A Formalization of the Lévy-Prokhorov Metric in Isabelle/HOL

Michikazu Hirata
School of Computing, Tokyo Institute of Technology, Japan

Abstract

The Lévy-Prokhorov metric is a metric between finite measures on a metric space. The metric was introduced to analyze weak convergence of measures. We formalize the Lévy-Prokhorov metric and prove Prokhorov's theorem in Isabelle/HOL. Prokhorov's theorem provides a condition for the relative compactness of sets of finite measures and plays essential roles in proofs of the central limit theorem, Sanov's theorem in large deviation theory, and the existence of optimal coupling in transportation theory. Our formalization includes important results in mathematics such as the Riesz representation theorem, which is a theorem in functional analysis and used to prove Prokhorov's theorem. We also apply the Lévy-Prokhorov metric to show that the measurable space of finite measures on a standard Borel space is again a standard Borel space.

2012 ACM Subject Classification Mathematics of computing → Probability and statistics; Mathematics of computing → Mathematical analysis

Keywords and phrases formalization of mathematics, measure theory, metric spaces, topology, Lévy-Prokhorov metric, Prokhorov's theorem, Isabelle/HOL

Digital Object Identifier 10.4230/LIPIcs.ITP.2024.21

Supplementary Material *Software (Formalization in the Archive of Formal Proofs)*: `https://www.isa-afp.org/entries/Levy_Prokhorov_Metric.html` [9]
Software (Formalization in the Archive of Formal Proofs): `https://www.isa-afp.org/entries/Riesz_Representation.html` [10]

Funding This work is supported by JSPS Research Fellowships for Young Scientists and JSPS KAKENHI Grant Number 23KJ0905.

Acknowledgements I would like to thank Yasuhiko Minamide and Tetsuya Sato for commenting on a draft of this paper. I also thank anonymous reviewers for their helpful comments and suggestions.

1 Introduction

The Lévy-Prokhorov metric is a mathematical tool to analyze asymptotic behaviors of distributions or measures in terms of weak convergence. Such analysis is one of the important aspects of probability theory and a foundation of statistics because the knowledge on asymptotic behaviors provides insights of what will be likely to happen when we collect large data.

Our motivation of formalizing the Lévy-Prokhorov metric is to prove that the measurable space of finite measures on a *standard Borel space* is again a standard Borel space, where a standard Borel space is a measurable space with certain good properties. Standard Borel spaces are often used in modern probability theory. The disintegration theorem, which guarantees the existence of *conditional probability kernels*, requires the underlying space to be a standard Borel space. Standard Borel spaces are also a theoretical basis for the theory of *quasi-Borel spaces*, a denotational model for higher-order probabilistic programs [8]. We formalize the Lévy-Prokhorov metric because we need to give a metric on finite measures in order to show that the space of finite measures is a standard Borel space. Another metric which metrizes weak convergence is the *Wasserstein metric*. The Wasserstein metric is applied in transportation theory and machine learning. We chose to formalize the Lévy-Prokhorov metric because the Wasserstein metric may fail to be a metric in the mathematical sense when the underlying metric is not bounded.

During the proof of our goal, we also prove important mathematical theorems in related areas. Our work is divided into three parts.

Weak Convergence and the Lévy-Prokhorov Metric. We first formalize the notion of weak convergence including the Portmanteau theorem, equivalent conditions of weak convergence, and the topology of weak convergence. We define the notion of weak convergence using *filters* as convergence in Isabelle/HOL. We then formalize the Lévy-Prokhorov metric. We prove the equivalence of the topology of weak convergence and the topology induced by the Lévy-Prokhorov metric. The proof is different from the common textbook proofs (e.g. [3, 4]). We obtain a simpler proof thanks to the generalization of weak convergence by filters.

Prokhorov's Theorem. We show Prokhorov's theorem using the Lévy-Prokhorov metric. Prokhorov's theorem states that a set of (uniformly bounded) finite measures is relatively compact if and only if it is *tight*. Prokhorov's theorem plays essential roles in the proofs of the central limit theorem, Sanov's theorem, and the existence of the optimal coupling in transportation theory. In order to formalize Prokhorov's theorem, we also prove (a special case of) Alaoglu's theorem and the Riesz representation theorem. The Riesz representation theorem is an important result in functional analysis. While its proof, including related lemmas, consists of around nine pages in Rudin's book [22], our formalization takes more than 2,100 lines of proofs.

Measurable Spaces of Finite Measures. One often considers the measurable space of measures on some measurable space. Such spaces are used in stochastic processes and semantics of probabilistic programming. The measurable space of measures is defined independently from metrics or topologies. We prove that the measurable space of finite measures is generated from the Lévy-Prokhorov metric. As a consequence, we obtain that the measurable space of finite measures on a standard Borel space is a standard Borel space.

Our formalization is mainly based on the lecture notes by Gaans [27]. The lecture notes includes detailed proofs about the Lévy-Prokhorov metric on probability measures. We extend their definitions and proofs for finite measures.

Related Work

Avigad et al. formalized the notion of weak convergence of probability measures on $\mathbb{R}$ and a special case of Prokhorov's theorem during the proof of the central limit theorem in Isabelle/HOL [1]. Compared to their work, our formalization of weak convergence treats finite measures on any metric spaces, and convergence is generalized by filters. While there is a simpler proof for the special case of Prokhorov's theorem that they formalized, Prokhorov's theorem that we formalize needs tools in functional analysis, such as the Riesz representation theorem, and thus requires more effort.

