Milan Merkle

TOPICS IN WEAK CONVERGENCE OF PROBABILITY MEASURES

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

CONTENTS

1. Introduction

Let $\mathcal{X}$ be a topological space and let P_n $(n = 1, 2, \dots)$ and P be probability measures defined on the Borel sigma field generated by open subsets of $\mathcal{X}$. We say that the sequence $\{P_n\}$ converges *weakly* to P, in notation $P_n \implies P$ if

$$\lim_{n \to +\infty} \int_{\mathcal{X}} f(x)\, dP_n(x) = \int_{\mathcal{X}} f(x)\, dP(x) \tag{1}$$

for every continuous and bounded real valued function $f : \mathcal{X} \mapsto \boldsymbol{R}$. In terms of random variables, let X_n $(n = 1, 2, \dots)$ and X be $\mathcal{X}$-valued random variables defined on a common probability space and let P_n and P be corresponding distributions, that is, $P_n(B) = \mathrm{Prob}(X_n \in B)$, where B is a Borel set in $\mathcal{X}$. Then we say that the sequence X_n converges weakly to X and write $X_n \implies X$ if and only if $P_n \implies P$.

As we shall see in Section 4, there are many stronger convergence concepts than the introduced one. However, the weak convergence is a very powerful tool in Probability Theory, partly due to its comparative simplicity and partly due to its natural behavior in some typical problems. The weak convergence appears in Probability chiefly in the following classes of problems.

- Knowing that $P_n \implies P$ we may replace P_n by P for n large enough. A typical example is the Central Limit Theorem (any of its versions), which enables us to conclude that the properly normalized sum of random variables has approximately a unit Gaussian law.
- Conversely, if $P_n \implies P$ then we may approximate P with P_n, for n large enough. A typical example of this sort is the approximation of Dirac's delta function (understood as a density of a point mass at zero) by, say triangle-shaped functions.
- In some problems, like stochastic approximation procedures, we would like to have a strong convergence result $X_n \to X$. However, the conditions required to prove the strong convergence are usually very complex and the proofs are difficult and very involved. Then, one usually replaces the strong convergence with some weaker forms; one is often satisfies with $X_n \implies X$.
- It is not always easy to construct a measure with specified properties. If we need to show just its existence, sometimes we are able to construct a sequence (or a net) of measures which can be proved to be weakly convergent and that its limit satisfy the desired properties. For example, this procedure is usually applied to show the existence of the Wiener measure.

The concept of weak convergence is so well established in Probability Theory that hardly any textbook even mention its topological heritage. It, indeed, is not too important in many applications, but a complete grasp of the definition of the weak convergence is not possible without understanding its rationale. The first part of this paper (Sections 2 and 3) is an introduction to weak convergence of probability measures from the topological point of view. Since the set of probability measures is not closed under weak convergence (as we shall see, the limit of a net of probability measures need not be a probability measure), for a full understanding of the complete concept, one has to investigate a wider structure, which turns out to be the set of all finitely additive Radon measures. In this context we present results concerning the Baire field and sigma field, which are usually omitted when discussing probability measures. In Section 4 we consider weak convergence of probability measures and present classical results regarding metrics of weak convergence. In Section 5 we show that the set of probability measures is not closed and effectively show the existence of a finitely, but not countably additive measure in the closure of the set of probability measures. Section 6 deals with the famous Prohorov's theorem on metric spaces. In Section 7 we consider weak convergence of probability measures on Hilbert spaces. Here we observe a separable Hilbert space equipped with weak and strong topology and in both cases we give necessary and sufficient conditions for relative compactness of a set of probability measures.

2. Weak convergence in topology

2.1. Topology induced by a subset of algebraic dual. Let $\mathcal{X}$ be a vector space over a field F, where F stands for $\boldsymbol{R}$ or $\boldsymbol{C}$. Let $\mathcal{X}'$ be the set of all linear maps $\mathcal{X} \mapsto F$ (so called algebraic dual space). Let $Y \subset \mathcal{X}'$ be a subspace such that Y separates points in $\mathcal{X}$, i.e., if $\varphi(x) = \varphi(y)$ for all $\varphi \in Y$ then $x = y$. Define Y-topology on $\mathcal{X}$ by the sub-base

$$\{\varphi^{-1}(V) \mid \varphi \in Y,\ \ V \text{ open set in } F\}$$

The base of Y-topology is obtained by taking finite intersections of sub-base elements. Equivalently, a base at zero for the Y-topology is consisted of sets

$$O_{\varphi_1,\dots,\varphi_n} = \{x \in \mathcal{X} \mid \varphi_j(x) < 1 \ \text{ for } j = 1, \dots, n\},$$

where $\{\varphi_1, \dots, \varphi_n\}$ is an arbitrary finite set of elements in Y.

This topology is a Hausdorff one, since we assumed that Y separates point of $\mathcal{X}$. That is, if $x \neq y$ are points in $\mathcal{X}$, then there is a $\varphi \in Y$ so that $\varphi(x) \neq \varphi(y)$ and consequently there are disjoint open sets V_x and V_y in F so that $\varphi(x) \in V_x$ and $\varphi(y) \in V_y$, hence $\varphi^{-1}(V_x) \cap \varphi^{-1}(V_y) = \emptyset$.

The convergence in Y-topology may be described as

$$x_d \to x \iff \varphi(x_d) \to \varphi(x) \quad \text{for all } \varphi \in Y,$$

where $\{d\}$ is a directed set. It is important to know that Y-topology may not be metrizable, even in some simple cases, as we shall see later. So, sequences must not be used as a replacement for nets.

If $Y_1 \subset Y_2 \subset \mathcal{X}'$, then the Y_1 topology is obviously weaker (contains no more open sets) than the Y_2 topology. Therefore, if $x_d \to x$ in Y_2-topology, then it also converges in Y_1 topology, and the converse is not generally true.

2.2. Weak topology. Now we observe only locally convex Hausdorff (LC) topological vector spaces (TVS) $\mathcal{X}$, i.e., those that have a basis for the topology consisted of convex sets. Let $\mathcal{X}^*$ be the topological dual of $\mathcal{X}$, i.e., the space of all continuous linear functionals $\mathcal{X} \mapsto F$. By one version of the Hahn-Banach theorem, $\mathcal{X}^*$ separates points in $\mathcal{X}$, if $\mathcal{X}$ is a LC TVS. Then $\mathcal{X}^*$-topology on $\mathcal{X}$ is called the weak topology. Since for every $\varphi \in \mathcal{X}^*$ we have that

$$x_d \to x \text{ in the original topology of } \mathcal{X} \implies \varphi(x_d) \to x,$$

we see that the weak topology is weaker than the original (strong) topology of $\mathcal{X}$.

The space $\mathcal{X}$ equipped with the weak topology will be denoted by $\mathcal{X}_w$.

2.3. Example. Let $\mathcal{X}$ be a real separable infinitely dimensional Hilbert space, with the inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \|$. Then x_n converges weakly to x if and only if $\langle y, x_n \rangle \to \langle y, x \rangle$ for any $y \in \mathcal{X}$. Let $x_n = e_n$ be an orthonormal base for $\mathcal{X}$. Then since $\|y\|^2 = \sum \langle y, e_n \rangle^2 < +\infty$, we see that $\langle y, e_n \rangle \to 0$ for any $y \in \mathcal{X}$ and so the sequence e_n converges weakly to 0. However, since $\|e_n - e_m\|^2 = 2$, this sequence does not converge in the norm topology of $\mathcal{X}$. $\square$

On finitely dimensional TVS, the weak and the strong topology coincide. However, on infinitely dimensional spaces, the weak topology exhibits some peculiar properties, as we shall see in the next subsection.

2.4. How weak is the weak topology? Let us firstly grasp some clues to understand the weak topology. We start with kernels of linear functionals and we prove the following theorem.

Theorem. *If* $\dim \mathcal{X} > 1$, *then there is no linear functional* $\varphi \in \mathcal{X}'$ *with* $\ker \varphi = \{0\}$.

Proof. Suppose that $\ker \varphi = \{0\}$ and let x_1, x_2 be arbitrary elements in $\mathcal{X}$, $x_1, x_2 \neq 0$. Then let $\lambda = \varphi(x_1)/\varphi(x_2)$, which is well defined, since $\varphi(x_2) \neq 0$ by assumptions. Let $y = x_1 - \lambda x_2$. Then $\varphi(y) = \varphi(x_1) - \lambda \varphi(x_2) = 0$, hence $y = 0$, i.e., $x_1 = \lambda x_2$. Since x_1, x_2 are arbitrary, the dimension of $\mathcal{X}$ is 1. $\square$

Let $\operatorname{sp} A$ denote the set of all finite linear combinations of elements of the set A.

2.5. Theorem. *Let* $\mathcal{X}$ *be a vector space over* F *and let* $\varphi_1, \ldots, \varphi_n \in \mathcal{X}'$. *Then (i) and (ii) below are equivalent:*

(i) $\varphi \in \operatorname{sp}\{\varphi_1, \ldots, \varphi_n\}$ *(ii)* $\bigcap_{i=1}^n \ker \varphi_i \subset \ker \varphi$

Proof. Suppose that (i) holds, that is, $\varphi(x) = \sum_{i=1}^n \alpha_i \varphi_i(x)$. Then clearly, $\varphi_i(x) = 0$ for all i implies that $\varphi(x) = 0$, which proves (ii). Conversely, assume that (ii) holds. Define a mapping $T : \mathcal{X} \mapsto F^n$ by $T(x) = (\varphi_1(x), \ldots, \varphi_n(x))$ and define $S(T(x)) = \varphi(x)$. Then S is well defined on the range of T, since if $T(x) = T(y)$ then $x - y$ is in $\ker \varphi_i$ for all i, hence $x - y$ is in $\ker \varphi$ and so

$S(x) = S(y)$. Clearly, S is a linear map and its extension to F^n must be of the form $F(t_1, \dots, t_n) = \alpha_1 t_1 + \cdots + \alpha_n t_n$, which means that

$$\varphi(x) = S(\varphi_1(x), \dots, \varphi_n(x)) = \alpha_1\varphi_1(x) + \cdots + \alpha_n\varphi_n(x),$$

which was to be proved. □

From Theorem 2.5 we get an immediate generalization of Theorem 2.4:

2.6. Corollary. *If $\mathcal{X}$ is an infinitely dimensional TVS and if $\varphi_1, \dots, \varphi_n$ are arbitrary linear functionals, then $\bigcap_{i=1}^n \ker\varphi_i \neq \{0\}$.*

Proof. Suppose that $\bigcap_{i=1}^n \ker\varphi_i = \{0\}$. Then for any $\varphi \in \mathcal{X}'$ we have that $\{0\} \subset \ker\varphi$ and then, by Theorem 2.5, $\varphi \in \mathrm{sp}\,\{\varphi_1, \dots, \varphi_n\}$, which means that $\mathcal{X}'$ is finite dimensional, and so is $\mathcal{X}$, which contradicts the assumption. □

The next theorem describes a fundamental weakness of the weak topology.

2.7. Theorem. *If $\mathcal{X}$ is an infinitely dimensional TVS then each weakly open set contains a non-trivial subspace.*

Proof. Let $U \subset \mathcal{X}$ be a weakly open set. Without loss of generality, assume that $0 \in U$ (otherwise, do a translation). Then U must contain a set of the form

$$O_{\varphi_1,\dots,\varphi_n} = \{x \in \mathcal{X} \mid \varphi_1(x) < 1, \dots, \varphi_n(x) < 1\},$$

for some $\varphi_1, \dots, \varphi_n \in \mathcal{X}^*$. Then clearly $\bigcup_{i=1}^n \ker\varphi_i \subset O_{\varphi_1,\dots,\varphi_n} \subset U$ and according to Corollary 2.6 $\bigcup_{i=1}^n \ker\varphi_i$ is a non-trivial subspace.

2.8. Corollary. *Let $\mathcal{X}$ be an infinitely dimensional normed space. Then an open ball of $\mathcal{X}$ is not weakly open.*

Proof. Let B be an open ball in $\mathcal{X}$. If it were open in the weak topology, then (by Theorem 2.7) it would have contained a nontrivial subspace, which is not possible (for instance, it is not possible that $\|\alpha x\| < r$ for all scalars α). □

So, the next theorem may come as a surprise.

2.9. Theorem. *Let $\mathcal{X}$ be a LC TVS. Then $\mathcal{X}$ and $\mathcal{X}_w$ have the same closed convex sets. For each convex $S \subset \mathcal{X}$ we have that $\bar{S}^w = \bar{S}$, where $\bar{S}^w$ is the closure of S in the weak topology.* □

2.10. Example. Let $\mathcal{X}$ be a separable metric space. Denote by $\mathcal{B}_s$ the Borel sigma field generated by norm-open sets and let $\mathcal{B}_w$ be the Borel sigma field generated by weakly open sets. Since any weakly open set is also norm-open, we generally have that $\mathcal{B}_w \subset \mathcal{B}_s$, but not conversely. In this special case, each strongly open set is a countable union of closed balls, which are, by Theorem 2.9 also weakly closed. So, $\mathcal{B}_s \subset \mathcal{B}_w$, which gives that, in a separable metric space, $\mathcal{B}_s = \mathcal{B}_w$. □

From Theorem 2.9 it follows that a closed ball in a normed space $\mathcal{X}$ is also weakly closed. But from Theorem 2.7 we see that the weak interior of any ball in an infinitely dimensional normed space is an empty set!

2.11. Weak star topology on a dual space. We are now going to introduce a yet weaker than the weak topology. Let $\mathcal{X}$ be a LC TVS and let $\mathcal{X}^*$ be its topological dual. Define a mapping $\Phi : \mathcal{X} \mapsto (\mathcal{X}^*)'$ by $\Phi_x(\varphi) = \varphi(x)$, where $x \in \mathcal{X}$ and $\varphi \in \mathcal{X}^*$. This map is linear and one-to one (the one-to one property follows from the fact that $\mathcal{X}^*$ separates points). Now we can observe the $\Phi(\mathcal{X})$-topology on $\mathcal{X}^*$. It is customary to identify $\Phi(\mathcal{X})$ with $\mathcal{X}$ itself (especially in the case when $\mathcal{X}$ is a normed space, since then the natural topologies on $\mathcal{X}$ and $\Phi(\mathcal{X})$ coincide). So, the $\mathcal{X}$-topology on $\mathcal{X}^*$ is called the weak-$*$ (weak star) topology. In fact, this is the topology of pointwise convergence of functionals, since

$$\varphi_d \to \varphi \quad (w - *) \iff \varphi_d(x) \to \varphi(x) \quad \text{for every } x \in \mathcal{X}.$$

2.12. Three topologies on duals of normed spaces. Let $\mathcal{X}$ be a normed space. Then its topological dual $\mathcal{X}^*$ is also normed, with $\|\varphi\| = \sup_{\|x\| \leq 1} |\varphi(x)|$. This norm defines the strong topology of $\mathcal{X}^*$. Further, the weak topology on $\mathcal{X}^*$ is defined as $\mathcal{X}^{**}$-topology and the weak star is $\mathcal{X}$-topology on $\mathcal{X}^*$. Since $\mathcal{X} \subset \mathcal{X}^{**}$, the weak star topology is weaker than the weak one, which is in turn weaker than the strong topology. Due to the order between topologies, it is not possible that a sequence (or a net) converges to one limit in one of mentioned topologies and to another limit in other topology. So, for instance, if a sequence converges to some x in the, say, weak star topology, then in the strong topology it either converges to x or does not converge at all.

