Topology and its Applications 158 (2011) 836–852

Contents lists available at ScienceDirect

Topology and its Applications

www.elsevier.com/locate/topol

Approach theory meets probability theory ☆

B. Berckmoes *, R. Lowen, J. Van Casteren

Department of Mathematics and Computer Science, University of Antwerp, Belgium

ARTICLE INFO

MSC:
60A05
60G05
60G07
62P99
65D99

Keywords:
Approach space
Distance
Weak topology
Polish space
Probability measure
Law
Random variable
Weak compactness
Tightness
Prokhorov metric
Ky-Fan metric
Atsuji space

ABSTRACT

In this paper we reconsider the basic topological and metric structures on spaces of probability measures and random variables, such as e.g. the weak topology and the total variation metric, replacing them with more intrinsic and richer approach structures. We comprehensibly investigate the relationships among, and basic facts about these structures, and prove that fundamental results, such as e.g. the portmanteau theorem and Prokhorov's theorem, can be recaptured in a considerably stronger form in the new setting.

1. Introduction

In this paper we reconsider the basic topological and metric structures on spaces of probability measures and spaces of random variables. It turns out that many topological structures and metrics used in the literature [4,6,16,19,23] are merely respectively the topological coreflections and the metric coreflections of more natural and structurally richer approach structures. For the sake of simplicity, throughout the paper we will consider a Polish space S and the set $\mathcal{P}(S)$ of probability measures on it. When dealing with random variables we will consider a fixed probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and the set of S-valued random variables $\mathcal{R}(S)$ defined thereon. As far as the topological structures are concerned we will deal in particular with the weak topology on $\mathcal{P}(S)$ and with convergence in probability on $\mathcal{R}(S)$. The important classical metrics which we will encounter are the Prokhorov metric ρ and the total variation metric d_{TV}, the former of which, as is well known, metrizes the weak topology on probability measures, and in the case of random variables the Ky-Fan metric K [4] and the indicator metric d_I [23].

☆ The editors of Topology and its Applications and the Journal of Mathematical Analysis and Applications have learned that there are significant similarities in several parts of published work by B. Berckmoes, B. Lowen, and J. Van Casteren. The exact references are: Approach theory meets probability theory, Topology and its Applications 158 (7) (2011) 836–852, and Distances on probability measures and random variables, Journal of Mathematical Analysis and Applications 374 (2) (2011) 412–428. The editors of both journals are satisfied that this unfortunate situation was an innocent mistake caused by inattention by the authors. This occurrence is deeply regretted by all parties.

* Corresponding author.
E-mail addresses: ben.berckmoes@ua.ac.be (B. Berckmoes), bob.lowen@ua.ac.be (R. Lowen), jan.vancasteren@ua.ac.be (J. Van Casteren).

0166-8641/$ – see front matter
doi:10.1016/j.topol.2011.01.004

On $\mathcal{P}(S)$ we propose an approach structure, in particular a non-metric distance δ_w which is more closely related to, and certainly is strongly inspired by, the weak topology and which we hence also call the *weak approach structure*. This approach also has a unifying effect in the sense that δ_w has several equivalent expressions, obtained by judiciously putting together the building blocks either from the weak topology, or somewhat surprisingly, also from the Prokhorov metric or from the total variation metric. On $\mathcal{R}(S)$ too we propose a natural approach structure which is similarly linked to convergence in probability [4,14] and to the indicator metric [23].

The systematic use of this distance, and the preservation of its numerical information, allows for an isometric or quantitative study of various important concepts in stochastic theories. In this paper, apart from a study of the new structures, their basic properties and interrelation we restrict ourselves to weak compactness and tightness. For compactness we fall back on the well-known concept of measure of non-compactness [3] as used also in functional analysis [22] and which was canonically recaptured in the setting of approach structures in [10]. Not to confuse with (probability) measures we will however refer to such quantitative values differently and speak of index of (relative, sequential) compactness rather than measure of (relative, sequential) compactness. In the same philosophy and along similar lines, we will introduce a weak and a strong index of tightness and we will prove a quantitative version of Prokhorov's theorem.

For more background information on measures of compactness see [3,7], on probabilistic concepts see [11,17,20,21] and on topological concepts see [5].

2. Preliminaries

There are many papers and a basic reference work [9] on approach theory and we would like to refer the interested reader to those for more indepth information. In these preliminaries we restrict our attention to the basic concepts required for the paper.

Approach spaces can be introduced in various equivalent ways, here we will use the characterizations using gauges, distance and limit operator. Note that we do not suppose that pseudo-quasi-metrics (shortly called *pq*-metrics) are finite.

An approach space is a set equipped with any of the above mentioned structures and if needed we will make clear on which set any of these structures is considered by using the set as an index.

2.1. A *gauge* $\mathcal{G}$ on X is a collection of *pq*-metrics that is closed under the formation of finite suprema and locally saturated, meaning that the following condition is satisfied

(G) Whenever e is a *pq*-metric such that $\forall x \in X,\ \forall \epsilon > 0,\ \forall \omega < \infty,\ \exists d \in \mathcal{G}\colon\ e(x,\cdot) \wedge \omega \leqslant d(x,\cdot) + \epsilon$ then $e \in \mathcal{G}$.

If X and Y are approach spaces and $f : X \to Y$ is a function, then f is called a *contraction* if

$$\forall d \in \mathcal{G}_Y\colon \quad d \circ (f \times f) \in \mathcal{G}_X.$$

For practical reasons, one often works with a *gauge basis* instead of with the entire gauge. By definition this is nothing more than an ideal basis. We recall that a set $\mathcal{H}$ of *pq*-metrics is called an ideal basis if for any $d, e \in \mathcal{H}$ there exists $c \in \mathcal{H}$ such that $d \vee e \leqslant c$. We then say that such a basis $\mathcal{H}$ generates a gauge $\mathcal{G}$ if saturating $\mathcal{H}$ according to the saturation condition (G) gives the entire collection $\mathcal{G}$. An approach space is called *uniform* if the gauge has a basis of *p*-metrics.

2.2. A *distance* on a set X is a function $\delta : X \times 2^X \to [0,\infty]$ with the following axioms:

(D1) $\forall x \in X\colon\ \delta(x,\{x\}) = 0$,
(D2) $\forall x \in X\colon\ \delta(x,\emptyset) = \infty$,
(D3) $\forall x \in X,\ \forall A, B \in 2^X\colon\ \delta(x, A \cup B) = \min\{\delta(x,A), \delta(x,B)\}$,
(D4) $\forall x \in X,\ \forall A \in 2^X,\ \forall \epsilon \in [0,\infty]\colon\ \delta(x,A) \leqslant \delta(x, A^{(\epsilon)}) + \epsilon$ where $A^{(\epsilon)} = \{x \mid \delta(x,A) \leqslant \epsilon\}$.

The transitions between gauges and distances is given by the formulas

$$\delta(x,A) = \sup_{d \in \mathcal{G}} \inf_{a \in A} d(x,a) \quad \text{and} \quad \mathcal{G} = \Big\{ d \ \Big|\ \forall A \subset X\colon \inf_{a \in A} d(\cdot,a) \leqslant \delta(\cdot,A) \Big\},$$

these transitions being of course inverse to each other. In terms of distances, a function $f : X \to Y$ is a contraction if

$$\forall x \in X,\ \forall A \subset X\colon \quad \delta_Y\big(f(x), f(A)\big) \leqslant \delta_X(x,A).$$

2.3. Given a subset $A \subset X$, we will use the notation $\mathbb{F}(\mathbb{A})$ (resp. $\mathbb{U}(\mathbb{A})$) for the collection of all filters (resp. ultrafilters) on X containing the set A, and $\dot{x}$ for the principal filter $\{A \subset X \mid x \in A\}$. A function $\lambda : \mathbb{F}(\mathbb{X}) \to [0,\infty]^X$ is called a *limit operator* if it satisfies

(L1) $\forall x \in X\colon\ \lambda\dot{x}(x) = 0$.
(L2) For any family $(\mathcal{F}_i)_{i \in I}$ of filters on X, $\lambda(\bigcap_{i \in I} \mathcal{F}_i) = \sup_{i \in I} \lambda\mathcal{F}_i$.

(L3) For any $\mathcal{F} \in \mathbb{F}(\mathbb{X})$ and any selection of filters $(\mathcal{S}(x))_{x\in X}$ on X,

$$\lambda\big(\mathcal{D}(\mathcal{S},\mathcal{F})\big) \leqslant \lambda\mathcal{F} + \sup_{x\in X} \lambda\big(\mathcal{S}(x)\big)(x).$$

In condition (L3) $\mathcal{D}(\mathcal{S},\mathcal{F})$ stands for the so-called *diagonal filter* $\mathcal{D}(\mathcal{S},\mathcal{F}) := \bigcup_{F\in\mathcal{F}} \bigcap_{y\in F} \mathcal{S}(y)$.

The transition between limits and distances is given by the following formulas

$$\lambda\mathcal{F}(x) = \sup_{A\in \sec\mathcal{F}} \delta(x,A) \quad \text{and} \quad \delta(x,A) = \inf_{\mathcal{U}\in\mathbb{U}(\mathbb{A})} \lambda\mathcal{U}(x),$$

with $\sec\mathcal{F} := \{A \subset X \mid \forall F \in \mathcal{F}\colon\ F \cap A \neq \emptyset\}$, the transitions again being inverse to each other.

In terms of limit operators a function $f : X \to Y$ is a contraction if

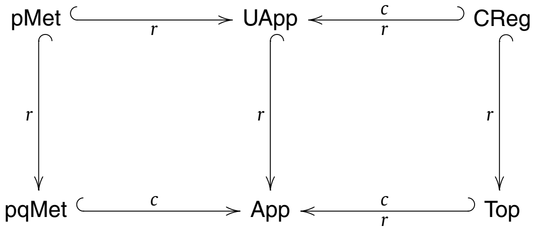
$$\forall \mathcal{F} \in \mathbb{F}(\mathbb{X})\colon \quad \lambda_Y\big(\text{stack}\, f(\mathcal{F})\big) \circ f \leqslant \lambda_X\mathcal{F}$$

where $\text{stack}\,\mathcal{F} := \{H \subset X \mid \exists F \in \mathcal{F}\colon\ F \subset H\}$.