The Lean mathematical library, mathlib [25], includes ongoing formalization of the weak convergence and the Lévy-Prokhorov metric by Kytölä [15]. Their definition of the weak convergence is also generalized by filters and treats not only probability measures but also finite measures. They showed that the Lévy-Prokhorov metric on the set of finite measures on a pseudo metric space is a pseudo metric. They proved the equivalence of the topology of weak convergence and the topology induced by the Lévy-Prokhorov metric on the space of probability measures. Our work contains more results than their work such as the equivalence of convergence with respect to the Lévy-Prokhorov metric and weak convergence (Theorem 10. 3) and Prokhorov's theorem (Theorem 20).

The Riesz representation theorem in its original form as given by Riesz is formalized by Narkawicz in PVS [19] and by Narita et al. in Mizar [18].

Paper Outline

In Section 2, we review the basic notions and theorems of topological spaces, metric spaces, and measurable spaces. In Section 3, we define the weak convergence of measures, the topology of weak convergence, and the Lévy-Prokhorov metric. We then show their properties. In Section 4, we explain Prokhorov's theorem and lemmas used in the proof of Prokhorov's theorem. In Section 5, we discuss the measurable space of finite measures.

We do not show Isabelle source code in this paper except for the definition of the topology of weak convergence and the Lévy-Prokhorov metric in Section 3.4. The definitions and statements in Isabelle/HOL are almost direct translations from the mathematical notation; therefore, printing them here would not provide any additional insights.

2 Preliminaries

In this section, we review basic definitions and theorems related to topology, metric spaces, and measure theory. Most of the results in this section are included in Isabelle/HOL's standard library.

2.1 Topology

Topology is a way of expressing *nearness* of points in a set. Let X be a set and $\mathcal{O}_X$ a set of subsets of X. The pair $(X, \mathcal{O}_X)$ is called a topological space when $\varnothing \in \mathcal{O}_X$, $X \in \mathcal{O}_X$, and $\mathcal{O}_X$ is closed under finite intersections and arbitrary unions. We sometimes write only X for $(X, \mathcal{O}_X)$, when the structure is obvious from the context. We follow the standard definitions of topology, such as,

- $U \subseteq X$ is an open set of $X \overset{\text{def}}{\iff} U \in \mathcal{O}_X$
- $C \subseteq X$ is a closed set of $X \overset{\text{def}}{\iff} X - C$ is open
- $f : X \to Y$ is a continuous map $\overset{\text{def}}{\iff} \forall U \in \mathcal{O}_Y.\, f^{-1}(U) \in \mathcal{O}_X$

for topological spaces X and Y.

2.2 Metric Spaces

While topological spaces express nearness in abstract way, metric spaces specify concrete distances. Let X be a set and $d : X \times X \to \mathbb{R}$. The pair (X, d) is called a *metric space* if the following holds.

- For all $x, y \in X$, $d(x, y) \geq 0$.
- For all $x, y \in X$, $d(x, y) = d(y, x)$.
- For all $x, y \in X$, $d(x, y) = 0 \iff x = y$.
- For all $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.

We sometimes write only X for (X, d). Let (X, d) be a metric space, $x \in X$ and $\varepsilon > 0$. The set $ball_X(x, \varepsilon) = \{y \in X \mid d(x, y) < \varepsilon\}$ is called an open ball with center x and radius ε. The set $cball_X(x, \varepsilon) = \{y \in X \mid d(x, y) \leq \varepsilon\}$ is called a closed ball with center x and radius ε. We assume that $\mathbb{R}$ is equipped with the standard distance $d(x, y) = |x - y|$ in this presentation.

A metric space X induces the topological space $(X, \mathcal{O}_d)$, where $\mathcal{O}_d$ consists of arbitrary unions of open balls. We call a topological space X *metrizable* if there exists a metric d on X, which induces X.

2.3 Filter and Convergence

In Isabelle/HOL's library, the notion of convergence is formalized in a general way using *filters*. A filter on I is a set of subsets of I satisfying certain conditions. Filters can, among other things, describe the set of all elements that are "sufficiently large" or "sufficiently close to a". We do not explain the detail of filters, which can be found in [13]. Let I be a set, F a filter on I, X a topological space, $\{x_i\}_{i\in I} \subseteq X$, and $x \in X$. The notion "$\{x_i\}_{i\in I}$ converges to x in X with respect to F", denoted by $(x_i \longrightarrow x)\ F$ in X (*limitin* in Isabelle/HOL), is defined by

$$(x_i \longrightarrow x)\ F \text{ in } X \iff \text{For every open neighborhood } U \text{ of } x,\ \textit{eventually } x_i \in U \text{ w.r.t. } F.$$

Intuitively, $(x_i \longrightarrow x)\ F$ in X means that x_i is eventually close to x in X. We call x the limit if $(x_i \longrightarrow x)\ F$ in X. When the topology is obvious from the context, we omit the topology and write $(x_i \longrightarrow x)\ F$ for $(x_i \longrightarrow x)\ F$ in X. For instance, there are filters F_{seq} on $\mathbb{N}$ and (at a) on $\mathbb{R}$ corresponding to "for sufficiently large n" and "for x sufficiently close to a," respectively. Convergences with respect to these filters have the same meaning as the usual definitions.

$$\begin{aligned}
\lim_{n\to\infty} x_n = x &\iff (x_n \longrightarrow x)\ F_{\text{seq}} \text{ in } \mathbb{R} \\
&\iff \forall\varepsilon > 0.\, \exists N.\, \forall n \geq N.\, |x_n - x| < \varepsilon \\
\lim_{x\to a} f(x) = L &\iff (f \longrightarrow L)\ (\text{at } a) \text{ in } \mathbb{R} \\
&\iff (\forall\varepsilon > 0.\, \exists\delta > 0.\, \forall x.\, x \neq a \wedge |x - a| < \delta \Longrightarrow |f(x) - L| < \varepsilon)
\end{aligned}$$

In addition to limit, limit inferior and limit superior are also generalized by filter in Isabelle/HOL. Limit inferior and limit superior with respect to F are denoted by Liminf_F and Limsup_F, respectively.

