

THE LINEAR TOPOLOGY ASSOCIATED WITH WEAK CONVERGENCE OF PROBABILITY MEASURES

LIANG HONG

ABSTRACT. This expository note aims at illustrating weak convergence of probability measures from a broader view than a previously published paper. Though the results are standard for functional analysts, this approach is rarely known by statisticians and our presentation gives an alternative view to most standard probability textbooks. In particular, this functional approach clarifies the underlying topological structure of weak convergence. We hope this short note is helpful for those who are interested in weak convergence as well as instructors of measure theoretic probability.

1. INTRODUCTION

Weak convergence of probability measures is often defined as follows [2, 4].

Definition 1.1. *Let X be a metrizable space. A sequence $\{P_n\}$ of probability measures on X is said to **converge weakly** to a probability measure P if*

$$\lim_{n \rightarrow \infty} \int_X f(x) dP_n(dx) = \int_X f(x) P(dx)$$

for every bounded continuous function f on X .

It is natural to ask the following questions:

1. Why the type of convergence defined in Definition 1.1 is called “weak convergence”?
2. How weak it is?
3. Is there any connection between weak convergence of probability measures and weak convergence in functional analysis?
4. Why we do not use a metric to describe it?

Answers to these questions are not obvious from Definition 1.1. Varadarajan [5] gives answers to Questions 3 and 4 under the assumption that X is either compact or separable. In addition, Varadarajan [5] derives several important properties of the space of probability measures.

Theorem 1.2. [5] *Suppose X is a metrizable space and $\mathcal{P}(X)$ is the space of probability measures on X . Then the following statements hold.*

- (1) X is separable if and only if $\mathcal{P}(X)$ is separable.
- (2) X is compact if and only if $\mathcal{P}(X)$ is compact.
- (3) X is Polish if and only if $\mathcal{P}(X)$ is Polish.

Remark. To be precise, the topology on $\mathcal{P}(X)$ is the weak* topology $\sigma(\mathcal{P}(X), C_b(X))$. We shall define this in Section 2.

In the next section, we answer Questions 1-4 using the theory of topological vector spaces. This approach has certain advantages: (1) it clearly illustrates the underlying topological structure of weak convergence; (2) it may lead to simple and clean proofs; (3) it allows one to work under minimal hypotheses. As a result, our framework is more general than that of Varadarajan [5] since we only require X to be a metrizable space, that is, X need not be compact or separable.

2. WEAK CONVERGENCE OF PROBABILITY MEASURES: A TOPOLOGICAL VECTOR SPACE POINT OF VIEW

Varadarajan [5] explains weak convergence of probability measures from the Banach space point of view. In this section, we attempt to illustrate weak convergence from the topological vector space point of view. Since a topological vector space need not be a normed space, our setup is more general than that of Varadarajan [5]. In particular, we do not assume that X is compact or separable. We will mainly follow Aliprantis and Border [1] and Bourbaki [3]. At this point, we recommend that readers consult Chapter II, pp. 40-43 of Bourbaki [3] before proceeding further. Unless otherwise stated, X denotes a metrizable space, \mathcal{B}_X denotes the Borel σ -algebra on X , $\Psi(X)$ denotes the space of all finite signed measures on (X, \mathcal{B}_X) , $\mathcal{P}(X)$ denotes the space of all probability measures on (X, \mathcal{B}_X) , $C_b(X)$ denotes the space of all bounded continuous functions on X , and $C_u(X)$ denotes the space of all bounded d -uniformly continuous functions on X , where d is a compatible metric. Define a map from $(C_b(X), \Psi(X))$ to R by

$$(f, \mu) \mapsto \langle f, \mu \rangle \equiv \int_X f d\mu.$$

It is evident that the above map is linear in each variable separately, i.e., it is a bilinear form. Moreover, the following theorem shows that the space $C_u(X)$ and $\Psi(X)$ each separate points of the other. Since each bounded continuous function on X can be approximated pointwisely by a sequence of bounded d -Lipschitz continuous functions [1, Corollary 3.15], $C_b(X)$ and $\Psi(X)$ each separate points of the other.

Theorem 2.1. [5] *For any two finite signed measures μ_1 and μ_2 on a metrizable space X , the following two conditions are equivalent:*

- (1) $\mu_1 = \mu_2$.
- (2) $\int_X f d\mu_1 = \int_X f d\mu_2$ for every $f \in C_u(X)$.

Therefore, $(C_b(X), \Psi(X))$ is a dual pair and $f \mapsto \langle f, \mu \rangle = \int_X f d\mu$ defines a functional from $\Psi(X)$ to R [3, Chapter II, pp. 40–41]. Hence, we may identify $\Psi(X)$ as a subspace of $R^{C_b(X)}$ (equipped with the product topology) and obtain a Hausdorff locally convex topology on $\Psi(X)$ [3, Proposition 2, Chapter II, p. 43], that is, the **weak* topology** $\sigma(\Psi(X), C_b(X))$. Since $\mathcal{P}(X) \subset \Psi(X)$, the weak* topology $\sigma(\Psi(X), C_b(X))$ induces a topology on $\mathcal{P}(X)$. This induced topology is often denoted by $\sigma(\mathcal{P}(X), C_b(X))$. In the literature of probability and statistics, $\sigma(\mathcal{P}(X), C_b(X))$ is often called the **weak topology** or **topology of convergence in distribution**; but precisely it is the induced weak* topology. From Chapter II, p. 42 of [3], we know that a neighborhood base at $Q \in \mathcal{P}(x)$ is given by all sets of the form