In terms of a gauge or gauge base $\mathcal{B}$ the limit is given by the formula

$$\lambda\mathcal{F}(x) = \sup_{d\in\mathcal{B}} \inf_{F\in\mathcal{F}} \sup_{y\in F} d(x,y).$$

2.4. Approach spaces together with contractions form the topological construct App. The relation among this category and several well-known other categories is depicted in the following diagram where r (resp. c) next to an arrow means the embedding is concretely reflective (resp. coreflective).



We mention the following salient facts about the above diagram. The interested reader who wants more information on it is referred to [9].

(1) If $(X,\mathcal{T})$ is a topological space then it is embedded into App by associating with it the approach space $(X,\mathcal{G}(\mathcal{T}))$ where $\mathcal{G}(\mathcal{T})$ is the gauge consisting of all pq-metrics which generate topologies coarser than $\mathcal{T}$, i.e.

$$\mathcal{G}(\mathcal{T}) := \{e\ pq\text{-metric} \mid \mathcal{T}_e \subset \mathcal{T}_d\}.$$

The embedding of Top in App becomes especially elegant when characterized with distances or limit operators. For a topological space, we have that the associated distance is either 0 or ∞, more precisely $\delta(x,A) = 0$ if $x \in \overline{A}$ and $\delta(x,A) = \infty$ otherwise, and likewise the limit operator is either 0 or ∞ depending on whether the filter converges or not.

(2) Top is embedded in App simultaneously concretely reflectively and coreflectively. If X is an approach space, then the Top coreflection is determined by the closure operator which states that $x \in \overline{A}$ if and only if $\delta(x,A) = 0$ which means in terms of the gauge that $\forall\epsilon > 0$, $\forall d \in \mathcal{G}$, $\exists y \in A\colon\ d(x,y) < \epsilon$, i.e. this coreflection is determined by the neighborhood system with subbase $\{B_d(x,\epsilon) \mid d \in \mathcal{G}, \epsilon > 0\}$.

(3) pqMet is embedded in App by associating with any pq-metric space (X,d) the approach space $(X,\mathcal{G}(d))$ where the gauge $\mathcal{G}(d)$ is generated by $\{d\}$. In this case this is simply the principal ideal $\mathcal{G}(d) := \{e\ pq\text{-metric} \mid e \leqslant d\}$.

(4) pqMet (resp. pMet) is concretely coreflectively embedded in App. If $(X,\mathcal{G})$ is an approach space, then the pqMet (resp. pMet) coreflection is determined by the pq-metric (resp. p-metric)

$$d(x,y) := \sup_{e\in\mathcal{G}} e(x,y) \quad \text{resp.} \quad d(x,y) := \sup_{e\in\mathcal{G}} e(x,y) \vee e(y,x)$$

where $\mathcal{G}$ may be replaced by any base, or in terms of the distance $d(x,y) := \delta(x,\{y\})$ (respectively $d(x,y) := \delta(x,\{y\}) \vee \delta(y,\{x\})$).

(5) The category UApp of the so-called uniform approach spaces is the full subcategory of App with objects those spaces which have a gauge generated by p-metrics. Hence it is the epireflective hull of pMet in App. It is also easily seen that UApp is actually concretely reflective in App analogously to the way CReg is embedded in Top.

(6) Given a source $(f_j : X \to X_j)_{j\in J}$ where all X_j are approach spaces the initial gauge is given by the base

$$\Big\{\sup_{j\in K} d_j \circ (f_j \times f_j) \;\Big|\; K \subset J \text{ finite},\ d_j \in \mathcal{G}_j\Big\}$$

and the initial limit is given by

$$\lambda\mathcal{F} = \sup_{j\in J} \lambda_j f_j(\mathcal{F}) \circ f_j.$$

2.5. Given an approach space X its *measure of compactness* is defined as

$$\mu_c(X) := \sup_{\mathcal{U} \in \mathbb{U}(\mathbb{X})} \inf_{x \in X} \lambda\mathcal{U}(x).$$

An approach space X is called 0-*compact* if $\mu_c(X) = 0$. A topological approach space is 0-compact if and only if it is compact, and a pseudometric approach space is 0-compact if and only if it is totally bounded [9,2,10].

In general, however, compactness of the topological coreflection is strictly stronger and we refer to this property as being *compact*.

2.6. In an approach space X, a filter $\mathcal{F}$ is said to be *Cauchy* if $\inf_{x \in X} \lambda\mathcal{F}(x) = 0$. For p-metric spaces it is easily seen that this notion coincides with the usual notion of a Cauchy filter, and for topological spaces this simply means the filter is convergent. This concept allows for a completeness theory for uniform approach spaces which extends the completeness theory for p-metric spaces. An approach space is said to be *complete* if every Cauchy filter converges (in the topological coreflection). An important result says that an approach space is complete if and only if its pseudometric coreflection is complete (we note that in case of non-completeness the completions on the other hand are in general totally different).

3. Spaces of probability measures

Let S be a Polish space with a fixed complete metric d, with topology $\mathcal{T}$, and with Borel sets $\mathcal{B}$. $\mathcal{P}(S)$ will denote the set of all probability measures on $\mathcal{B}$. One of the most important and most widely used structures on $\mathcal{P}(S)$ is the so-called *weak topology*, which we denote by $\mathcal{T}_w$ [4,14]. Although this topology is called the weak topology, from the point of view of functional analysis it would better have been called the weak* topology, but we will adhere to the usual term. We consider the Banach space of all continuous bounded real-valued functions $\mathcal{C}_b(S)$ equipped with the supremum norm and consider its continuous dual $\mathcal{C}_b(S)'$. $\mathcal{P}(S)$ is embedded in $\mathcal{C}_b(S)'$ by the assignment

$$\mathcal{P}(S) \to \mathcal{C}_b(S)' : P \mapsto \left(f \mapsto \int f\,dP\right)$$

and as such is identified with the dual unit sphere. Thus it inherits the weak* topology induced on $\mathcal{C}_b(S)'$ by $\mathcal{C}_b(S)$ via restriction. This weak* topology is a locally convex topology generated by the collection of seminorms $\{p_f \mid f \in \mathcal{C}_b(S),\ 0 \leqslant f \leqslant 1\}$ where

$$p_f(P) := \left|\int f\,dP\right|$$

and the restriction to $\mathcal{P}(S)$ is called the weak topology on probability measures. The above collection of seminorms however generates a collection of pseudometrics

$$d_f(P, Q) = \left|\int f\,dP - \int f\,dQ\right|$$

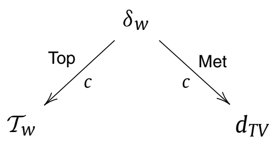
and it is immediately clear that this collection of pseudometrics is a subbasis for a gauge in the sense of 2(1), precisely

$$\left\{ \sup_{f \in \mathcal{H}} d_f \;\middle|\; \mathcal{H} \subset \mathcal{C}_b(S) \text{ finite},\ 0 \leqslant f \leqslant 1\ \forall f \in \mathcal{H} \right\}$$

is an ideal basis which generates a unique and canonical approach structure on $\mathcal{P}(S)$ which we refer to as the *weak approach structure* on probability measures. All associated structures will be denoted by the index w. Thus the gauge generated by the above base will be denoted $\mathcal{G}_w$.

We begin by verifying what are the topological and metric coreflections of the weak approach structure. As we are dealing with concrete coreflections, in the following theorem we only mention the structures since the underlying sets remain invariant.

Proposition 3.1. *The Top-coreflection of a space with the weak approach structure is determined by the weak topology and the pMet-coreflection is determined by the total variation metric*



Proof. For the topological coreflection this is an immediate consequence of 2(4)(2) and the definition of the weak topology (see e.g. below or [14]). For the metric coreflection this follows from 2(4)(4) and the definition of the total variation metric (again see e.g. below or [15]). □

We recall that the weak topology [4,14] has various different but equivalent bases $\mathcal{B}_w^i$, $i = 1, \ldots, 4$, for the neighborhoods. Let $P \in \mathcal{P}(S)$. Then $\mathcal{B}_w^1$ consists of the sets $V^1(P, \mathcal{G}, \varepsilon)$ where $\mathcal{G}$ is a finite collection of open sets, $\varepsilon > 0$ and

$$V^1(P, \mathcal{G}, \varepsilon) := \{Q \in \mathcal{P}(S) \mid \forall G \in \mathcal{G}\colon\ Q(G) > P(G) - \varepsilon\}.$$

$\mathcal{B}_w^2$ consists of the sets $V^2(P, \mathcal{F}, \varepsilon)$ where $\mathcal{F}$ is a finite collection of closed sets, $\varepsilon > 0$ and

$$V^2(P, \mathcal{F}, \varepsilon) := \{Q \in \mathcal{P}(S) \mid \forall F \in \mathcal{F}\colon\ Q(F) < P(F) + \varepsilon\}.$$

$\mathcal{B}_w^3$ consists of the sets $V^3(P, \mathcal{H}, \varepsilon)$ where $\mathcal{H}$ is a finite collection of continuous (resp. uniformly continuous or Lipschitz) functions taking values in $[0, 1]$, $\varepsilon > 0$ and

$$V^3(P, \mathcal{H}, \varepsilon) := \left\{Q \in \mathcal{P}(S) \,\middle|\, \forall f \in \mathcal{H}\colon\ \left|\int f\,dP - \int f\,dQ\right| < \varepsilon\right\}.$$

$\mathcal{B}_w^4$ consists of the sets $V^2(P, \mathcal{E}, \varepsilon)$ where $\mathcal{E}$ is a finite collection of P-continuity sets, $\varepsilon > 0$ and

$$V^2(P, \mathcal{E}, \varepsilon) := \{Q \in \mathcal{P}(S) \mid \forall E \in \mathcal{E}\colon\ |P(E) - Q(E)| < \varepsilon\}.$$

We also recall that the total variation metric [15] is defined as

$$d_{TV}(P, Q) := \sup_{B \in \mathcal{B}} |P(B) - Q(B)|$$

and is equally well given by various formulas analogous to the various bases for the weak topology given above, and notably by

$$d_{TV}(P, Q) = \sup_{f \in \mathcal{C}_b(S),\, 0 \leqslant f \leqslant 1} \left|\int f\,dP - \int f\,dQ\right|.$$

From 3.1 it follows that the distance δ_w, in a natural way, "distancizes" the weak topology. Note that whereas there are a variety of metrics which metrize the weak topology, the canonicity of the weak distance comes from the fact that it is built, as we will now see (especially in 3.3) by exactly the same "building-blocks" as the weak topology itself.