A Characterization of Closed Sets by Limits. There is a characterization of closed sets using convergence with respect to *nets* (Exercise A.48 [7]). We formalize the following characterization of closed sets by limit with respect to filters because nets and filters are equally expressive in terms of convergence (Section 4 [23]).

► **Lemma 1.** *Let X be a topological space and $C \subseteq X$. Then, the following are equivalent.*
1. *C is closed in X.*
2. *For all sets I, filters F on I, $\{x_i\}_{i\in I} \subseteq C$, $x \in X$ such that $\varnothing \notin F$ and $(x_i \longrightarrow x)\ F$ in X, we have $x \in C$.*

If X is first-countable, then these are also equivalent to the following.

3. *For all $\{x_n\}_{n\in\mathbb{N}} \subseteq C$, $x \in X$ such that $(x_i \longrightarrow x)\ F_{\text{seq}}$ in X, we have $x \in C$.*

The implication that 1 implies 2 (and 3) is already included in Isabelle/HOL's library. We prove the other implications. The last condition of the above equivalence has already been formalized for metric spaces. Since metric spaces are first-countable, our result is a relaxed version of the existing result. There is also a characterization of open sets by limit with respect to filters. The characterization is easily derived from that of closed sets.

From the characterization of closed sets, we obtain a condition to decide whether two topological spaces are equal using limit with respect to filters because topological spaces are determined by closed sets.

▶ **Corollary 2.** *Let* $(X, \mathcal{O}_X)$ *and* $(X, \mathcal{O}'_X)$ *be topological spaces.*

- *If* $(x_i \longrightarrow x)\ F$ *in* $(X, \mathcal{O}_X) \iff (x_i \longrightarrow x)\ F$ *in* $(X, \mathcal{O}'_X)$ *for all* I, F, $\{x_i\}_{i\in I}$, *and* x, *then* $\mathcal{O}_X = \mathcal{O}'_X$.
- *If* $(x_n \longrightarrow x)\ F_{\mathrm{seq}}$ *in* $(X, \mathcal{O}_X) \iff (x_n \longrightarrow x)\ F_{\mathrm{seq}}$ *in* $(X, \mathcal{O}'_X)$ *for all* $\{x_n\}_{n\in\mathbb{N}}$ *and* x, *and both of* $(X, \mathcal{O}_X)$ *and* $(X, \mathcal{O}'_X)$ *are first-countable, then* $\mathcal{O}_X = \mathcal{O}'_X$.

▶ Remark 3. In Isabelle/HOL, we cannot quantify filters as "for any filter F" due to Isabelle/HOL's type system. For instance, when showing $P \longleftrightarrow (\forall F :: \square\ \mathit{filter}.\ Q\ F)$, we need to specify some type $\square$ on which filters are defined. We state Lemma 1 and Corollary 2 by quantifying[1] filters as the type $F :: {'a}\ \mathit{set}\ \mathit{filter}$ when the topology is $X :: {'a}\ \mathit{topology}$ because we use a filter on $V(x)$ (the set of all open neighbourhoods of x) to prove the lemmas. Details of the filter are found in the lecture notes by Heil [7].

Finally, we define the Cauchy sequence and related notions.

▶ **Definition 4.**
- *A sequence* $\{x_n\}_{n\in\mathbb{N}}$ *on a metric space* X *is called a Cauchy sequence if* $\forall \varepsilon > 0.\ \exists N.\ \forall n, m \geq N.\ d(x_n, x_m) < \varepsilon$.
- *A metric space is complete if every Cauchy sequence has a limit.*
- *A topological space* X *is called a completely metrizable space if there exists a complete metric on* X, *which induces* X.
- *A topological space* X *is called a Polish space if* X *is separable and completely metrizable.*

2.4 Measure Theory

The current measure theory library in Isabelle/HOL was first formalized by Hölzl and Heller [12] and has been extended by several other works [1, 5]. Let M be a set and Σ_M a set of subsets of M. A pair (M, Σ_M) is called a *measurable space* if Σ_M is non-empty and closed under complements and countable unions. We sometimes write M for a measurable space (M, Σ_M). A member $A \in \Sigma_M$ is called a *measurable set*. A function f from a measurable space M to a measurable space N is *measurable* if $f^{-1}(A) \in \Sigma_M$ for all $A \in \Sigma_N$. Let M be a measurable space, $\mu : \Sigma_M \to [0, \infty]$ is a *measure* on M if $\mu(\varnothing) = 0$ and $\mu(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n=0}^{\infty} \mu(A_n)$ for any disjoint family $\{A_n\}_{n\in\mathbb{N}} \subseteq \Sigma_M$. A measure μ on M is called a *finite measure* if $\mu(M) < \infty$, a *sub-probability measure* if $\mu(M) \leq 1$, and a *probability measure* if $\mu(M) = 1$. For a measure μ on M and a measurable function $f : M \to \mathbb{R}$, $\int f \mathrm{d}\mu$ denotes the Lebesgue integral of f with respect to μ.

A topological space $(X, \mathcal{O}_X)$ induces the measurable space $(X, \sigma[\mathcal{O}_X])$, where $\sigma[\mathcal{O}_X]$ is the least σ-algebra including all open sets of X. The measurable space $(X, \sigma[\mathcal{O}_X])$ is called the Borel space. Notice that a metric space is also treated as a measurable space since it induces a topological space. The Borel space induced by a metric space (X, d) is denoted by (X, Σ_d).