2.13. Example. Let c_0 be the set of all real sequences converging to zero, with the norm $\|x\| = \sup_n |x_n|$. Then it is well known that $c_0^* = l_1$ and $c_0^{**} = l_1^* = l_\infty$, where l_1 is the space of sequences with the norm $\|x\|_{l_1} = \sum |x_n| < +\infty$ and l_∞ is the space of bounded sequences with $\|x\|_\infty = \sup_n |x_n|$. Linear maps are realized via so called duality pairing $\langle x, y \rangle$, acting like inner products with one component from $\mathcal{X}$ and the other one from $\mathcal{X}^*$. Observe a sequence in l_1, $x_n = \{x_{k,n}\}$ and let $y = \{y_k\}$ be an element in l_1. Then x_n converges to y:

- Strongly, if $\|x_n - y\| = \sup_k |x_{k,n} - y_k| \to 0$ as $n \to +\infty$.
- Weakly, if $\langle \xi, x_n \rangle = \sum_k \xi_k x_{k,n} \to \sum_k \xi_k y_k$, for any $\xi = \{\xi_k\} \in l_\infty$.
- Weak-star, if $\sum_k \xi_k x_{k,n} \to \sum_k \xi_k y_k$, for any $\xi = \{\xi_k\} \in c_0$.

Now observe the sequence $e_n = (0, 0, \dots, 0, 1, 0, \dots) \in l_1$ (with 1 as the n-th component). Then $\langle \xi, e_n \rangle = \xi_n$ and if $\xi \in c_0$ then $\langle \xi, e_n \rangle \to 0$, hence e_n converges to 0 weak star. However, if $\xi \in l_\infty$, then $\langle \xi, e_n \rangle$ need not converge, so e_n does not converge in the weak topology. Further, in the norm topology e_n does not converge to zero, because $\|e_n\| = 1$ for all n; therefore, $\{e_n\}$ is not convergent in the strong topology of l_1.

2.14. Canonical injections. Let $\mathcal{X}$ be a normed space, let $\mathcal{X}^*$ be its topological dual space and let $\mathcal{X}^{**} = (\mathcal{X}^*)^*$ be its second dual. If $\|x\|$ is a norm on $\mathcal{X}$, then the norm on $\mathcal{X}^*$ is defined by $\|\varphi\| = \sup_{\|x\| \leq 1} \|\varphi(x)\|$. The norm on $\mathcal{X}^{**}$ is then defined by $\|\Phi\| = \sup_{\|\varphi\| \leq 1} |\Phi(\varphi)|$. Observe the canonical mapping $\mathcal{X} \mapsto \mathcal{X}^{**}$ which is defined, as in 2.11 by

$$\Phi_x(\varphi) = \varphi(x).$$

Then for each $x \in \mathcal{X}$, Φ_x is a continuous linear functional defined on $\mathcal{X}^*$, and so it is a member of $\mathcal{X}^{**}$ with the norm

(2)
$$\|\Phi_x\| = \sup_{\|\varphi\|\le 1} |\Phi_x(\varphi)| = \sup\Big\{\Big|\varphi\Big(\frac{x}{\|x\|}\Big)\Big| \cdot \|x\| \;\Big|\; \varphi \in \mathcal{X}^*, \sup_{\|x\|\le 1} |\varphi(x)| \le 1\Big\} \le \|x\|.$$

On the other hand, by one version of the Hahn-Banach theorem, if $\mathcal{X}$ is a normed space, for each $x \in \mathcal{X}$ there exists $\varphi_0 \in \mathcal{X}^*$ with $\|\varphi_0\| = 1$ and $\varphi_0(x) = \|x\|$. Therefore,

$$\|\Phi_x\| = \sup_{\|\varphi\|\le 1} |\Phi_x(\varphi)| \ge \Phi_x(\varphi_0) = \varphi_0(x) = \|x\|. \tag{3}$$

From (2) and (3) it follows that $\|\Phi_x\| = \|x\|$. So, the canonical mapping $x \mapsto \Phi_x$ is bicontinuous, i.e.,

$$x_n \to x \iff \Phi_{x_n} \to \Phi_x$$

Further, as we already observed in 2.11, this mapping is linear and one-to one injection from $\mathcal{X}$ to $\mathcal{X}^{**}$. Since $\Phi(\mathcal{X})$ is isomorphic and isometric to $\mathcal{X}$, it can be identified with $\mathcal{X}$ in the algebraic and topological sense. This fact is usually denoted as $\mathcal{X} \subset \mathcal{X}^{**}$. If $\Phi(\mathcal{X}) = \mathcal{X}^{**}$, we say that $\mathcal{X}$ is a reflexive space, usually denoted as $\mathcal{X} = \mathcal{X}^{**}$.

2.15. Example. Each Hilbert space is reflexive. Due to the Riesz representation theorem, any linear functional in a Hilbert space H is of the form $\varphi_y(x) = \langle y, x\rangle$, where $y \in H$ and also $\|\varphi_y\| = \|y\|$. Hence, we may identify φ_y with y and write $H^* = H$. This equality means, in fact, that there exists a canonical injection (in fact, bijection) $H \mapsto H^*$ realized by the mapping $y \mapsto \varphi_y$.

From $H^* = H$ it follows that $H^{**} = (H^*)^* = H$, i.e., H is reflexive.

The space c_0 introduced in the example 2.13 is not reflexive, since $c_0^{**} = l_\infty$. However, $c_0 \subset l_\infty$. □

2.16. Inclusions. Now suppose that $\mathcal{X}_1 \subset \mathcal{X}_2$ are vector spaces with the same norm $\|\cdot\|$. Let φ be a continuous linear functional defined on $\mathcal{X}_2$. Then clearly, the restriction of φ to $\mathcal{X}_1$ is a continuous linear functional on $\mathcal{X}_1$ and therefore we have that $\mathcal{X}_2^* \subset \mathcal{X}_1^*$. For the second duals we similarly find that $\mathcal{X}_1^{**} \subset \mathcal{X}_2^{**}$. Hence,

$$\mathcal{X}_1 \subset \mathcal{X}_2 \implies \mathcal{X}_1^* \supset \mathcal{X}_2^* \implies \mathcal{X}_1^{**} \subset \mathcal{X}_2^{**}.$$

A paradoxical situation may arise if we have two Hilbert spaces $H_1 \subset H_2$. Then by canonical injection we have $H_1^* = H_1$ and $H_2^* = H_2$, which would lead to $H_2 \subset H_1$! This example shows that we have to be cautious while using equality as a symbol for canonical injection.

2.17. Weak star compact sets. For investigation of convergence, it is important to understand the structure of compact sets. Let $\mathcal{X}$ be a normed space. It is well known that a closed ball of $\mathcal{X}$ is compact in the strong topology if and only if $\mathcal{X}^*$ is finitely dimensional. Since $\mathcal{X}^*$ is also a normed space, the same holds for $\mathcal{X}^*$. However, in the weak star topology, we have the following result.

2.17. Theorem (Banach–Alaoglu). *Let $\mathcal{X}$ be an arbitrary normed space. A closed ball of $\mathcal{X}^*$ is weak star compact.*

Proof. Without a loss of generality, observe a closed unit ball of $\mathcal{X}^*$, call it B. Hence, B contains all linear continuous mappings φ from $\mathcal{X}$ to F such that $|\varphi(x)| \le \|x\|$ for all $x \in \mathcal{X}$. For any $x \in \mathcal{X}$, define $D_x = \{t \in F \mid |t| \le \|x\|\}$ and $K = \prod_{x\in\mathcal{X}} D_x$, with a product topology on K. If f is an element of K and $f(x)$ its co-ordinate in K, then f is a function $f : \mathcal{X} \mapsto F$. The product topology is the topology of pointwise convergence: $f_d \to f$ if and only if $f_d(x) \to f(x)$ for any $x \in \mathcal{X}$. So, B with the weak topology on it is a subset of K. Since each D_x is compact, Tychonov's theorem states that K is also compact, so we just need to show that B is closed in K. To this end, let φ_d be a net in B which converges to some $f \in K$. Then it is trivial to show that f must be linear; then by $|\varphi_d(x)| \le \|x\|$ it follows that f is also continuous and that $\|f\| \le 1$. Therefore, $f \in B$ and B is closed, hence compact. □

2.19. Remark. Tychonov's theorem states that the product space $\prod_i \mathcal{X}_i$ in the product topology as explained above, is compact if and only if each of $\mathcal{X}_i$ is compact. The proof of Banach-Alaoglu theorem relies on Tychonov's theorem, and the proof of the latter, in the part which is used here, relies on the Axiom of Choice (more precisely, Zorn's lemma, cf. [**9, 15, 35, 36**]). □

This theorem implies that any bounded sequence in $\mathcal{X}^*$ must have a convergent subnet. Unfortunately, such a subnet need not be a sequence, since the weak star topology on $\mathcal{X}^*$ need not be metrizable. However, the next theorem claims that in one special case we can introduce a metric.

2.20. Theorem. *Let $\mathcal{X}$ be a separable normed vector space. Then the $w-*$ topology on a closed ball of $\mathcal{X}^*$ is metrizable.*

Proof. Assume, without a loss of generality that B is the closed unit ball (centered at the origin) of the dual $\mathcal{X}^*$ of a separable normed vector space $\mathcal{X}$. The metrization of B can be realized, for instance, as follows. Let $\varphi_1, \varphi_2 \in B$, so

$$\sup_{\|x\|\le 1} \varphi_i(x) \le 1, \qquad i = 1, 2.$$

Let $\{x_n\}$ be a dense countable set in the unit ball of $\mathcal{X}$. Define

$$d(\varphi_1, \varphi_2) = \sum_n \frac{|\varphi_1(x_n) - \varphi_2(x_n)|}{2^n}.$$

Then $|\varphi_1(x_n) - \varphi_2(x_n)| \le \|\varphi_1 - \varphi_2\| \cdot \|x_n\| \le 2$ and the series converges, so d is a well defined function (even on the whole space $\mathcal{X}^*$). It is now a matter of an exercise to show that the d-topology on B coincides with the $w-*$ topology.

2.21. Corollary. *Let $\mathcal{X}$ be a separable normed vector space and let $\mathcal{X}^*$ be its topological dual space. Then every bounded sequence $\{\varphi_n\} \in \mathcal{X}^*$ has a weak star convergent subsequence.*

Proof. Every bounded sequence is contained in some closed ball B, which is, by Theorem 2.18, weak star compact. By Theorem 2.20, the weak star topology on B is metrizable, i.e., there is a metric d such that

$$\varphi_d \to \varphi \quad (w-*) \iff d(\varphi_d, \varphi) \to 0.$$

In a metric space, compactness is equivalent to sequential compactness, so, any sequence in B has a convergent subsequence. □

2.22. Example. Let H be a separable Hilbert space. Since it is reflexive, weak and weak star topology coincide. Define a metric

$$d(x,y) = \sum_n \frac{|\langle x-y, e_n\rangle|}{2^n}$$

where $\{e_n\}$ is an orthonormal base in H. We shall prove that this metric also generates the weak topology on the unit ball of H. Suppose that $x_n \to x$ weakly, i.e., $\langle x_n, y\rangle \to \langle x, y\rangle$ for any $y \in H$, where $\|x_n\| \leq 1, \|x\| \leq 1$. Since

$$|\langle x_n - x, e_k\rangle| \leq \|x_n - x\| \cdot \|e_k\| \leq 2,$$

the series

$$d(x_n, x) = \sum_k \frac{|\langle x_n, e_k\rangle - \langle x, e_k\rangle|}{2^k}$$

converges uniformly in n and so, by evaluating limits under the sum, we conclude that

$$\lim_{n\to+\infty} d(x_n, x) = 0.$$

Conversely, let $d(x_n, x) \to 0$ as $n \to +\infty$, where $\|x_n\| \leq 1$ and $\|x\| \leq 1$. Then it follows that $\langle x_n - x, e_k\rangle \to 0$ for every k. Now for any $y \in H$,

$$\langle x_n, y\rangle - \langle x, y\rangle = \sum_k \langle x_n - x, e_k\rangle\langle y, e_k\rangle.$$

By Cauchy-Schwarz inequality,

$$\begin{aligned}
\Big|\sum_{k=m}^{+\infty} \langle x_n - x, e_k\rangle\langle y, e_k\rangle\Big| &\leq \sum_{k=m}^{+\infty} |\langle x_n - x, e_k\rangle| \cdot |\langle y, e_k\rangle| \\
&\leq \Big(\sum_{k=m}^{+\infty} \langle x_n - x, e_k\rangle^2 \sum_{k=m}^{+\infty} \langle y, e_k\rangle^2\Big)^{1/2} \\
&\leq \|x_n - x\| \cdot \Big(\sum_{k=m}^{+\infty} \langle y, e_k\rangle^2\Big)^{1/2} \\
&\leq 2\Big(\sum_{k=m}^{+\infty} \langle y, e_k\rangle^2\Big)^{1/2}
\end{aligned}$$

and therefore, the series $\sum_k \langle x_n - x, e_k\rangle\langle y, e_k\rangle$ converges uniformly with respect to n. Hence,

$$\lim_{n\to+\infty}\langle x_n, y\rangle - \langle x, y\rangle = \sum_k \lim_{n\to+\infty}\langle x_n - x, e_k\rangle\langle y, e_k\rangle = 0.$$

So, $\{x_n\}$ converges weakly to x.

However, the metric described here does not generate the weak topology on the whole H. To see this, let $x_n = ne_n$. Then $d(x_n, 0) \to 0$ as $n \to +\infty$, but $\langle x_n, y\rangle = n\langle e_n, y\rangle$, which need not converge.