As in the case of the weak topology, the weak approach structure, notably its gauge $\mathcal{G}_w$ has various different bases. For any finite collection $\mathcal{G}$ of open sets, we let

$$d_1^{\mathcal{G}} : \mathcal{P}(S) \times \mathcal{P}(S) \longrightarrow [0, \infty] : (P, Q) \mapsto \sup_{G \in \mathcal{G}} \big(P(G) - Q(G)\big) \vee 0$$

and we put $\mathcal{D}_1 := \{d_1^{\mathcal{G}} \mid \mathcal{G}$ finite collection of open sets$\}$. For any finite collection $\mathcal{F}$ of closed sets, we let

$$d_2^{\mathcal{F}} : \mathcal{P}(S) \times \mathcal{P}(S) \longrightarrow [0, \infty] : (P, Q) \mapsto \sup_{F \in \mathcal{F}} \big(Q(F) - P(F)\big) \vee 0$$

and we put $\mathcal{D}_2 := \{d_2^{\mathcal{F}} \mid \mathcal{F}$ finite collection of closed sets$\}$. For any finite collection $\mathcal{H}$ of continuous maps, with range $[0, 1]$, we let

$$d_3^{\mathcal{H}} : \mathcal{P}(S) \times \mathcal{P}(S) \longrightarrow [0, \infty] : (P, Q) \mapsto \sup_{f \in \mathcal{H}} \left|\int f\,dP - \int f\,dQ\right|$$

and we put $\mathcal{D}_3 := \{d_3^{\mathcal{H}} \mid \mathcal{H} \subset C_b(S)$ finite, $\forall f \in \mathcal{H}\colon\ 0 \leqslant f \leqslant 1\}$. As before, continuous maps may be replaced by uniformly continuous maps or Lipschitz maps. For any finite collection $\mathcal{E}$ of P-continuity sets we put

$$d_4^{\mathcal{E}} : \mathcal{P}(S) \times \mathcal{P}(S) \longrightarrow [0, \infty] : (P, Q) \mapsto \sup_{E \in \mathcal{E}} |P(E) - Q(E)|$$

and we put $\mathcal{D}_4 := \{d_4^{\mathcal{E}} \mid \mathcal{E}$ finite set of P-continuity sets$\}$. For any $\alpha > 0$ we let

$$d_5^{\alpha} : \mathcal{P}(S) \times \mathcal{P}(S) \longrightarrow [0, \infty] : (P, Q) \mapsto \sup_{A \in \mathcal{B}} \big(P(A) - Q\big(A^{(\alpha)}\big)\big) \vee 0$$

and we put $\mathcal{D}_5 := \{d_5^{\alpha} \mid \alpha > 0\}$.

The collections $\mathcal{D}_i$, $i \in \{1, 2\}$, consist of pseudo-quasi-metrics, whereas the collections $\mathcal{D}_i$, $i \in \{3, 4\}$, consist of pseudo-metrics. The mappings d_5^{α} do not individually satisfy the triangle inequality, however, as is easily verified, they do satisfy the combined inequality

$$d_5^{\alpha}(P, Q) \leqslant d_5^{\alpha/2}(P, R) + d_5^{\alpha/2}(R, Q)$$

for any $\alpha > 0$ and any $P, Q, R \in \mathcal{P}(S)$, which is sufficient to generate a distance [9]. This last collection is inspired by the so-called Prokhorov metric [4] which is defined as

$$\rho(P, Q) := \inf\{\alpha > 0 \mid \forall A \in \mathcal{B}_S\colon\ P(A) \leqslant Q\big(A^{(\alpha)}\big) + \alpha\}.$$

Actually all collections generate the same distance, namely δ_w. In order to prove this we collect the main technical arguments in the following preliminary lemma which will be used several times.

Lemma 3.2. *The following hold.*

(1) *For each $P \in \mathcal{P}(S)$, $\epsilon > 0$ and $\alpha > 0$ there exists a finite collection $\mathcal{G}$ of open sets in S such that for every $Q \in \mathcal{P}(S)$*

$$\sup_{A \in \mathcal{B}_S} \big(P(A) - Q\big(A^{(\alpha)}\big)\big) \leqslant \sup_{G \in \mathcal{G}} \big(P(G) - Q(G)\big) + \epsilon.$$

(2) *For each $P \in \mathcal{P}(S)$, $\epsilon > 0$ and $F \subset S$ closed there exists an $\alpha > 0$ such that for every $Q \in \mathcal{P}(S)$*

$$Q(F) - P(F) \leqslant \sup_{A \in \mathcal{B}_S} \big(P(A) - Q\big(A^{(\alpha)}\big)\big) + \epsilon.$$

(3) *For each $P \in \mathcal{P}(S)$, $\epsilon > 0$ and $F \subset S$ closed there exists $f \in \mathcal{C}_b(S)$ with $0 \leqslant f \leqslant 1$ such that for all $Q \in \mathcal{P}(S)$*

$$Q(F) - P(F) \leqslant \left| \int f\,dP - \int f\,dQ \right| + \epsilon.$$

(4) *For each $f \in \mathcal{C}_b(S)$ such that $0 \leqslant f \leqslant 1$ and $\epsilon > 0$ there exists a finite set of closed sets $\mathcal{F}$ such that for all $P, Q \in \mathcal{P}(S)$*

$$\left| \int f\,dP - \int f\,dQ \right| \leqslant \sup_{F \in \mathcal{F}} \big(Q(F) - P(F)\big) + \epsilon.$$

Proof. (1) By separability we can choose a finite collection of open balls $(B_i)_{i=1}^{j}$ with radii $\alpha/4$ such that $P(S \setminus \bigcup_{i=1}^{j} B_i) \leqslant \epsilon$. Then the collection

$$\mathcal{G} := \big\{(B_{i_1} \cup \cdots \cup B_{i_k})^{(\alpha/2)} \;\big|\; 1 \leqslant i_1 < \cdots < i_k \leqslant j\big\}$$

satisfies the requirement. Indeed, take a probability measure Q in $\mathcal{P}(S)$ and a Borel set A in S. Let I be the set of those natural numbers $1 \leqslant i \leqslant j$ for which $B_i \cap A \neq \emptyset$ and put $B := \bigcup_{i \in I} B_i$. Then we have

$$\begin{aligned} P(A) &\leqslant P(B) + P\left(S \Big\backslash \bigcup_{i=1}^{j} B_i\right) \\ &\leqslant P\big(B^{(\alpha/2)}\big) + \epsilon \\ &\leqslant \sup_{G \in \mathcal{G}} \big(P(G) - Q(G)\big) + Q\big(B^{(\alpha/2)}\big) + \epsilon. \end{aligned}$$

In view of the fact that $B^{(\alpha/2)} \subset A^{(\alpha)}$, we conclude that

$$P(A) \leqslant \sup_{G \in \mathcal{G}} \big(P(G) - Q(G)\big) + Q\big(A^{(\alpha)}\big) + \epsilon$$

whence the claim.

(2) We can choose $\alpha > 0$ such that $P(F^{(\alpha)}) \leqslant P(F) + \epsilon$. For any probability measure Q in $\mathcal{P}(S)$ we then have

$$\begin{aligned} Q(F) - P(F) &\leqslant \big(Q(F) - P\big(F^{(\alpha)}\big)\big) + \epsilon \\ &\leqslant \sup_{A \in \mathcal{B}_S} \big(Q(A) - P\big(A^{(\alpha)}\big)\big) + \epsilon. \end{aligned}$$

(3) Again we can choose $\alpha > 0$ such that $P(F^{(\alpha)}) \leqslant P(F) + \epsilon$. Then the function f defined by

$$f(x) := \left(1 - \frac{1}{\alpha} d(x, F)\right) \vee 0, \quad \forall x \in S$$

satisfies the requirement.

(4) Choose $k \in \mathbb{N}_0$ such that $\frac{1}{k} \leqslant \epsilon$ and, for all $i \in \{1, \ldots, k\}$, let

$$F_i := \left\{ \frac{i}{k} \leqslant f \right\}$$

and consider the collection $\mathcal{F} := \{F_i \mid i \in \{1, \ldots, k\}\}$. Then for any $P \in \mathcal{P}$,

$$\frac{1}{k} \sum_{F \in \mathcal{F}} P(F) \leqslant \int f\,dP \leqslant \frac{1}{k} + \frac{1}{k} \sum_{F \in \mathcal{F}} P(F)$$

from which it is easily seen that the collection $\mathcal{F}$ satisfies the requirement. □

Proposition 3.3 *(Distance portmanteau theorem). All collections $\mathcal{D}_i$, $i \in \{0, \dots, 5\}$, are bases for $\mathcal{G}_w$ and hence generate the same distance δ_w. Writing out this distance explicitly then gives the following expressions where $P \in \mathcal{P}(S)$ and $\Gamma \subset \mathcal{P}(S)$*

$$\delta_w(P, \Gamma) = \sup_{\mathcal{G}} \inf_{Q \in \Gamma} \sup_{G \in \mathcal{G}} \big(P(G) - Q(G)\big) \vee 0 \tag{1}$$

$$= \sup_{\mathcal{F}} \inf_{Q \in \Gamma} \sup_{F \in \mathcal{F}} \big(Q(F) - P(F)\big) \vee 0 \tag{2}$$

$$= \sup_{\mathcal{E}} \inf_{Q \in \Gamma} \sup_{E \in \mathcal{E}} \big|P(E) - Q(E)\big| \tag{3}$$

$$= \sup_{\mathcal{C}} \inf_{Q \in \Gamma} \sup_{f \in \mathcal{C}} \left|\int f\, dP - \int f\, dQ\right| \tag{4}$$

$$= \sup_{\alpha > 0} \inf_{Q \in \Gamma} \sup_{A \in \mathcal{B}} \big(P(A) - Q\big(A^{(\alpha)}\big)\big) \vee 0 \tag{5}$$

where the suprema in (1)–(4) *are respectively taken over all finite collections $\mathcal{G}$* (*resp.* $\mathcal{F}$) *of open* (*resp. closed*) *sets, $\mathcal{E}$ of P-continuity sets, $\mathcal{C}$ of continuous* (*resp. uniformly continuous or Lipschitz*) *with range* [0, 1].