[1] The direction 1 implies 2 of Lemma 1, which is already included in Isabelle/HOL's library, has been proved for all filters of any type using Isabelle's polymorphism. That is, $P \Longrightarrow Q\ (F :: {'b}\ \mathit{filter})$ can be stated in Isabelle.

3 The Lévy-Prokhorov Metric

Historically, Lévy first introduced a metric, known as the Lévy metric, between cumulative distribution functions [17]. Later, Prokhorov defined the Lévy-Prokhorov metric between finite measures analogous to the Lévy metric [21]. In this section, we review the notion of weak convergence and the Lévy-Prokhorov metric. At the end of this section, we discuss our formalization of the topology of weak convergence and the Lévy-Prokhorov metric. For a measurable space X, $\mathcal{P}(X)$ denotes the set of all finite measures on X. Note that X can be a metric space or a topological space since they both induce a measurable space.

3.1 Weak Convergence

The existing formalization of weak convergence in Isabelle/HOL's standard library is restricted to sequences on $\mathbb{N}$ of probability measures on $\mathbb{R}$. We define the notion of weak convergence which treats finite measures on any topological spaces. The convergence in our formalization is generalized by filters.

▶ **Definition 5** (Weak Convergence). *Let X be a topological space, I a set, F a filter on I, $\{\mu_i\}_{i\in I} \subseteq \mathcal{P}(X)$, and $\mu \in \mathcal{P}(X)$. We say that $\{\mu_i\}_{i\in I}$ converges weakly to μ with respect to F, denoted by $(\mu_i \Rightarrow_{\mathrm{wc}} \mu)\ F$, if $\left(\int f \mathrm{d}\mu_i \longrightarrow \int f \mathrm{d}\mu\right)\ F$ for all $f \in \mathrm{C_b}(X)$, where $\mathrm{C_b}(X)$ is the set of all bounded continuous functions from X to $\mathbb{R}$.*

The notion of weak convergence has several equivalent statements when X is a metric space.

▶ **Theorem 6** (The Portmanteau Theorem). *Let X be a metric space, I a set, F a filter on I, $\{\mu_i\}_{i\in I} \subseteq \mathcal{P}(X)$, and $\mu \in \mathcal{P}(X)$. Then, the following are equivalent.*

1. $(\mu_i \Rightarrow_{\mathrm{wc}} \mu)\ F$.
2. *For all $f \in \mathrm{UC_b}(X)$, $\left(\int f \mathrm{d}\mu_i \longrightarrow \int f \mathrm{d}\mu\right)\ F$.*
3. *$(\mu_i(X) \longrightarrow \mu(X))\ F$ and for every closed set C, $\mathrm{Limsup}_F\{\mu_i(C)\}_{i\in I} \leq \mu(C)$.*
4. *$(\mu_i(X) \longrightarrow \mu(X))\ F$ and for every open set U, $\mathrm{Liminf}_F\{\mu_i(U)\}_{i\in I} \geq \mu(U)$.*
5. *For every measurable set $A \in \Sigma_X$ such that $\mu(\partial A) = 0$, $(\mu_i(A) \longrightarrow \mu(A))\ F$.*

The set $\mathrm{UC_b}(X)$ denotes the set of all bounded uniform continuous functions $f : X \to \mathbb{R}$.

The Portmanteau theorem is commonly stated for probability measures rather than finite measures. Notice that we require the condition $(\mu_i(X) \longrightarrow \mu(X))\ F$ in 3 and 4. This condition does not appear in the Portmanteau theorem for probability measures. In the proof for probability measures, we use $\mu_i(X) = \mu(X) = 1$. For finite measures, $\mu_i(X)$ is not equal to $\mu(X)$ in general. Hence, we use the condition $(\mu_i(X) \longrightarrow \mu(X))\ F$ instead of $\mu_i(X) = \mu(X) = 1$ in order to approximate $\mu_i(X)$ to $\mu(X)$ during the proof.

3.2 Topology of Weak Convergence

Let X be a topological space. Topology of weak convergence on X, denoted by $\mathcal{O}_{\mathrm{WC}_X}$, is the coarsest topology on $\mathcal{P}(X)$ which makes $(\lambda\mu.\ \int f \mathrm{d}\mu) : \mathcal{P}(X) \to \mathbb{R}$ continuous for all $f \in \mathrm{C_b}(X)$. As the name suggests, convergence in the topology of weak convergence is equal to weak convergence.

▶ **Lemma 7.** *Let X be a topological space, I a set, F a filter on I, $\{\mu_i\}_{i\in I} \subseteq \mathcal{P}(X)$, and $\mu \in \mathcal{P}(X)$. Then,*

$$(\mu_i \longrightarrow \mu)\ F \text{ in } (\mathcal{P}(X), \mathcal{O}_{\mathrm{WC}_X}) \iff (\mu_i \Rightarrow_{\mathrm{wc}} \mu)\ F.$$

3.3 The Lévy-Prokhorov Metric

In the lecture notes by Gaans, they only treat the case when $\mathcal{P}(X)$ is the set of all probability measures on X. We generalize their definitions and proofs to the set of all finite measures.

▶ **Definition 8** (Lévy-Prokhorov Metric). *For a metric space (X, d), the Lévy-Prokhorov metric $d_{\mathcal{P}(X)}$ is a metric on $\mathcal{P}(X)$ defined by*

$$d_{\mathcal{P}(X)}(\mu, \nu) = \inf\{\alpha > 0 \mid \forall A \in \Sigma_X.\, \mu(A) \leq \nu(A^\alpha) + \alpha \wedge \nu(A) \leq \mu(A^\alpha) + \alpha\},$$

where $A^\alpha = \bigcup_{x \in A} ball_X(x, \alpha)$.