3. Finitely additive measures and Radon integrals

3.1. Spaces of measures as dual spaces. In general, it might be very hard to find the dual space of a given space, i.e., to represent it (via canonical injections) in terms of some well known structure. We are particularly interested in spaces of measures; it turns out that they can be viewed as dual spaces of some spaces of functions. The functionals on spaces of functions are expressed as integrals:

$$\varphi(f) = \int f(t)\, d\mu(t)$$

where μ is a measure which determines a functional. Then by a canonical injection, we can identify functionals and corresponding measures. There are several results in various levels of difficulty, depending on assumptions that one imposes on the underlying space X on which we observe measures. In this section we will present the most general result [**1**] regarding an arbitrary topological space. It turns out that finitely additive measures are the key notion in this general setting.

Although a traditional probabilist works solely with countably additive measures on sigma fields, their presence in Probability has a purpose to make mathematics simpler and is by no means natural. As Kolmogorov [**19**, p. 15] points out, "*dots* in describing any observable random process we can obtain only finite fields of probability. Infinite fields of probability occur only as idealized models of real random processes". Finitely additive measures have recently arose an increasing interest in Probability, so the exposition which follows may be interesting in its own rights.

3.2. Fields and sigma fields. Let X be a set and $\mathcal{F}$ a class of its subsets such that

1) $X \in \mathcal{F}$,
2) $B \in \mathcal{F} \implies B' \in \mathcal{F}$,
3) $B_1, B_2 \in \mathcal{F} \implies B_1 \cup B_2 \in \mathcal{F}$ Then we say that $\mathcal{F}$ is a field. If 3) is replaced by stronger requirement

3') $B_1, B_2, \ldots \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} B_i \in \mathcal{F}$, then we say that $\mathcal{F}$ is a sigma field.

It is easy to see that a field is closed under finitely many set operations of any kind. Further, let $\mathcal{F}_i$, $i \in I$, be fields on X. Then $\mathcal{F} = \bigcap_{i\in I} \mathcal{F}_i$ is also a field, where

I is any collection of indices. This follows trivially by verification of conditions 1)–3) above.

Given any collection of sets $\mathcal{A}$ which are subsets of X, there is a field which contains $\mathcal{A}$: it is the family of all subsets of X. The intersection of all fields that contain $\mathcal{A}$ is called the field generated by $\mathcal{A}$. Obviously, a field $\mathcal{F}$ generated by $\mathcal{A}$ is the smallest field that contains $\mathcal{A}$, in the sense that there is no field which is properly contained in $\mathcal{F}$ and contains $\mathcal{A}$.

A sigma field is closed under countably many set operations. We define a sigma field generated by a collection of sets in much the same way as in the case of fields.

3.3. Borel field. Let X be a topological space. The field generated by the collection of all open sets is called the Borel field. Since the complement of an open set is a closed set, the Borel field is also generated by the collection of all closed sets.

Borel sigma field is the sigma field generated by open or closed sets. In separable metric spaces, the Borel sigma field is also generated by open or closed balls, since any open set can be expressed as a countable union of such balls.

Specifically, on the real line, Borel sigma field is generated by open and closed intervals of any kind. However, Borel field *is not generated by intervals*, since an arbitrary open set need not be represented as a finite union of intervals.

3.4. Baire field. Let X be a topological space and let $\boldsymbol{C}(X)$ be the collection of all bounded and continuous real valued functions defined on X. The Baire field is the field generated by the collection of sets

$$\mathcal{A} = \{Z \subset X \mid Z = f^{-1}(C)\},$$

where f is any function in $\boldsymbol{C}(X)$ and C is any closed set of real numbers.

Boundedness of functions in $\boldsymbol{C}(X)$ is not relevant, but is assumed here for the purposes of this paper. Indeed, for any continuous function $f : X \mapsto \boldsymbol{R}$, the function $g(x) = \operatorname{arctg} f(x)$ is a continuous bounded function defined on X and the collection of all $g^{-1}(C)$ coincides with the collection of all $f^{-1}(C)$, where C runs over closed subsets of $\boldsymbol{R}$.

It is well known that for any closed set $C \subset \boldsymbol{R}$ there is a continuous bounded function g_C such that $g_C^{-1}(\{0\}) = C$ (this is a consequence of a more general result that holds, for instance, on metric spaces, see [**6**, Theorem 1.2]. For an $f \in \boldsymbol{C}(X)$ and a closed set $C \subset \boldsymbol{R}$, define $F(x) = g_C(f(x))$. Then $F \in \boldsymbol{C}(X)$ and $F^{-1}(\{0\}) = f^{-1}(C)$. Therefore, we may think of the Baire field as being generated by sets of the form $f^{-1}(\{0\})$, for $f \in \boldsymbol{C}(X)$.

Let us recall that $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$ and $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$; also $f^{-1}(A') = (f^{-1}(A))'$ if the complement is taken with respect to the domain of f. Hence, we have:

$$f^{-1}(C_1) \cup f^{-1}(C_2) = f^{-1}(C_1 \cup C_2); \quad (f^{-1}(C))' = f^{-1}(C')$$

and also $X = f^{-1}(\boldsymbol{R})$, for any f. Therefore, the Baire field is also generated by the collection of the sets $f^{-1}(O)$, where O is an open set of real numbers and

$f \in \boldsymbol{C}(X)$. Further,

$$(f^{-1}(C))' = f^{-1}(O), \quad \text{where } O = C' \text{ is an open set.}$$

Now it is clear that if A belongs to the Borel field in $\boldsymbol{R}$ and if $f \in \boldsymbol{C}(X)$, then $f^{-1}(A)$ belongs to the Baire field in X.

From now on, sets of the form $f^{-1}(C)$, where C is closed in $\boldsymbol{R}$, will be called Z-sets, and the sets of the form $f^{-1}(O)$, with O being an open set in $\boldsymbol{R}$, will be called U-sets.

Since the inverse image (with a continuous function) of any open (resp. closed) set is again an open (resp. closed) set, we see that Z-sets are closed and U-sets are open in X. Hence, the Baire field is a subset of Borel field; the same relation holds for the sigma fields. The converse is not generally true, since a closed set need not be a Z-set. With some restrictions on topology of X, the converse becomes true, for instance, in metric spaces. In general, in normal spaces in which every closed set can be represented as a countable intersection of open sets (so called G_δ set), every closed set is a Z-set (cf. [**15**, Corollary 1.5.11]) and so the Baire and the Borel field coincide.

3.5. Theorem. *The family of Z-sets is closed under finite unions and countable intersections. The family of U-sets is closed under countable unions and finite intersections.*

Proof. By 3.4, a set is a Z-set if and only if it is of the form $f^{-1}(\{0\})$ for some $f \in \boldsymbol{C}(X)$. So, let $Z_1 = f_1^{-1}(\{0\})$, $Z_2 = f_2^{-1}(\{0\})$. If $g(x) = f_1(x)f_2(x)$, then $g^{-1}(\{0\}) = Z_1 \cup Z_2$, so the union of two Z-sets is again a Z-set. Let $Z_1, Z_2, \ldots$ be Z-sets. Then there are continuous and bounded functions $f_1, f_2, \ldots$ such that $Z_n = f_n^{-1}(\{0\}), n = 1, 2, \ldots$. Define the function

$$F(x) = \sum_{n=1}^{+\infty} \frac{f_n^2(x)}{2^n \|f_n\|^2},$$

where $\|f_n\| = \sup_{x \in X} |f(x)|$. Since the above series is uniformly convergent on X, F is a continuous and bounded function; moreover, $F(x) = 0$ if and only if $f_n(x) = 0$ for all $n \geq 1$. Hence $F^{-1}(\{0\}) = \bigcap_{n=1}^{\infty} Z_n$, which proves that any countable intersection of Z-sets is a Z-set.

Statements about U-sets can be proved by taking complements.

3.6. Measures and regularity. Let X be a topological space. Let μ be a non-negative and finitely additive set function on some field or a sigma field $\mathcal{F}$ of subsets of X, with values in $[0, +\infty]$ (allowing $+\infty$ if not specified otherwise). Such a function will be called a measure.

We say that a set $A \in \mathcal{F}$ is μ-regular if

$$\mu(A) = \sup\{\mu(Z) \mid Z \subset A\} = \inf\{\mu(U) \mid A \subset U\}, \tag{4}$$

where Z and U are generic notations for Z-sets and U-sets respectively.

If all sets in $\mathcal{F}$ are μ-regular, we say that the measure μ is regular.

Note that a prerequisite for regularity is that all Z-sets and U-sets must be measurable, which is the case if $\mathcal{F}$ contains the Baire field. In the next theorem we give alternative conditions for regularity.

3.7. Theorem. *Let X be a topological space, $\mathcal{F}$ a field which contains the Baire field and μ a measure on $\mathcal{F}$. A set $A \in \mathcal{F}$ with $\mu(A) < +\infty$ is μ-regular if and only if either of the following holds:*

(i) For each $\varepsilon > 0$ there exists a Z-set Z_ε and an U-set U_ε so that

$$Z_\varepsilon \subset A \subset U_\varepsilon \quad \text{and} \quad \mu(U_\varepsilon \setminus Z_\varepsilon) < \varepsilon. \tag{5}$$

(ii) There are Z-sets $Z_1, Z_2, \dots$ and U-sets $U_1, U_2, \dots$ such that

$$Z_1 \subset Z_2 \subset \cdots \subset A \subset \cdots U_2 \subset U_1$$

and

$$\mu(A) = \lim_{n\to+\infty} \mu(Z_n) = \lim_{n\to+\infty} \mu(U_n).$$

Proof. (i) is straightforward, using properties of the infimum and the supremum. (ii) Suppose that A is μ-regular. Then for each n there is a Z-set Z_n^* such that $Z_n^* \subset A$ and $\mu(A) - 1/n < \mu(Z_n^*) < \mu(A)$. Let $Z_n = Z_1^* \cup \cdots \cup Z_n^*$, for $n = 1, 2, \dots$. Then Z_n are Z-sets by Theorem 3.5. Further, $Z_1 \subset Z_2 \subset \cdots \subset A$ and $\mu(A) - 1/n < \mu(Z_n) < \mu(A)$, hence $\lim \mu(Z_n) = \mu(A)$. The part regarding U_n can be proved similarly.

Conversely, if there exist Z_n and U_n as in the statement of the theorem, then for a fixed $\varepsilon > 0$ there is an n such that $Z_n \subset A$ and $0 < \mu(A) - \mu(Z_n) < \varepsilon$, hence $\mu(A)$ is the least upper bound for $\mu(Z)$ over all Z-subsets of A. Similarly, it follows that $\mu(A)$ is the greatest lower bound for $\mu(U)$, over all U-sets that contain A.

3.8. Remark. The previous theorem does not imply either the countable additivity or continuity of μ. Also it holds regardless whether μ is defined on a sigma field or just on a field.

3.9. Theorem. *Suppose that μ is a countably additive measure defined on a sigma field $\mathcal{F}$ which contains the Baire field. Then a set $A \in \mathcal{F}$, $\mu(A) < +\infty$, is μ-regular if and only if there are Z-sets $Z_1, Z_2, \dots$ and U-sets $U_1, U_2, \dots$ such that*

$$Z_1 \subset Z_2 \subset \cdots \subset A \subset \cdots U_2 \subset U_1$$

and

$$\mu\Big(A \setminus \bigcup_{n=1}^{+\infty} Z_n\Big) = 0, \quad \mu\Big(\bigcap_{n=1}^{+\infty} U_n \setminus A\Big) = 0.$$

Proof. By the previous theorem, A is μ-regular if and only if $\mu(A) = \lim \mu(Z_n) = \lim \mu(U_n)$; by continuity property of sigma additive measures we have that $\lim \mu(Z_n) = \mu(\bigcup_n Z_n)$ and $\lim \mu(U_n) = \mu(\bigcap_n U_n)$, which ends the proof.

3.10. Theorem *Let X be a topological space and $\mathcal{F}$ a field which contains the Baire field. Let μ be a measure on $\mathcal{F}$, with $\mu(X) < +\infty$. Then the family $\mathcal{R}$ of all μ-regular sets in $\mathcal{F}$ is a field.*

Proof. Since $X = f^{-1}(\boldsymbol{R})$ and $\boldsymbol{R}$ is open and closed, it follows that both conditions in (4) hold and so $X \in \mathcal{R}$.

Suppose that $A \in \mathcal{R}$. Then for a fixed $\varepsilon > 0$ there are sets Z_ε and U_ε such that (5) holds. Taking complements we get

$$U'_\varepsilon \subset A' \subset Z'_\varepsilon, \quad Z'_\varepsilon \setminus U'_\varepsilon = U_\varepsilon \setminus Z_\varepsilon,$$

which implies that A' is also μ-regular.

Finally, suppose that $A_1, A_2, \dots A_n \in \mathcal{R}$. By Theorem 3.7(i), for any given $\varepsilon > 0$, there are Z-sets Z_i and U-sets U_i such that

$$Z_i \subset A_i \subset U_i \quad \text{and} \quad \mu(U_i \setminus Z_i) < \frac{\varepsilon}{2^i}, \quad i = 1, 2, \dots$$

Let $A = \bigcup_{i=1}^n A_i$, $Z = \bigcup_{i=1}^n Z_i$ and $U = \bigcup_{i=1}^n U_i$. Then Z is a Z-set and U is a U-set and we have

$$Z \subset A \subset U \quad \text{and} \quad \mu(U \setminus Z) \le \sum_i \mu(U_i \setminus Z_i) \le \varepsilon, \tag{6}$$

so $A \in \mathcal{R}$.

3.11. Theorem. *Let X be a topological space, $\mathcal{F}$ a sigma field that contains the Baire field. Let μ be a countably additive measure on $\mathcal{F}$, with $\mu(X) < +\infty$. Then the family $\mathcal{R}$ of all μ-regular sets in $\mathcal{F}$ is a sigma field.*

Proof. In the light of Theorem 3.10, we need to prove only that a countable union of μ-regular sets is μ-regular.

Let $A_1, A_2, \dots$ be μ-regular sets; for any $\varepsilon > 0$, there are Z-sets Z_i and U-sets U_i such that

$$Z_i \subset A_i \subset U_i \quad \text{and} \quad \mu(U_i \setminus Z_i) < \frac{\varepsilon}{2^i}.$$

Let $A = \bigcup_{i=1}^\infty A_i$, $Z = \bigcup_{i=1}^\infty Z_i$ and $U = \bigcup_{i=1}^\infty U_i$. Then U is a U-set (Theorem 3.5) and Z can be approximated by a finite union $Z^{(n)} = \bigcup_{i=1}^n Z_i$, where n is chosen in such a way that $\mu(Z \setminus Z^{(n)}) < \varepsilon$ (continuity of the countably additive measure). So, we have that

$$\mu(U \setminus Z^{(n)}) \le \mu(U \setminus Z) + \mu(Z \setminus Z^{(n)}) < 2\varepsilon,$$

which ends the proof.