Proof. In order to see that all collections are equivalent bases for the same gauge it suffices to use the foregoing lemma and the saturation condition (G) for gauges. The formulas (1)–(5) then follow immediately from the definition of a distance generated by a gauge (see 2.2) and from the foregoing Lemma 3.2. □

There is another interesting way to characterize δ_w. In [19] Topsøe showed that the weak topology on $\mathcal{P}(S)$ is the initial topology for the source

$$\big(\omega_G : \mathcal{P}(S) \longrightarrow \big([0, \infty], \mathcal{T}_r\big) : P \mapsto P(G)\big)_{G \in \mathcal{T}},$$

where $\mathcal{T}_r$ stands for the so-called right-limit topology on $[0, \infty]$ which is generated by the open sets of type $]a, \infty]$. This however is nothing else but the topology of the topological coreflection of the space $\mathbb{P} := ([0, \infty], \delta_{\mathbb{P}})$ where $\delta_{\mathbb{P}}(x, A) := (x - \sup A) \vee 0$. $\mathbb{P}$ is an initially dense object in App [9]. The next result shows that Topsøe's theorem also holds in our setting.

Proposition 3.4. *The weak distance δ_w on $\mathcal{P}$ is initial for the source*

$$\big(\omega_G : \mathcal{P}(S) \longrightarrow \mathbb{P} : P \mapsto P(G)\big)_{G \in \mathcal{T}}.$$

Proof. Since the maps ω_G only attain finite values and since a subbase for the initial gauge is given by

$$\big\{d \circ (\omega_G \times \omega_G) \;\big|\; G \text{ open}, d \in \mathcal{G}_{\mathbb{P}}\big\}$$

where $\mathcal{G}_{\mathbb{P}}$ stands for the gauge of $\mathbb{P}$, this is an immediate consequence of the fact that a base for the gauge of $\mathbb{P}$ is given by the pseudo-quasi-metrics

$$[0, \infty] \times [0, \infty] \to [0, \infty] : (x, y) \mapsto (x \wedge a - y \wedge a) \vee 0$$

where $a < \infty$ [9]. Note however that all values which come into play are bounded by 1 and hence it suffices to consider the pseudo-quasi-metric $d(x, y) = (x - y) \vee 0$. The above subbase for the initial gauge hence generates the base $\mathcal{D}_1$. □

Theorem 3.5 *(Convergence portmanteau theorem). Given a sequence $(P_n)_n$ and P in $\mathcal{P}(S)$ we have*

$$\lambda_{\delta_w}(P_n)(P) = \sup_{G} \limsup_{n} \big(P(G) - P_n(G)\big) \vee 0 \tag{6}$$

$$= \sup_{F} \limsup_{n} \big(P_n(F) - P(F)\big) \vee 0 \tag{7}$$

$$= \sup_{A} \limsup_{n} \big|P(A) - P_n(A)\big| \tag{8}$$

$$= \sup_{f} \limsup_{n} \left|\int f\, dP - \int f\, dP_n\right| \tag{9}$$

$$= \sup_{\alpha > 0} \limsup_{n} \sup_{A \in \mathcal{B}} \big(P(A) - P_n\big(A^{(\alpha)}\big)\big) \vee 0 \tag{10}$$

the suprema respectively running over all open sets, closed sets, P-continuity sets in S, and all continuous (*or uniformly continuous, or Lipschitz*) *functions f from S to* [0, 1].

Proof. We only prove the first equality, the formulas (6)–(10) then follow from this upon applying once again Lemma 3.2. From 3.4 and the formula for an initial limit [9] it follows that

$$\lambda(P_n)(P) = \sup_{G\in\mathcal{T}} \lambda_{\mathbb{P}}\big(P_n(G)\big)\big(P(G)\big).$$

Now it suffices to remark that all values are finite (less than 1) and hence, as in 3.4 the structure of $\mathbb{P}$ which comes into play is only the structure in finite points (the interval $[0,1]$) and there the structure is simply a pseudo-quasi-metric ($d(x,y) = (x-y)\vee 0$). Hence it follows from the formula for the limit in a pseudo-quasi-metric space [9] that

$$\lambda_{\mathbb{P}}\big(P_n(G)\big)\big(P(G)\big) = \inf_n \sup_{k\geqslant n}\big(P(G) - P_k(G)\big) \vee 0$$

and the formula follows. □

By explicitly writing down when the expressions in the foregoing result become zero, one obtains all characterizations of weak convergence in the classic portmanteau theorem.

Corollary 3.6 *(Classic portmanteau theorem). ([4,14]) A sequence $(P_n)_n$ in $\mathcal{P}(S)$ converges weakly to $P\in\mathcal{P}(S)$ if and only if any of the following equivalent properties hold*

$$\forall G \text{ open}:\quad P(G) \leqslant \liminf P_n(G), \tag{11}$$

$$\forall F \text{ closed}:\quad \limsup_n P_n(F) \leqslant P(F), \tag{12}$$

$$\forall P\text{-continuity set } A:\quad \lim_n P_n(A) = P(A), \tag{13}$$

$$\forall f\in\mathcal{F}\big(S,[0,1]\big):\quad \lim_n \int f\,dP_n = \int f\,dP \tag{14}$$

where $\mathcal{F}(S,[0,1])$ stands for all continuous (or uniformly continuous, or Lipschitz) functions from S to $[0,1]$.

Another fundamental fact about the weak topology is that if $f: X\to Y$ is a continuous function then its canonical extension $\hat{f}:\mathcal{P}(X)\to\mathcal{P}(Y)$ defined by $\hat{f}(P)(B) := P(f^{-1}(B))$, for all $B\in\mathcal{B}(Y)$, is continuous with respect to the weak topologies. The result here is considerably stronger.

Proposition 3.7. *If $f: X\to Y$ is a continuous function and we equip $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ with the weak approach structures, then $\hat{f}:\mathcal{P}(X)\to\mathcal{P}(Y)$ is a contraction.*

Proof. Let $\mathcal{G}$ be a finite collection of open sets in Y, and let $P, Q\in\mathcal{P}(X)$; then

$$\begin{aligned} d_0^{\mathcal{G}}\big(\hat{f}(P), \hat{f}(Q)\big) &= \sup_{G\in\mathcal{G}}\big(P\big(f^{-1}(G)\big) - Q\big(f^{-1}(G)\big)\big)\vee 0 \\ &= d_0^{\mathcal{H}}(P,Q), \end{aligned}$$

where $\mathcal{H} := \{f^{-1}(G) \mid G\in\mathcal{G}\}$. This proves that

$$d_0^{\mathcal{G}}\circ \hat{f}\times\hat{f} = d_0^{\mathcal{H}} \in \mathcal{G}_w,$$

which by 2.1 proves our claim. □

Corollary 3.8. *([19]) If $f: X\to Y$ is continuous then $\hat{f}:\mathcal{P}(X)\to\mathcal{P}(Y)$ is continuous with respect to the weak topologies.*

Corollary 3.9. *If $f: X\to Y$ is continuous then $\hat{f}:\mathcal{P}(X)\to\mathcal{P}(Y)$ is nonexpansive with respect to the total variation metrics.*

If we put Pol for the category of completely metrizable separable topological spaces (Polish spaces) and continuous maps then it follows from 3.7 that

$$\mathsf{Pol}\to\mathsf{App}: \begin{cases} S\to(\mathcal{P}(S),\delta_w), \\ f\to\hat{f} \end{cases}$$

is functorial.

The relation between the various "quantitative" structures which came into play is given in the following result. Note that simple examples show that in general the inequalities are strict. We recall that ρ stands for the Prokhorov metric and d_{TV} for the total variation metric [4] (see Section 2 for the definitions). We also note that $\rho\leqslant d_{TV}$ is a known inequality [23].

Proposition 3.10. *The following inequalities* (*as approach distances*) *hold*: $\rho \leqslant \delta_w \leqslant d_{TV}$.

Proof. Fix a probability measure P and a collection of probability measures Γ on S. To prove that $\rho \leqslant \delta_\omega$ let $\delta_w(P, \Gamma) < \gamma$. Then by definition we can find a measure $Q_\gamma \in \Gamma$ such that

$$\sup_{A \in \mathcal{B}_S} \left(P(A) - Q_\gamma\left(A^{(\gamma)}\right)\right) < \gamma$$

and hence

$$\rho(P, \Gamma) = \inf_{Q \in \Gamma} \inf\left\{\alpha > 0 \,\middle|\, \forall A \in \mathcal{B}_S\colon\ P(A) \leqslant Q\left(A^{(\alpha)}\right) + \alpha\right\} \leqslant \gamma.$$

This, by the arbitrariness of γ, implies that $\rho(P, \Gamma) \leqslant \delta_w(P, \Gamma)$. The second inequality is an immediate consequence of the fact that d_{TV} is the metric coreflection of the weak approach structure. □

It is known that the weak topology is completely metrizable. However whereas this requires the choice of a new "external" structure (a complete compatible metric) the weak approach structure does not require this, it is complete itself.

Theorem 3.11. *The weak approach structure is complete.*

Proof. From [9] (see also 2.6) we know that in order to verify that an approach space is complete it is sufficient to verify that its metric coreflection is complete. It is however a well-known fact that d_{TV} is a complete metric [8]. □

Proposition 3.12. *The weak approach structure is first countable in the sense that for each $P \in \mathcal{P}(S)$ the localized gauge*

$$\mathcal{G}_w(P) := \left\{d(P, \cdot) \,\middle|\, d \in \mathcal{G}_w\right\}$$

has a countable base, consequently by and large, sequences suffice.

Proof. Although by definition the weak approach structure was constructed making use of the (non-separable) Banach space $\mathcal{C}_b(S)$, when considering the various bases for the weak gauge $\mathcal{G}_w$, in particular $\mathcal{D}_3$ we mentioned that we could also restrict ourselves to uniformly continuous maps $S \to [0, 1]$. Now since S is a separable metrizable space it can be embedded in a countable product of unit intervals and consequently there exists an equivalent totally bounded metrization of S. The completion $\widehat{S}$ of S under this metric hence is compact and the Banach spaces $\mathcal{C}(\widehat{S})$ and $\mathcal{U}(S)$ of uniformly continuous maps are isomorphic. $\widehat{S}$ being compact, $\mathcal{C}(\widehat{S})$ and hence also $\mathcal{U}(S)$ are separable. Then it follows that also the space of uniformly continuous maps $S \to [0, 1]$ is separable (always under the supremum-norm). If $\mathcal{E}$ is a countable dense subset then it follows immediately from the definition of $\mathcal{D}_3$ (with uniformly continuous maps) that an alternative equivalent base for $\mathcal{G}_w$ is given by

$$\mathcal{D}_3' := \left\{d_3^{\mathcal{H}} \,\middle|\, \mathcal{H} \subset \mathcal{E} \text{ finite, } \forall f \in \mathcal{H}\colon\ 0 \leqslant f \leqslant 1\right\}.$$

As this base is countable, so are the localized bases $\mathcal{D}_3'(P) = \{d(P, \cdot) \mid d \in \mathcal{D}_3^s\}$. □

4. Spaces of random variables

In this section we consider spaces of random variables. Let $(\Omega, \mathcal{A}, P)$ be a fixed probability space and let $\mathcal{R}(S)$ be the set of all S-valued random variables on Ω. In this context an important topology is given by the topology $\mathcal{T}_p$ of convergence in probability and a natural metric is the so-called indicator metric [23] d_I where

$$d_I(X, Y) := P\left(\left\{\omega \,\middle|\, d\left(X(\omega), Y(\omega)\right) > 0\right\}\right) = P\left(\left\{\omega \,\middle|\, X(\omega) \neq Y(\omega)\right\}\right).$$

Note that $d_I(X, Y) = 0$ if and only if X and Y are equal almost everywhere.

