in $\mathbb{E}$.

We will be specially interested in the convexity properties of distance functions between elements of $\mathbb{E}$. The following result [40, Lemma 4.1] ensures that $\mathbb{E} \times \mathbb{E}$ is a convex combination space.

Lemma 2.1. *If $\mathbb{E}$ is a convex combination space, then $\mathbb{E} \times \mathbb{E}$ is a convex combination space with the convex combination operation*

$$[\lambda_i, (u_i, v_i)]_{i=1}^{n} = ([\lambda_i, u_i]_{i=1}^{n}, [\lambda_i, v_i]_{i=1}^{n})$$

and the metric

$$d_{\max}((u_1, v_1), (u_2, v_2)) = \max\{d(u_1, u_2), d(v_1, v_2)\}.$$

A complete separable convex combination space satisfies the following properties.

Lemma 2.2. *Let $\mathbb{E}$ be a complete separable convex combination space. Let $\rho : \mathbb{E} \times \mathbb{E} \to [0, \infty)$ be a quasimetric on $\mathbb{E}$ which is midpoint convex and lower semicontinuous as a bivariate function. Let X, Y be integrable random elements of $\mathbb{E}$. Then*

1. *There exists a sequence of measurable functions $\phi_k : \mathbb{E} \to \mathbb{E}$ (which does not depend on X) such that each $\phi_k(X)$ is a simple random element, $d(\phi_k(X), X) \searrow 0$ almost surely and $E[d(\phi_k(X), X)] \to 0$.*
2. $\rho(E[X], E[Y]) \le E[\rho(X, Y)]$.

Proof. The first part is [41, Proposition 4.1]. The second one comes from the fact that $\mathbb{E} \times \mathbb{E}$ is a convex combination space (Lemma 2.1) and an application of Jensen's inequality [40, Theorem 3.1] to the function ρ. □

We denote by $\mathcal{K}(\mathbb{E})$ the space of all non-empty compact subsets of $\mathbb{E}$ and by $\mathcal{K}_c(\mathbb{E})$ the space of all non-empty compact convex subsets of $\mathbb{E}$.

Definition 2.8. The *Hausdorff metric* in $\mathcal{K}(\mathbb{E})$ is defined by

$$d_H(K, L) = \max(\sup_{v_1 \in K} \inf_{v_2 \in L} d(v_1, v_2), \sup_{v_2 \in L} \inf_{v_1 \in K} d(v_1, v_2))$$

for every $K, L \in \mathcal{K}(\mathbb{E})$.

As is shown in [41, Theorem 6.2], the space $\mathcal{K}(\mathbb{E})$ is again a convex combination space.

Lemma 2.3. *Endow $\mathcal{K}(\mathbb{E})$ with the convex combination operation*

$$[\lambda_i, K_i]_{i=1}^n = \{[\lambda_i, v_i]_{i=1}^n : v_i \in K_i \ \forall i \in \{1, \ldots, n\}\}.$$

Then $(\mathcal{K}(\mathbb{E}), d_H)$ is a convex combination space. Moreover, if $(\mathbb{E}, d)$ is separable and complete, then $(\mathcal{K}(\mathbb{E}), d_H)$ is so as well.

Definition 2.9. A *random compact set* in a convex combination space $\mathbb{E}$ is a random element of $(\mathcal{K}(\mathbb{E}), d_H)$. A *random compact convex set* in $\mathbb{E}$ is a random element of $(\mathcal{K}_c(\mathbb{E}), d_H)$.

Definition 2.10. A set F of random variables is called *uniformly integrable* if given $\varepsilon > 0$ there exists an $M \geq 0$ such that

$$E[|\xi| \cdot I_{\{|\xi|>M\}}] \leq \varepsilon$$

for all $\xi \in F$.

To prove Vitali's convergence theorem for convex combination spaces, we will use this version for real random variables.

Lemma 2.4. *Let ξ_n, ξ be random variables such that $\{\xi_n\}_n$ is uniformly integrable. If $\xi_n \to \xi$ almost surely, then $E[\xi_n] \to E[\xi]$.*

Next, the extension of this concept for convex combination spaces is as follows.

Definition 2.11. A set F of random elements of a convex combination space $(\mathbb{E}, d)$ will be called *uniformly integrable in d* if given $\varepsilon > 0$ there exists an $M \geq 0$ such that, for some $v \in \mathbb{E}$,

$$E[d(X, v) \cdot I_{\{d(X,v)>M\}}] \leq \varepsilon$$

for all $X \in F$.

Weak convergence of random elements in a topological space is defined as follows (see [6, Chapter 1] for further details).

Definition 2.12. Let X_n, X be random elements in a topological space $\mathbb{E}$. Then $\{X_n\}_n$ *converges weakly* to X if $E[f(X_n)] \to E[f(X)]$ for every continuous bounded function $f : \mathbb{E} \to \mathbb{R}$.

For $\mathbb{E} = \mathbb{R}^d$, weak convergence is the same thing as convergence in distribution. Hence this is a generalization of convergence in distribution to more general spaces.

Definition 2.13. Let $(\mathbb{E}, \rho)$ be a quasimetric space and let $X_n, X : (\Omega, \mathcal{A}, P) \to \mathbb{E}$. We will say that $\{X_n\}_n$ *converges almost surely* to X if each $\rho(X_n, X)$ is a random variable and $\{\rho(X_n, X) \not\to 0\}$ is P-null. We will say that $\{X_n\}_n$ *converges in probability* if each $\rho(X_n, X)$ is a random variable and $P(\rho(X_n, X) < \varepsilon) \to 1$ for each $\varepsilon > 0$.

Definition 2.14. A sequence $\{X_n\}_n$ of random elements in a topological space is *tight* if for any $\varepsilon > 0$ there exists a compact subset $K_\varepsilon \subseteq \mathbb{E}$ such that $P(X_n \in K_\varepsilon) > 1 - \varepsilon$ for each $n \in \mathbb{N}$.

At some points we will work with X_n and X which may fail to be Borel measurable (random elements) with respect to the topology of a quasimetric ρ but satisfy weaker measurability properties (specifically, see assumption 2 in Theorem 3.1) which still allow one to use the notions of weak convergence and tightness, and to ensure that $\rho(X_n, X)$ is a random variable.

Finally, another important tool that will be used in our first result is Jakubowski's almost sure Skorokhod representation [16, Theorem 2], which allows one to obtain a sequence of random elements with the same distribution as a subsequence of tight random elements.

Lemma 2.5 *(Jakubowski). Let $(\mathbb{E}, \tau)$ be a topological space, let $X_n, X : (\Omega, \mathcal{A}, P) \to (\mathbb{E}, \tau)$ be random elements of $\mathbb{E}$. Assume*

1. *There exists a countable set of continuous functions which separates points in $\mathbb{E}$.*
2. *$\{X_n\}_n$ is tight.*

Then there exists a subsequence $\{X_{n'}\}_n$ of $\{X_n\}_n$ and random elements $Y_{n'}, Y : ([0,1], \mathcal{B}_{[0,1]}, \ell) \to (\mathbb{E}, \tau)$ such that

(a) *$P_{X_{n'}} = \ell_{Y_{n'}}$ for each $n \in \mathbb{N}$.*
(b) *$Y_{n'}(t) \to Y(t)$ for each $t \in [0,1]$.*

Recall that a set of functions $\{f_i\}_{i \in I}$ is said to *separate points* of a space $\mathbb{E}$ if, for any points $x \neq y$ there is some $i \in I$ for which $f_i(x) \neq f_i(y)$. The requirement that the functions take on values in $[-1, 1]$ is not restrictive, since $[-1, 1]$ can be replaced by $\mathbb{R}$ in the statement. Indeed, if $\{f_n : \mathbb{E} \to \mathbb{R}\}_{n \in \mathbb{N}}$ separates points in $\mathbb{E}$, taking

$$f_{n,m} = m^{-1} \cdot \max(-m, \min(f_n, m))$$

one obtains a countable family which still separates points: $f_n(x) \neq f_n(y)$ implies $f_{n,m}(x) \neq f_{n,m}(y)$ for $m > \max\{|f_n(x)|, |f_n(y)|\}$.

Definition 2.15. A topological space $\mathbb{E}$ is *Polish* if its topology is generated by some complete separable metric, *Lusin* if it is the continuous image of a Polish space by a bijective mapping, *Suslin* if it is the continuous image of a Polish space, and *Radon* if $P(A) = \sup_{K \subseteq A} P(K)$ for every probability measure P on $\mathcal{B}_{\mathbb{E}}$ (where K ranges over compact sets).

Definition 2.16. A probability measure P in a measurable space $(\Omega, \mathcal{A})$ is *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 $A_1 \subseteq A \subseteq A_2$ and $P(X \in A_2 \setminus A_1) = 0$.

Definition 2.17. A measurable space $(\Omega, \mathcal{A})$ is *perfect* if every probability measure defined in $\mathcal{A}$ is perfect.

The next result appears, e.g., in [1, Lemma 4.4].

Lemma 2.6. *Every Polish space, endowed with its Borel σ-algebra, is perfect.*

Consider the following spaces of fuzzy sets of $\mathbb{R}^d$:

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

where

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

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

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

Let us consider the following metrics in $\mathcal{F}(\mathbb{R}^d)$ and $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$, respectively:

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

Thus $\mathcal{F}_c(\mathbb{R}^d)$ is a proper subset of both $\mathcal{F}(\mathbb{R}^d)$ and $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$.

Definition 2.18. A *fuzzy random variable* is a mapping $X : (\Omega, \mathcal{A}, P) \to (\mathcal{F}(\mathbb{R}^d), d_p)$ such that for every $\alpha \in [0, 1]$, the α-cut mapping X_α given by $X_\alpha(\omega) = X(\omega)_\alpha$ is a random compact set.

The support function allows one to characterize convex fuzzy sets via certain real functions.

Definition 2.19. For each $U \in \widehat{\mathcal{F}}_{c,1}(\mathbb{R}^d)$ its *support function* is defined by

$$s_U : \mathbb{S}^{d-1} \times (0,1] \to \mathbb{R}$$
$$(r, \alpha) \mapsto s_U(r, \alpha) = \max_{x\in U_\alpha} \langle r, x\rangle.$$

For $U \in \mathcal{F}_c(\mathbb{R}^d)$, the support function extends naturally to $\mathbb{S}^{d-1} \times [0, 1]$ with $s_U(r, 0) = \max_{x\in U_0}\langle r, x\rangle$.
For $p \in [1, \infty)$, we consider the metric ([10, (1.14), p. 53]) in $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$ given by

$$\rho_p(U, V) = \left(\int_{(0,1]\times\mathbb{S}^{d-1}} |s_U(r, \alpha) - s_V(r, \alpha)|^p dr\, d\alpha \right)^{1/p}.$$

By Theorems 3 and 4 in [42], $(\mathcal{F}_c(\mathbb{R}^d), d_p)$ and $(\mathcal{F}_c(\mathbb{R}^d), d_\infty)$ are convex combination spaces, the former being separable but not complete and the latter being complete but not separable.

3. Statement and discussion of a Vitali convergence theorem

In this section, we will state and discuss the abstract Vitali theorem for (subspaces of) convex combination spaces, in sufficient generality for ρ assumed to be a quasimetric, which is not necessarily a metric. In Section 5, further variants will be presented.