Note that $d_{\mathcal{P}(X)}(\mu, \nu) < \infty$ because $\infty \neq \max(\mu(X), \nu(X)) \in \{\alpha > 0 \mid \forall A \in \Sigma_X.\, \mu(A) \leq \nu(A^\alpha) + \alpha \wedge \nu(A) \leq \mu(A^\alpha) + \alpha\}$. The Lévy-Prokhorov metric is also expressed using open sets, closed sets, and compact sets.

▶ **Lemma 9.**

$$\begin{aligned} d_{\mathcal{P}(X)}(\mu, \nu) &= \inf\{\alpha > 0 \mid \forall U\colon \textit{open}.\, \mu(U) \leq \nu(U^\alpha) + \alpha \wedge \nu(U) \leq \mu(U^\alpha) + \alpha\} \\ &= \inf\{\alpha > 0 \mid \forall C\colon \textit{closed}.\, \mu(C) \leq \nu(C^\alpha) + \alpha \wedge \nu(C) \leq \mu(C^\alpha) + \alpha\}. \end{aligned}$$

If X is separable and complete, then

$$d_{\mathcal{P}(X)}(\mu, \nu) = \inf\{\alpha > 0 \mid \forall K\colon \textit{compact}.\, \mu(K) \leq \nu(K^\alpha) + \alpha \wedge \nu(K) \leq \mu(K^\alpha) + \alpha\}.$$

The convergence with respect to the Lévy-Prokhorov metric is equivalent to the weak convergence when X is separable.

▶ **Theorem 10** (Theorem 4.1 and 4.2 [27]). *The following hold.*

1. $(\mathcal{P}(X), d_{\mathcal{P}(X)})$ *is a metric space.*

Let I be a set, F a filter on I, $\{\mu_i\}_{i \in I} \subseteq \mathcal{P}(X)$ and $\mu \in \mathcal{P}(X)$.

2. $(\mu_i \longrightarrow \mu)$ F *in* $(\mathcal{P}(X), \mathcal{O}_{d_{\mathcal{P}(X)}})$ *implies* $(\mu_i \Rightarrow_{\mathrm{wc}} \mu)$ F.
3. *If X is separable,* $(\mu_i \longrightarrow \mu)$ F *in* $(\mathcal{P}(X), \mathcal{O}_{d_{\mathcal{P}(X)}})$ *if and only if* $(\mu_i \Rightarrow_{\mathrm{wc}} \mu)$ F.

The proofs are similar to the one when $\mathcal{P}(X)$ is the set of all probability measures and $F = F_{\mathrm{seq}}$ (i.e., the convergence is not generalized by filters). The Lévy-Prokhorov metric metrizes the topology of weak convergence when X is separable.

▶ **Corollary 11.** *If X is separable, the Lévy-Prokhorov metric metrizes the topology of weak convergence, i.e., $\mathcal{O}_{\mathrm{WC}_X} = \mathcal{O}_{d_{\mathcal{P}(X)}}$.*

The generalization by filters of weak convergence and Theorem 10 enables us to prove this lemma easily.

Proof. The metrizability is shown from the equivalence of convergences. From Lemma 7 and Theorem 10, convergences in $(\mathcal{P}(X), \mathcal{O}_{\mathrm{WC}_X})$ and $(\mathcal{P}(X), \mathcal{O}_{d_{\mathcal{P}(X)}})$ are equivalent for all filters. Hence, we have $\mathcal{O}_{\mathrm{WC}_X} = \mathcal{O}_{d_{\mathcal{P}(X)}}$ from Corollary 2. ◀

Even though Corollary 11 is a well-known result, only a few books include its proof. We found two books showing Corollary 11. In the book by Billingsley [3], they directly prove the equivalence by examining neighborhoods. In the book by Deuschel and Stroock [4], they prove the equivalence by using the equivalence of convergence with respect to the filter F_{seq} (i.e., sequences are defined on $\mathbb{N}$ such as $\{\mu_n\}_{n \in \mathbb{N}}$). As we stated in Corollary 2, their proof requires the assumption that $(\mathcal{P}(X), \mathcal{O}_{\mathrm{WC}_X})$ is first-countable. They use the fact that $(\mathcal{P}(X), \mathcal{O}_{\mathrm{WC}_X})$ is second-countable (and thus also first-countable) without providing any proof that it is second-countable. If we follow their proof, we will need additional efforts to show the first countability of $(\mathcal{P}(X), \mathcal{O}_{\mathrm{WC}_X})$. In our proof, we do not need the first countability because we generalized the notion of weak convergence and equivalence of convergence by filters.

Thanks to Corollary 11, we identify $(\mathcal{P}(X), \mathcal{O}_{d_{\mathcal{P}(X)}})$ with $(\mathcal{P}(X), \mathcal{O}_{\mathrm{WC}_X})$, when X is a separable metric space.

▶ **Proposition 12** (Proposition 4.4 [27]). *If X is a separable metric space, then $\mathcal{P}(X)$ is also a separable metric space.*

The proof is similar to the one when $\mathcal{P}(X)$ is the set of all probability measures on X. If $\{a_n\}_{n\in\mathbb{N}}$ is a dense subset of X, then

$$\bigcup_{k\in\mathbb{N}} \{r_0\delta_{a_0} + \cdots + r_k\delta_{a_k} \mid r_0, \ldots, r_k \in \mathbb{Q} \cap [0, \infty)\}$$

is a countable dense subset of $\mathcal{P}(X)$, where δ_a denotes the Dirac measure centered at a.