$\mathcal{R}(S)$ is naturally endowed with an approach structure as follows. We consider the functions φ^a, $a > 0$, determined by

$$\varphi^a(X, Y) = P\left(\left\{\omega \in \Omega \,\middle|\, d\left(X(\omega), Y(\omega)\right) \geqslant a\right\}\right), \quad X, Y \in \mathcal{R}(S).$$

Each function, for a fixed a, gives the probability that the random variables X and Y lie at a distance larger than or equal to a. Again, as in the case of the base $\mathcal{D}_5$ for $\mathcal{G}_w$ these functions do not satisfy the triangle inequality. However, again, they too satisfy a combined triangle inequality.

Lemma 4.1. *For any $a, b > 0$ and $X, Y, Z \in \mathcal{R}(S)$ we have*

$$\varphi^{a+b}(X, Z) \leqslant \varphi^a(X, Y) + \varphi^b(Y, Z).$$

Proof. This follows from

$$\begin{aligned}\varphi^{a+b}(X,Z) &= P\big(\big\{\omega\in\Omega \,\big|\, d\big(X(\omega),Z(\omega)\big)\geqslant a+b\big\}\big)\\ &\leqslant P\big(\big\{\omega\in\Omega \,\big|\, d\big(X(\omega),Y(\omega)\big)\geqslant a\big\}\cup\big\{\omega\in\Omega \,\big|\, d\big(Y(\omega),Z(\omega)\big)\geqslant b\big\}\big)\\ &\leqslant P\big(\big\{\omega\in\Omega \,\big|\, d\big(X(\omega),Y(\omega)\big)\geqslant a\big\}\big)+P\big(\big\{\omega\in\Omega \,\big|\, d\big(Y(\omega),Z(\omega)\big)\geqslant b\big\}\big)\\ &= \varphi^a(x,y)+\varphi^b(y,z). \qquad \square\end{aligned}$$

Hence it follows again from [9] that this collection generates a gauge $\mathcal{G}_p$. We will denote the distance generated by $\mathcal{G}_p$ by δ_p and refer to the approach structure as the *c.i.p. approach structure* (c.i.p. for convergence in probability).

As was the case for the weak approach structure in the foregoing section, here too we can find alternative bases, one of which is particularly interesting. For any $a>0$ put

$$K_a(X,Y) := \inf\big\{\theta \,\big|\, \varphi^{\theta a}(X,Y)\leqslant\theta\big\}.$$

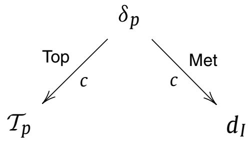
Then it follows immediately from 4.1 that the maps K_a are pseudometrics. Actually K_1 is nothing else than the so-called Ky-Fan metric [4]. Let us denote $\mathcal{B}_1 := \{\varphi^a \mid a>0\}$ and $\mathcal{B}_2 := \{K_a \mid a>0\}$.

Proposition 4.2. *Both $\mathcal{B}_1$ and $\mathcal{B}_2$ generate the c.i.p. approach structure on random variables.*

Proof. By definition $\mathcal{B}_1$ generates δ_p. Let δ stand for the distance generated by $\mathcal{B}_2$. Since for any a and θ we have that $\varphi^{\theta a}(X,Y)<\theta$ implies $K_a(X,Y)\leqslant\theta$ we immediately have $\delta\leqslant\delta_p$. Conversely, if $0<\theta<\delta_p(X,\Sigma)$, then there exists $a>0$ such that $\theta<\inf_{Y\in\Sigma}\varphi^a(X,Y)$. Letting $b:=a\theta^{-1}$ it follows that $\theta\leqslant\inf_{Y\in\Sigma}K_b(X,Y)$ which proves that $\delta_p\leqslant\delta$. □

Again we first look at the topological and metric coreflections.

Proposition 4.3. *The* Top *coreflection of a space with the c.i.p. approach structure is determined by the topology of convergence in probability and the* pMet *coreflection is determined by the indicator metric*



Proof. It suffices to notice that, for any $X\in\mathcal{R}(S)$, a basis for the neighborhoods of X in the topological coreflection is given by the collection of sets

$$\big\{Y\in\mathcal{R}(S) \,\big|\, P\big(\big\{\omega\in\Omega \,\big|\, d\big(X(\omega),Y(\omega)\big)\geqslant a\big\}\big)<\varepsilon\big\},\quad a\in[0,\infty[,\ \varepsilon>0,$$

and that this collection is also precisely a basis for the topology of convergence in probability. As for the metric coreflection it suffices to notice that, for any $X,Y\in\mathcal{R}(S)$, we have

$$\sup_{a>0} P\big(\big\{\omega\in\Omega \,\big|\, d\big(X(\omega),Y(\omega)\big)\geqslant a\big\}\big) = P\big(\big\{\omega\in\Omega \,\big|\, d\big(X(\omega),Y(\omega)\big)>0\big\}\big) = d_I(X,Y). \qquad \square$$

Proposition 4.4. *The limit operator λ_p on $\mathcal{R}(S)$ (for sequences of random variables) is given by the following formula*:

$$\lambda_p\big((X_n)_n\big)(X) = \sup_{a>0}\limsup_{n\to\infty} P\big(\big\{\omega\in\Omega \,\big|\, d\big(X(\omega),X_n(\omega)\big)\geqslant a\big\}\big).$$

Proof. Since the basis for the gauge given by the functions φ^a is increasing with decreasing a this follows at once from the formula for a limit derived from a gauge (basis) as given in 2.3. □

Of course it also follows at once from this formula that the topological coreflection is given by convergence in probability since the limit operator will produce a zero value exactly if $(X_n)_n$ converges to X in probability, i.e.

$$\forall a>0:\quad \lim_n P\big(\big\{\omega\in\Omega \,\big|\, d\big(X(\omega),X_n(\omega)\big)\geqslant a\big\}\big)=0.$$

In analogy with the results for spaces of probability measures here too our construction is functorial, however of a more metric nature, which is to be expected from the prominent role played by the metric in the definition of the maps φ^a.

Proposition 4.5. *Suppose S and T are Polish spaces with fixed metrics d_S and d_T. If $f:S\to T$ is a contraction and we equip $\mathcal{R}(S)$ and $\mathcal{R}(T)$ with the c.i.p. approach structures then $\tilde{f}:\mathcal{R}(S)\to\mathcal{R}(T)$ is a contraction where $\tilde{f}$ is defined by $\tilde{f}(X):=f\circ X$.*

Proof. We denote by φ_S^a and φ_T^a the maps (made by means of the metrics d_S and d_T) which constitute bases for the c.i.p. approach gauges on $\mathcal{R}(S)$ and $\mathcal{R}(T)$ respectively. It suffices now to note that for any $a > 0$ and any $X, Y \in \mathcal{R}(S)$

$$\varphi_T^a(f \circ X, f \circ Y) \leqslant \varphi_S^a(X, Y). \qquad \square$$

If we put Pol_m for the category of complete separable metric spaces and contractions then it follows from 4.5 that

$$\mathsf{Pol}_m \to \mathsf{App}: \begin{cases} S \to (\mathcal{R}(S), \delta_p), \\ f \to \tilde{f} \end{cases}$$

is functorial.

In analogy to Theorem 3.10 we have the following result where K stands for the Ky-Fan metric [4], which we recall is given by

$$K(X, Y) := \inf\bigl\{\alpha > 0 \bigm| P\bigl\{\omega \bigm| d\bigl(X(\omega), Y(\omega)\bigr) \geqslant \alpha\bigr\} \leqslant \alpha\bigr\}.$$

In the following proposition the inequality $K \leqslant d_I$ is known [23].

Proposition 4.6. *The following inequalities (as approach distances) hold:* $K \leqslant \delta_p \leqslant d_I$.

Proof. This is an easy consequence of the definitions and the fact that the gauge of the c.i.p. approach structure is generated by the collection $\mathcal{B}_2 = \{K_\gamma \mid \gamma > 0\}$ (see 4.2). $\square$

Again, as was the case for the weak approach structure, the c.i.p. approach structure turns out to be complete, and this irrespective of whether the original metric on S was complete or not.

Theorem 4.7. *The c.i.p. approach structure is complete.*

Proof. From [9] (see also 2.6) we know that in order to verify that an approach space is complete it is sufficient to verify that its metric coreflection is complete. So all we have to do is show that de indicator metric is complete. Let $(X_n)_n$ be d_I-Cauchy and choose a subsequence $(X_{k_n})_n$ with the property that for each n, $P(X_{k_n} \neq X_{k_{n+1}}) \leqslant \frac{1}{2^n}$. By the Borel–Cantelli lemma we now have that $P(A) = 0$ if we set $A := \bigcap_m \bigcup_{n \geqslant m} \{X_{k_n} \neq X_{k_{n+1}}\}$. Observe that for each $\omega \in \Omega \setminus A$ the sequence $X_{k_n}(\omega)$ is eventually constant. We denote this constant value by $X(\omega)$. Now X is an almost everywhere defined random variable and it is obvious that $(X_{k_n})_n$ converges almost everywhere, and hence also in probability, to X. We claim that even $d_I(X_{k_n}, X)$ converges to 0 as n tends to ∞.