3.12. Theorem. *Let X be a topological space, $\mathcal{F}$ the Baire sigma field and μ a countably additive measure on $\mathcal{F}$, with $\mu(X) < +\infty$. Then μ is regular.*

Proof. By Theorem 3.11, all μ-regular sets make a sigma field $\mathcal{R}$. We need to show that $\mathcal{R} = \mathcal{F}$, which will be accomplished if we show that each Z-set is

μ-regular. So, let Z be a Z-set. Then there is a function $f \in \boldsymbol{C}(X)$ such that $Z = f^{-1}(0)$. Let $O_n = (-1/n, 1/n)$ and $U_n = f^{-1}(O_n)$. Then $U_1 \supset U_2 \supset \cdots \supset Z$ and $\bigcap_n U_n = Z$. By continuity of countably additive measure μ we have that $\mu(Z) = \lim_n \mu(U_n)$, so the condition of Theorem 3.7(ii) holds (with $Z_n = Z$ for all n), hence Z is μ-regular.

3.13. Remark. Theorem 3.12 implies that only non-countably additive measures may be non-regular. The condition of regularity as defined here obviously turns out to be natural for Baire fields. However, in Borel fields, one often uses a different concept of regularity, which is the approximation by closed sets rather than by sets of the form $f^{-1}(C)$. In spaces in which any closed set is G_δ, any countably additive measure is regular (on Borel sigma field) in the latter sense, cf. [**27**].

3.14. Radon measures and Radon integrals. Let X be a topological space and let $\mathcal{F}$ be the Baire field on X. Let $\mathcal{M}^+(X)$ be the set of all non-negative, finite, finitely additive and regular measures on $\mathcal{F}$. A generalized measure (or a Radon finitely additive measure) is any set function on $\mathcal{F}$ which can be represented as $m(A) = m_1(A) - m_2(A)$, where $m_1, m_2 \in \mathcal{M}^+(X)$. The set of all generalized measures will be denoted by $\mathcal{M}(X)$. It is a linear vector space; a norm can be introduced by the so called total variation of a measure:

$$|m| = m^+(X) + m^-(X), \tag{7}$$

where $m^+(X) = \sup\{m(B) \mid B \in \mathcal{F}\}$, $m^-(X) = -\inf\{m(B) \mid B \in \mathcal{F}\}$. $\mathcal{M}(X)$ with the norm (7) is a Banach space.

We are now ready to define an integral of a bounded function with respect to a generalized measure. Let f be an $\mathcal{F}$-measurable function and suppose that $\|f\| = K < +\infty$. Let $A_1, A_2, \ldots, A_n$ be any partition of the interval $[-K, K]$ into disjoint intervals (or, in general, sets from the Borel field on $\boldsymbol{R}$) and let $B_i = f^{-1}(A_i)$. In each A_i choose a point y_i and make the integral sum

$$S_d = \sum_{i=1}^{n} y_i m(B_i), \quad \text{where } d = (A_1, \ldots, A_n, y_1, \ldots, y_n).$$

If we direct the set $\{d\}$ in a usual way, saying that $d_1 \prec d_2$ if the partition in d_2 is finer than the one in d_1, then we can prove that S_d is a Cauchy net, hence there is a finite limit, which is the integral of f with respect to the finitely additive measure m, $\int f(x)\,dm(x)$.

3.15. Theorem (Aleksandrov [**1**]). *For an arbitrary topological space X, any linear continuous functional on $\boldsymbol{C}(X)$ is of the form*

$$\varphi(f) = \langle f, m\rangle = \int f(x)\,dm(x), \tag{8}$$

Moreover,

$$\sup_{\|f\| \le 1} \left| \int f(x)\,dm(x) \right| = |m|.$$

There is an isometrical, isomorphical and one to one mapping between the space of all continuous linear functionals on $\boldsymbol{C}(X)$ *and the space* $\mathcal{M}(X)$*; in that sense we write* $\boldsymbol{C}(X)^* = \mathcal{M}(X)$. $\square$

In some special cases, $\boldsymbol{C}(X)^*$ has a simpler structure. For example, if X is a compact topological space, then $\boldsymbol{C}(X)^*$ can be identified with the set of all Baire countably additive $\boldsymbol{R}$-valued measures on the Baire sigma-field of X. If, in addition, X is a compact metric space, then $\boldsymbol{C}(X)$ is a separable normed vector space and $\boldsymbol{C}(X)^*$ is the set of all Borel $\boldsymbol{R}$-valued countably additive measures on X.

4. Weak convergence in probability

4.1. Convergence of probability measures. Let now X be a metric space and let $\mathcal{B}$ be the sigma field of Borel (= Baire) subsets of X. Let $\mathcal{M}_1(X)$ be the set of all probability measures on X. Then according to 3.15, $\mathcal{M}_1(X)$ is a subset of the unit ball in $\boldsymbol{C}(X)^*$. The structure of the second dual $\boldsymbol{C}(X)^{**}$ is too complex, but it is well known that $B(X)$ - the set of all bounded Borel-measurable functions is a subset of $\boldsymbol{C}(X)^{**}$. So, we have the following inclusions:

$$\begin{gathered}\text{Original space: } \boldsymbol{C}(X)\\ \text{Dual space: } \mathcal{M}(X); \qquad \mathcal{M}_1(X) \subset \mathcal{M}(X)\\ \text{Second dual: } \boldsymbol{C}(X) \subset B(X) \subset \boldsymbol{C}(X)^{**}.\end{gathered}$$

Let $\langle f, \mu \rangle$ be defined as in (8). On $\mathcal{M}(X)$ we may observe the following topologies:

- The uniform topology [struk], with the norm $\sup |\langle f, \mu \rangle|$, where the supremum is taken over the unit ball in $B(X)$.
- The strong topology, defined by $\sup |\langle f, \mu \rangle|$, where the supremum is taken over the unit ball in $\boldsymbol{C}(X)$.
- The weak topology defined by $\langle f, \mu \rangle$, for $f \in \mathcal{M}^*(X) = \boldsymbol{C}(X)^{**}$.
- The $B(X)$-topology, defined by $\langle f, \mu \rangle$, for $f \in B(X)$.
- The weak star topology, defined by $\langle f, \mu \rangle$, where $f \in \boldsymbol{C}(X)$.

First four topologies are too strong, and they do not respect a topological structure of X, as the following example shows.

Example. Let δ_x, δ_y be point masses at x and y respectively. Then $\langle f, \delta_x \rangle - \langle f, \delta_y \rangle = f(x) - f(y)$. If x and y are close in X, then $f(x)$ and $f(y)$ need not be close unless f is continuous. So, in this example, the weak or $B(X)$-topology are inadequate, but the weak star topology preserves the closedness of x and y. $\square$

The convergence in the weak star topology is usually called the weak convergence in the probabilistic literature. This does not lead to a confusion, since the true weak convergence is never studied.

If μ_d converges weakly to μ, we write $\mu_d \implies \mu$.

The weak star convergence of probability measures is well investigated. We shall firstly give equivalent bases for weak star topology on the whole set $\mathcal{M}^+(X)$. So, the next theorem is not restricted to probability measures.

Let us recall that we say that A is a continuity set for a measure μ on a Borel algebra $\mathcal{B}$ if $\mu(\partial A) = 0$, or, equivalently, if $\mu(A) = \mu(\bar{A}) = \mu(A^\circ)$, where ∂A is the boundary, $\bar{A}$ is the closure and A° is the interior of A. On a Baire algebra we will say that A is a continuity set for μ if there is an U-set U and a Z-set Z such that $U \subset A \subset Z$ and $\mu(Z \setminus U) = 0$.

4.2. Theorem [**34**, p. 56]. *Let W be the weak star topology on $\mathcal{M}^+(X)$, where X is a topological space. Then the following families of sets make a local base of W around some measure $\mu_0 \in \mathcal{M}^+(X)$:*

$$B_0 = \{\mu \mid |\langle f_i, \mu\rangle - \langle f_i, \mu_0\rangle| < \varepsilon,\ i = 1, \dots k\}, \quad f_i \in \boldsymbol{C}(X)$$
$$B_1 = \{\mu \mid \mu(F_i) < \mu_0(F_i) + \varepsilon,\ i = 1, \dots, k\}, \quad F_i \textit{ are } Z\textit{-sets in } X$$
$$B_2 = \{\mu \mid \mu(G_i) > \mu_0(G_i) - \varepsilon,\ i = 1, \dots, k\}, \quad G_i \textit{ are } U\textit{-sets in } X$$
$$B_3 = \{\mu \mid |\mu(A_i) - \mu_0(A_i)| < \varepsilon,\ i = 1, \dots, k\}, \quad A_i \textit{ are continuity sets for } \mu_0,$$

If X is a metric space, then we deal with the Borel algebra and so F_i above can be taken to be closed and G_i to be open sets.

As a straightforward consequence, we get the following

4.3. Theorem. *Measures with a finite support are dense in $\mathcal{M}^+(X)$.*

Proof. Let $\mu_0 \in \mathcal{M}^+(X)$ and let $B(\mu_0)$ be its neighborhood of the form

$$B(\mu_0) = \{\mu \mid \mu(F_i) < \mu_0(F_i) + \varepsilon,\ i = 1, \dots, k\},$$

where F_i are fixed Z-sets. The family of sets F_i together with their intersections and the complement of their union defines a finite partition of X. In each set B of this partition choose a point x_B and define $\mu_1(x_B) = \mu_0(B)$. The measure μ_1 is with a finite support (hence, countably additive!) and clearly $\mu_1(F_i) = \mu_0(F_i)$; so $\mu_1 \in B(\mu_0)$.

4.4. Theorem. *Let μ_d be a net of measures in $\mathcal{M}^+(X)$ and let $\mu_0 \in \mathcal{M}^+(X)$. The following statements are equivalent [**6**, **30**, **34**]:*

(i) $\mu_d \implies \mu_0$, *i.e.,* $\lim_d \int f\, d\mu_d = \int f\, d\mu_0$, *for each* $f \in \boldsymbol{C}(X)$.
(ii) $\overline{\lim}\, \mu_d(F) \le \mu_0(F)$ *for any Z-set $F \subset X$ and* $\lim \mu_d(X) = \mu_0(X)$.
(iii) $\underline{\lim}\, \mu_d(G) \ge \mu_0(G)$ *for each U-set $G \subset X$ and* $\lim \mu_d(X) = \mu_0(X)$.
(iv) $\lim \mu_d(A) = \mu_0(A)$ *for each continuity set for* μ_0.

In a special case when we have probability measures on a metric space X, there is a richer structure that yields additional equivalent conditions. To proceed we need some facts on semicontinuous functions.

4.5. Semicontinuous functions. Let X be a metric space. A function $f : X \mapsto \boldsymbol{R}$ is called upper semicontinuous if $\overline{\lim}\, f(x_n) \le f(x)$ for each sequence $\{x_n\}$ such that $x_n \to x$. The function f is lower semicontinuous if $\underline{\lim}\, f(x_n) \ge f(x)$ for each sequence $x_n \to x$.

An important property of semicontinuous functions is that for each $M \in \boldsymbol{R}$ the set $\{x \mid f(x) < M\}$ is open for an upper semicontinuous function and the set $\{x \mid f(x) > M\}$ is open for a lower semicontinuous functions.

4.6. Theorem. *Let μ_d be a net of probability measures on a metric space X and let μ_0 be a probability measure on X. The following statements are equivalent [**6, 30**]:*

(i) $\mu_d \implies \mu_0$, *i.e.,* $\lim_d \int f\, d\mu_d = \int f\, d\mu_0$, *for each* $f \in \boldsymbol{C}(X)$.

(ii) $\lim_d \int f\, d\mu_d = \int f\, d\mu_0$ *for each* $f \in \boldsymbol{C}_u(X)$ *(uniformly continuous and bounded functions).*

(iii) $\overline{\lim}\, \mu_d(F) \le \mu_0(F)$ *for any closed set* $F \subset X$.

(iv) $\underline{\lim}\, \mu_d(G) \ge \mu_0(G)$ *for each open set* $G \subset X$.

(v) $\lim \mu_d(A) = \mu_0(A)$ *for each continuity set for* μ_0.

(vi) $\overline{\lim} \int f\, d\mu_d \le \int f\, d\mu_0$ *for each upper semicontinuous and bounded from above function* $f : X \mapsto \boldsymbol{R}$.

(vii) $\underline{\lim} \int f\, d\mu_d \ge \int f\, d\mu_0$ *for each lower semicontinuous and bounded from below function* $f : X \mapsto \boldsymbol{R}$.

(viii) $\lim \int f\, d\mu_d = \int f\, d\mu_0$ *for each* μ_0 a.e. *continuous function* $f : X \mapsto \boldsymbol{R}$.

4.7. Vague convergence. In [**32**], a concept of so called vague convergence is introduced as follows. Let $K(X)$ be the set of all continuous functions with a compact support defined on X. Then we say that μ_d converges vaguely to μ if $\langle f, \mu_d \rangle \to \langle f, \mu \rangle$ for each $f \in K(X)$. This kind of convergence is clearly weaker than the weak star convergence. For example, the sequence δ_n converges vaguely to 0, although it does not converge in the weak star sense.

4.8. Metrics of weak convergence. By Theorem 2.20, the weak star topology on the closed unit ball of $\mathcal{M}(X)$ is metrizable if $C(X)$ is a separable metric space, which is the case if and only if X is a compact space. However, even if the weak star topology of the unit ball of $\mathcal{M}$ is not metrizable, this topology on the set of all probability measures may be metrizable; as a matter of fact, it probably is always metrizable, as we shall see in the subsequent discussion.

4.9. Theorem. *Let X be a separable metric space. Then the weak star topology on $\mathcal{M}_1(X)$ is metrizable by the metric*

$$d(P,Q) = \inf\{\varepsilon > 0 \mid Q(B) \le P(B^\varepsilon) + \varepsilon,\ \ P(B) \le P(Q^\varepsilon) + \varepsilon,\ \ B \in \mathcal{B}\}, \tag{9}$$

where $B^\varepsilon = \{x \in S \mid d(x,B) < \varepsilon\}$, and $\mathcal{B}$ is the Borel sigma algebra. □

The metric (9) is known as Lévy's metric or Prohorov's metric [**6, 30**]. Although the proof of Theorem 4.9 relies on separability of X, it has to be noted that the metrizability of $\mathcal{M}_1(X)$ is related to the so called problem of measure [**6, 12**] and that the examples of non-metrizable $\mathcal{M}_1(X)$ are not known. So, there is a strongly founded conjecture that for any metric space X, the topology of the weak star convergence on the set $\mathcal{M}_1(X)$ is metrizable and one metric is given by (9).

Moreover, it is known that, if X is a complete separable metric space (Polish space), then so is $\mathcal{M}_1(X)$.

There is another metric of weak star convergence [**30**, p. 117], similar to the one introduced in Theorem 2.20.