Theorem 3.1. *Let $(\mathbb{E}, d)$ be a separable complete convex combination space. Let C be a convex subset of $\mathbb{E}$. Let ρ be a midpoint convex, lower semicontinuous quasimetric on C. Let X_n and X be random elements in $(\mathbb{E}, d)$ taking on values in C such that there exist $v_1, v_2 \in C$ verifying*

1. *X_n and X are integrable, and $E[X_n], E[X] \in C$.*
2. *There exists a countable family $\{f_i : \mathbb{E} \to [-1, 1]\}_{i\in\mathbb{N}}$ of ρ-continuous functions which separates points in $\mathbb{E}$, and the X_n and X are measurable with respect to the σ-algebra generated by the $\{f_i\}_i$, i.e., the smallest σ-algebra which makes the mappings $\{f_i\}_i$ measurable.*

3. *$\{X_n\}_n$ is ρ-tight and X is ρ-tight.*
4. *$E[\rho(v_1, X)] < \infty$.*
5. *$\{\rho(X_n, v_2)\}_n$ is uniformly integrable.*

If $X_n \to X$ weakly in ρ, then $\rho(E[X_n], E[X]) \to 0$.

Moreover, if ρ is quasisymmetric then assumptions (2) and (3) can be replaced by the following:

2'. *X_n, X are random elements with respect to ρ.*
3'. *X takes on values almost surely in a ρ-separable subset of C.*

This abstract, general theorem has a large number of assumptions which are not (or, rather, are not visible) in the traditional statement for random variables in $\mathbb{R}$ (Lemma 2.4). That makes it convenient to proceed to a careful discussion for the reader's benefit.

1. **The quasimetric ρ.** There are two ways of using this theorem. The first is as a convergence theorem in abstract convex combination spaces, which is achieved by taking the special case that $C = \mathbb{E}$ and $\rho = d$ (see Section 5). This includes, as a particular case, the Vitali theorem for Bochner expectations in a separable Banach space, since the linear space operations and the metric induced by the norm define a separable complete convex combination space. Thus it contributes to showing the viability of convex combination spaces as an abstract framework for spaces where some of the key properties of a normed linear space are not satisfied.

The theorem can also be used as a *generator of Vitali theorems* in concrete spaces (or types of spaces). A convergence theorem for the expectation has two components: the structure of the carrier space, and the type of convergence. The first component is necessary to define which functions are measurable (random elements), which are integrable and how the integral (expectation) is calculated. For instance, in a separable Banach space the Bochner integral is constructed using limits of simple functions. A convergence theorem for the Bochner integral in an abstract Banach space will then also use convergence in the norm as the second component, because that is the natural convergence provided by the Banach space structure.

But, when working in a concrete space whose elements are not naturally elements of a linear space, the typical situation is that there is not a unique way of calculating the distance, and that those distance functions have a varying degree of mathematical niceness. In the specific situation of fuzzy sets, there are several (if not many) metrics: some are not separable, some are not complete, some do not define a convex combination space. Then the threefold requirement of separability, completeness, and the convex combination axioms is still quite strong. Thus, following the idea of the second author in [40, Theorem 4.2] both components are explicitly split apart:

(i) We use a nice metric d for the structure-building. It defines measurability, integrability and expectation, for which it is assumed to be separable, complete and to define a convex combination space.
(ii) Then the convergence theorem is obtained in a different distance function ρ which is asked to satisfy only weaker properties. It can fail to be separable or complete, to be a metric and to define a convex combination space. The X_n may fail to converge in the nice metric d, and even to be random elements with respect to ρ. We still obtain the conclusion '$X_n \to X$ weakly in ρ implies $E[X_n] \to E[X]$ in ρ'.

The nice structure of the space with the metric d can be used to *generate* convergence theorems in specific spaces for different choices of ρ. The price to pay in the abstract theorem is the addition of appropriate assumptions on the relationship between ρ and d, and between the X_n and ρ.

2. **The subspace C.** When developing the applications in Section 6, we realized the convenience of not working in $\mathbb{E}$ but in a certain subset C. The rationale was that considering both d and ρ leads to situations in which ρ cannot be defined in the whole of $\mathbb{E}$ without using infinite values. Then C is used to restrict adequately the domain of ρ.

If C were d-closed, since it is convex in $\mathbb{E}$ it would be a separable complete convex combination space itself. Thus, in the more interesting applications C will be non-closed or even non-measurable in $\mathbb{E}$ (e.g., if $\mathbb{E}$ is the completion of C then C may not be measurable in $\mathbb{E}$). In absence of measurability, it is strictly necessary to distinguish between C and $\mathbb{E}$ since measurability of X_n, X as C-valued functions would not be equivalent to their measurability as $\mathbb{E}$-valued functions. This explains why the theorem cannot be stated using only one space. Also note that applying Theorem 3.1

as a 'generator' will produce a Vitali convergence theorem in (C, ρ), not $(\mathbb{E}, d)$. Different choices of ρ may come with different choices of C and thus produce convergence theorems in different spaces.

3. **Integrability of X.** In the Vitali theorem for real-valued random variables (Lemma 2.4), the limit is not assumed to be integrable since it follows from the other conditions. If $\rho = d$, the integrability of X can be proved using similar ideas combined with the key fact that X is the (weak) limit of X_n in d. That breaks down in the general case since the assumption is $X_n \to X$ weakly in ρ, not d. In that case it becomes necessary to check the integrability of X, but since one also has to prove $E[X] \in C$, it is not an additional burden to prove that $E[X]$ exists.

4. **The existence of countably many ρ-continuous separating functions.** This assumption is needed to apply Jakubowski's variant of the Skorokhod representation theorem. Applying Theorem 3.1 will typically not require looking for a countable sequence $\{f_n\}_n$ of separating functions. In many cases, ρ will be a metric and the second part of the theorem will apply. For the case that condition (2') is not met, it is not hard to prove the following:

(i) If ρ is a separable metric then the condition is satisfied: take a countable dense subset $\{v_n\}_n$ and set $f_n = \rho(v_n, \cdot)/(1 + \rho(v_n, \cdot))$.
(ii) If a space satisfies the condition then any finer topology in it does so as well: the f_n themselves are still continuous in the finer topology.

For instance, in the case of fuzzy sets the metric d_∞ is not separable but the weaker metric d_1 is, meaning that this condition is satisfied for $\rho = d_\infty$.

With respect to the measurability requirement, it is equivalent to require each $f_i(X_n)$ and $f_i(X)$ to be a random variable. This is enough to ensure that tightness and weak convergence are well defined in (C, ρ), see the discussion in [16, pp. 169–170].

If the X_n and X are random elements with respect to ρ, that is immediately satisfied. If ρ defines a finer topology than d (e.g., if $\rho = d_\infty$, $d = d_1$), by the discussion above the requirement is satisfied. Also, if ρ defines a coarser topology than d, since X_n, X are random elements with respect to d it is satisfied as well.

5. **The ρ-tightness of $\{X_n\}_n$ and X.** This condition is necessary to apply Jakubowski's theorem as well. Let us briefly comment on some special cases.

If ρ is a separable complete metric (e.g., in the important case $\rho = d$), by Prokhorov's theorem the assumption that X_n converges weakly already implies that this condition is satisfied. More generally, in a Prokhorov space (see [7, Definition 8.10.8]; that includes, e.g., all locally compact Hausdorff spaces) this condition can be eliminated or simplified, depending on the extra properties of the topology of ρ.

If ρ defines a metrizable topology, by the results in [7, Section 8.10.(ii)] the tightness of each X_n and X individually, together with the weak convergence of X_n, ensure that the condition holds. In particular, if additionally (C, ρ) is a Radon space (or, more specially, a Lusin or Suslin space) then the condition is always satisfied.

In non-metrizable spaces, useful ways to simplify this condition can be found in [45, Section 7].

6. **Integrability conditions with respect to ρ.** Since X is integrable, the random variable $d(v, X)$ is integrable for each $v \in \mathbb{E}$. That is not enough to ensure $\rho(v, X)$ is integrable as well (such is the case, for instance, in spaces of fuzzy sets when $d = d_1$, $\rho = d_\infty$, $v = I_{\{0\}}$). However, as follows from the proof, $\rho(v, X)$ and $\rho(X, v)$ are always random variables, so the measurability of each $\rho(X_n, v_2)$ is granted and not part of the assumption.

Notice that uniform integrability is required for a variable of the form $\rho(X_n, v)$, not $\rho(v, X_n)$ (recall that ρ is not required to be a symmetric function). Hence we prefer to use two points v_1, v_2 in the statement (although the proof is reduced to the case $v_1 = v_2$) since, in the absence of symmetry, working with distances *from* X_n and *to* X_n may have to be done differently and it may be that the points v_1, v_2 for which checking the conditions is simpler, are different.

4. Proof of Theorem 3.1

We start by stating some results which will be used in the proof.

Lemma 4.1. *Let X, Y be integrable random elements of a complete separable convex combination space. If $P_X = P_Y$ then $E[X] = E[Y]$.*

Proof. Assume first X and Y are simple. Since they are identically distributed, they must take on the same values $\{v_i\}_{i=1}^k$ with the same probabilities p_i. By definition,

$$E[X] = E[Y] = [p_i, \mathbf{K}(v_i)]_{i=1}^k.$$

For the general case, by Lemma 2.2 (1) there exists a sequence $\{\phi_k\}_k$ of measurable functions such that $\phi_k(X)$ and $\phi_k(Y)$ are simple random elements and $\phi_k(X) \to X$ and $\phi_k(Y) \to Y$ almost surely. Therefore

$$d(E[X], E[Y]) \leq d(E[X], E[\phi_k(X)]) + d(E[\phi_k(X)], E[\phi_k(Y)]) + d(E[\phi_k(Y)], E[Y]).$$

On one hand, $d(E[X], E[\phi_k(X)]) \to 0$, by Lemma 2.2 (1) and (2). Analogously, $d(E[Y], E[\phi_k(Y)]) \to 0$. On the other hand, since $\phi_k(X)$ and $\phi_k(Y)$ are simple random elements with the same distribution, $d(E[\phi_k(X)], E[\phi_k(Y)]) = 0$. Then $d(E[X], E[Y]) = 0$, namely $E[X] = E[Y]$. □

The following characterization of uniform integrability can be found, e.g., in [3, Theorem 2.4.5, p. 57] (Proposition 2.4.12 in that book ensures that their definition of uniform integrability and ours are equivalent).

Lemma 4.2 *(Dunford–Pettis theorem). Let $(\Omega, \mathcal{A}, P)$ be a probability space. Then a subset of $L^1(\Omega, \mathcal{A}, P)$ is uniformly integrable if and only if it is relatively compact in the weak topology of $L^1(\Omega, \mathcal{A}, P)$.*

We will also use a recent metrization theorem [26, Theorem 3.26.(12)]. The following lemma contains part of the content of the theorem.