In order to prove this claim, fix $\epsilon > 0$ and let n_0 be such that for all $m \geqslant n \geqslant n_0$ we have $d_I(X_{k_n}, X_{k_m}) \leqslant \epsilon$. Now since for all $k \in \mathbb{N}_0$ and all $m \geqslant n \geqslant n_0$ we have

$$\begin{aligned} P\left(d(X_{k_n}, X) > \frac{1}{k}\right) &\leqslant P\left(d(X_{k_n}, X_{k_m}) > \frac{1}{2k}\right) + P\left(d(X_{k_m}, X) > \frac{1}{2k}\right) \\ &\leqslant \epsilon + P\left(d(X_{k_m}, X) > \frac{1}{2k}\right), \end{aligned}$$

we get, after letting first $m \to \infty$ and then $k \to \infty$, that $P(d(X_{k_n}, X) > 0) \leqslant \epsilon$ for all $n \geqslant n_0$ whence the claim, and we are done. $\square$

Proposition 4.8. *The c.i.p. approach structure is first countable in the sense that for each $P \in \mathcal{P}(S)$ the localized gauge*

$$\mathcal{G}_p(P) := \bigl\{d(P, \cdot) \bigm| d \in \mathcal{G}_p\bigr\}$$

has a countable base, consequently by and large, sequences suffice.

Proof. This is an immediate consequence of the definition of either the base $\mathcal{B}_1$ or the base $\mathcal{B}_2$ as in both cases the indices of the functions in the base may be restricted to range e.g. over the sequence $\{\frac{1}{n} \mid n \geqslant 1\}$. $\square$

We end this section with some important relations between the structures which we have introduced on $\mathcal{P}(S)$ and $\mathcal{R}(S)$ respectively. We recall that if X is a random variable, then its *law* is the probability measure P_X in $\mathcal{P}(S)$ defined by $P_X(B) := P(X^{-1}(B))$, for all $B \in \mathcal{B}$ (the so-called *image measure*).

It is well known that convergence in probability of a sequence of random variables implies weak convergence of their laws. Since δ_p has the topology of convergence in probability as topological coreflection and δ_w has the weak topology as topological coreflection it is natural to ask what becomes of the above property in the new setting.

Proposition 4.9. *The function*

$$L : \big(\mathcal{R}(S), \delta_p\big) \longrightarrow \big(\mathcal{P}(S), \delta_w\big) : X \longrightarrow P_X$$

is a contraction, and consequently for any sequence of random variables $(X_n)_n$ and any random variable X, we have

$$\lambda_w\big(\big(L(X_n)\big)_n\big)\big(L(X)\big) \leqslant \lambda_p\big((X_n)_n\big)(X).$$

Proof. Let $X \in \mathcal{R}(S)$ and $\mathcal{A} \subset \mathcal{R}(S)$. We use the following expressions

$$\delta_w\big(L(X), L(\mathcal{A})\big) = \sup_{\mathcal{H}} \inf_{Y \in \mathcal{A}} d_3^{\mathcal{H}}\big(L(X), L(Y)\big)$$

and

$$\delta_p(X, \mathcal{A}) = \sup_{a>0} \inf_{Y \in \mathcal{A}} \varphi^a(X, Y)$$

where $\mathcal{H}$ ranges over finite sets of uniformly continuous maps with range $[0, 1]$. Let $\mathcal{H}$ be such a set and let $\varepsilon > 0$ be fixed. For all $f \in \mathcal{H}$, choose $\theta_f > 0$ such that, for all $x, y \in S$, $d(x, y) \leqslant \theta_f \Rightarrow |f(x) - f(y)| \leqslant \varepsilon$ and put $\theta := \min_{f \in \mathcal{H}} \theta_f$. Then it will suffice to prove that, for all $Y \in \mathcal{A}$, we have $d_3^{\mathcal{H}}(L(X), L(Y)) \leqslant \varphi^\theta(X, Y) + \varepsilon$. This follows from

$$\begin{aligned} d_3^{\mathcal{H}}\big(L(X), L(Y)\big) &= \sup_{f \in \mathcal{H}} \left| \int f \circ X \, dP - \int f \circ Y \, dP \right| \\ &\leqslant \sup_{f \in \mathcal{H}} \left(\int_{\{d(X,Y)<\theta\}} |f \circ X - f \circ Y| \, dP + \int_{\{d(X,Y) \geqslant \theta\}} |f \circ X - f \circ Y| \, dP \right) \\ &\leqslant \sup_{f \in \mathcal{H}} \big(\varepsilon + P\big(\big\{\omega \mid d\big(X(\omega), Y(\omega)\big) \geqslant \theta\big\}\big)\big) \\ &= \varepsilon + \varphi^\theta(X, Y). \qquad \square \end{aligned}$$

Corollary 4.10. *([4]) If a sequence of random variables $(X_n)_n$ converges in probability to a random variable X, then it also converges in law to X.*

Corollary 4.11. *If a sequence of random variables $(X_n)_n$ converges to a random variable X in the topology generated by the indicator metric, then their laws converge to the law of X in the total variation metric.*

A converse to 4.10 also holds in case the limit random variable is constant. This result too has the appropriate generalization.

Proposition 4.12. *If $x \in S$ and $(X_n)_n$ is a sequence of random variables on S, then we have that $\lambda_p((X_n)_n)(x) \leqslant \lambda_w((L(X_n))_n)(P_x)$.*

Proof. Suppose that $0 < a < \lambda_p((X_n)_n)(x)$; then it follows from 4.4 that we can find $b > 0$ such that

$$\forall n \exists m \geqslant n: \quad P\big(\big\{d(X_m, x) \geqslant b\big\}\big) \geqslant a.$$

Now define

$$f : M \longrightarrow [0, 1] : y \longrightarrow \frac{d(x, y)}{b} \wedge 1.$$

Then f is a continuous map on S and

$$\begin{aligned} \left| \int f \, dP_{X_m} - \int f \, dP_x \right| &= \left| \int f \, dP_{X_m} \right| \\ &\geqslant \left| \int_{\{d(X_m, x) \geqslant b\}} f \circ X_m \, dP \right| \\ &\geqslant a, \end{aligned}$$

and consequently $a \leqslant \lambda_w(((L(X_n))_n))(P_x)$, which proves our claim. $\square$

Corollary 4.13. *([4]) If $x \in S$ and $(X_n)_n$ is a sequence of random variables which converges in law to P_x, then it also converges in probability to the random variable with constant value x.*

The combined results of 3.7, 4.5 and 4.9 show that if $f : S \to T$ is a contraction then the following is a commutative diagram of contractions for the weak and c.i.p. approach structures. As an immediate consequence, applying the topological and metric coreflections, the diagram is also a commutative diagram of continuous maps for the weak topologies and the topologies of convergence in probability and a commutative diagram of contractions for the total variation metrics and the indicator metrics.

$$\begin{array}{ccc} \mathcal{R}(S) & \xrightarrow{\tilde{f}} & \mathcal{R}(T) \\ \downarrow L & & \downarrow L \\ \mathcal{P}(S) & \xrightarrow[\hat{f}]{} & \mathcal{P}(T) \end{array}$$

5. A version of Prokhorov's theorem

In analogy with the index of compactness (see [9,10,2] and 2.5), for a set $\Gamma \subset \mathcal{P}(S)$ we define its *index of relative sequential compactness* (w.r.t. δ_w) as the number

$$c_{\delta_w}(\Gamma) := \sup_{(P_n)_n} \inf_{(P_{k_n})_n} \inf_{P \in \mathcal{P}(S)} \lambda_{\delta_w}(P_{k_n})(P)$$

the supremum being taken over all sequences $(P_n)_n$ in Γ and the first infimum over all subsequences $(P_{k_n})_n$. We have the following important theorem.

Theorem 5.1. *A collection of probability measures $\Gamma \subset \mathcal{P}(S)$ is weakly relatively sequentially compact if and only if $c_{\delta_w}(\Gamma) = 0$.*

Proof. The 'only if'-part follows immediately from the definition of $c_{\delta_w}(\Gamma)$. The 'if'-part however requires a more technical argument and, maybe surprisingly, involves the fact that δ_w (or equivalently d_{TV}) is complete. Let Γ be a collection of probability measures on S such that $c_{\delta_w}(\Gamma) = 0$ and consider a sequence $(P_n)_n$ in Γ. Choose a subsequence $(P_{k_2(n)})_n$ and a probability measure $Q_2 \in \mathcal{P}(S)$ such that

$$\lambda_{\delta_w}(P_{k_2(n)})(Q_2) \leqslant 1/2.$$

Now choose a further subsequence $(P_{k_2 \circ k_3(n)})_n$ and a probability measure $Q_3 \in \mathcal{P}(S)$ such that

$$\lambda_{\delta_w}(P_{k_2 \circ k_3(n)})(Q_3) \leqslant 1/3.$$

We may continue this procedure ending up for each $m \geqslant 2$ with a sequence $(P_{k_2 \circ \cdots \circ k_m(n)})_n$ and a probability measure $Q_m \in \mathcal{P}(S)$ such that

$$\lambda_{\delta_w}(P_{k_2 \circ \cdots \circ k_m(n)})(Q_m) \leqslant 1/m. \tag{15}$$

For simplicity in notation, for any m we will let $k^m := k_2 \circ \cdots \circ k_m$ so that $(P_{k^m(n)})_n$ stands for the mth consecutive subsequence of $(P_n)_n$. We claim that $(Q_m)_m$ is a d_{TV}-Cauchy sequence. Indeed, fix $q > p \geqslant 2$, then for each n and $\alpha > 0$ we have the estimate

$$\sup_{A \in \mathcal{B}_S} \bigl(Q_p(A) - Q_q\bigl(A^{(2\alpha)}\bigr)\bigr) \leqslant \sup_{A \in \mathcal{B}_S} \bigl(Q_p(A) - P_{k^q(n)}\bigl(A^{(\alpha)}\bigr)\bigr) + \sup_{A \in \mathcal{B}_S} \bigl(Q_q(A) - P_{k^q(n)}\bigl(A^{(\alpha)}\bigr)\bigr)$$

entailing that

$$\begin{aligned} &\sup_{A \in \mathcal{B}_S} \bigl(Q_p(A) - Q_q\bigl(A^{(2\alpha)}\bigr)\bigr) \\ &\quad \leqslant \limsup_n \sup_{A \in \mathcal{B}_S} \bigl(Q_p(A) - P_{k^q(n)}\bigl(A^{(\alpha)}\bigr)\bigr) + \limsup_n \sup_{A \in \mathcal{B}_S} \bigl(Q_q(A) - P_{k^q(n)}\bigl(A^{(\alpha)}\bigr)\bigr) \\ &\quad \leqslant \limsup_n \sup_{A \in \mathcal{B}_S} \bigl(Q_p(A) - P_{k^p(n)}\bigl(A^{(\alpha)}\bigr)\bigr) + \limsup_n \sup_{A \in \mathcal{B}_S} \bigl(Q_q(A) - P_{k^q(n)}\bigl(A^{(\alpha)}\bigr)\bigr) \end{aligned}$$

which, applying inequality (15) twice, finally leads to

$$\begin{aligned} d_{TV}(Q_p, Q_q) &= \sup_{\alpha > 0} \sup_{A \in \mathcal{B}_S} \bigl(Q_p(A) - Q_q\bigl(A^{(\alpha)}\bigr)\bigr) \\ &\leqslant \lambda_{\delta_w}(P_{k^p(n)})(Q_p) + \lambda_{\delta_w}(P_{k^q(n)})(Q_q) \\ &\leqslant 1/p + 1/q. \end{aligned}$$