3.4 Implementation in Isabelle/HOL

We explain our implementation of the topology of weak convergence and the Lévy-Prokhorov metric. We sometimes use usual mathematical symbols in source code for readability.

Topology of Weak Convergence

We define the topology of weak convergence by combining existing constants which generate topological spaces. Let f be a bounded continuous function on X and $\mathcal{O}_f$ the least topology on $\mathcal{P}(X)$, which makes $(\lambda N.\ \int x.\ f\ x\ \partial N)$ continuous[2]. Then, $(\mathcal{P}(X), \mathcal{O}_f)$ is written in Isabelle/HOL as follows:

$$(\mathcal{P}(X), \mathcal{O}_f) = \textit{pullback-topology}\ \mathcal{P}(X)\ (\lambda N.\ \int x.\ f\ x\ \partial N)\ \mathbb{R},$$

where

$$\textit{pullback-topology} :: {'a}\ \textit{set} \Rightarrow ({'a} \Rightarrow {'b}) \Rightarrow {'b}\ \textit{topology} \Rightarrow {'a}\ \textit{topology}$$
$$\textit{pullback-topology}\ A\ f\ Y = \text{The least topology on } A \text{ which makes } f\colon A \to Y \text{ continuous.}$$

The set of all open sets $\mathcal{O}_f$ is extracted as follows:

$$\mathcal{O}_f = \textit{Collect}\ (\textit{openin}\ (\textit{pullback-topology}\ \mathcal{P}(X)\ (\lambda N.\ \int x.\ f\ x\ \partial N)\ \mathbb{R})),$$

where

$$\textit{openin} :: {'a}\ \textit{topology} \Rightarrow {'a}\ \textit{set} \Rightarrow \textit{bool}, \quad \textit{openin}\ X\ U \iff U \text{ is an open set of } X.$$
$$\textit{Collect} :: ({'a} \Rightarrow \textit{bool}) \Rightarrow {'a}\ \textit{set}, \quad \textit{Collect}\ P = \{x.\ P\ x\}.$$

Finally, we define the topology of weak convergence $(\mathcal{P}(X), \mathcal{O}[\bigcup_{f\in \mathrm{C_b}(X)} \mathcal{O}_f])$.

```
definition weak-conv-topology :: 'a topology ⇒ 'a measure topology where
weak-conv-topology X ≡ topology-generated-by
  (⋃f∈{f. continuous-map X ℝ f ∧ (∃B. ∀x∈topspace X. |f x| ≤ B)} .
    Collect (openin (pullback-topology 𝒫(X) (λN. ∫x. f x ∂N) ℝ)))
```

[2] In Isabelle/HOL, the Lebesgue integral of f with respect to N is denoted by $\int x.\ f\ x\ \partial N$.

The term *continuous-map* X $\mathbb{R}$ f means that f is a continuous map from X to $\mathbb{R}$ and *topology-generated-by* receives a set of sets and returns the least topology, including the received set. The topological space *weak-conv-topology* X meets the requirements of the topology of weak convergence.

lemma *continuous-map-weak-conv-topology*:
  assumes *continuous-map* $X\ \mathbb{R}\ f$ **and** $\bigwedge x.\ x \in \mathit{topspace}\ X \Longrightarrow |f\ x| \le B$
  shows *continuous-map* $(\mathit{weak\text{-}conv\text{-}topology}\ X)\ \mathbb{R}\ (\lambda N.\ \int x.\ f\ x\ \partial N)$

lemma *weak-conv-topology-minimal*:
  assumes $\mathit{topspace}\ Y = \mathcal{P}(X)$
   and $\bigwedge f\ B.\ \mathit{continuous\text{-}map}\ X\ \mathbb{R}\ f \Longrightarrow (\bigwedge x.\ x \in \mathit{topspace}\ X \Longrightarrow |f\ x| \le B)$
      $\Longrightarrow \mathit{continuous\text{-}map}\ Y\ \mathbb{R}\ (\lambda N.\ \int x.\ f\ x\ \partial N)$
  shows $\mathit{openin}\ (\mathit{weak\text{-}conv\text{-}topology}\ X)\ U \Longrightarrow \mathit{openin}\ Y\ U$

The first lemma guarantees that *weak-conv-topology* X makes $(\lambda N.\ \int x.\ f\ x\ \partial N)$ continuous and the second lemma states that *weak-conv-topology* X is the least topology in such topological spaces.

From Lemma 7, weak convergence and convergence in the topology of weak convergence are equivalent. Thus, we define the notion of weak convergence as an abbreviation of the convergence in the topology of weak convergence. Then, the usual definition of weak convergence (Definition 5) is shown as a lemma.

abbreviation *weak-conv-on* :: $('a \Rightarrow 'b\ \mathit{measure}) \Rightarrow 'b\ \mathit{measure} \Rightarrow 'a\ \mathit{filter} \Rightarrow 'b\ \mathit{topology} \Rightarrow \mathit{bool}$
 where $\mathit{weak\text{-}conv\text{-}on}\ \mathit{Ni}\ N\ F\ X \equiv \mathit{limitin}\ (\mathit{weak\text{-}conv\text{-}topology}\ X)\ \mathit{Ni}\ N\ F$

lemma *weak-conv-on-def'*:
  assumes $\bigwedge i.\ \mathit{Ni}\ i \in \mathcal{P}(X)$ **and** $N \in \mathcal{P}(X)$
  shows $\mathit{weak\text{-}conv\text{-}on}\ \mathit{Ni}\ N\ F\ X \longleftrightarrow$
      $(\forall f.\ \mathit{continuous\text{-}map}\ X\ \mathbb{R}\ f \longrightarrow (\exists B.\ \forall x{\in}\mathit{topspace}\ X.\ |f\ x| \le B)$
        $\longrightarrow ((\lambda i.\ \int x.\ f\ x\ \partial \mathit{Ni}\ i) \longrightarrow (\int x.\ f\ x\ \partial N))\ F)$

The term *limitin* (*weak-conv-topology* X) $\mathit{Ni}\ N\ F$ denotes $(\mathit{Ni} \longrightarrow N)\ F$ in $(\mathcal{P}(X), \mathcal{O}_{\mathrm{WC}_X})$ in our presentation.