4.10. Theorem. *Let X be a separable metric space. Then there is a countable set $\{f_1, f_2, \dots\}$ of uniformly continuous bounded real valued functions with values* *in* $[0,1]$ *so that the span of this set is dense in the set of all uniformly continuous and bounded functions on* X*. Now define*

$$\rho(P,Q) = \sum_{n=1}^{+\infty} \frac{|\langle f_n, P\rangle - \langle f_n, Q\rangle|}{2^n}.$$

Then ρ *is a metric on* $\mathcal{M}_1(X)$*, which is topologically equivalent to Prohorov's metric.*

This theorem can be proved by noticing that the condition $\rho(P_d, Q) \to 0$ is equivalent to the condition (ii) of Theorem 4.6 applied to P_d and Q. Since the condition (ii) holds (as an equivalent condition to weak star convergence) only for countably additive measures, we conclude that the metric of Theorem 4.10 can not generally be extended to the unit ball of $\mathcal{M}(X)$; hence, the unit ball of $\mathcal{M}(X)$ generally is not metrizable.

5. Finitely, but not countably additive measures in the closure of the set of probability measures

In this section we discuss topics of relative weak star compactness and closedness of the set of all probability measures. We will show that in a non-compact topological space X, under slight additional assumptions (say, if X is a metric s-pace) the set of probability measures $\mathcal{M}_1$ is not closed under the weak star limits. We actually show the existence of an additive, but not countably additive measure in the closure of $\mathcal{M}_1$. The fact that $\mathcal{M}_1$ is not closed is the main rationale for Prohorov's theorem, which will be presented in the next section.

5.1. Nets and filters. Nets and filters are introduced in Mathematics as generalizations of sequences. Nets were defined and discussed in papers of Moore [moore] in a context of determining a precise meaning of the limit of integral sums; early developments of nets can be found in papers [**8, 18, 25, 26, 33**]. Filters were introduced by Cartan [**10, 11**] in the second decade of 20th century. The theory of both filters and nets was completed by the mid of 20th century. We will give here a brief account of basic definitions and theorems, largely taken from [**35**, Sections 11 and 12].

A set D is called a *directed set* if there is a relation $\leq$ on D such that

(i) $x \leq x$ for all $x \in D$
(ii) If $x \leq y$ and $y \leq z$ then $x \leq z$
(iii) For any $x, y \in D$ there is a $z \in D$ so that $x \leq z$ and $y \leq z$.

A *net* in a set X is any mapping of a directed set D into X, usually denoted by x_d, say, like sequences.

Let D and E be directed sets and let φ be a function $E \to D$ such that:

(i) $a \leq b \implies \varphi(a) \leq \varphi(b)$ for each $a, b \in E$;
(ii) For each $d \in D$ there is an $e \in E$ so that $d \leq \varphi(e)$.

Then $x_{\varphi(e)}$ is a *subnet* of the net x_d; more often denoted by x_{d_e}.

Let X be a topological space. We say that a net $x_d, d \in D$ converges to some point $x \in X$ if for each neighborhood U of x there is a $d_0 \in D$ so that $x_d \in U$ whenever $d \geq d_0$.

Let $S \subset X$. We say that $x_d, d \in D$ is *eventually* (or residually) in S if there is a $d_0 \in D$ so that $x_d \in S$ whenever $d \geq d_0$. Hence, a net x_d converges to x iff it is eventually in every neighborhood of x.

A net x_d is called an *ultranet* if for each $S \subset X$ it is eventually in S or eventually in S'.

Let S be any nonempty set. A collection $\mathcal{F}$ of non-empty subsets of S is called a *filter* if

(i) $S \in \mathcal{F}$.

(ii) If $B_1, B_2 \in \mathcal{F}$ then $B_1 \cap B_2 \in \mathcal{F}$.

(iii) If $B_1 \in \mathcal{F}$ and $B_1 \subset B_2 \subset S$, then $B_2 \in \mathcal{F}$.

A subcollection $\mathcal{F}_0 \subset \mathcal{F}$ is a *base* for the filter $\mathcal{F}$ if $\mathcal{F} = \{B \subset S \mid B \supset B_0 \text{ for some } B_0 \in \mathcal{F}_0\}$, that is, if $\mathcal{F}$ is consisted of all supersets of sets in $\mathcal{F}_0$. Any collection $\mathcal{F}_0$ can be a base for some filter $\mathcal{F}$ provided that given any two sets $A, B \in \mathcal{F}_0$ there is a $C \in \mathcal{F}_0$ so that $C \subset A \cap B$.

In a topological space X, the set of all neighborhoods of some fixed point x is a filter, called the *neighborhood filter*. Its base is the neighborhood base at x.

A filter $\mathcal{F}_1$ is *finer* than the filter $\mathcal{F}_2$ if $\mathcal{F}_1 \supset \mathcal{F}_2$.

We say that a filter $\mathcal{F}$ in a topological space X *converges* to $x \in X$ if $\mathcal{F}$ is finer than the neighborhood filter at x.

A filter $\mathcal{F}$ is called *principal* or *fixed* if $\bigcap_{B \in \mathcal{F}} B \neq \emptyset$; otherwise it is called *non-pricipal* or *free*.

A filter $\mathcal{F}$ on S is called an *ultrafilter* if there no filter on S which is strictly finer than $\mathcal{F}$. It can be shown [**35**, Theorem 12.11] that a filter $\mathcal{F}$ is an ultrafilter iff for any $B \in S$ either $B \in \mathcal{F}$ or $B' \in \mathcal{F}$. For example, the family of all sets that contain a fixed point $x \in X$ is an ultrafilter on X.

5.2. Relation between nets and filters. Both nets and filters are used to describe convergence and related notions. In fact, there is a close relationship between nets and filters.

Let $x_d, d \in D$ be a net in X. The sets $B_{d_0} = \{x_d \mid d \geq d_0\}$ make a base for a filter $\mathcal{F}$; we say that the filter $\mathcal{F}$ *is generated* by the net x_d.

Conversely, let $\mathcal{F}$ be a filter on a set S. Let D be the set of all pairs (x, F), where F runs over $\mathcal{F}$ and $x \in F$. Define the order by $(x_1, F_1) \leq (x_2, F_2) \iff F_1 \supset F_2$. Then the mapping $(x, F) \mapsto x$ is a *net based on* $\mathcal{F}$.

5.3. Conditions for compactness. A topological space X is called *compact* if every open cover has a finite subcover. The following conditions are equivalent [**35**, Theorem 17.4]:

a) X is compact

b) each family of closed subsets of X with the finite intersection property has an non-empty intersection,

c) for each filter in X there is a finer convergent filter,

d) each net in X has a convergent subnet,

e) each ultrafilter in X is convergent,

f) each ultranet in X is convergent.

5.4. Space of probability measures is not closed in $\mathcal{M}$. Since $\mathcal{M}(X)$, the space of generalized measures introduced in 3.14, is the dual space of the normed space $\boldsymbol{C}(X)$, by Theorem 2.18 its unit ball $B_1 = \{m \in \mathcal{M} \mid |m| = 1\}$ is compact in the weak star topology. If $\mathcal{M}_1(X)$, the space of all probability measures, were closed in $\mathcal{M}(X)$, then it would have been also compact, being a subset of B_1. Then (at least if X is a separable metric space), since the topology on $\mathcal{M}_1(X)$ is metrizable, any sequence of probability measures would have had a weak star convergent subsequence and Prohorov's theorem and the notion of tightness (Section 6) would not be of any interest. However, this is not generally true. For example, for $X = \boldsymbol{R}$, the sequence $\{P_n\}$ of point masses at $n = 1, 2, \dots$ clearly does not have any weak star convergent subsequence. However, it must have a convergent subnet, and the limiting measure is in $\mathcal{M} \setminus \mathcal{M}_1$.

Here we give a rather general result [**20**] which proves the existence of a measure in the closure of $\mathcal{M}_1(X)$, which is not a probability measure (not countably additive). Before we proceed, we need a lemma concerning normal spaces. Recall that a topological space X is called normal if for any two disjoint closed sets A and B in X there are disjoint open sets U and V such that $A \subset U$ and $B \subset V$. Equivalently, a space X is normal if and only if for any two disjoint sets A and B there is an $f \in \boldsymbol{C}(X)$ such that $f(A) = \{0\}$, $f(B) = \{1\}$ and $0 \le f(x) \le 1$ for all $x \in X$ (Urysohn's lemma).

5.5. Lemma. *Let X be a normal space which contains an infinite sequence $S = \{x_1, x_2, \dots\}$ with no cluster points. Then for any infinite proper subset $S_0 \subset S$ there is a function $f \in \boldsymbol{C}(X)$ such that $f(x) = 0$ if $x \in S_0$ and $f(x) = 1$ if $x \in S \setminus S_0$.*

Proof. Let S_0 be any infinite proper subset of S and let $S_1 = S \setminus S_0$. Then S_0 and S_1 are closed sets (no cluster points), hence by normality, the desired function exists.

5.6. Theorem. *Let X be a normal topological space and suppose that it contains a countable subset $S = \{x_1, x_2, \dots\}$ with no cluster points. Let P_n be point masses at x_n, that is, $P_n(B) = 1$ if $x_n \in B$ and $P_n(B) = 0$ otherwise. Then there exists a $w - *$ limit of a subnet P_{n_d} of the sequence of point masses P_n. Any such limit ψ satisfies:*

(i) For any set $B \subset X$ it holds either $\psi(B) = 0$ or $\psi(B) = 1$, with $\psi(S) = 1$.
(ii) ψ is a finitely (but not countably) additive set function
(iii) For every finite or empty set $B \subset X$, $\psi(B) = 0$.

The corresponding subnet n_d is the net based on the filter of sets of ψ-measure 1.

Proof. From the previous considerations it follows that $\{P_n\}$ has a cluster point. Clearly, we must have a directed set D and a net $x_d \in S$ such that

$$\lim f(x_d) = \int f(x)\, d\psi(x), \tag{10}$$

for some measure ψ in the unit ball of $\mathcal{M}(X)$ and for all $f \in \boldsymbol{C}(X)$. Then ψ is additive; further, it is a straightforward consequence of (10) that ψ has to be concentrated on S, i.e., $\psi(S) = 1$ and also that $\psi(B) = 0$ for any finite set B (otherwise, one could modify f in such a way that the right hand side of (10) changes without affecting the left hand side). Since $\psi(S) = 1$ and $\psi(x_n) = 0$ for each n, ψ is not countably additive. Hence, (ii) and (iii) are proved. It remains to show that $\psi(B)$ must be 0 or 1 for any B. Suppose contrary, that there exists a set B with $\psi(B) = q$, $0 < q < 1$. Then $S \cap B$ is neither a finite nor a cofinite set. For a d_0 being fixed, infinitely many x_d with $d \geq d_0$ belong either to B or to B'. By Lemma 5.5, there is an $f \in \boldsymbol{C}(X)$ which takes value 1 at $S \cap B$ and 0 at $S \cap B'$. Then the right hand side of (10) equals $\psi(B) = q$ and for any $\varepsilon > 0$ there is a d_0 so that for each $d \geq d_0$, $|f(x_d) - q| < \varepsilon$. Now suppose that B contains infinitely many x_d's for $d \geq d_0$, then we'd have $|1 - q| < \varepsilon$; otherwise $|q| < \varepsilon$ which are both impossible.

Therefore, we proved that any ψ which is a $w - *$ limit of a subnet $\{P_n\}$ satisfies conditions (i)–(iii). An analysis of the construction of Radon integral with respect to ψ, in 3.14, reveals that the subnet of convergence is the net based on the ultrafilter $\mathcal{F}$ consisted of sets with $\psi(B) = 1$.

5.7. Remarks. In any at least countable set X there exists a measure ψ which satisfies conditions (i)–(iii). This can be shown using theory of filters and the Axiom of Choice. The family of all sets $B \subset X$ with $\psi(B) = 1$ makes a non-principal ultrafilter. The existence of non-principal ultrafilters can be proved, but there is no concrete example of such a filter.

Theorem 5.6 holds, for instance, in any non-compact metric space.

6. Tightness and Prohorov's theorem

Although the notion of tightness can be defined in a more general context, in this section we observe only probability measures on metric spaces. Hence, X will denote a metric space, $\mathcal{B}$ a Borel sigma field and $\mathcal{M}_1(X)$ the set of all probability measures.

6.1. Definition. Let $\mathcal{P}$ be a set of probability measures on X. We say that $\mathcal{P}$ is *tight* if for any $\varepsilon > 0$ there is a compact set $K \subset X$ such that $\mu(K') \leq \varepsilon$.

The notion of tightness makes sense even if $\mathcal{P}$ is a singleton. In this case we have the following result.

6.2. Theorem. *If X is a complete and separable metric space, then each probability measure is tight.*

Proof. By separability of X, for each n there is a sequence $A_{n_1}, A_{n_2}, \ldots$ of open balls of radii $1/n$ that cover X. Choose i_n so that

$$P\Big(\bigcup_{i \leq i_n} A_{n_i}\Big) > 1 - \frac{\varepsilon}{2^n}$$

and let

$$K = \bigcap_{n \geq 1} \bigcup_{i \leq i_n} A_{n_i}.$$

Since $K \subset \bigcup_{i \le i_n} A_{i_n}$ for each n, K is totally bounded set in a complete metric space and hence $\bar{K}$ is compact. Further,

$$P(K') \le \sum_{n=1}^{+\infty} P\Big(\Big(\bigcup_{i \le i_n} A_{n_i}\Big)'\Big) < \varepsilon \sum_{i=1}^{+\infty} \frac{1}{2^n} = \varepsilon,$$

hence $P(\bar{K}') \le P(K') < \varepsilon$.

6.3. Definition. We say that a set $\mathcal{P}$ of probability measures is *relatively compact* if any sequence of probability measures $P_n \in \mathcal{P}$ contains a subsequence P_{n_k} which converges weak star to a probability measure in $\mathcal{M}_1(X)$. □

A precise topological term for relative compactness would be *relative sequential compactness in* $\mathcal{M}_1(X)$.

6.4. Remark. If X is compact, then from the previous section it follows that any set of probability measures is relatively compact. Otherwise, we need some conditions which are easier to check. One such condition is given in the next theorem. The proof presented here relies on the material of the previous section and departs from a classical presentation.