$$\left\{ P \in \mathcal{P}(X) : \left| \int f_k dP - \int f_k dQ \right| \leq \epsilon, k = 1, 2, \dots, n \right\}, \quad (2.1)$$

where n is a positive integer, $f_1, \dots, f_n \in C_b(X)$, and $\epsilon > 0$. From now on, $\mathcal{P}(X)$ will always be equipped with this topology unless stated otherwise. Indeed, we have given answers to Questions 1 and 3 in Section 1. Definition 2 on Chapter II, p. 42 of [3] implies the answer to Question 2: weak convergence of probability measures corresponds to the weakest topology on $\mathcal{P}(X)$ that makes all the maps $f \mapsto \int f dP$, $P \in \mathcal{P}(X)$ continuous. Next, we answer Question 4.

In general, a topological vector space need not be metrizable. For a topological vector space, the metrizability condition is given by the following theorem [3, Chapter II, p. 16].

Theorem 2.2. *A Hausdorff topological vector space is metrizable if and only if it has a countable neighborhood base at zero.*

This theorem might be a little bit too general since we have noticed that the induced weak* topology $\sigma(\mathcal{P}(X), C_b(X))$ is Hausdorff and locally convex. Indeed, the metrizability of a Hausdorff locally convex space can also be characterized [3, Corollary, Chapter II, p. 24].

Theorem 2.3. *A Hausdorff locally convex space (X, \mathcal{T}) is metrizable if and only if \mathcal{T} is generated by a sequence $\{q_n\}$ of seminorms; in this case, \mathcal{T} is generated by the metric d given by*

$$d(x, y) = \sum_{n=1}^{\infty} \frac{q_n(x - y)}{2^n [1 + q_n(x - y)]}.$$

LINEAR TOPOLOGY OF PROBABILITY MEASURES

Clearly, the sufficient condition in Theorem 2.3 need not hold for the weak* topology $\sigma(\mathcal{P}(X), C_b(X))$. Therefore, in general we can only use neighborhood bases or subbases to describe weak topology. This answers Question 4 in Section 1.

Following the same argument in Section 2, the weak* topology

$$\sigma(\Psi(X), C_u(X))$$

is well-defined; it induces a relative topology on $\mathcal{P}(X)$, which is often denoted as $\sigma(\mathcal{P}(X), C_u(X))$. Since $C_u(X)$ is a proper subset of $C_b(X)$ unless X is compact, $\sigma(\Psi(X), C_b(X))$ is generally stronger than $\sigma(\Psi(X), C_u(X))$ [3, Corollary 3, Chapter II, p. 43]. Surprisingly, the following theorem [1, Theorem 15.2] shows that they induce the same topology on $\mathcal{P}(X)$.

Theorem 2.4. *Let X be a metrizable space and d be a compatible metric. Then $\sigma(\mathcal{P}(X), C_b(X)) = \sigma(\mathcal{P}(X), C_u(X))$.*

Finally, we remark that the $\sigma(\Psi(X), C_b(X))$ admits a uniformity \mathcal{D} on $\Psi(X)$ since a topological vector space is a commutative topological group. As a result, $\mathcal{P}(X)$ carries the corresponding relative uniformity \mathcal{U} . It is clear that the base for \mathcal{D} consists of all the sets of the form

$$\{(P_1, P_2) \mid P_1 - P_2 \in U\},$$

where U is a neighborhood at zero in the space $\Psi(X)$, $\psi(\Psi(X), C_b(X))$ and that

$$\sigma(\mathcal{P}(X), C_b(X))$$

is the topology on the uniform space $(\mathcal{P}(X), \mathcal{U})$. If $\{P_n\}$ is a sequence in $\mathcal{P}(X)$ and $\lim_n \int f dP_n$ exists for every $f \in C_b(X)$, then $\{P_n\}$ is \mathcal{U} -Cauchy. However, there is no guarantee that $\{P_n\}$ converges in $\sigma(\mathcal{P}(X), C_b(X))$ even if $\mathcal{P}(X)$ is Polish; because different uniformities may generate the same topology.

ACKNOWLEDGMENTS

The author is grateful to the anonymous referee whose comments and suggestions led to several improvements in this article. The author benefited from some discussions with Patrik J. Fitzsimmons, Lutz Mattner, and Pietro Rigo on a this topic. Thanks are also due to Christoph Lendenmann who read an earlier draft and caught some typos and obscurities.

REFERENCES

- [1] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, (third edition), Springer, Berlin, New York, 2006.
- [2] P. Billingsley, *Convergence of Probability Measures*, (second edition), Wiley, New York, 1999.

L. HONG

- [3] N. Bourbaki, *Elements of Mathematics: Topological Vector Spaces*, Chapters 1–5, Springer, Berlin, New York, 1987.
- [4] K. R. Parthasarathy, *Probability Measure on Metric Spaces*, American Mathematical Society, Providence, Rhode Island, 1967.
- [5] V. S. Varadarajan, *Weak convergence of measures on separable metric spaces*, *Sankhyā*, **19** (1958), 15–22.

MSC2010: 00-01, 60F05, 60A10

Key words and phrases: Weak convergence, probability measure, topological vector space, dual pair

DEPARTMENT OF MATHEMATICS, ROBERT MORRIS UNIVERSITY, MOON TOWNSHIP, PA 15108

E-mail address: hong@rmu.edu