The Lévy-Prokhorov Metric

To formalize the Lévy-Prokhorov metric in Isabelle/HOL, we use the set-based metric space library, which has recently appeared in the standard distribution since Isabelle 2023. The library was ported from HOL Light by Paulson [20]. Another metric space library that has been used is based on type classes [13]. While set-based metric space enable us to treat metric spaces with arbitrary carrier sets, type-based metric spaces only work for an entire type. For each type, there can only be one *metric-space* instance. This works well for situations where there is a "canonical" metric space for a type, but it lacks the flexibility to describe, for instance, the set of all metric spaces with a given carrier set. The library based on type classes is unsuitable for our use because we use the set of finite measures on a measurable space, which is not the universe of the type.

In Isabelle/HOL's library, the set-based metric space is defined with the **locale** command.

locale *Metric-space* =
  fixes $M :: 'a\ \mathit{set}$ **and** $d :: 'a \Rightarrow 'a \Rightarrow \mathit{real}$
  assumes *nonneg*: $\bigwedge x\ y.\ 0 \le d\ x\ y$
  assumes *commute*: $\bigwedge x\ y.\ d\ x\ y = d\ y\ x$
  assumes *zero*: $\bigwedge x\ y.\ [\![x \in M;\ y \in M]\!] \Longrightarrow d\ x\ y = 0 \longleftrightarrow x{=}y$
  assumes *triangle*: $\bigwedge x\ y\ z.\ [\![x \in M;\ y \in M;\ z \in M]\!] \Longrightarrow d\ x\ z \le d\ x\ y + d\ y\ z$

The **locale** command introduces a context. In this case, a set M and a function d are fixed and the four assumptions hold, i.e., (M, d) forms a metric space in the context of *Metric-space*. Notice that the non-negativity and commutativity must hold on not only M but the whole type universe. These assumptions make it easier to use non-negativity and commutativity in proofs, and do not change the essential structure of the metric space. Owing to these assumptions, we need to take care of non-negativity and commutativity even outside the carrier set when we define a metric space.

We introduced a new locale *Levy-Prokhorov* which is logically equivalent to *Metric-space*.

```
locale Levy-Prokhorov = Metric-space
```

Remember that the Lévy-Prokhorov metric is defined as follows.

$$d_{\mathcal{P}(X)}(\mu, \nu) = \inf\{\alpha > 0 \mid \forall A \in \Sigma_X.\, \mu(A) \le \nu(A^\alpha) + \alpha \wedge \nu(A) \le \mu(A^\alpha) + \alpha\},$$
$$\text{where } A^\alpha = \bigcup_{x \in A} ball_X(x, \alpha).$$

Hence, we define the Lévy-Prokhorov metric in the context of *Levy-Prokhorov* as follows:

```
definition 𝒫 ≡ {N. sets N = sets (borel-of mtopology) ∧ finite-measure N}

definition LPm :: 'a measure ⇒ 'a measure ⇒ real where
LPm N L ≡
  if N ∈ 𝒫 ∧ L ∈ 𝒫 then
    (⨅ {e. e > 0 ∧ (∀A∈sets (borel-of mtopology).
                          measure N A ≤ measure L (⋃a∈A. mball a e) + e ∧
                          measure L A ≤ measure N (⋃a∈A. mball a e) + e)})
  else 0
```

In the definition of $\mathcal{P}$, the projection function *sets* receives a measure and returns the σ-algebra on which the measure is defined. The constant *mtopology* denotes the topological space induced by (M, d), and *borel-of mtopology* denotes the Borel space generated from *mtopology*. In the definition of *LPm*, *measure N A* corresponds to *N (A)* in usual mathematics notation. Notice that *LPm* returns *0* when one of the arguments is not a member of $\mathcal{P}$ because the set to which we apply infimum might be empty when *LPm* receives an infinite measure. In Isabelle/HOL, the infimum operator on real numbers does not return ∞ nor any specific value when applied to the empty set; i.e., the value of $\bigsqcap \varnothing$ is unknown. This is a problem because *LPm* needs to be a non-negative function on the whole type universe due to the definition of *Metric-space*.

We then prove that $(\mathcal{P}, LPm)$ is a metric space in the context of *Levy-Prokhorov*.

```
sublocale LPm: Metric-space 𝒫 LPm
```

The reader might wonder why we define a new locale *Levy-Prokhorov*, which is logically equivalent to *Metric-space*, rather than using *Metric-space* directly. If we try to define the Lévy-Prokhorov metric in the context of *Metric-space* without introducing a new locale, it does not work.

```
context Metric-space
begin

definition 𝒫 ≡ {N. sets N = sets (borel-of mtopology) ∧ finite-measure N}
definition LPm ≡ ...

sublocale LPm: Metric-space 𝒫 LPm

end
```

The problem is that we try to instantiate *Metric-space* inside the context of *Metric-space*. This causes Isabelle to build an infinite chain; thus, Isabelle does not terminate. This workaround is explained in the Isabelle tutorial on locales [2].