Lemma 4.3 *(Mitrea). Let $\mathbb{E}$ be a set and $\rho : \mathbb{E} \times \mathbb{E} \to [0, \infty)$ a function for which constants $C_0, C_1 > 0$ exist such that, for all $x, y, z \in \mathbb{E}$,*

(i) $\rho(x, y) = 0 \Leftrightarrow x = y$,
(ii) $\rho(y, x) \leq C_0 \cdot \rho(x, y)$,
(iii) $\rho(x, z) \leq C_1 \cdot \max\{\rho(x, y), \rho(y, z)\}$.

Then the topology $\tau(\rho)$ is metrizable by a metric $\tilde{d}$ such that

$$C_1^{-2} \cdot \rho(x, y) \leq \tilde{d}(x, y)^{\log_2 C_1} \leq C_0 \cdot \rho(x, y). \tag{1}$$

Proof. The original result has $\max\{1, C_0\}$ instead of C_0 in (1). But if $C_0 < 1$ holds true, a double application of (ii) gives $\rho(x, y) \leq C_0^2 \cdot \rho(x, y)$ (i.e., $\rho(x, y) = 0$) for arbitrary x, y. In view of (i), either $C_0 \geq 1$ or $\mathbb{E}$ has a single point, and in both cases the inequality holds with C_0 replacing $\max\{1, C_0\}$. □

Note that the lower semicontinuity of ρ with respect to d, by (1), passes on to $\tilde{d}$; but the midpoint convexity of ρ will not be preserved by $\tilde{d}$ except in special cases.

We proceed now to the proof of Theorem 3.1.

Proof of Theorem 3.1. Since integrability does not depend on the chosen point, there exists a $v_3 \in C$ satisfying both conditions (4) and (5). Reasoning by contradiction, assume that $\rho(E[X_n], E[X])$ does not converge to 0. Then there exist a subsequence $\{X_{n'}\}_n$ of $\{X_n\}_n$ and a neighborhood U of $E[X_{n'}]$ such that $X_{n'} \notin U$ for each $n \in \mathbb{N}$.

By Lemma 2.5, for an appropriate subsequence $\{X_{n''}\}_n$ of $\{X_{n'}\}_n$, there exist random elements $Z_{n''}, Z : ([0, 1], \mathcal{B}_{[0,1]}, \ell) \to (\mathbb{E}, \rho)$ such that $\ell_{Z_{n''}} = P_{X_{n''}}$, $\ell_Z = P_X$ and $Z_{n''}(t) \to Z(t)$ for every $t \in [0, 1]$, with respect to the topology generated by ρ. Now, notice that $\mathcal{L}_{[0,1]}$ contains $\mathcal{B}_{[0,1]}$, so $Z_{n''}, Z : ([0, 1], \mathcal{L}_{[0,1]}, \mathbb{P}) \to (\mathbb{E}, \rho)$, with $\mathbb{P}$ being the completion of ℓ, are random elements.

Although it is tempting to claim that the probability that $Z_{n''}$ and Z are in C is 1, the possible non-measurability of C requires the following argument. Since

$$\{\omega \in \Omega : X_{n''}(t) \in C\} = \Omega,$$

it is in particular a measurable set. The metric space $(\mathbb{E}, d)$ is complete and separable, so it is Polish. By Lemma 2.6, $(\mathbb{E}, \mathcal{B}_{\mathbb{E}})$ is perfect, whence there exists a $C_n^* \in \mathcal{B}_{\mathbb{E}}$ such that $C_n^* \subseteq C$ and

$$P(X_{n''} \in C_n^*) = P(X_{n''} \in C) = 1.$$

Since $Z_{n''}$ and Z have the same distributions as $X_{n''}$ and X, and $C_n^* \in \mathcal{B}_{\mathbb{E}}$,

$$\mathbb{P}(\{t \in [0,1] : Z_{n''}(t) \in C_n^*\}) = P(\{\omega \in \Omega : X_{n''}(\omega) \in C_n^*\}) = 1.$$

Thus $\{t \in [0,1] : Z_{n''}(t) \notin C_n^*\}$ is an ℓ-null set which contains $\{t \in [0,1] : Z_{n''}(t) \notin C\}$, therefore the latter is ℓ-null too. Since the Lebesgue σ-algebra $\mathcal{L}_{[0,1]}$ is complete, $\{t \in [0,1] : Z_{n''}(t) \in C\}$ is measurable. In conclusion,

$$\mathbb{P}(\{t \in [0,1] : Z_{n''}(t) \in C\}) = \mathbb{P}(\{t \in [0,1] : Z_{n''}(t) \in C^*\}) = 1.$$

Accordingly, we define random elements $Y_{n''}$ as follows:

$$Y_{n''}(t) = \begin{cases} Z_{n''}(t) & \text{if } Z_{n''}(t) \in C \\ v_3 & \text{if } Z_{n''}(t) \notin C. \end{cases}$$

We similarly check that $\{t \in [0,1] : Z(t) \notin C\}$ is a null Lebesgue measurable set. Thus the set

$$N = \{t \in [0,1] : Z(t) \notin C\} \cup \bigcup_{n \in \mathbb{N}} \{t \in [0,1] : Z_{n''}(t) \notin C\}$$

is so as well. Let us check that the $Y_{n''}$ are random elements. For any $B \in \mathcal{B}_{\mathbb{E}}$,

$$Y_{n''}^{-1}(B) = (N \cap Y_{n''}^{-1}(B)) \cup (N^c \cap Z_{n''}^{-1}(B))$$

where $N \cap Y_{n''}^{-1}(B)$ is null (hence Lebesgue measurable) and $N^c \cap Z_{n''}^{-1}(B)$ is measurable.

Let us show now that $Y_{n''}$ has the same distribution as $Z_{n''}$. For $B \in \mathcal{B}_{\mathbb{E}}$,

$$\begin{aligned}\mathbb{P}_{Y_{n''}}(B) &= \mathbb{P}(\{t \in [0,1] : Y_{n''}(t) \in B\}) = \mathbb{P}(\{t \in [0,1] : t \in Y_{n''}^{-1}(B)\}) \\ &= \mathbb{P}((N \cap Y_{n''}^{-1}(B)) \cup (N^c \cap Z_{n''}^{-1}(B))) = \mathbb{P}(N \cap Y_{n''}^{-1}(B)) + \mathbb{P}(N^c \cap Z_{n''}^{-1}(B)) \\ &= \mathbb{P}(\{t \in N^c : t \in Z_{n''}^{-1}(B)\}) = \mathbb{P}(\{t \in [0,1] : t \in Z_{n''}^{-1}(B)\}) = \mathbb{P}_{Z_{n''}}(B).\end{aligned}$$

Analogously, one can define

$$Y(t) = \begin{cases} Z(t) & \text{if } Z(t) \in C \\ v_3 & \text{if } Z(t) \notin C \end{cases}$$

and verify that it is a random element which takes values in C and is distributed as X.

Next, by Lemma 2.1, $(Y_{n''}, Y) : \Omega \to \mathbb{E} \times \mathbb{E}$ is Borel measurable in $d_{\max}$ (because the Borel σ-algebra of $d_{\max}$ equals the product σ-algebra $\mathcal{B}_{\mathbb{E}} \otimes \mathcal{B}_{\mathbb{E}}$). Since $\rho : \mathbb{E} \times \mathbb{E} \to \mathbb{R}$ is lower semicontinuous, the set

$$\rho^{-1}((-\infty, t]) = \{(x, y) \mid \rho(x, y) \le t\}$$

is closed for each $t \in \mathbb{R}$, hence ρ is Borel measurable as a bivariate function. Therefore $\rho(Y_{n''}, Y)$, being the composition of measurable functions, is a random variable.

Let us show that $\{\rho(Y_{n''}, Y)\}_n$ is uniformly integrable. Since $Y_{n''}$ has the same distribution as $X_{n''}$, for each $n \in \mathbb{N}$ and $M \ge 0$ it follows that $\rho(Y_{n''}, v_3) \cdot I_{\{\rho(Y_{n''}, v_3) > M\}}$ has the same distribution as $\rho(X_{n''}, v_3) \cdot I_{\{\rho(X_{n''}, v_3) > M\}}$. Hence

$$E\left[\rho(Y_{n''}, v_3) \cdot I_{\{\rho(Y_{n''}, v_3) > M\}}\right] = E\left[\rho(X_{n''}, v_3) \cdot I_{\{\rho(X_{n''}, v_3) > M\}}\right].$$

By an analogous reasoning, $E[\rho(v_3, Y)] = E[\rho(v_3, X)]$, so $\rho(v_3, Y)$ is an integrable random variable.

Set

$$\begin{aligned} T : L^1([0,1], \mathcal{L}_{[0,1]}, \mathbb{P}) &\to L^1([0,1], \mathcal{L}_{[0,1]}, \mathbb{P}) \\ f &\mapsto T(f) = f + \rho(v_3, Y). \end{aligned}$$

Since $L^1([0,1], \mathcal{L}_{[0,1]}, \mathbb{P})$ is a topological vector space and $\rho(v_3, Y)$ is integrable, T is well defined and continuous. By Lemma 4.2, the set $\{\rho(Y_{n''}, v_3)\}_n$ is relatively compact in the weak topology of $L^1([0,1], \mathcal{L}_{[0,1]}, \mathbb{P})$. Since the continuous image of a relatively compact set is relatively compact (e.g., [13, Theorem 6.8, p. 254]), $\{T(\rho(Y_{n''}, v_3))\}_n$ is also relatively weakly compact. Again by Lemma 4.2, the sequence $\{\rho(Y_{n''}, v_3) + \rho(v_3, Y)\}_n$ is uniformly integrable.

Then for $\varepsilon > 0$, there exists an $M^* > 0$ such that

$$E\left[(\rho(Y_{n''}, v_3) + \rho(v_3, Y)) \cdot I_{\{\rho(Y_{n''},v_3)+\rho(v_3,Y)>M^*\}}\right] \leq \varepsilon/R,$$

where $R > 0$ is the constant in the relaxed triangle inequality for the quasimetric ρ. Letting $M = R \cdot M^*$,

$$\begin{aligned} E\left[\rho(Y_{n''}, Y) \cdot I_{\{\rho(Y_{n''},Y)>M\}}\right] &\leq E\left[R \cdot (\rho(Y_{n''}, v_3) + \rho(v_3, Y)) \cdot I_{\{\rho(Y_{n''},Y)>M\}}\right] \\ &\leq E\left[R \cdot (\rho(Y_{n''}, v_3) + \rho(v_3, Y)) \cdot I_{\{R\cdot(\rho(Y_{n''},v_3)+\rho(v_3,Y))>M\}}\right] \\ &\leq R \cdot E\left[(\rho(Y_{n''}, v_3) + \rho(v_3, Y)) \cdot I_{\{(\rho(Y_{n''},v_3)+\rho(v_3,Y))>M^*\}}\right] \leq \varepsilon. \end{aligned}$$

Thus, $\{\rho(Y_{n''}, Y)\}_n$ is uniformly integrable. Recall that, by construction, $Y_{n''} \to Y$ in ρ almost surely. By Lemma 2.4, $E[\rho(Y_{n''}, Y)] \to 0$. By Lemma 2.2 (2), $\rho(E[Y_{n''}], E[Y]) \leq E[\rho(Y_{n''}, Y)]$, so it follows that

$$\rho(E[Y_{n''}], E[Y]) \to 0.$$

By Lemma 4.1, $E[X] = E[Y]$ and $E[X_{n''}] = E[Y_{n''}]$ for every $n'' \in \mathbb{N}$. Therefore

$$\rho(E[X_{n''}], E[X]) \to 0.$$

This contradicts the fact that $X_{n''} \notin U$ for each $n \in \mathbb{N}$ (recall that U is a neighborhood of $E[X]$). Accordingly, it is false that $E[X_n]$ does not ρ-converge to $E[X]$. That proves the first part.