6.5. Theorem (Prohorov [**28**]). *Let X be an arbitrary metric space and let $\mathcal{P}$ be a tight set of measures. Then $\mathcal{P}$ is relatively compact.*

Proof. Let X be a metric space and let $\mathcal{P}$ be a tight set of Borel probability measures on it. Then for each $n \in \boldsymbol{N}$, let K_n be a compact subset of X such that $P(K_n) > 1 - 1/n$ for all $P \in \mathcal{P}$; we may assume that $K_1 \subset K_2 \subset \cdots$. A unit ball in any of spaces $\boldsymbol{C}(K_n)^*$ is compact and metrizable. For a given sequence $\{P_k\}$ of probability measures in $\mathcal{P}$, its restriction to a compact space K_n has a convergent subsequence. Then we can use a diagonal argument to show that there is a subsequence $P_{k'}$ such that

$$P_{k'} \implies P^{(n)} \quad \text{on } K_n, \quad n = 1, 2, \ldots \tag{11}$$

for some measures $P^{(n)}$ on K_n. Since K_n are increasing sets, the restriction of $P^{(n)}$ to K_{n-1} must coincide with $P^{(n-1)}$. Since $P^{(n)}$ is in the dual space of $\boldsymbol{C}(K_n)$, it is countably additive, and by (11), $P^{(n)}(K_n) \ge 1 - \varepsilon$.

Now if B is a Borel subset of X, define

$$P(B) = \lim_{n \to +\infty} P^{(n)}(B \cap K_n).$$

The limit here exists because of

$$P^{(n)}(B \cap K_n) \ge P^{(n)}(B \cap K_{n-1}) = P^{(n-1)}(B \cap K_{n-1}),$$

hence the sequence $\{P^{(n)}(B \cap K_n)\}$ is increasing and clearly is bounded from above by 1. To show that $P_{k'} \implies P$, we use the characterization of Theorem 4.6(iii). Let F be any closed set in X. From (11) we have that

$$\overline{\lim}\, P_{k'}(F \cap K_n) \le P^{(n)}(F \cap K_n) \quad \text{for each } n = 1, 2, \ldots$$

Further,

$$P_{k'}(F) \le P_{k'}(F \cap K_n) + \frac{1}{n}$$

and so

$$\overline{\lim}\, P_{k'}(F) \le \overline{\lim}\, P_{k'}(F \cap K_n) + \frac{1}{n} \le P^{(n)}(F \cap K_n) + \frac{1}{n}.$$

Letting now $n \to +\infty$ we get that

$$\overline{\lim}\, P_{k'}(F) \le P(F),$$

that is, $P_{k'} \implies P$.

6.6. Theorem. *Let X be a complete separable metric space. If $\mathcal{P}$ is a relatively compact set of probability measures on X, then $\mathcal{P}$ is tight.*

Proof. By Theorem 4.9, the weak star topology on $\mathcal{M}_1$ is metrizable, so relative sequential compactness of $\mathcal{P}$ as defined in 6.3 becomes the topological compactness of $\bar{\mathcal{P}}$, that is, any open cover of $\bar{\mathcal{P}}$ has a finite subcover. Here we understand that the closure of $\mathcal{P}$ is in the metric space $\mathcal{M}_1$. Without loss of generality, we may and will assume that $\mathcal{P}$ itself is compact in $\mathcal{M}_1$.

Fix $\varepsilon > 0$ and $\delta > 0$. If $P \in \mathcal{P}$, then by Theorem 6.2 it is tight, so there is a compact set K_P such that $P(K_P) > 1-\varepsilon/2$. Being compact, K_P is totally bounded, that is, it can be covered with finitely many open δ-balls $B_{P,i}$, $i = 1, 2, \dots, k_P$. Let $G_P = \bigcup_{i=1}^{k_P} B_{P,i}$. By Theorem 4.2, there is a neighborhood of P (in the weak star topology of $\mathcal{M}(X)$) of the form

$$U_P = \{\mu \mid \mu(G_P) > P(G_P) - \varepsilon/2\}$$

The family $\{U_P\}_{P\in\mathcal{P}}$ makes an open cover of $\mathcal{P}$ and hence there is a finite subcover, say $U_{P_1}, \dots, U_{P_m}$. Then let $K_\delta = \bigcup_{j=1}^{m} G_{P_j}$. For any $Q \in \mathcal{P}$ we have that

$$Q(G_P) > P(G_P) + \varepsilon/2 \ge P(K_P) - \varepsilon/2 > 1 - \varepsilon,$$

which implies that also $Q(K_\delta) > 1-\varepsilon$. Let now K be the closure of the intersection of all $K_{1/n}$; it is a closed and totally bounded set, hence compact, and we have that $Q(K) > 1 - \varepsilon$ for all $P \in \mathcal{P}$.

7. Weak convergence of probability measures on Hilbert spaces

In this section we firstly review basic fact related to the weak convergence of probability measures on finite dimensional vector spaces. The simple characteristic function technique which is usually applied there, becomes more complex on infinite dimensional Hilbert spaces.

7.1. Weak convergence of probability measures on $\boldsymbol{R}^k$. On finite dimensional spaces, the notion of weak convergence of probability measures coincides with the notion of convergence of distributions (see [**6**], for example). If F_n and

F are distribution functions of k-dimensional random variables X_n and X respectively, and if $\lim F_n(\boldsymbol{x}) = F(\boldsymbol{x})$ in each point $\boldsymbol{x} \in \boldsymbol{R}^k$ where F is continuous, then we say that the corresponding sequence X_n converges to X in distribution. This occurs if and only if $P_n \implies P$, where P_n and P are probability measures on $\boldsymbol{R}^n$ - distributions of X_n and X respectively.

A useful tool for investigation of weak convergence is the notion of a characteristic function. If P is a probability measure on $\boldsymbol{R}^k$, then its characteristic function is defined by

$$\varphi(\boldsymbol{t}) = \int_{\boldsymbol{R}^k} e^{i\langle \boldsymbol{t}, \boldsymbol{x}\rangle}\, dP(\boldsymbol{x}) \tag{12}$$

where

$$\boldsymbol{x} = (x_1, \dots, x_k), \quad \boldsymbol{t} = (t_1, \dots, t_k), \quad \langle \boldsymbol{t}, \boldsymbol{x}\rangle = \sum_{i=1}^{k} t_i x_i.$$

It is a well known fact (Bochner's theorem) that a function φ defined on $\boldsymbol{R}^k$ is a characteristic function of some probability measure if and only if it is positive definite, continuous at origin and $\varphi(0) = 1$.

It is also a basic fact that $P_n \implies P$ if and only if $\lim_n \varphi_n(\boldsymbol{t}) = \varphi(\boldsymbol{t})$, where φ_n and φ are the corresponding characteristic functions. This is indeed a very strong result, since it says that it suffices to test the condition (1) with only two (classes of) functions, $x \mapsto \cos\langle \boldsymbol{t}, \boldsymbol{x}\rangle$ and $x \mapsto \sin\langle \boldsymbol{t}, \boldsymbol{x}\rangle$.

7.2. Positive definite functions. Let X be any linear vector space. A complex valued function φ defined on X is said to be positive (or non-negative) definite if for any finite $A = (a_1, \dots, a_n) \in \boldsymbol{C}^n$ and $x = (x_1, \dots, x_n) \in X^n$ the following holds:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \bar{a}_j \varphi(x_i - x_j) \geq 0.$$

Positive definite functions have some interesting properties, which can be proved directly from the above definition, using an appropriate choice of A and x. We list some of these properties (see [**21**] for proofs):

$$\varphi(0) \geq 0 \tag{i}$$

$$\bar{\varphi}(x) = \varphi(-x) \tag{ii}$$

$$|\varphi(x)| \leq \varphi(0) \tag{iii}$$

$$|\varphi(x) - \varphi(y)|^2 \leq 2\varphi(0)(\varphi(0) - \operatorname{Re}\varphi(x - y)) \tag{iv}$$

$$\varphi(0) - \operatorname{Re}\varphi(2x) \leq 4(\varphi(0) - \operatorname{Re}\varphi(x)) \tag{v}$$

From (iv) it immediately follows that a positive definite function is uniformly continuous on X with respect to any metric topology if and only if its real part is continuous at zero.

7.3. Characteristic functions on Hilbert spaces. Let H be a real separable Hilbert space. Let $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ be the inner product and the norm which defines the topology on H and let $\{e_i\}$ be an orthonormal basis.

Let P be a probability measure on H. The corresponding characteristic function is defined formally in the same way as in finite dimensional spaces:

$$\varphi(x) = \int_H e^{i\langle x,y\rangle}\, dP(y), \qquad x \in H. \tag{13}$$

A characteristic function uniquely determines the corresponding measure [**27**].

It is easy to see that the function defined by (13) is positive definite and continuous at zero (with respect to the given norm). However, these properties are not sufficient for a function to be a characteristic function, as in finite dimensional cases. In order to proceed further, we need some facts about Hilbertian seminorms.

7.4. Hilbertian seminorms. A real valued function p defined on a vector space X is called a Hilbertian seminorm if for all $x, x_1, x_2 \in X$ and $a \in \boldsymbol{R}$:

(i) $p(x) \geq 0$
(ii) $p(ax) = |a|p(x)$
(iii) $p(x_1 + x_2) \leq p(x_1) + p(x_2)$
(iv) $p(x) > 0$ for some $x \in X$
(v) $p^2(x_1 + x_2) + p^2(x_1 - x_2) = 2(p^2(x_1) + p^2(x_2))$

Due to (v), for a Hilbertian seminorm p one can define the corresponding inner product:

$$p(x, y) = \frac{1}{4}(p^2(x_1 + x_2) - p^2(x_1 - x_2)). \tag{14}$$

Let now H be a Hilbert space. Besides its norm, one can define various Hilbertian seminorms on H. One such seminorm is, for example,

$$p_n(x) = \sqrt{\sum_{i=1}^{n} \langle x, e_i\rangle^2}, \qquad n \in \boldsymbol{N}, \tag{15}$$

where $\{e_i\}$ is an orthonormal basis with respect to the original norm. The inner product which corresponds to the seminorm (15) is given by

$$p_n(x, y) = \sum_{i=1}^{n} \langle x, e_i\rangle\langle y, e_i\rangle.$$

Let Π denotes the set of all Hilbertian seminorms p that satisfy

$$p(x) \leq C\|x\| \quad \text{for some } C > 0 \tag{16}$$

$$\sum_{i=1}^{+\infty} p^2(e_i) < +\infty, \tag{17}$$

for an orthonormal (with respect to the original norm) basis $\{e_i\}$. It can be shown that the quantity in (17) does not depend on the choice of an orthonormal basis in H (see [**22**] for the proof and more details). Seminorms p_n defined by (15) belong to Π, since $p_n(x) \leq \|x\|$ and $\sum p_n^2(e_i) = n$.

It is easy to see that

$$p \in \Pi \implies cp \in \Pi \quad \text{for any } c > 0 \tag{18}$$

and

$$p_1, \dots, p_n \in \Pi \implies \sqrt{p_1^2 + p_2^2 + \cdots + p_n^2} \in \Pi. \tag{19}$$

Now denote by $\mathcal{I}$ a topology on H defined by the following basis of neighborhoods at zero:

$$\{x \in H \mid p_1(x) < \varepsilon_1, \dots, p_n(x) < \varepsilon_n\},$$

where $p_1, \dots, p_n \in \Pi$, $n \in \boldsymbol{N}$, $\varepsilon_i > 0$. Equivalently, by (18) and (19), a basis of neighborhoods at zero for the $\mathcal{I}$-topology is given by

$$\{x \in H \mid p(x) < \varepsilon\}, \qquad p \in \Pi,\ \varepsilon > 0.$$

Then a sequence $\{x_n\}$ converges in the $\mathcal{I}$ - topology to x if and only if $\lim_n p(x_n - x) = 0$ for any $p \in \Pi$. The $\mathcal{I}$-topology is stronger than the norm topology. If a function φ defined on H is continuous in the $\mathcal{I}$-topology it must be norm continuous, but the converse does not hold.

7.5. Theorem. *Let φ be the characteristic function of a probability measure P on H. Then for any $\varepsilon > 0$ there is a seminorm $p_\varepsilon \in \Pi$ such that for all $x \in H$,*

$$1 - \operatorname{Re}\varphi(x) \leq p_\varepsilon^2(x) + \varepsilon \tag{20}$$

and φ is $\mathcal{I}$-continuous on H.

Proof. Since H is a complete separable normed space, by Theorem 6.2 the probability measure P is tight. That is, for a given $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset H$ such that $P(K') \leq \varepsilon/2$. So, we have that

$$\begin{aligned} 1 - \operatorname{Re}\varphi(x) = \int (1 - \cos\langle x, y\rangle)\, dP(y) &\leq \int_{K_\varepsilon} (1 - \cos\langle x, y\rangle)\, dP(y) + \varepsilon \\ &\leq \frac{1}{2} \int_{K_\varepsilon} \langle x, y\rangle^2\, dP(y) + \varepsilon \end{aligned}$$

Since K_ε is compact and $y \mapsto \langle x, y\rangle^2$ is a continuous function, then it is bounded on K_ε and we may define

$$p_\varepsilon = \Big(\frac{1}{2} \int_{K_\varepsilon} \langle x, y\rangle^2\, dP(y)\Big)^{1/2}. \tag{21}$$

It is now easy to show that p_ε is a Hilbertian seminorm which satisfies (16) and (17), hence (20) is proved. From the inequality (iv) in 7.2 and (20) we have that

$$|\varphi(x) - \varphi(y)|^2 \leq 2(p_\varepsilon^2(x-y) + \varepsilon),$$

which implies the uniform continuity of φ in $\mathcal{I}$ topology.

7.6. Example. Consider the function $f(x) = e^{-\|x\|^2/2}$. It is norm continuous and $f(0) = 1$. Using Schoenberg's theorem [**5**] it can be shown that f is a positive definite function. Suppose that f is $\mathcal{I}$-continuous. Then the norm is also $\mathcal{I}$-continuous, which implies that there is a $p \in \Pi$ and a $\delta > 0$ so that $p(x) < \delta \implies \|x\| < 1/2$. For such a p we have that $\sum p^2(e_i) < +\infty$, hence there is an e_i such that $p(e_i) < \delta$ and so $\|e_i\| < 1/2$, which is a contradiction.

By Theorem 7.5, f is not a characteristic function on H. □

The next theorem is proved by Sazonov [**29**].

7.7. Theorem. *A function $\varphi : H \mapsto \boldsymbol{C}$ is the characteristic function of a probability measure if and only if it is positive definite, $\mathcal{I}$-continuous at zero and $\varphi(0) = 1$.*

7.8. Weak convergence on H via characteristic functions. Contrary to finite dimensional cases, the convergence of characteristic functions alone is not sufficient for weak convergence of probability measures. Here is where relative compactness of probability measures plays a key role.