Hence, $(Q_m)_m$ is a d_{TV}-Cauchy sequence. The completeness of d_{TV} allows us to conclude that $(Q_m)_m$ must converge in total variation to a probability measure P. Now consider the diagonal sequence $(P'_n := P_{k^n(n)})_n$, which is a subsequence of $(P_n)_n$. We will prove that $(P'_n)_n$ converges weakly to P, demonstrating the fact that Γ is weakly relatively sequentially compact. Fix $\epsilon > 0$ and choose m such that $d_{TV}(P, Q_m) < \epsilon/2$ and $1/m < \epsilon/2$. Then for each n and $\alpha > 0$ we have the estimate

$$\sup_{A\in\mathcal{B}_S}\big(P(A) - P'_n\big(A^{(2\alpha)}\big)\big) \leqslant \sup_{A\in\mathcal{B}_S}\big(P(A) - Q_m\big(A^{(\alpha)}\big)\big) + \sup_{A\in\mathcal{B}_S}\big(Q_m(A) - P'_n\big(A^{(\alpha)}\big)\big)$$

which, applying inequality (15), entails that

$$\begin{aligned}
&\limsup_n \sup_{A\in\mathcal{B}_S}\big(P(A) - P'_n\big(A^{(2\alpha)}\big)\big)\\
&\quad\leqslant \sup_{A\in\mathcal{B}_S}\big(P(A) - Q_m\big(A^{(\alpha)}\big)\big) + \limsup_n \sup_{A\in\mathcal{B}_S}\big(Q_m(A) - P'_n\big(A^{(\alpha)}\big)\big)\\
&\quad\leqslant \sup_{A\in\mathcal{B}_S}\big(P(A) - Q_m\big(A^{(\alpha)}\big)\big) + \limsup_n \sup_{A\in\mathcal{B}_S}\big(Q_m(A) - P_{k^m(n)}\big(A^{(\alpha)}\big)\big)\\
&\quad\leqslant \epsilon/2 + 1/m < \epsilon.
\end{aligned}$$

Since the foregoing inequality holds for all $\epsilon > 0$ and $\alpha > 0$, we see that $\lambda_{\delta_w}(P'_n)(P) = 0$, which implies the weak convergence of $(P'_n)_n$ to P. □

We recall that a collection Γ of probability measures on S is said to be *tight* iff for every $\epsilon > 0$ there exists a compact set $K \subset S$ such that for all $P \in \Gamma$ we have $P(S \setminus K) < \epsilon$. We generalize this notion in two ways. For a collection $\Gamma \subset \mathcal{P}(S)$ we define its *weak index of tightness* as the number

$$t_w(\Gamma) := \sup_{\mathcal{G}} \inf_{\mathcal{G}_0} \sup_{P\in\Gamma} P\Big(X \Big\backslash \bigcup \mathcal{G}_0\Big)$$

where $\mathcal{G}$ ranges over all open covers of S and $\mathcal{G}_0$ over all finite subcollections of $\mathcal{G}$.

Theorem 5.2. *For a metric d metrizing S and $\Gamma \subset \mathcal{P}(S)$ we have*

$$t_w(\Gamma) = \sup_{\delta_x} \inf_{Y} \sup_{P\in\Gamma} P\Big(S \Big\backslash \bigcup_{x\in Y} B_d(x, \delta_x)\Big),$$

the first supremum ranging over all choices $\delta_x > 0$, $x \in S$, and the infimum over all finite sets Y in S.

Proof. Let us denote the right hand side by $b(\Gamma)$.

$t_w(\Gamma) \leqslant b(\Gamma)$: Fix $\epsilon > 0$ and an open cover $\mathcal{G}$ of S and assume that $\mathcal{G}$ consists of countably many G_n increasing to S. For each $x \in S$ we let n_x be the smallest natural number for which $x \in G_{n_x}$ and we choose $\delta_x > 0$ so small that $B_d(x, \delta_x) \subset G_{n_x}$. Now pick a finite set Y in S so that $P(S \setminus \bigcup_{x\in Y} B(x, \delta_x)) \leqslant b(\Gamma) + \epsilon$ for all $P \in \Gamma$. Observe that since Y is finite, it must be contained in a set G_{n_0} belonging to $\mathcal{G}$. Furthermore, for each $x \in Y$ we have $B(x, \delta_x) \subset G_{n_x} \subset G_{n_0}$, by construction of n_x. It follows that $P(S \setminus G_{n_0}) \leqslant P(S \setminus \bigcup_{x\in Y} B(x, \delta_x)) \leqslant b(\Gamma) + \epsilon$ for all $P \in \Gamma$. Hence we infer that $t_w(\Gamma) \leqslant b(\Gamma)$.

$b(\Gamma) \leqslant t_w(\Gamma)$: Fix $\epsilon > 0$, $\delta_x > 0$ for all $x \in S$ and let $\mathcal{G}$ be the open cover consisting of all balls $B_d(x, \delta_x)$. Since we can pick finitely many x_i such that $P(S \setminus \bigcup_i B_d(x_i, \delta_{x_i})) \leqslant t_w(\Gamma) + \epsilon$, it easily follows that $b(\Gamma) \leqslant t_w(\Gamma)$. □

We recall the definition of an Atsuji space. A metric space is called an Atsuji space if any pair of nonempty disjoint closed subsets lie at a strictly positive distance from each other or equivalently if any open cover has a Lebesgue number [1,13] which is why they are also called Lebesgue spaces. Typical extreme examples are a compact space on the one hand and a discrete space, e.g. $\mathbb{N}$, on the other.

Proposition 5.3. *If (S, d) is (moreover) an Atsuji space then it is possible to replace the choice of radii $(\delta_x)_x$ in the foregoing result by a fixed choice for all points, i.e. then*

$$t_w(\Gamma) = \sup_{\delta>0} \inf_{Y} \sup_{P\in\Gamma} P\Big(S \Big\backslash \bigcup_{x\in Y} B_d(x, \delta)\Big)$$

the infimum ranging over all finite sets Y in S.

Proof. This is an immediate consequence of the fact that in an Atsuji space every open cover has a Lebesgue number. □

We define the *strong index of tightness* of Γ as the number

$$t_s(\Gamma) := \inf_K \sup_{P \in \Gamma} P(S \setminus K)$$

the infimum being taken over all compact sets $K \subset S$. Observe that the inequality $t_w(\Gamma) \leqslant t_s(\Gamma)$ always holds true. The following theorem illustrates the fact that both indices indeed generalize the classical notion of tightness.

Theorem 5.4. *A collection Γ of probability measures on a complete separable metric space S is tight if and only if $t_w(\Gamma) = 0$ if and only if $t_s(\Gamma) = 0$.*

Proof. We restrict the proof to the only non-trivial assertion, namely that $t_w(\Gamma) = 0$ implies tightness of Γ. Fix $\epsilon > 0$. Choose a countable dense subset $\{x_i \mid i \in \mathbb{N}\}$ then for any $m \geqslant 1$ the family of balls $(B(x_i, 1/m))_i$ is an open cover and thus there exists a finite subset $(B(x_i, 1/m))_{i=0,\dots,n_m}$ such that

$$\forall P \in \Gamma: \quad P\left(X \Big\backslash \bigcup_{i=0}^{n_m} B(x_i, 1/m)\right) \leqslant \frac{\epsilon}{2^n}.$$

Put

$$K := \bigcap_{m=1}^{\infty} \bigcup_{i=0}^{n_m} \overline{B(x_i, 1/m)}$$

then K is compact and for all $P \in \Gamma$, $P(X \setminus K) \leqslant \epsilon$. □

That the indices of compactness and tightness also produce meaningful non-zero values is shown by the following simple example.

Example 5.5. Consider the real line with the usual Borel σ-algebra, fix $\alpha > 0$ and let Γ be the set of all probability measures

$$P_n := (1 - \alpha)\delta_0 + \alpha\delta_n$$

where δ_x stands for the Dirac measure at x and where n is any natural number $\geqslant 1$. Then it is easily verified that both the weak and strong index of tightness and the index of relative sequential compactness are equal to α.

Theorem 5.7 will provide us with important inequalities generalizing Prokhorov's theorem. For its proof some preparation is required.

For a collection Γ of probability measures on a separable metric space S and $\epsilon > 0$ we will consider the set

$$\Gamma(\epsilon) := \left\{\left(1 - \epsilon'\right)P + \epsilon' Q \;\middle|\; P \in \Gamma,\ Q \in \mathcal{P}(X),\ 0 \leqslant \epsilon' \leqslant \epsilon\right\}, \tag{16}$$

see e.g. [12] for the use of these types of "contaminated" sets in robust statistics. The following lemma furnishes an estimate for the index of relative sequential compactness of such sets.