The proof of the second part is very similar. Lemma 4.3 provides a topologically equivalent metric $\tilde{d}$. Indeed, ρ satisfies the assumptions in the lemma, with $C_1 = 2R$ in (iii), since the sum of two terms is bounded above by twice their maximum. Observe then that both weak convergence and almost sure convergence only depend on the topology, so a Skorokhod representation theorem for $\tilde{d}$ yields a Skorokhod representation for ρ. That allows us to use a Skorokhod theorem for metric instead of nonmetric spaces. The assumptions in Wichura's version [46, Theorem 1] are (2') and (3') instead of (2) and (3). The remainder of the proof is analogous, using ρ. As discussed before, since $\tilde{d}$ may fail to be midpoint convex it cannot replace ρ in the whole of the proof. Note that, in [46, Theorem 1], the measurable space is not guaranteed to be [0, 1]; the role of the Lebesgue σ-algebra in the proof is played by the completion of the σ-algebra of that space. □

5. Dominated convergence theorem and variants of the Vitali convergence theorem

In this section we present alternative forms of Theorem 3.1 and derive similar versions of the dominated convergence theorem.

Let us begin by stating the Vitali theorem in the special case, when the convergence is in the metric given by the structure of convex combination space (i.e., $C = \mathbb{E}$ and $\rho = d$).

Corollary 5.1. *Let $(\mathbb{E}, d)$ be a separable complete convex combination space. Let X_n and X be random elements in $\mathbb{E}$ such that $\{X_n\}_n$ is uniformly integrable in d.*

If $X_n \to X$ weakly in d, then $d(E[X_n], E[X]) \to 0$.

Proof. We will apply the second part of Theorem 3.1. If $\rho = d$, obviously the assumptions on ρ, as well as (2') and (3'), are satisfied. From the uniform integrability follows the finiteness of $E[d(X_n, v)]$, i.e. the integrability of the X_n. The integrability of X in (1) follows then reasoning like in the real case, with an application of Skorokhod's representation theorem to obtain an almost surely converging sequence with the same distributions. Assumption (4) is just the same thing as integrability, whereas (5) is the uniform integrability assumed in the statement of the corollary. □

When particularized to $\mathbb{R}$, this corollary is still stronger than the form of the Vitali theorem used in the proof (Lemma 2.4).

Using the equivalence between uniform integrability in $\mathbb{E}$ and uniform integrability of a sequence of distances, from the Vitali theorem one can obtain a dominated convergence theorem as follows.

Corollary 5.2. *Let $(\mathbb{E}, d)$ be a separable complete convex combination space. Let X_n and X be random elements in $\mathbb{E}$ such that, for some $v_2 \in \mathbb{E}$, there exists a $g \in L^1(\Omega, \mathcal{A}, P)$ with $d(X_n, v_2) \leq g$ for all $n \in \mathbb{N}$.*

If $X_n \to X$ weakly in ρ, then $\rho(E[X_n], E[X]) \to 0$.

Proof. By hypothesis, the sequence of random variables $\{\rho(X_n, v_2)\}_n$ is dominated by a function in $L^1(\Omega, \mathcal{A}, P)$. By [8, Remark 3.13.(b), p. 72], every dominated sequence is uniformly integrable. Then, for all $\varepsilon > 0$, there exists an $M \geq 0$ such that

$$E\left[\rho(X_n, v_2) \cdot I_{\{\rho(X_n, v_2) > M\}}\right] \leq \varepsilon,$$

that is, $\{\rho(X_n, v_2)\}_n$ is uniformly integrable. For the result, it suffices to apply Corollary 5.1. □

The assumptions involved in the Skorokhod theorems are not necessary if the weak convergence is suitably strenghtened.

Theorem 5.3. *Let $(\mathbb{E}, d)$ be a separable complete convex combination space. Let C be a convex subset of $\mathbb{E}$. Let ρ be a midpoint convex, lower semicontinuous quasimetric in C. Let X_n and X be random elements in $\mathbb{E}$ taking on values in C such that there exist $v_1, v_2 \in C$ verifying*

1. *X_n and X are integrable, and $E[X_n], E[X] \in C$.*
4. *$E[\rho(v_1, X)] < \infty$.*
5. *$\{\rho(X_n, v_2)\}_n$ is uniformly integrable.*

If $X_n \to X$ in probability in ρ, then $\rho(E[X_n], E[X]) \to 0$.

Proof. We begin by dispelling the concern whether it makes sense to say that $X_n \to X$ in probability in ρ, since no measurability assumption with respect to ρ remains. That convergence in probability is, by definition, the same thing as saying that $\rho(X_n, X) \to 0$ in probability. And in the proof of Theorem 3.1 it is shown that $\rho(Y_{n''}, Y)$ is a random variable; the same argument applies to $\rho(X_n, X)$.

To obtain the theorem under the assumption that $X_n \to X$ in probability, consider the subsequence $\{X_{n'}\}_n$ taken in the first paragraph of the proof of Theorem 3.1. Since $\rho(X_{n'}, X) \to 0$ in probability, there must be a further subsequence $\{\rho(X_{n''}, X)\}_n$ converging to 0 almost surely. Namely, $X_{n''} \to X$ almost surely in ρ. With this almost surely convergent subsequence, the proof of Theorem 3.1 carries on with $X_{n''}, X$ in place of $Y_{n''}, Y$. Since it becomes unnecessary to invoke the Skorokhod representation theorem, auxiliary variables $Z_{n''}, Z$ are not needed and assumptions (2) and (3) can be omitted. □

The corresponding dominated convergence theorem is as follows.

Corollary 5.4. *Let $(\mathbb{E}, d)$ be a separable complete convex combination space. Let C be a convex subset of $\mathbb{E}$. Let ρ be a midpoint convex, lower semicontinuous metric in C. Let X_n and X be random elements in $\mathbb{E}$ taking on values in C such that there exist $v_1, v_2 \in C$ verifying*

1. *X_n and X are integrable, and $E[X_n], E[X] \in C$.*
4. *$E[\rho(X, v_1)] < \infty$.*
5. *For some $v_2 \in \mathbb{E}$, there exists a $g \in L^1(\Omega, \mathcal{A}, P)$ such that $\rho(X_n, v_2) \leq g$ for all $n \in \mathbb{N}$.*

If $X_n \to X$ in probability in ρ, then $\rho(E[X_n], E[X]) \to 0$.

Finally, we will state dominated convergence theorems corresponding to Theorem 3.1. While Corollary 5.6 is valid for a quasisymmetric quasimetric, for ease of reference we state it for a metric.

Corollary 5.5. *Let $(\mathbb{E}, d)$ be a separable complete convex combination space. Let C be a convex subset of $\mathbb{E}$. Let ρ be a midpoint convex, lower semicontinuous quasimetric in C. Let X_n and X be random elements in $\mathbb{E}$ taking on values in C such that there exist $v_1, v_2 \in C$ verifying*

1. *X_n and X are integrable, and $E[X_n], E[X] \in C$.*
2. *There exists a countable family $\{f_i : \mathbb{E} \to [-1, 1]\}_{i\in\mathbb{N}}$ of ρ-continuous functions which separates points in $\mathbb{E}$, and the X_n and X are measurable with respect to the σ-algebra generated by the $\{f_i\}_i$.*
3. *$\{X_n\}_n$ is ρ-tight and X is ρ-tight.*
4. *$E[\rho(v_1, X)] < \infty$.*
5. *For some $v_2 \in \mathbb{E}$, there exists a $g \in L^1(\Omega, \mathcal{A}, P)$ such that $\rho(X_n, v_2) \le g$ for all $n \in \mathbb{N}$.*

If $X_n \to X$ weakly in ρ, then $\rho(E[X_n], E[X]) \to 0$.

Corollary 5.6. *If in Corollary 5.5 ρ is a metric and assumptions 2 and 3 are replaced by 2' and 3' below, its conclusion still holds.*

2'. *X_n, X are random elements with respect to ρ.*
3'. *X takes on values almost surely in a ρ-separable subset of C.*

6. Applications

In this section, we will present a number of theorems in specific spaces or types of spaces, which follow from Theorem 3.1. We will show that both results from the literature and results not in the literature can be reached with our approach. Our aim is to illustrate how to apply that theorem even in cases where the space is not a complete separable convex combination space. It should be understood that, in some cases, a direct proof is simpler than a proof through Theorem 3.1 or may not involve steps like establishing the lower semicontinuity of ρ with respect to d.

A detailed comparison with the literature will be performed in Section 7. Applications focus on spaces of crisp and fuzzy sets, as well as the space of probability distributions discussed in the Introduction. Further examples of convex combination spaces can be found, for instance, in [41,37].

For space reasons, we will not write explicitly the version of the dominated convergence theorem which follows from each instance of a Vitali theorem.

6.1. Space of probability distributions with finite variance

Let $\mathcal{P}_2(\mathbb{R})$ denote the space of all probability distributions in the real line having finite variance. This space is endowed with the convex combination operation

$$[\lambda_i, P_i]_{i=1}^n = \lambda_1 \cdot P_1 * \ldots * \lambda_n \cdot P_n,$$

i.e., $[\lambda_i, P_i]_{i=1}^n$ is the distribution of a random variable $\sum_{i=1}^n \lambda_i X_i$ where the X_i are independent, each having P_i as its distribution. With the Wasserstein metric w_2 (both the operation $*$ and the metric were defined in the Introduction), the space $\mathcal{P}_2(\mathbb{R})$ becomes a complete separable convex combination space [40, Lemma 6.2].

In that space, the expectation of a random distribution is always a degenerate distribution (which can be identified with a non-random number). This is related to the extreme scarceness of convex points in the space, and to the fact that only convex points can possibly be the limits in the law of large numbers: the only convex points are the degenerate distributions. An example will clarify that: the formula

$$[n^{-1}, \mathcal{N}(\mu, \sigma)]_{i=1}^n = \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}}) \to \delta_\mu$$

implies that the expectation of a degenerate random distribution which is constantly $\mathcal{N}(\mu, \sigma)$ is δ_μ (i.e., it is given by the expectation of the $\mathcal{N}(\mu, \sigma)$), not $\mathcal{N}(\mu, \sigma)$ itself like in a linear space. Due to this phenomenon, the law of large numbers in this space has the interesting corollary that a sequence $n^{-1}\sum_{i=1}^n X_i$ of averages *of random variables* will converge to a non-random value if the distribution of each X_i is chosen randomly.

For any integrable random element X of $\mathcal{P}_2(\mathbb{R})$, let μ_X be defined by the identity $E(X) = \delta_{\mu_X}$. As an application of Corollary 5.1, we have the following Vitali-type theorem.