Theorem. *Let $\{P_n\}$ be a sequence of probability measures on H and let φ_n be the corresponding characteristic functions. Let P and φ be a probability measure and its characteristic function. If $P_n \implies P$ then $\lim_n \varphi_n(x) = \varphi(x)$ for all $x \in H$.*

Conversely, if a sequence P_n of probability measures on H is relatively compact and $\lim_n \varphi_n(x) = \varphi(x)$ for all $x \in H$, then there exists a probability measure P such that φ is its characteristic function and $P_n \implies P$.

Proof. Since the mapping $x \mapsto e^{i\langle x,y\rangle}$ is norm-continuous, we have that $P_n \implies P$ implies $\varphi_n(x) \to \varphi(x)$ for all $x \in H$. To show the converse, assume that $\{P_n\}$ is relatively compact and that $\lim_n \varphi_n(x) = \varphi(x)$ for all $x \in H$, but $\{P_n\}$ does not converge weakly. Then there are two subsequences $\{P_{n'}\}$ and $\{P_{n''}\}$ with different limits, $P^{(1)}$ and $P^{(2)}$. Then characteristic functions $\varphi_{n'}$ and $\varphi_{n''}$ converge to different limits (i.e., to characteristic functions of $P^{(1)}$ and $P^{(2)}$ respectively), which is a contradiction to the assumption that $\{\varphi_n\}$ is a convergent sequence.

7.9. Example. Let P_n be point masses at e_n. The corresponding characteristic functions are $\varphi_n(x) = e^{ix_n}$, where $x_n = \langle x, e_n\rangle$. Then for every $x \in H$, $\lim_n \varphi_n(x) = 1$ and 1 is the characteristic function of the point mass at zero, P_0. But clearly, $\{P_n\}$ is not a weakly convergent sequence, assuming that H is equipped with the norm topology. To show that exactly, note that if $P_n \implies P$ in the norm topology, then P can only be P_0 because of convergence of characteristic functions. Now since H is a normed space, there is an $f \in \boldsymbol{C}(H)$ such that $f(B_{1/4}) = 1$ and $f(B'_{1/2}) = 0$, where B_r is the ball centered at zero with the radius r. For such a function we have

$$\int f(x)\,dP_n(x) = f(e_n) = 0 \quad \text{and} \quad \int f(x)\,dP_0(x) = f(0) = 0,$$

so P_n does not weakly converge to P.

7.10. Weak convergence on H with respect to strong and weak topology. In this item, we observe H with the strong (norm) topology (H_s) and with the weak topology as defined in 2.2 (H_w). Although, by 2.10, the Borel sets are the same in both cases, there is a difference in the concepts of weak convergence of measures, arising from the fact that $\boldsymbol{C}(H_w)$ is in general a proper subset of $\boldsymbol{C}(H_s)$. Hence, if $P_n \implies P$ in H_w, we need some additional requirements to conclude that $P_n \implies P$ in H_s, unless H is finite dimensional, in which case $H_w = H_s$. In the next two theorems we show that an additional necessary and sufficient condition is "uniform finite dimensional approximation", expressed by (22) below.

7.11. Theorem. *Let $\{P_n\}$ be a sequence of probability measures on H. If $P_n \implies P$ in H_w and for all $\varepsilon > 0$,*

$$\lim_N \sup_n P_n\Big(\sum_{i=N}^{+\infty} \langle x, e_i\rangle^2 \geq \varepsilon\Big) = 0, \tag{22}$$

then $P_n \implies P$ in H_s.

Proof. By Theorem 4.6(ii), it suffices to show that, under the above assumptions,

$$\int f(x)\,dP_n(x) \to \int f(x)\,dP(x), \tag{23}$$

for every uniformly norm continuous and bounded function f. The idea of the proof is to approximate f by a function which is continuous and bounded in H_w. For $x \in H$ let $g_N(x) = \sum_{i=1}^{N-1} \langle x, e_i\rangle e_i$. Then g_N is a linear operator $H \mapsto H$, $\|g_N(x)\| \leq (N-1)\|x\|$ for all $x \in H$ and hence $\|g_N(x) - g_N(y)\| \leq (N-1)\|x - y\|$, so g_N is uniformly continuous. Let now f be any norm continuous real valued function on H_s, with $\|f\| = M_f$. The function $x \mapsto f(g_N(x))$ is continuous in H_w. Consider the difference $d_N(x) = f(x) - f(g_N(x))$ and fix a $\delta > 0$. By uniform continuity of f, there is an $\varepsilon > 0$ so that

$$|d_N(x)| < \delta \quad \text{whenever} \quad \|x - g_N(x)\|^2 = \sum_{i=N}^{+\infty} \langle x, e_i\rangle^2 < \varepsilon.$$

By (22), for such an ε we can find N_0 so that for each $N \geq N_0$ we have

$$P_n(A_N) < \delta \quad \text{for every } n, \ \text{ where } A_N = \Big\{x \in H \ \Big|\ \sum_{i=N}^{+\infty} \langle x, e_i\rangle^2 \geq \varepsilon\Big\}. \tag{24}$$

Now for any $N \geq N_0$ we have

$$\begin{aligned} \Big|\int f(x)\,dP_n(x) - \int f(g_N(x))\,dP_n(x)\Big| &= \Big|\int d_N(x)\,dP_N(x)\Big| \\ &\leq \Big|\int_{A'_N} d_N(x)\,dP_n(x)\Big| + \Big|\int_{A_N} d_N(x)\,dP_n(x)\Big| \\ &\leq \delta + 2M_f\delta = \delta(1+2M_f). \end{aligned} \tag{25}$$

Further, by continuity of f, the fact that $g_N(x) \to x$ as $N \to \infty$ and the dominated convergence theorem, for $N \geq N_1$ we have

$$\Big|\int f(g_N(x))\,dP(x) - \int f(x)\,dP(x)\Big| \leq \delta \tag{26}$$

Finally, since $f(g(\cdot)) \in \boldsymbol{C}(H_w)$, we have that

$$\Big|\int f(g_N(x))\,dP_N(x) - \int f(g_N(x))\,dP(x)\Big| \leq \delta, \tag{27}$$

where $N \geq \max(N_0, N_1)$. The weak convergence of P_n in H_s follows now from (25)–(27).

7.12. Theorem. *Any weak star convergent sequence of probability measures* $\{P_n\}$ *in* H_s *satisfies (22).*

Proof. Suppose that $P_n \implies P$ in H_s. For an $\varepsilon > 0$, let A_N be defined by (24). Since A_N is closed in the norm topology, by Theorem 4.6(iv) we have that, for any fixed N,

$$\overline{\lim}\, P_n(A_N) \leq P(A_N). \tag{28}$$

Since $\{A_N\}$ is a decreasing sequence of sets with $\bigcap_{N=1}^{+\infty} A_N = \emptyset$, by continuity of probability measures we have that

$$\lim_{N\to+\infty} P(A_N) = 0. \tag{29}$$

Now fix a $\delta > 0$ and choose N_0 large enough so that $P(A_{N_0}) \leq \delta$ for $N \geq N_0$. By (28), there are only finitely many measures, say $P_{n_1}, \ldots, P_{n_k}$ such that $P_{n_i}(A_{N_0}) > \varepsilon$; however, by continuity, there is an integer $N_1 > N_0$ such that $P_{n_i}(A_N) \leq \varepsilon$ for all $N \geq N_1$. Hence, for $N \geq N_1$ we have that

$$\sup_n P_n(A_N) \leq \delta,$$

which is equivalent with (22).

7.13. Theorem. *Let $\mathcal{P}$ be a relatively compact set of probability measures in H_w. Then it is relatively compact in H_s if and only if*

$$\lim_N \sup_{P\in\mathcal{P}} P\Big(\sum_{i=N}^{+\infty}\langle x,e_i\rangle^2 \geq \varepsilon\Big) = 0, \tag{30}$$

for any $\varepsilon > 0$.

Proof. By Theorem 7.11, (30) is a sufficient additional condition for relative compactness in H_s. Conversely, suppose that $\mathcal{P}$ is relatively compact in H_s but (30) does not hold. Then there is an $\varepsilon > 0$ and a $\lambda > 0$ such that for each n there is an $N \geq n$ and $P_n \in \mathcal{P}$ so that

$$P_n\Big(\sum_{i=N}^{+\infty}\langle x,e_i\rangle^2 \geq \varepsilon\Big) > \lambda. \tag{31}$$

By weak compactness in H_s, there is a weakly convergent subsequence of $\{P_n\}$, which together with (31) contradicts Theorem 7.12. Therefore, (30) holds.

7.14. Theorem (Prohorov's theorem in H_w). *For $r = 1, 2, \dots$, let $B_r = \{x \in H \mid \|x\| \leq r\}$. A set of probability measures $\mathcal{P}$ is relatively compact on H_w if and only if for every $\varepsilon > 0$ there is an integer $r \geq 1$ such that $P(B_r') \leq \varepsilon$ for all $P \in \mathcal{P}$.*

Proof. A key point in the proof is the observation that H_w may be represented as the union of balls $B_r = \{x \in H \mid \|x\| \leq r\}$, $r = 1, 2, \dots$, which are compact sets in the weak topology (Theorem 2.18) and the weak topology on B_r is metrizable. So, the proof of the "if" part goes in the same way as the proof of Theorem 6.5.

For the converse, it suffices to prove that if P_n is a weakly convergent sequence in H_w then for each $\varepsilon > 0$ there is a ball B_r such that $P_n(B_r') \leq \varepsilon$ for all n. Indeed, assuming that we proved such a claim, suppose that $\mathcal{P}$ is relatively compact and that there is an $\varepsilon > 0$ such that for each positive integer n there is a measure $P_n \in \mathcal{P}$ with $P(B_n') > \varepsilon$. Then there is a subsequence $P_{n'}$ which is weakly convergent in H_w to some probability measure P, and we have that $P_{n'}(B_{n'}') > \varepsilon$. Since $n' \to +\infty$, this is a contradiction with the assumed claim.

So, let $P_n \implies P$ in H_w, where P_n and P are probability measures. Then by continuity of P, for any $\varepsilon > 0$ there is a ball B_r such that $P(B_r) \geq 1 - \varepsilon/2$. Consider now the open sets (in fact, U-sets) in H_w:

$$G_{k,m}^{(r)} = \Big\{x \in H \;\Big|\; \sum_{i=1}^{k}\langle x,e_i\rangle^2 < r^2 + \frac{1}{m}\Big\}, \quad k, m = 1, 2, \dots$$

Then it is easy to see that the sets $G_{k,m}^{(r)}$ are decreasing as k and m increase. Moreover,

$$B_r = \bigcap_{k=1}^{+\infty}\bigcap_{m=1}^{+\infty} G_{k,m}^{(r)}.$$

By Theorem 4.4, for each k and m we have

$$\underline{\lim}\, P_n(G_{k,m}^{(r)}) \geq P(G_{k,m}^{(r)}) \geq 1 - \frac{\varepsilon}{2},$$

hence there is an n_0 such that

$$P_n(G_{k,m}^{(r)}) \geq 1 - \varepsilon \quad \text{for all } n \geq n_0.$$

For a fixed n, letting here $k \to +\infty$ and $m \to +\infty$, we get

$$P_n(B_r) \geq 1 - \varepsilon \quad \text{for all } n \geq n_0.$$

Now, for each of measures P_i $(1 \leq i \leq n_0 - 1)$ there is a ball B_{r_i} such that $P_i(B_{r_i}) \geq 1 - \varepsilon$. Let $R = \max\{r, r_1, r_2, \dots, r_{n_0-1}\}$. Then $P_n(B_R) \geq 1 - \varepsilon$ for all $n \geq 1$, which was to be proved.

7.15. Example. Let P_n be unit masses at e_n, as in Example 7.9. We will show that P_n is relatively compact in H_w. Indeed, all P_n are concentrated in B_1, hence by Theorem 7.14, the sequence $\{P_n\}$ is relatively compact. Moreover, since $e_n \to 0$ in the weak topology (Example 2.3), then $f(e_n) \to f(0)$ for any $f \in \boldsymbol{C}(H_w)$. Hence, $\int f(x)\, dP_n(x) = f(e_n) = \int f(x)\, dP_0(x)$, where P_0 is the unit mass at 0. So, $\{P_n\}$ is a weakly convergent sequence in H_w, $P_n \implies P_0$, which is also consistent with Example 7.9.

7.16. Relative compactness via characteristic functions. In the next two theorems, we give conditions for relative compactness in H_w and H_s in terms of characteristic functions. Recall that by Theorem 7.5, to each characteristic function φ and an $\varepsilon > 0$ there corresponds a Hilbertian seminorm p_ε such that (20) holds. Let $\mathcal{P} = \{P_\alpha\}$ be a set of probability, where α belongs to an index set A. A seminorm which corresponds to the characteristic function φ_α of a given P_α with an $\varepsilon > 0$ in the sense of (20), will be denoted by $p_{\alpha,\varepsilon}$. Let us note that $p_{\alpha,\varepsilon}$ are not uniquely determined. One natural choice is given by (21).

In the proofs of the next two theorems, a key role is played by the integration of a Hilbertian seminorm with respect to a finite dimensional Gaussian measure. Let p be a Hilbertian seminorm and $p(\cdot,\cdot)$ a corresponding inner product as in (14). Suppose $\mathcal{G}$ is an N-dimensional Gaussian measure which is concentrated on $\boldsymbol{R}^N$ spanned by $\{e_1, e_2, \dots, e_N\}$, as a product of N coordinate measures $\mathcal{N}(0, \sigma^2)$. Then

$$\begin{aligned}
\int p^2(x)\, d\mathcal{G}(x) &= \int_{\boldsymbol{R}^N} p\Big(\sum_{i=1}^N x_i e_i, \sum_{j=1}^N x_j e_j\Big)\, d\mathcal{G}(x) \\
&= \sum_{i=1}^N \sum_{j=1}^N \int_{\boldsymbol{R}^N} x_i x_j p(e_i, e_j)\, d\mathcal{G}(x) \\
&= \sum_{i=1}^N \int_{\boldsymbol{R}^N} x_i^2 p(e_i, e_i)\, d\mathcal{G}(x) \\
&= \sigma^2 \sum_{i=1}^N p^2(e_i). \qquad (32)
\end{aligned}$$

7.17. Theorem *A set of probability measures $\mathcal{P} = \{P_\alpha \mid \alpha \in A\}$ is relatively compact in H_w if and only if for every $\varepsilon > 0$ there is a set of seminorms $\{p_{\alpha,\varepsilon}\}_{\alpha \in A}$ such that*

$$\sup_{\alpha \in A} \sum_{i=1}^{+\infty} p_{\alpha,\varepsilon}^2(e_i) < +\infty. \tag{33}$$

Proof. By Theorem 7.14, we need to show that the condition (33) is equivalent to the condition that for any $\varepsilon > 0$ there is a ball B_r such that

$$P_\alpha(B_r') \le \varepsilon \qquad \text{for all } P_\alpha \in \mathcal{P}. \tag{34}$$

For a given $\varepsilon > 0$, assume that (34) holds for $\varepsilon/2$ in place of ε and with some B_r. Then, as in the proof of Theorem 7.5, we show that

$$1 - \operatorname{Re} \varphi_\alpha(x) \le p_{\alpha,\varepsilon} + \varepsilon,$$

where φ_α is the characteristic function of P_α and

$${}^2_{\alpha,\varepsilon}(x) = \frac{1}{2} \int_{B_r} \langle x, y \rangle^2 \, dP_\alpha(y). \tag{35}$$

Then we have that

$$\sum_{i=1}^{+\infty} p_{\alpha,\varepsilon}^2(e_i) = \frac{1}{2} \int_{B_r} \|y\| \, dP_\alpha(y) \le r^2,$$

so (33) holds.