Lemma 5.6. *For a set $\Gamma \subset \mathcal{P}(S)$ and $\epsilon > 0$ we have the estimate*

$$c_{\delta_w}\left(\Gamma(\epsilon)\right) \leqslant c_{\delta_w}(\Gamma) + \epsilon.$$

Proof. Take $\Gamma \subset \mathcal{P}(S)$ and $\epsilon > 0$. Then for a sequence $(R_n)_n$ where

$$R_n := (1 - \epsilon_n)P_n + \epsilon_n Q_n$$

in $\Gamma(\epsilon)$ and $\delta > 0$ we find a subsequence (P_{k_n}) of (P_n) and probability measure P such that

$$\sup_{\alpha > 0} \limsup_n \sup_{A \in \mathcal{B}_S} \left(P(A) - P_{k_n}\left(A^{(\alpha)}\right)\right) \leqslant c_{\delta_w}(\Gamma) + \delta.$$

Now the inequality

$$\begin{aligned}
&\sup_{\alpha > 0} \limsup_n \sup_{A \in \mathcal{B}_S} \left(P(A) - R_{k_n}\left(A^{(\alpha)}\right)\right) \\
&\quad = \sup_{\alpha > 0} \limsup_n \sup_{A \in \mathcal{B}_S} \left(P(A) - P_{k_n}\left(A^{(\alpha)}\right) + \epsilon_{k_n}(P_{k_n} - Q_{k_n})\left(A^{(\alpha)}\right)\right) \\
&\quad \leqslant \sup_{\alpha > 0} \limsup_n \sup_{A \in \mathcal{B}_S} \left(P(A) - P_{k_n}\left(A^{(\alpha)}\right)\right) + \epsilon \\
&\quad \leqslant \left(c_{\delta_w}(\Gamma) + \delta\right) + \epsilon
\end{aligned}$$

establishes the desired result. □

The reason for introducing both a weak and strong index of tightness will become clear in our general form of a Prokhorov theorem for distances, as they turn out to provide respectively a lower and an upper bound for the index of weak compactness, which in consequence allows us to derive some further interesting results.

Theorem 5.7 *(Prokhorov for distances). For every collection Γ of probability measures on a complete separable metric space S the following inequalities are valid*

$$t_w(\Gamma) \leqslant c_{\delta_w}(\Gamma) \leqslant t_s(\Gamma).$$

Proof. $t_w(\Gamma) \leqslant c_{\delta_w}(\Gamma)$: Suppose that $c_{\delta_w}(\Gamma) < \gamma$ and choose $\epsilon > 0$ such that $c_{\delta_w}(\Gamma) < \gamma - \epsilon$. Take a countable open cover $\mathcal{G} := \{G_n \mid n \in \mathbb{N}\}$ and suppose that for all $n \in \mathbb{N}$ there exists $P_n \in \Gamma$ such that

$$P_n\left(\bigcup_{i=0}^{n} G_i\right) < 1 - \gamma.$$

Since $c_{\delta_w}(\Gamma) < \gamma - \epsilon$ there exists a subsequence $(P_{k_n})_n$ and a $P \in \mathcal{P}(X)$ such that

$$\lambda_{\delta_w}(P_{k_n} \to P) < \gamma - \epsilon.$$

This implies that for all n

$$\begin{aligned} P\left(\bigcup_{i=0}^{n} G_i\right) &\leqslant \sup_m \inf_{l \geqslant m} P_{k_l}\left(\bigcup_{i=0}^{n} G_i\right) + \gamma - \epsilon \\ &\leqslant \sup_{m, k_m \geqslant n} \inf_{l \geqslant m} P_{k_l}\left(\bigcup_{i=0}^{k_l} G_i\right) + \gamma - \epsilon \\ &\leqslant 1 - \gamma + \gamma - \epsilon = 1 - \epsilon. \end{aligned}$$

However, since $\bigcup_{i=0}^{n} G_i \uparrow X$ this is impossible. Hence there exists a finite subset $\mathcal{G}_0 \subset \mathcal{G}$ such that for all $P \in \Gamma$ we have $P(X \setminus \bigcup \mathcal{G}_0) \leqslant \gamma$, and thus $t_w(\Gamma) \leqslant \gamma$.

$c_{\delta_w}(\Gamma) \leqslant t_s(\Gamma)$: Fix $\epsilon > 0$. Now we are allowed to pick a compact set $K \subset S$ such that the inequality $P(S \setminus K) \leqslant t_s(\Gamma) + \epsilon$ is valid for every probability measure $P \in \Gamma$. If we put $\Gamma(\cdot \mid K) := \{P(\cdot \mid K) \mid P \in \Gamma\}$, then the relation $P = P(K)P(\cdot \mid K) + P(S \setminus K)P(\cdot \mid S \setminus K)$ shows that $\Gamma \subset \Gamma(\cdot \mid K)(t_s(\Gamma) + \epsilon)$ (see 16). Now note that if there exists a compact set $K \subset S$ containing the support of every probability measure P in a set $\Lambda \subset \mathcal{P}$ (S then Λ) is weakly relatively sequentially compact. (Indeed, since the Daniell–Stone theorem allows us to identify the space of probability measures on S whose support is contained in K with the space of positive linear functionals λ on $\mathcal{C}(K)$ for which $\lambda(1) = 1$, provided with the weak* topology, it suffices to observe that the latter is a closed subspace of the closed dual unit ball of $\mathcal{C}(K)$, and that this ball is compact and metrizable due to the Banach–Alaoglu theorem, see e.g. [18].) Applying this, Lemma 5.6 and Theorem 5.1, we conclude that

$$\begin{aligned} c_{\delta_w}(\Gamma) &\leqslant c_{\delta_w}\big(\Gamma(\cdot \mid K)\big(t_s(\Gamma) + \epsilon\big)\big) \\ &\leqslant c_{\delta_w}\big(\Gamma(\cdot \mid K)\big) + t_s(\Gamma) + \epsilon \\ &= t_s(\Gamma) + \epsilon \end{aligned}$$

whence the desired inequality. □

Corollary 5.8. *([4,14]) Let Γ be a collection of probability measures on a complete separable metric space S. Then Γ is weakly relatively sequentially compact if and only if it is tight.*

Proof. Let Γ be weakly relatively compact, then by Theorem 5.1 $c_{\delta_w}(\Gamma) = 0$, and by Theorem 5.7 $t_w(\Gamma) = 0$. Now Theorem 5.4 allows us to conclude that Γ is tight.

Conversely, let Γ be tight. Then again by Theorem 5.4 we see that $t_s(\Gamma) = 0$, and by Theorem 5.7 $c_{\delta_w}(\Gamma) = 0$. Now Theorem 5.1 implies that Γ is weakly relatively compact. □

Although at present the precise situation with regard to the weak and strong indices of tightness is not yet completely understood we do have the following results.

Theorem 5.9. *Assume that there exists a sequence $(U_n)_n$ of relatively compact open sets which increases to S. Then for $\Gamma \subset \mathcal{P}(S)$ we have*

$$t_w(\Gamma) = c_{\delta_w}(\Gamma) = t_s(\Gamma).$$

Proof. It suffices to show that in this case $t_s(\Gamma) \leqslant t_w(\Gamma)$. Let $\epsilon > 0$. Now it is possible to find a U_n such that $\sup_{P \in \Gamma} P(S \setminus U_n) \leqslant t_w(\Gamma) + \epsilon$. Let K be the compact set $\overline{U_n}$ and observe that, since $U_n \subset K$, we have $\sup_{P \in \Gamma} P(S \setminus K) \leqslant \sup_{P \in \Gamma} P(S \setminus U_n) \leqslant t_w(\Gamma) + \epsilon$. We conclude that $t_s(\Gamma) \leqslant t_w(\Gamma)$. $\square$

Theorem 5.9 has the following obvious corollary for Euclidean spaces.

Corollary 5.10. *For $\Gamma \subset \mathcal{P}(\mathbb{R}^d)$ we have $t_w(\Gamma) = c_{\delta_w}(\Gamma) = t_s(\Gamma)$.*

References

[1] M. Atsuji, Uniform continuity of continuous functions of metric spaces, Pacific J. Math. 8 (1958) 11–16.
[2] R. Baekeland, R. Lowen, Measures of compactness in approach spaces, Comment. Math. Univ. Carolin. 36 (2) (1995) 327–345.
[3] J. Banaś, K. Goebel, Measures of Noncompactness in Banach Spaces, Lecture Notes in Pure and Applied Mathematics, vol. 60, Marcel Dekker, New York, 1980.
[4] P. Billingsley, Convergence of Probability Measures, John Wiley and Sons, 1968.
[5] J. Dugundji, Topology, Allyn and Bacon Series in Advanced Mathematics, Allyn and Bacon, Inc., Boston, MA–London–Sydney, 1978 (reprinting of the 1966 original).
[6] R.M. Dudley, Distances of probability measures and random variables, Ann. Math. Statist. 39 (5) (1968) 1563–1572.
[7] In-Sook Kim, Martin Väth, Some remarks on measures of noncompactness and retractions onto spheres, Topology Appl. 154 (2007) 3056–3069.
[8] S.D. Jacka, G.O. Roberts, On strong forms of weak convergence, Stochastic Process. Appl. 67 (1997) 41–53.
[9] R. Lowen, Approach Spaces: The Missing Link in the Topology–Uniformity–Metric Triad, Oxford Mathematical Monographs, Oxford University Press, 1997.
[10] R. Lowen, Kuratowski's measure of non-compactness revisited, Quart. J. Math. Oxford (2) 39 (1988) 235–254.
[11] L. Meziani, Tightness of probability measures on function spaces, J. Math. Anal. Appl. 354 (2009) 202–206.
[12] S. Morgenthaler, A survey of robust statistics, Stat. Methods Appl. 15 (2007) 271–293.
[13] S. Nadler, T. West, A note on Lebesgue spaces, Topology Proc. 6 (1981) 363–369.
[14] K.R. Parthasarathy, Probability Measures on Metric Spaces, Academic Press, 1967.
[15] K.R. Parthasarathy, T. Steerneman, A tool in establishing total variation convergence, Proc. Amer. Math. Soc. 94 (1985) 626–630.
[16] D. Pollard, Convergence of Stochastic Processes, Springer-Verlag, 1984.
[17] Y.V. Prokhorov, Convergence of random processes and limit theorems in probability theory, Theory Probab. Appl. 1 (1956) 157–214.
[18] W. Rudin, Real and Complex Analysis, McGraw–Hill Series in Higher Mathematics, McGraw–Hill, 1966.
[19] F. Topsøe, Topology and Measure, Lecture Notes in Mathematics, vol. 133, Springer-Verlag, 1970.
[20] V.S. Varadarajan, Convergence of stochastic processes, Bull. Amer. Math. Soc. 67 (1961) 276–280.
[21] W. Whitt, Convergence of probability measures on the function space $\mathcal{C}[0, \infty)$, Ann. Math. Statist. 41 (1970) 939–944.
[22] A. Wiśnicki, J. Wośko, On relative Hausdorff measures of noncompactness and relative Chebyshev radii in Banach spaces, Proc. Amer. Math. Soc. 124 (1996) 2465–2474.
[23] V.M. Zolotarev, Probability metrics, Teor. Veroyatn. Primen. 28 (2) (1983) 264–287 (in Russian).