Corollary 6.1. *Let X_n and X be random elements of $\mathcal{P}_2(\mathbb{R})$ such that $\{X_n\}_n$ is uniformly integrable in w_2.*
If $X_n \to X$ weakly in w_2 then $\mu_{X_n} \to \mu_X$.

Proof. By Corollary 5.1,

$$|\mu_{X_n} - \mu_X| = w_2(E[X_n], E[X]) \to 0. \quad \square$$

6.2. Fuzzy sets with the d_p-metric

Consider the metric space $(\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d), d_p)$, with $p \in [0, \infty)$. From [19, Corollary 3.3], it is the completion of $(\mathcal{F}_c(\mathbb{R}^d), d_p)$ whence it is separable and complete. Being that completion, it is a convex combination space as well [42, Section 3]. Therefore, Corollary 5.1 applies.

Corollary 6.2. *Let $X_n, X : \Omega \to \widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$ be random elements such that $\{X_n\}_n$ is uniformly integrable in d_p. If $X_n \to X$ weakly in d_p, then $d_p(E[X_n], E[X]) \to 0$.*

We will now consider non-convex fuzzy sets in the space $(\mathcal{F}(\mathbb{R}^d), d_p)$.

Proposition 6.3. *Let $X_n, X : \Omega \to \mathcal{F}(\mathbb{R}^d)$ be fuzzy random variables such that $\{X_n\}_n$ is uniformly integrable in d_p. If $X_n \to X$ weakly in d_p, then $d_p(E[X_n], E[X]) \to 0$.*

Proof. By [2, Theorem 3.1], $\mathcal{F}_c(\mathbb{R}^d)$ is a measurable subset of $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$. Next, notice that $\operatorname{co} X_n$ is a random element in $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$: for any $B \in \mathcal{B}_{(\mathcal{F}_c(\mathbb{R}^d), d_p)}$,

$$(\operatorname{co} X)^{-1}(B) = (\operatorname{co} X)^{-1}(B \cap \mathcal{F}_c(\mathbb{R}^d)) \in \mathcal{A}.$$

Next, let $\varepsilon > 0$ and $M \geq 0$ be such that

$$E[d(X_n, v) \cdot I_{\{d(X_n,v)>M\}}] \leq \varepsilon$$

for every $n \in \mathbb{N}$. Then, by [41, Lemma 6.1] for every $K, L \in \mathcal{K}(\mathbb{R}^d)$ we have

$$d_H(\operatorname{co} K, \operatorname{co} L) \leq d_H(K, L).$$

Then

$$E[d_p(\operatorname{co} X_n, v) \cdot I_{\{d_p(\operatorname{co} X_n,v)>M\}}] \leq E[d_p(X_n, v) \cdot I_{\{d_p(X_n,v)>M\}}] \leq \varepsilon.$$

By the continuous mapping theorem in [2, Theorem 5.1], $\operatorname{co} X_n \to \operatorname{co} X$ weakly in d_p. Corollary 5.1 allows us to conclude $E[\operatorname{co} X_n] \to E[\operatorname{co} X]$ in d_p. Finally, by [39, Proposition 17], $E[\operatorname{co} X] = E[X]$, hence $E[X_n] \to E[X]$ in d_p. $\square$

Remark 6.1. From [42, Theorem 5], $(E[X]_\alpha) = E[X_\alpha]$ for each $\alpha \in (0, 1]$, i.e., the expectation in $\mathcal{F}(\mathbb{R}^d)$ as a convex combination space is consistent with the expectation of each α-cut in $\mathcal{K}(\mathbb{R}^d)$ as a convex combination space (just like the Puri–Ralescu expectation [32] is consistent with the Aumann expectation of each α-cut).

Under mild conditions, but not always, both expectations in $\mathcal{K}(\mathbb{R}^d)$ are equal. More specifically, $E[X]$ will be the same fuzzy set defined by Puri and Ralescu whenever X takes on convex values or the space $(\Omega, \mathcal{A}, P)$ is nonatomic.

6.3. Fuzzy sets with the d_∞-metric

The metric d_∞ is not separable, whence it is not possible to apply, like in the preceding examples, the Vitali theorem for the metric of a complete separable convex combination space. Correspondingly, we will consider $d = d_1$ and $\rho = d_\infty$ in Theorem 3.1. That requires showing that d_∞, as a bivariate function, is lower semicontinuous with respect to d_1. The known fact that $d_\infty(X, Y)$ is a random variable for any fuzzy random variables X, Y with values in $\mathcal{F}(\mathbb{R}^d)$ implies that d_∞ is *measurable* (just take X, Y to be the coordinate projections in $\mathcal{F}(\mathbb{R}^d) \times \mathcal{F}(\mathbb{R}^d)$). Thus the following proposition sharpens measurability to lower semicontinuity.

Proposition 6.4. *Let $p \in [1, \infty)$. The metric d_∞ is lower semicontinuous as a function on $\mathcal{F}_c(\mathbb{R}^d) \times \mathcal{F}_c(\mathbb{R}^d)$ when $\mathcal{F}_c(\mathbb{R}^d)$ is endowed with the metric d_p.*

Proof. First, notice that if a metric δ is lower semicontinuous with respect to δ', then it is also lower semicontinuous in any finer topology. Thus it suffices to consider the case $p = 1$. Further, the metrics d_1 and ρ_1 are equivalent, whence lower semicontinuity with respect to d_1 and ρ_1 is the same thing.

Since any real function f is lower semicontinuous if and only if the sets $\{f \le t\}$ are closed for each $t \in \mathbb{R}$, the proof will be complete if we show that

$$\begin{cases} \rho_1(U_n, U) \to 0 \\ \rho_1(V_n, V) \to 0 \\ d_\infty(U_n, V_n) \le t \end{cases}$$

imply $d_\infty(U, V) \le t$.

As $\rho_1(U_n, U) \to 0$, using Fatou's lemma

$$0 \le \int\limits_{[0,1]\times\mathbb{S}^{d-1}} \liminf_n |s_{U_n}(r,\alpha) - s_U(r,\alpha)| dr\, d\alpha$$
$$\le \liminf_n \int\limits_{[0,1]\times\mathbb{S}^{d-1}} |s_{U_n}(r,\alpha) - s_U(r,\alpha)| dr\, d\alpha = \liminf_n \rho_1(U_n, U) = 0.$$

Thus

$$\int\limits_{[0,1]\times\mathbb{S}^{d-1}} \liminf_n |s_{U_n}(r,\alpha) - s_U(r,\alpha)| dr\, d\alpha = 0.$$

Since $\liminf_n |s_{U_n} - s_U|$ is a non-negative function, there exists a null set $N_1 \subseteq [0, 1] \times \mathbb{S}^{d-1}$ such that

$$\liminf_n |s_{U_n}(r,\alpha) - s_U(r,\alpha)| = 0$$

for each $(r, \alpha) \notin N_1$. Analogously, there exists a null set N_2 such that

$$\liminf_n |s_{V_n}(r,\alpha) - s_V(r,\alpha)| = 0$$

for each $(r, \alpha) \notin N_2$. Fix for now some arbitrary $(r, \alpha) \notin N_1 \cup N_2$. Also fix an arbitrary $\varepsilon > 0$. Then there exists a subsequence $\{n'\}_n$ such that

$$|s_{U_{n'}}(r,\alpha) - s_U(r,\alpha)| \to 0.$$

Next, there exists a further subsequence $\{n''\}_n$ of $\{n'\}_n$ such that

$$|s_{V_{n''}}(r,\alpha) - s_V(r,\alpha)| \to 0.$$

Accordingly, there exists an $m \in \mathbb{N}$ such that $|s_{U_m}(r, \alpha) - s_U(r, \alpha)| < \varepsilon$ and $|s_{V_m}(r, \alpha) - s_V(r, \alpha)| < \varepsilon$. Notice that m depends on the choice of (r, α) but this will not be an obstacle for the proof. Since

$$d_\infty(U_m, V_m) = \sup_{(r,\alpha)\in\mathbb{S}^{d-1}\times[0,1]} |s_{U_m}(r,\alpha) - s_{V_m}(r,\alpha)|,$$

we have

$$|s_U(r,\alpha) - s_V(r,\alpha)| \le |s_U(r,\alpha) - s_{U_m}(r,\alpha)| + |s_{U_m}(r,\alpha) - s_{V_m}(r,\alpha)| + |s_{V_m}(r,\alpha) - s_V(r,\alpha)| < t + 2\varepsilon.$$

Therefore

$$\sup_{(r,\alpha)\in(\mathbb{S}^{d-1}\times[0,1])\setminus(N_1\cup N_2)} |s_U(r,\alpha) - s_V(r,\alpha)| \le t + 2\varepsilon.$$

Now let π_i be the projection over the ith coordinate $(i = 1, 2)$ of $\mathbb{S}^{d-1} \times [0, 1]$. Take any $(\alpha, r) \in N_1 \cup N_2$. If $\alpha > 0$ then there exist sequences $\{\alpha_k\}_k \subseteq \pi_2((\mathbb{S}^{d-1} \times [0, 1]) \setminus (N_1 \cup N_2))$ such that $\alpha_k \nearrow \alpha$ and $\{r_l\}_l \subseteq \pi_1((\mathbb{S}^{d-1} \times [0, 1]) \setminus$

$(N_1 \cup N_2))$ such that $r_l \to r$. And if $\alpha = 0$ then there exist sequences $\{\alpha_k\}_k \subseteq \pi_2((\mathbb{S}^{d-1} \times [0,1]) \setminus (N_1 \cup N_2))$ such that $\alpha_k \searrow 0$ and $\{r_l\}_l$ defined as in the previous case such that $r_l \to r$. By the triangle inequality,

$$|s_U(r,\alpha) - s_V(r,\alpha)| \le |s_U(r,\alpha) - s_U(r_l,\alpha_k)| + |s_U(r_l,\alpha_k) - s_V(r_l,\alpha_k)| + |s_V(r_l,\alpha_k) - s_V(r,\alpha)|$$

For the first summand,

$$|s_U(r,\alpha) - s_U(r_l,\alpha_k)| \le |s_U(r,\alpha) - s_U(r,\alpha_k)| + |s_U(r,\alpha_k) - s_U(r_l,\alpha_k)|.$$

Since $s_U(r,\cdot)$ is left continuous (e.g., [25, Theorem 3.1]) for all $r \in \mathbb{S}^{d-1}$ and all $U \in \mathcal{F}_c(\mathbb{R}^d)$, it follows that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for $\alpha' \in [\alpha - \delta, \alpha]$, it holds that $|s_U(r,\alpha) - s_U(r,\alpha')| < \varepsilon$. For $\varepsilon > 0$, it suffices to take $k_1^* \in \mathbb{N}$ such that $\alpha_{k_1^*} \in [\alpha - \delta, \alpha]$.

By an analogous reasoning,

$$|s_V(r,\alpha) - s_V(r_l,\alpha_k)| \le |s_V(r,\alpha) - s_V(r,\alpha_k)| + |s_V(r,\alpha_k) - s_V(r_l,\alpha_k)|$$

and there exists a $k_2^* \in \mathbb{N}$ such that $|s_V(r,\alpha) - s_V(r,\alpha_{k_2^*})| < \varepsilon$. Finally, set $k^* = \max\{k_1^*, k_2^*\}$.