Conversely, fix an $\varepsilon > 0$ and assume that (33) holds for some family of seminorms $\{p_{\alpha,\varepsilon}\}$. Let

$$A_{r,N} = \Big\{ y \in H \;\Big|\; \sum_{i=1}^{N} \langle y, e_i \rangle^2 > r^2 \Big\}, \quad N = 1, 2, \ldots$$

Note that $A_{r,1} \subset A_{r,2} \subset \cdots$ and $\bigcup_{N=1}^{+\infty} A_{r,N} = B_r'$, so

$$\lim_{N \to +\infty} P_\alpha(A_{r,N}) = P_\alpha(B_r') \tag{36}$$

For an $y \in A_{r,N}$ we have that

$$1 - \exp\Big(-\frac{1}{2r^2} \sum_{i=1}^{N} \langle y, e_i \rangle^2 \Big) > 1 - e^{-1/2} > \frac{1}{3}$$

and so, for every $P_\alpha \in \mathcal{P}$,

$$\frac{1}{3} P_\alpha(A_{r,N}) < \int_{A_{r,N}} \Big(1 - \exp\Big(-\frac{1}{2r^2}\sum_{i=1}^N y_i^2\Big)\Big)\, dP_\alpha(y)$$

$$< 1 - \int_H \exp\Big(-\frac{1}{2r^2}\sum_{i=1}^N y_i^2\Big)\, dP_\alpha(y). \tag{37}$$

Let $\mathcal{G}$ be a Gaussian measure on $\boldsymbol{R}^N$, defined as the product of coordinate Gaussian measures $\mathcal{N}(0, 1/r^2)$. Its characteristic function is $y \mapsto \exp\Big(-\sum_{i=1}^N y_i^2/2r^2\Big)$ and we have

$$\int_H \exp\Big(-\frac{1}{2r^2}\sum_{i=1}^N y_i^2\Big)\, dP_\alpha(y) = \int_{\boldsymbol{R}^N}\int_H \exp\Big(i\sum_{i=1}^N y_i x_i\Big)\, dP_\alpha(y)\, d\mathcal{G}(x). \tag{38}$$

Let

$$\varphi_{\alpha,N}(x) = \int_H \exp\Big(i\sum_{i=1}^N y_i x_i\Big)\, dP_\alpha(y).$$

For $x = \sum_{i=1}^N x_i e_i$ we have that $\varphi_{\alpha,N}(x) = \varphi_\alpha(x)$ and so $\operatorname{Re}\varphi_{\alpha,N}(x) \ge 1 - p^2_{\alpha,\varepsilon}(x) - \varepsilon$. Then from (38) we get

$$\int \exp\Big(-\frac{1}{2r^2}\sum_{i=1}^N y_i^2\Big)\, dP_\alpha(y) \ge 1 - \int_{\boldsymbol{R}^N} p^2_{\alpha,\varepsilon}(x)\, d\mathcal{G}(x) - \varepsilon$$

$$= 1 - \frac{1}{r^2}\sum_{i=1}^N p^2_{\alpha,\varepsilon}(e_i) - \varepsilon. \tag{39}$$

From (37) and (39) we find that

$$P_\alpha(A_{r,N}) < \frac{3}{r^2}\sum_{i=1}^N p^2_{\alpha,\varepsilon}(e_i) + 3\varepsilon.$$

Now let $N \to +\infty$ and use (36) to get

$$P_\alpha(B'_r) \le \frac{3}{r^2}\sum_{i=1}^{+\infty} p^2_{\alpha,\varepsilon}(e_i) + 3\varepsilon,$$

which proves that (33) implies (34).

7.18. Theorem. *Suppose that $\mathcal{P} = \{P_\alpha \mid \alpha \in A\}$ is a relatively compact set of probability measures in H_w. Then the following conditions are equivalent:*

(i) For any $\varepsilon > 0$ *there is a choice of* $\{p_{\alpha,\varepsilon}\}_{\alpha\in A}$ *so that*

$$\lim_{N\to+\infty} \sup_{\alpha\in A} \sum_{i=N}^{+\infty} p^2_{\alpha,\varepsilon}(e_i) = 0,$$

(ii) For any $\varepsilon > 0$,

$$\lim_{N\to+\infty} \sup_{\alpha\in A} P_\alpha\Big(\sum_{i=N}^{+\infty} \langle x, e_i\rangle^2 > \varepsilon\Big) = 0.$$

Proof. Assume that (i) holds. From (20) it follows that

$$\operatorname{Re} \int \exp(i\langle x, y\rangle)\, dP_\alpha(x) \geq 1 - \varepsilon - p^2_{\alpha,\varepsilon}(y).$$

Put here $y = \sum_{j=N}^{S} a_j e_j$ and integrate with respect to the product of coordinate Gaussian $\mathcal{N}(0,1)$ measures to get

$$\int \exp\Big(-\frac{1}{2}\sum_{j=N}^{S} \langle x, e_j\rangle^2\Big)\, dP_\alpha(x) \geq 1 - \varepsilon - \sum_{j=N}^{S} p^2_{\alpha,\varepsilon}(e_j).$$

Now letting $S \to +\infty$ and using the monotone convergence theorem, we obtain

$$\int \exp\Big(-\frac{1}{2}\sum_{j=N}^{+\infty} \langle x, e_j\rangle^2\Big)\, dP_\alpha(x) \geq 1 - \varepsilon - \sum_{j=N}^{+\infty} p^2_{\alpha,\varepsilon}(e_j). \tag{40}$$

Introduce the notations:

$$\sum_{j=N}^{+\infty} p^2_{\alpha,\varepsilon}(e_j) = S_{\alpha,\varepsilon}(N), \quad \frac{1}{2}\sum_{j=N}^{S} \langle x, e_j\rangle^2 = X(N).$$

Then for any $\lambda > 0$, (40) yields:

$$\begin{aligned} 1 - \varepsilon - S_{\alpha,\varepsilon}(N) &\leq \int \exp(-X(N))\, dP_\alpha(x) \\ &= \int_{X(N)<\lambda} \exp(-X(N))\, dP_\alpha(x) + \int_{X(N)\geq\lambda} \exp(-X(N))\, dP_\alpha(x) \\ &\leq P_\alpha(X(N) < \lambda) + e^{-\lambda} P_\alpha(X(N) \geq \lambda) \\ &= 1 - (1 - e^{-\lambda}) P_\alpha(X(N) \geq \lambda), \end{aligned}$$

wherefrom we get

$$\sup_\alpha P_\alpha(X(N) \geq \lambda) < \frac{\varepsilon + \sup_\alpha S_{\alpha,\varepsilon}(N)}{1 - e^{-\lambda}}.$$

Letting here $N \to +\infty$ and then $\varepsilon \to 0$ and assuming that (i) holds, we obtain (ii) (with 2λ in place of ε). Note that this part is independent of the assumption that $\mathcal{P}$ is relatively compact in H_w.

To prove the opposite direction, assume that $\mathcal{P}$ is relatively compact in H_w and that (ii) holds. Let $p_{\alpha,\varepsilon}$ be seminorms defined by (35). Then we have

$$2\sum_{i=N}^{+\infty} p_{\alpha,\varepsilon}^2(e_i) = \int_{B_r} \sum_{i=N}^{+\infty} \langle x, e_i\rangle^2 \, dP_\alpha(x) \le \lambda + r^2 P_\alpha\Big(\sum_{i=N}^{+\infty} \langle x, e_i\rangle^2 \ge \lambda\Big).$$

Taking the supremum with respect to α, letting $N \to +\infty$ and $\lambda \to 0$, we obtain (i).

7.19. Theorem. *A set $\mathcal{P} = \{P_\alpha \mid \alpha \in A\}$ of probability measures on H_s is relatively compact if and only if for every $\varepsilon > 0$ there is a set of seminorms $\{p_{\alpha,\varepsilon}\}_{\alpha\in A}$, related to P_α as in (20), such that the following two conditions hold:*

(i) For every $\varepsilon > 0$,

$$\sup_{\alpha\in A} \sum_{i=1}^{+\infty} p_{\alpha,\varepsilon}^2(e_i) < +\infty.$$

(ii) For every $\varepsilon > 0$,

$$\lim_{N\to+\infty} \sup_{\alpha\in A} \sum_{i=N}^{+\infty} p_{\alpha,\varepsilon}^2(e_i) = 0.$$

Proof. Directly from theorems 7.13, 7.17 and 7.18.

7.20. Remarks. Let us remark that if the above conditions on seminorms hold for one choice of the family $p_{\alpha,\varepsilon}$, they need not hold for some other choice. For instance, suppose that $\alpha = 1, 2, \ldots$ and let $p_{n,\varepsilon}$ be a family of seminorms related to characteristic functions via (20) and satisfying (33). Then the family of seminorms $q_{n,\varepsilon}$ defined by $q_{n,\varepsilon}^2(x) = np_{n,\varepsilon}^2(x)/\sum_{i=1}^{+\infty} p_{n,\varepsilon}^2(e_i)$ also satisfies (20), but not (33).

Theorem 7.19 is proved in [**27**] by different means. The analysis of a relationship between weak convergence on H_w and H_s is adopted from [**23**]. Separate conditions in H_w may be useful since they are easier to check.

Acknowledgements. My understanding of the material presented in Section 5 is much improved after communications with Sheldon Axler [**3**] and several mathematicians of the Internet based Real Analysis group [**14**]. The presentation in Section 2 is influenced by Sheldon Axler's brilliant lectures of Functional Analysis at Michigan State University [**2**].

References

[1] A.D. Alexandroff, *Additive set functions in abstract spaces I–III*, Mat. Sb. **8(50)** (1940), 307–348,

[2] S. Axler, *Functional analysis*, course notes (1983–1984), Michigan State University, unpublished.

[3] S. Axler, private communication, 1997

[4] R.G. Bartle, N. Dunford, J. Schwartz, *Weak compactness and vector measures*, Canad. J. Math. **7** (1955), 289–305.

[5] C. Berg, J. Christensen, P. Ressel, *Harmonic Analysis on Semigroups*, Springer-Verlag, New York, 1984.

[6] P. Billingsley, *Convergence of Probability Measures*, Wiley, 1968.

[7] P. Billingsley, *Weak convergence of measures: Applications in Probability*, CBMS-NSF regional conference series in Applied Mathematics, SIAM, Philadelphia, 1971.

[8] G. Birkoff, *Moore-Smith convergence in general topology*, Ann. Math. **38** (1937), 39–56.

[9] N. Bourbaki, *Topologie Géneérale*, Hermann, Paris; Russian translation, Nauka, Moskva, 1968.

[10] H. Cartan, *Théorie des Filtres*, C.R. Acad. Sci. Paris **205** (1937), 595–598.

[11] H. Cartan, *Filtres et Ultrafiltres*, C.R. Acad. Sci. Paris **205** (1937), 777–779.

[12] C.C. Chang, H.J. Keisler, *Model Theory*, North-Holland, 1973.

[13] N. Dunford, J.T. Schwartz, *Linear Operators, Part I: General Theory*, Interscience, New York, London, 1958; Russian translation: Moskva, 1962.

[14] Gerald A. Edgar, John Hagood, David Ross, Eric Schechter, Gord Sinnamon, Contributions to Real-Analysis Discussion Group, December 1998.

[15] R. Engelking, *General Topology*, second edition, PWN, Warszaw, 1985.

[16] H. Heyer, *Probability measures on locally compact groups*, Springer-Verlag, 1977.

[17] K. Ito, *Regularization of linear random functionals*, Probability Theory and Mathematical Statistics (Tbilisi), Proceedings, 257–267, Lecture Notes in Mathematics, Vol. 1021, Springer-Verlag, New York–Berlin (1982).

[18] J.L. Kelley, *Convergence in Topology*, Duke Math. J. **17** (1950), 277–283.

[19] A.N. Kolmogorov, *Foundations of the Theory of Probability*, Second English Edition, Chelsea, New York, 1956.

[20] M. Merkle, *Towards a linear limes*, unpublished manuscript, 1997.

[21] M. Merkle, *On positive definite functions defined on vector spaces*, Univ. Beograd, Publ. Elektrotehn. Fak. Ser. Mat **1** (1990), 35–40.

[22] M. Merkle, *Multi-Hilbertian spaces and their duals*, Techn. Report No 291, Center for Stochastic Processes, Department of Statistics, University of North Carolina at Chapel Hill, 1990.

[23] M. Merkle, *On weak convergence of measures on Hilbert spaces*, J. Multivariate Analysis **29**:2 (1989), 252–259.

[24] E.H. Moore, *Definition of limit in general integral analysis*, Proc. Nat. Acad. Sci. **1** (1915), 628.

[25] E.H. Moore, E. H., *General Analysis I, part II*, Mem. Amer. Philos. Soc., Philadelphia, 1939.

[26] E.H. Moore, H.L. Smith, *A general theory of limits*, Amer. J. Math. **44** (1922), 102–121.

[27] K.R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, 1967.

[28] Yu.V. Prohorov, *Shodimost slučainih processov i predelynie teoremi teorii veroyatnostei*, Teor. ver. primenen. **1** (1956), 177–238.

[29] V. Sazonov, *A remark on characteristic functionals*, Theory Probab. Appl. **3**:2 (1958), 188–192.

[30] D. Stroock, *Probability Theory, an Analytic View*, Cambridge University Press, 1993.

[31] G.P. Tolstov, *Measure and Integral*, Nauka, Moscow, 1976. (in Russian)

[32] T. Tjur, *Probability Based on Radon Measures*, Wiley, 1980.

[33] J.W. Tukey, *Convergence and Uniformity in Topology*, Ann. of Math. Studies 2, Princeton University Press, 1940.

[34] V.S. Varadarajan, *Meri na topologičeskih prostranstvah*, Matem. sb. **55(97)** (1961), 35–100.

[35] S. Willard, S., *General Topology*, Addison-Wesley, 1970

[36] K. Yosida, *Functional Analysis*, Academic Press, New York 1965.