Next, $s_U(\cdot,\alpha_{k^*})$ is continuous for all $U \in \mathcal{F}_c(\mathbb{R}^d)$, so there exists an l^* such that $|s_U(r,\alpha_{k^*}) - s_U(r_{l^*},\alpha_{k^*})| < \varepsilon$ and $|s_V(r,\alpha_{k^*}) - s_V(r_{l^*},\alpha_{k^*})| < \varepsilon$. Then

$$|s_U(r,\alpha) - s_V(r,\alpha)| < t + 4\varepsilon.$$

In conclusion,

$$d_\infty(U,V) = \sup_{(r,\alpha)\in\mathbb{S}^{d-1}\times[0,1]} |s_U(r,\alpha) - s_V(r,\alpha)| \le t + 4\varepsilon,$$

so $d_\infty(U,V) \le t$, as wished, by the arbitrariness of ε. □

Remark 6.2. Since $\mathbb{S}^{d-1} \times [0,1]$ is a separable set, the existence of the sequences mentioned in the preceding result is guaranteed. For the sequence $\{\alpha_k\}_k$ (if $\alpha > 0$), it suffices to take any element $\alpha_k \in [(\alpha - k^{-1})_+, \alpha] \cap \pi_2(\mathbb{S}^{d-1} \times [0,1])$. The case $\alpha = 0$ is analogous. For $\{r_l\}_l$, notice that every set $M \subseteq \mathbb{S}^{d-1}$ such that $\mathbb{S}^{d-1} \setminus M$ is null, is dense. Reasoning by contradiction, suppose that M is not dense, then there exists an $x \notin M$ such that $d(x,M) > \varepsilon$, hence $B = \{y \in \mathbb{S}^{d-1} : d(x,y) < \varepsilon\}$ is an open set such that $B \cap M = \emptyset$ with nonzero measure.

There follows that d_∞ is lower semicontinuous but cannot be upper semicontinuous.

Corollary 6.5. *d_∞ is not upper semicontinuous with respect to d_p.*

Proof. If d_∞ were upper semicontinuous in d_p, it would be continuous and both metrics would induce the same topology, which is not true. □

We finally reach the Vitali theorem for d_∞.

Theorem 6.6. *Let X_n and X be integrably bounded fuzzy random variables in $\mathcal{F}(\mathbb{R}^d)$ such that $\{X_n\}_n$ is uniformly integrable in d_∞. Assume further that one of the following sets of conditions hold:*

(a) *$\operatorname{co} X_n, \operatorname{co} X$ are random elements with respect to d_∞, and $\operatorname{co} X$ takes on values almost surely in a d_∞-separable subset of $\mathcal{F}_c(\mathbb{R}^d)$.*
(b) *$\{X_n\}_n$ is d_∞-tight, and X is d_∞-tight.*

If $X_n \to X$ weakly in d_∞, then $d_\infty(E[X_n], E[X]) \to 0$.

Proof. First, suppose that X_n and X take on values in $\mathcal{F}_c(\mathbb{R}^d)$. Since X_n and X are integrably bounded, it follows that they are integrable. Next, $E[X_n], E[X] \in \mathcal{F}_c(\mathbb{R}^d)$. Take a countable dense set $\{U_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F}_c(\mathbb{R}^d)$ and set

$$f_i = \frac{d_p(\cdot, U_i)}{1 + d_p(\cdot, U_i)}.$$

Since d_∞ is stronger than d_p, the f_i are continuous and for $U \neq V$ there exists some U_i such that $f_i(U) < f_i(V)$ since some U_i are closer to U than to V because $\{U_i\}_{i\in\mathbb{N}}$ is dense. Besides, since X_n and X are Borel measurable in d_p we can conclude that $f_i(X_n)$ and $f_i(X)$ are random variables. Moreover,

$$E[d_\infty(I_{\{0\}}, X)] = E[d_H(\{0\}, X_0)] < \infty$$

since X is integrably bounded. Then we can apply Theorem 3.1 with $\mathbb{E} = \widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$, $C = \mathcal{F}_c(\mathbb{R}^d)$, $d = d_1$, $\rho = d_\infty$ and $v_1 = v_2 = I_{\{0\}}$ and obtain $E[X_n] \to E[X]$ in d_∞.

For the general case, reasoning like in Proposition 6.3 one can conclude that $E[\operatorname{co} X_n] \to E[\operatorname{co} X]$ in d_∞ implies $E[X_n] \to E[X]$ in d_∞. □

Remark 6.3. The condition on the values of X which appears in (a) above can be checked with the characterization of d_∞-separability in [36, Theorem 1]. Namely, $\mathfrak{A} \subseteq \mathcal{F}(\mathbb{R}^d)$ is d_∞-separable if and only if the discontinuity set

$$D(\mathfrak{A}) = \{\alpha \in (0, 1] \mid \text{ the mapping } \alpha \mapsto U_\alpha \text{ is discontinuous at } \alpha \text{ for some } U \in \mathfrak{A}\}$$

is at most countable.

Since conditions (a) and (b) come, respectively, from (2'–3') and (3) in Theorem 3.1, and as shown in the previous section these are not necessary if weak convergence is strengthened to convergence in probability, we have the following corollary.

Corollary 6.7. *Let X_n and X be fuzzy random variables in $\mathcal{F}(\mathbb{R}^d)$ such that $\{X_n\}_n$ is uniformly integrable in d_∞. If $X_n \to X$ in probability in d_∞, then $d_\infty(E[X_n], E[X]) \to 0$.*

6.4. Compact sets with the Hausdorff metric

Our version for random compact sets is stated below, which is a consequence of Lemma 2.3 and Corollary 5.1.

Corollary 6.8. *Let $(\mathbb{E}, d)$ be a complete and separable convex combination space. Let X_n and X be integrable random compact sets in $\mathbb{E}$ such that $\{X_n\}_n$ is uniformly integrable in d_H. If $X_n \to X$ weakly in d_H, then $E[X_n] \to E[X]$ in d_H.*

Proof. By Lemma 2.3, $\mathcal{K}(\mathbb{E})$ is a complete and separable convex combination space, then Corollary 5.1 can be applied. □

6.5. Compact convex sets with the Bartels–Pallaschke metric

The Bartels–Pallaschke metric between compact convex sets in $\mathbb{R}^d$ is stronger than the Hausdorff metric. It is complete [11, Theorem 7.1] but not separable. It was introduced by Diamond et al. [11, p. 275], who named it after Bartels and Pallaschke since they had defined a norm in the space of differences of continuous sublinear functions [4] and the distance between two sets equals the norm of the difference of their support functions.

Definition 6.1. Let $K, L \in \mathcal{K}_c(\mathbb{R}^d)$. The *Bartels–Pallaschke metric* is defined by

$$d_{BP}(K, L) = \inf\{\max\{\|K'\|, \|L'\|\} : K + L' = L + K'\}.$$

One checks easily $d_{BP} \geq d_H$, whence the latter is continuous with respect to the former. We will show now that the former is lower semicontinuous with respect to the latter, as this is required to apply the abstract Vitali theorem.

Proposition 6.9. *The metric d_{BP} is lower semicontinuous as a function on $\mathcal{K}_c(\mathbb{R}^d) \times \mathcal{K}_c(\mathbb{R}^d)$ when $\mathcal{K}_c(\mathbb{R}^d)$ is endowed with the Hausdorff metric.*

Proof. Let $\{K_n\}_n, \{L_n\}_n \subseteq \mathcal{K}_c(\mathbb{R}^d)$ converge in d_H to K and L, respectively. Fix an arbitrary $\varepsilon > 0$. For each $n \in \mathbb{N}$ there exist $C_n, D_n \in \mathcal{K}_c(\mathbb{R}^d)$ such that $K_n + C_n = L_n + C_n$ and

$$\max\{\|C_n\|, \|D_n\|\} \leq d_{BP}(K_n, L_n) + \varepsilon.$$

Moreover, by definition $d_{BP}(K_n, L_n) \leq \|K_n\| + \|L_n\|$. Since both $\{\|K_n\|\}_n$ and $\{\|L_n\|\}_n$ are convergent sequences, the sequence $\{d_{BP}(K_n, L_n)\}_n$ is bounded. Therefore $\{d_{BP}(K_n, L_n)\}_n$ has at least one convergent subsequence.

Let $\{d_{BP}(K_{n'}, L_{n'})\}_n$ be any arbitrary convergent subsequence. Then $\{\|C_{n'}\|\}_n$ and $\{\|D_{n'}\|\}_n$ are bounded by a convergent sequence, there exist $m > 0$ such that all $C_{n'}, D_{n'}$ are in the closed ball $B(0, m)$, which is a compact set. By [10, Proposition 2.4.4], the space $(\mathcal{K}(B(0, m)), d_H)$ is a compact subset of $\mathcal{K}(\mathbb{R}^d)$. Then there exist subsequences $\{C_{n''}\}_n$ and $\{D_{n''}\}_n$ of $\{C_{n'}\}_n$ and $\{D_{n'}\}_n$ respectively, and compact sets C, D such that $C_{n''} \to C$ and $D_{n''} \to D$ in d_H.

Next, we wish to show $K + C = L + D$. There exists an $n_0 \in \mathbb{N}$ such that $d_H(C_{n''}, C), d_H(D_{n''}, D), d_H(K_{n''}, K)$, and $d_H(L_{n''}, L)$ are all strictly smaller than any arbitrary $\varepsilon' > 0$ for each $n'' \geq n_0$. Then

$$\begin{aligned} d_H(K + C, L + D) &\leq d_H(K + C, K_{n''} + C_{n''}) + d_H(K_{n''} + C_{n''}, L_{n''} + D_{n''}) + d_H(L_{n''} + D_{n''}, L + D) \\ &\leq d_H(K, K_{n''}) + d_H(C, C_{n''}) + 0 + d_H(L_{n''}, L) + d_H(D_{n''}, D) < 4\varepsilon', \end{aligned}$$

so $K + C = L + D$ by the arbitrariness of ε'. Furthermore, for the value $\varepsilon' = \varepsilon$ we obtain

$$\|C\| \leq d_H(C, C_{n''}) + \|C_{n''}\| < \|C_{n''}\| + \varepsilon.$$

Analogously, $\|D\| < \|D_{n''}\| + \varepsilon$. Thus

$$\max\{\|C\|, \|D\|\} < \max\{\|C_{n''}\|, \|D_{n''}\|\} + \varepsilon \leq d_{BP}(K_{n''}, L_{n''}) + 2\varepsilon.$$

Therefore,

$$d_{BP}(K, L) \leq \max\{\|C\|, \|D\|\} \leq d_{BP}(K_{n''}, L_{n''}) + 2\varepsilon.$$

Since $d_{BP}(K_{n''}, L_{n''})$ is a convergent subsequence, from the arbitrariness of ε we have

$$\lim_n d_{BP}(K_{n''}, L_{n''}) \geq d_{BP}(K, L).$$

Moreover, since $d_{BP}(K_{n'}, L_{n'})$ is convergent too, there follows

$$\lim_n d_{BP}(K_{n'}, L_{n'}) \geq d_{BP}(K, L).$$

In conclusion, the limit of every convergent subsequence of $\{d_{BP}(K_n, L_n)\}_n$ is at least $d_{BP}(K, L)$, so it follows that $\liminf_n \{d_{BP}(K_n, L_n)\} \geq d_{BP}(K, L)$. That proves that d_{BP} is lower semicontinuous. □

Reasoning as in Corollary 6.5 we are able to obtain a similar result, this time for d_{BP} and d_H.

Corollary 6.10. *The metric d_{BP} is not upper semicontinuous as a bivariate function.*

Since d_{BP} is stronger than d_H, there may be random compact sets which are not random elements with respect to d_{BP}. As a consequence of Proposition 6.9, however, we have the following result.

Corollary 6.11. *Let X and Y be random compact convex sets of $\mathbb{R}^d$. Then $d_{BP}(X, Y)$ is a random variable, regardless of whether X and Y are random elements with respect to d_{BP} or not.*

Proof. By Proposition 6.9, the set

$$\{(K, L) \in \mathcal{K}_c(\mathbb{R}^d) \times \mathcal{K}_c(\mathbb{R}^d) : d_{BP}(K, L) \leq a\}$$

is closed for every $a \in \mathbb{R}$. Then

$$(d_{BP}(X, Y))^{-1}((-\infty, a]) = \{\omega \in \Omega : d_{BP}(X(\omega), Y(\omega)) \leq a\}$$

is the preimage of a closed set, hence measurable. From the arbitrariness of a, we deduce that $d_{BP}(X, Y)$ is a random variable. □

To apply Theorem 3.1, we also need to show that d_{BP} is midpoint convex. This can be obtained from the sublinearity of the Bartels–Pallaschke norm but we present a proof closer to the properties of convex combination spaces. The following lemma is part of the proof but we state it separately for future reference.

Lemma 6.12. *Any function $\rho : \mathbb{E} \times \mathbb{E} \to \mathbb{R}$ satisfying property (CC4), i.e., negative curvature, is convex.*

Proof. From (CC4), the statement is true for convex combinations of $n = 2$ points:

$$\rho([\lambda, (x_1, y_1); 1-\lambda, (x_2, y_2)]) \le \rho([\lambda, x_1, (1-\lambda)x_2], [\lambda, y_1, (1-\lambda)y_2])$$

By induction, assume that the inequality is true for any convex combination of $n-1$ points. Using (CC2) and (CC4),

$$\begin{aligned}&\rho([\lambda_i, (x_i, y_i)]_{i=1}^n)\\ &\quad = \rho([\lambda_1, (x_1, y_1); \ldots; \lambda_{n-2}, (x_{n-2}, y_{n-2}); \lambda_{n-1} + \lambda_n, [\frac{\lambda_{n-2+j}}{\lambda_{n-1}+\lambda_n}, (x_{n-2+j}, y_{n-2+j})]_{j=1}^2])\\ &\quad \le \sum_{i=1}^{n-2} \lambda_i \rho(x_i, y_i) + (\lambda_{n-1}+\lambda_n)\rho([\frac{\lambda_{n-2+j}}{\lambda_{n-1}+\lambda_n}, (x_{n-2+j}, y_{n-2+j})]_{j=1}^2) \le \sum_{i=1}^{n} \lambda_i \rho(x_i, y_i). \quad \square\end{aligned}$$

Lemma 6.13. *The Bartels–Pallaschke metric is convex as a function on $\mathcal{K}_c(\mathbb{R}^d) \times \mathcal{K}_c(\mathbb{R}^d)$.*

Proof. First, let us show that that d_{BP} satisfies (CC4). By the triangle inequality,

$$\begin{aligned}&d_{BP}(\lambda K_1 + (1-\lambda)K_2, \lambda L_1 + (1-\lambda)L_2)\\ &\quad \le d_{BP}(\lambda K_1 + (1-\lambda)K_2, \lambda L_1 + (1-\lambda)K_2) + d_{BP}(\lambda L_1 + (1-\lambda)K_2, \lambda L_1 + (1-\lambda)L_2)\\ &\quad = d_{BP}(\lambda K_1, \lambda L_1) + d_{BP}((1-\lambda)K_2, (1-\lambda)L_2)\\ &\quad = \lambda d_{BP}(K_1, L_1) + (1-\lambda) d_{BP}(K_2, L_2).\end{aligned}$$

By Lemma 6.12, d_{BP} is convex. $\square$

Our Vitali convergence theorem for the Bartels–Pallaschke metric is as follows.

Theorem 6.14. *Let X_n and X be random compact convex sets in $\mathbb{R}^d$ such that $\{X_n\}_n$ is uniformly integrable in d_H. Assume further that one of the following sets of conditions hold:*

(a) *X_n, X are random elements with respect to d_{BP}, and X takes on values almost surely in a d_{BP}-separable subset of $\mathcal{K}_c(\mathbb{R}^d)$.*
(b) *$\{X_n\}_n$ is d_{BP}-tight, and X is d_{BP}-tight.*

If $X_n \to X$ weakly in d_{BP}, then $d_{BP}(E[X_n], E[X]) \to 0$.

Proof. We begin by observing that, for any $A \in \mathcal{K}_c(\mathbb{R}^d)$, the identity

$$d_{BP}(K, \{0\}) = \|K\|$$

holds (taking $K' = K$, $L' = \{0\}$ in the definition). With that, the proof is similar to that of Theorem 6.6 with $\mathbb{E} = C = \mathcal{K}_c(\mathbb{R}^d)$, $d = d_H$, $\rho = d_{BP}$ (by Proposition 6.9 and Lemma 6.13, d_{BP} satisfies the assumptions on ρ).

We need to take

$$f_i = \frac{d_H(\cdot, K_i)}{1 + d_H(\cdot, K_i)}$$

for a countable d_H-dense subset $\{K_i\}_{i\in\mathbb{N}} \subseteq \mathcal{K}_c(\mathbb{R}^d)$. Since d_H is weaker than d_{BP}, the f_i are d_{BP}-continuous, and if $K \neq L$ the density ensures that some K_i will be closer to K than to L, yielding $f_i(K) < f_i(L)$. The fact that each

$f_i(X_n)$, $f_i(X)$ is a random variable follows from the assumption that X_n, X are d_H-Borel functions and the f_i are d_H-continuous.

In order to apply Theorem 3.1, we need $\{d_{BP}(X_n, L)\}_n$ to be uniformly integrable for some $L \in \mathcal{K}_c(\mathbb{R}^d)$. Taking $L = \{0\}$, the identity

$$d_{BP}(X_n, \{0\}) = \|X_n\| = d_H(X_n, \{0\})$$

and the assumption of uniform integrability with respect to d_H ensure the d_{BP}-uniform integrability. □

For convergence in probability, we obtain the following.

Corollary 6.15. *Let X_n and X be random compact convex sets in $\mathbb{R}^d$ such that $\{X_n\}_n$ is uniformly integrable in d_H. If $X_n \to X$ in probability in d_{BP}, then $d_{BP}(E[X_n], E[X]) \to 0$.*

7. Comparison to the literature

There are some kinds of results which lie beyond the scope of Theorem 3.1.

(i) Results involving vector-valued integrals other than the Bochner integral. Since the expectation in a convex combination space is defined using approximations by simple functions, the Vitali convergence theorem for the Bochner integral of random elements in a Banach space is a particular case. But it does not include other integrals like Pettis, Henstock–Kurzweil, McShane, and others, or expectations of random sets or fuzzy random variables based on those integrals, e.g., [28, Theorem 2.1] (Pettis integral), [29, Theorem 3.7] (set-valued Denjoy–Pettis integral), [47, Theorem 3.8] (fuzzy-valued Kluvánek–Lewis integral), to name a few examples.

(ii) Results involving random sets with unbounded values. We are not aware of metrics on possibly unbounded sets which provide a convex combination space. Hence we do not subsume results like [27, Theorem 2.1.68.(i, ii)].

(iii) Results involving conditional expectations with respect to a σ-algebra. Theorem 3.1 is about ordinary expectations, whence it cannot include results like [43, Proposition 4.7], [27, Theorem 2.1.78], [14, Theorem 2.5], [12, Theorem 4.1].

Convex combination spaces. In [40, Theorem 4.2], a dominated convergence theorem under the assumption of convergence in probability was presented. It follows from Corollary 5.4, which also allows for the quasimetric ρ to be defined only in a subset C.

In [40, Remark 2], it is claimed that convergence in probability can be replaced by weak convergence if $\rho = d$, by an application of the Skorokhod representation theorem in complete separable metric spaces. But the Skorokhod theorem does not guarantee that the almost surely convergent sequence it provides still satisfies the assumption in the dominated convergence theorem, whence the argument given there by one of us is incorrect. As shown in the proof of Theorem 3.1, applying the Skorokhod theorem preserves uniform integrability, whence the dominated convergence theorem with weak convergence can subsequently be deduced from the Vitali convergence theorem. Thus the result suggested in [40, Remark 2] is correct (it follows from Corollary 5.1) even if its justification was incorrect.

Probability distributions. We are not aware of any convergence theorems for random distributions with the convolution operation and its associated expectation (this should not be confused with the asymptotic behavior of iterated convolutions in topological groups like in [5]).

Fuzzy random variables. There is an abundant literature on convergence theorems for fuzzy random variables since [32]. The applications of Theorem 3.1 in Sections 6.2 and 6.3 subsume what seem to be the most general results for $\mathcal{F}_c(\mathbb{R}^d)$-valued fuzzy random variables (Theorem 4.6 in [2] and Theorem 8.2 in [20]), as well as other previous results for $\mathcal{F}(\mathbb{R}^d)$-valued mappings.

Indeed, [2, Theorem 4.6] is a Vitali convergence theorem under weak convergence in $(\mathcal{F}_c(\mathbb{R}^d), d_p)$ while [20, Theorem 8.2] is a dominated convergence theorem under almost sure convergence in $(\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d), d_p)$ and $(\mathcal{F}_c(\mathbb{R}^d), d_\infty)$. We achieve a common derivation of the best of both results. Corollary 6.2 extends [2, Theorem 4.6] from $\mathcal{F}_c(\mathbb{R}^d)$-valued to $\widehat{\mathcal{F}}_{c,p}(\mathbb{R}^d)$-valued mappings. That weakens the assumptions in [20, Theorem 8.2] from almost sure convergence to weak convergence and from domination to uniform integrability. The reader is referred to the discussion before [2, Remark 4.2] for a careful comparison of both theorems being generalized.

In its turn, Proposition 6.3 extends [2, Theorem 4.6] from $\mathcal{F}_c(\mathbb{R}^d)$-valued to $\mathcal{F}(\mathbb{R}^d)$-valued mappings. As commented in Remark 6.1, in the non-convex case the Puri–Ralescu expectation and the expectation in the sense of convex combination spaces may differ but are identical whenever the probability space is nonatomic. Hence, in the nonatomic case Proposition 6.3 also generalizes the dominated convergence theorem for the d_1-metric in [18, Theorem 4.1].

As regards the d_∞-metric, Theorem 6.6 and Corollary 6.7 similarly generalize previous results, provided the values are convex or the measurable space is nonatomic. Theorem 6.6 provides a Vitali convergence theorem under weak d_∞-convergence; there seem to be no former results with that weakened assumption. Corollary 6.7, when specialized to dominated sequences covers the dominated convergence theorems in [32, Theorem 4.3] (almost sure convergence, nonatomic probability space; we write in brackets the features improved upon), [20, Theorem 8.2] (values in $\mathcal{F}_c(\mathbb{R}^d)$, almost sure convergence), [12, Theorem 3.1] (values in $\mathcal{F}_c(\mathbb{R}^d)$, random elements with respect to d_∞), and [23, Theorem 5.3.5, p. 183] in the nonatomic case (almost sure convergence).

Some results in the literature are valid in spaces more general than $\mathbb{R}^d$ (separable Banach spaces). In view of the fact that separable Banach spaces are complete separable convex combination spaces, and in view of Corollary 6.8 for random compact sets, it is plausible that fuzzy random variables with values in a separable Banach space or a more general convex combination space can be dealt with. However this would require redeveloping a number of instrumental results which were only available in $\mathbb{R}^d$, which exceeds the scope of this section.

Let us mention [15, Theorem 4.5] of Jang and Kwon, to show that it is subsumed by Corollary 6.7 provided the carrier space is $\mathbb{R}^d$. Since it requires a nonatomic probability space, the expectations used in both results are identical. It uses a generalization of d_∞,

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

where h is an increasing function with $0 < h \leq 1$. One easily checks

$$h(0) \cdot d_\infty(U, V) \leq d_h(U, V) \leq d_\infty(U, V)$$

whence it follows that both the assumptions and conclusion are equivalent whether one considers d_h or d_∞.

Random compact sets. Proposition 5.2 in [40] depends on [40, Remark 2] and, in view of the discussion above which justifies the validity of that remark, it still stands. It is a dominated convergence theorem for weakly convergent sequences of random compact sets in a convex combination space, which is subsumed by Corollary 6.8.

Under mild assumptions (convex values or nonatomic probability space), the expectation of a random compact set in a separable Banach space, regarded as a random element of a convex combination space, equals its Aumann expectation. Thence, [27, Theorem 2.1.70, p. 269] (for carrier space $\mathbb{R}^d$) follows from Corollary 6.8, at least if convergence in distribution is understood with respect to the Hausdorff metric.

Similarly, under those assumptions the Vitali and dominated convergence theorems for the Hausdorff metric in [27, Theorem 2.1.61 and Theorem 2.1.68.(iii)], which require at least convergence in probability, follow from Corollary 6.8. The latter theorem is the same as [14, Theorem 2.8.(4)].

Regarding the Bartels–Pallaschke metric, it seems that Theorem 6.14 is the first convergence theorem with respect to that metric.

8. Concluding remarks

In this paper, we have presented an abstract Vitali convergence theorem which allows one to obtain concrete convergence theorems in a space (C, ρ) provided one can find a convex combination space $\mathbb{E} \supseteq C$ endowed with an appropriate metric d. Thanks to some assumptions on the relationship of both X_n, X and d with ρ, the space $(\mathbb{E}, d)$ has the well-behaved structure that lets us obtain results for different choices of C and ρ.

Some questions which are prompted by the results in this paper are the following.

(1) Does the Jensen inequality in [40, Theorem 3.1] and [43, Proposition 4.6] hold under assumptions weaker than lower semicontinuity? If so, the assumption on ρ can be weakened accordingly. It seems from our results that establishing lower semicontinuity can be the most burdensome part of applying Theorem 3.1.
(2) Can the results for the Bartels–Pallaschke metric be extended to fuzzy sets by considering metrics analogous to d_∞ and d_p (which are based on the Hausdorff metric between α-cuts)?

(3) Is the dominated convergence theorem of Jonasson [17, Theorem 13] related to our results? In his paper, the space must be ordered but the negative curvature condition is replaced by a weaker one (loosely speaking, the convexity of the metric as a bivariate function is replaced by quasiconvexity).

Similarly, it would be interesting to know whether more spaces of fuzzy sets and spaces of probability measures are convex combination spaces.

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 and fellowships from Spain (PID2019-104486GB-I00), the Principality of Asturias (SV-PA-21-AYUD/2021/50897 and PA-21-PF-BP20-112), and the University of Oviedo (PAPI-20-PF-21). Their contribution is gratefully acknowledged.

References

[1] M. Alonso de la Fuente, P. Terán, Harmonizing two approaches to fuzzy random variables, Fuzzy Optim. Decis. Mak. 19 (2020) 177–189.
[2] M. Alonso de la Fuente, P. Terán, Some results on convergence and distributions of fuzzy random variables, Fuzzy Sets Syst. 435 (2022) 149–163.
[3] H. Attouch, G. Buttazzo, G. Michaille, Variational Analysis in Sobolev and BV Spaces, SIAM, Philadelphia, 2014.
[4] S.G. Bartels, D. Pallaschke, Some remarks on the space of differences of sublinear functions, Appl. Math. 22 (1994) 419–426.
[5] R. Bhattacharya, Speed of convergence of the n-fold convolution of a probability measure on a compact group, Z. Wahrscheinlichkeitstheor. Verw. Geb. 25 (1972) 1–10.
[6] P. Billingsley, Convergence of Probability Measures, Wiley, Nueva York, 1968.
[7] V.I. Bogachev, Measure Theory, vol. 2, Springer, Berlin, 2007.
[8] E. Çınlar, Probability and Stochastics, Springer, New York, 2011.
[9] M.M. Deza, E. Deza, Encyclopedia of Distances, third edition, Springer, Heidelberg, 2014.
[10] P. Diamond, P. Kloeden, Metric Spaces of Fuzzy Sets: Theory and Applications, World Scientific, Singapore, 1994.
[11] P. Diamond, P. Kloeden, A. Rubinov, A. Vladimirov, Comparative properties of three metrics in the space of compact convex sets, Set-Valued Anal. 5 (1997) 267–289.
[12] Y. Feng, Convergence theorems for fuzzy random variables and fuzzy martingales, Fuzzy Sets Syst. 103 (1999) 435–441.
[13] S.G. Georgiev, K. Zennir, Functional Analysis with Applications, De Gruyter, Berlin, 2019.
[14] F. Hiai, Convergences of conditional expectations and strong laws of large numbers for multivalued random variables, Transl. Am. Math. Soc. 291 (1985) 613–627.
[15] L.C. Jang, J.S. Kwon, Convergence of sequences of set-valued and fuzzy-set-valued functions, Fuzzy Sets Syst. 93 (1998) 241–246.
[16] A. Jakubowski, The almost sure Skorokhod representation for subsequences in nonmetric spaces, Theory Probab. Appl. 42 (1998) 167–174.
[17] J. Jonasson, On positive random objects, J. Theor. Probab. 11 (1999) 81–125.
[18] E.P. Klement, M.L. Puri, D.A. Ralescu, Limit theorems for fuzzy random variables, Proc. R. Soc. Ser. A 407 (1986) 171–182.
[19] V. Krätschmer, Some complete metrics on spaces of fuzzy subsets, Fuzzy Sets Syst. 130 (2002) 357–365.
[20] V. Krätschmer, Integrals of random fuzzy sets, Test 15 (2006) 433–469.
[21] C.C.A. Labuschagne, A.L. Pinchuck, Doob's decomposition of set-valued submartingales via ordered near vector spaces, Quaest. Math. 32 (2009) 247–264.
[22] C.C.A. Labuschagne, A.L. Pinchuck, C.J. Van Alten, A vector lattice version of Rådström's embedding theorem, Quaest. Math. 30 (2007) 285–308.
[23] S. Li, Y. Ogura, V. Kreinovich, Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables, Springer, Dordrecht, 2002.
[24] V. Lupulescu, D. O'Regan, A new derivative concept for set-valued and fuzzy-valued functions. Differential and integral calculus in quasilinear metric spaces, Fuzzy Sets Syst. 404 (2021) 75–110.
[25] M. Ming, On embedding problems of fuzzy number space: part 5, Fuzzy Sets Syst. 55 (1993) 313–318.
[26] D. Mitrea, I. Mitrea, M. Mitrea, S. Monniaux, Groupoid Metrization Theory, Birkhäuser, New York, 2013.
[27] I. Molchanov, Theory of Random Sets, Springer, London, 2017.
[28] K. Musiał, Pettis integral, Chapter 12 in: E. Pap (Ed.), Handbook of Measure Theory, Elsevier, Amsterdam, 2002, pp. 531–586.
[29] C.K. Park, Convergence theorems for set-valued Denjoy–Pettis integrable mappings, Commun. Korean Math. Soc. 24 (2009) 227–237.
[30] J.B. Prolla, Approximation of continuous convex-cone-valued functions by monotone operators, Stud. Math. 102 (1992) 175–192.

[31] J.B. Prolla, The Weierstrass-Stone theorem for convex-cone-valued functions, J. Comput. Appl. Math. 53 (1994) 171–183.
[32] M.L. Puri, D.A. Ralescu, Fuzzy random variables, J. Math. Anal. Appl. 114 (1986) 409–422.
[33] N.V. Quang, P.T. Nguyen, Some strong laws of large number for double array of random upper semicontinuous functions in convex combination spaces, Stat. Probab. Lett. 96 (2015) 85–94.
[34] N.V. Quang, N.T. Thuan, On the strong laws of large number for double arrays of random variables in convex combination spaces, Acta Math. Hung. 134 (2012) 543–564.
[35] H. Rådström, An embedding theorem for spaces of convex sets, Proc. Am. Math. Soc. 3 (1952) 165–169.
[36] P. Terán, On Borel measurability and large deviations for fuzzy random variables, Fuzzy Sets Syst. 157 (2006) 2558–2568.
[37] P. Terán, On a uniform law of large numbers for random sets and subdifferentials of random functions, Stat. Probab. Lett. 78 (2008) 42–49.
[38] P. Terán, Strong law of large numbers for t-normed arithmetics, Fuzzy Sets Syst. 159 (2008) 343–360.
[39] P. Terán, Algebraic, metric and probabilistic properties of convex combinations based on the t-normed extension principle: the strong law of large numbers, Fuzzy Sets Syst. 223 (2013) 1–25.
[40] P. Terán, Jensen's inequality for random elements in metric spaces and some applications, J. Math. Anal. Appl. 414 (2014) 756–766.
[41] 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.
[42] 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 Uncertainty Modelling, Springer, Berlin, 2006, pp. 153–160.
[43] N.T. Thuan, Approach for a metric space with a convex combination operation and applications, J. Math. Anal. Appl. 435 (2016) 440–460.
[44] N.T. Thuan, N.V. Quang, P.T. Nguyen, Complete convergence for arrays of rowwise independent random variables and fuzzy random variables in convex combination spaces, Fuzzy Sets Syst. 250 (2014) 52–68.
[45] F. Topsøe, Compactness and tightness in a space of measures with the topology of weak convergence, Math. Scand. 34 (1974) 187–210.
[46] M.L. Wichura, On the construction of almost uniformly convergent random variables with given weakly convergent image laws, Ann. Math. Stat. 41 (1970) 284–291.
[47] C.L. Zhou, X. Chen, A new fuzzy-valued integral and its convergence theorems, Filomat 33 (2019) 1877–1887.