4 Prokhorov's Theorem

One of the important results related to the Lévy-Prokhorov metric is Prokhorov's theorem. In a typical situation in probability theory or statistics, one may want to know whether a sequence of measures has a limit or at least has a converging subsequence. Prokhorov's theorem is applied to prove the existence of a converging subsequence. The theorem is used in proofs for various important results such as the central limit theorem, Sanov's theorem, and the existence of *optimal coupling*. The central limit theorem and Sanov's theorem are key concepts in probability theory. The central limit theorem states that under appropriate conditions, the distribution of normalized sample means converges weakly to the standard normal distribution. Sanov's theorem is an important result in the large deviation theory (e.g. Section 3.2 [4]). The theorem describes the asymptotic behavior of atypical samples and gives evidence why we use the relative entropy (Kullback-Leibler divergence) to evaluate estimated distributions. Both the central limit theorem and Sanov's theorem use Prokhorov's theorem. In transportation theory, a *coupling* is a *plan* how to move resources from supply areas to demand areas. A coupling is represented as a measure satisfying certain conditions. An optimal coupling is a coupling that minimizes the total *cost* of transporting resources. In the proof of the existence of an optimal coupling, Prokhorov's theorem is essential [28, 29].

In this section, we discuss Prokhorov's theorem and related topics.

4.1 Regular Measures

We define the notion of regular measures and tightness of measures. The regularity of measures gives ways to approximate a measured value $\mu(A)$ by open sets, closed sets, and compact sets. The tightness of measures is used to express a condition in Prokhorov's theorem.

▶ **Definition 13.** *Let X be a topological space. A measure μ on X is called:*
1. *inner regular if $\mu(A) = \sup\{\mu(C) \mid C \subseteq A,\ C \text{ is closed}\}$ for all measurable sets A,*
2. *outer regular if $\mu(A) = \inf\{\mu(U) \mid A \subseteq U,\ U \text{ is open}\}$ for all measurable sets A, and*
3. *regular if μ is inner regular and outer regular.*

▶ **Proposition 14.** *Let X be a metrizable space. Then, any finite measure on X is regular.*

▶ Remark 15. This definition of inner regular by Gaans is different from the standard definition. In general, a measure μ on X is called inner regular if
1'. $\mu(A) = \sup\{\mu(K) \mid K \subseteq A,\ K \text{ is compact}\}$ for all measurable sets A.
This definition is stronger than the condition 1 in Definition 13, when every compact set is closed (e.g. when X is metrizable). As we will see soon, Proposition 14 still holds even if we use the condition 1' as inner regularity when X is a Polish space (Corollary 19).

Proposition 14 has been already included in the standard Isabelle/HOL's library. They assume that X is a Polish space and use the condition 1' as the definition of inner regular. Their formalization is restricted to measures on the Borel space of topological space on type classes; thus, they treat only when X is the universal set such as $\mathbb{R}$. We formalize the general result when X is an arbitrary metrizable space or a Polish space.

Next, we define tightness.

▶ **Definition 16** (Tightness). *Let X be a topological space and $\Gamma \subseteq \mathcal{P}(X)$. We call Γ tight if for every $\varepsilon > 0$, there exists a compact set K of X such that $\mu(X - K) \leq \varepsilon$ for all $\mu \in \Gamma$. A measure μ on X is tight if $\{\mu\}$ is tight.*

The existing definition of tightness in Isabelle/HOL's library is restricted to when Γ is a sequence on $\mathbb{N}$ of probability measures on $\mathbb{R}$.

▶ **Lemma 17.** *If X is metrizable and μ is a tight measure on X, then $\mu(A) = \sup\{\mu(K) \mid K \subseteq A,\ K \text{ is compact}\}$ for all measurable sets A.*

▶ **Theorem 18.** *If X is a Polish space, then any finite measure on X is tight.*

▶ **Corollary 19.** *If X is a Polish space and μ is a finite measure on X, then $\mu(A) = \sup\{\mu(K) \mid K \subseteq A,\ K \text{ is compact}\}$ for all measurable sets A.*

4.2 Prokhorov's Theorem

We formalize Prokhorov's theorem. Let $\mathcal{P}_r(X) = \mathcal{P}(X) \cap \{\mu \mid \mu(X) \leq r\}$ for $r \geq 0$.

▶ **Theorem 20** (Prokhorov's Theorem). *Let X be a Polish space and $\Gamma \subseteq \mathcal{P}_r(X)$ for some $r \geq 0$. Then, the following are equivalent.*
1. *Γ is relatively compact.*
2. *Γ is tight.*

▶ Remark 21. Actually, the assumption $\Gamma \subseteq \mathcal{P}_r(X)$ is relaxed to $\Gamma \subseteq \mathcal{P}(X)$ in the proof that 1 implies 2. The completeness assumption is not required in the proof that 2 implies 1.
The following corollary is applied to show the existence of a converging subsequence.

▶ **Corollary 22.** *Let X be a separable metrizable space and $\{\mu_n\}_{n\in\mathbb{N}} \subseteq \mathcal{P}_r(X)$ for some $r \geq 0$. If $\{\mu_n\}_{n\in\mathbb{N}}$ is tight, then there exists a subsequence $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ and $\mu \in \mathcal{P}_r(X)$ such that $(\mu_{n_k} \Rightarrow_{\mathrm{wc}} \mu)\ F_{\mathrm{seq}}$.*

Avigad et al. formalized the above corollary when $\{\mu_n\}_{n\in\mathbb{N}}$ is a sequence of probability measures on $\mathbb{R}$ and applied it to prove the central limit theorem [1]. In the case of probability measures on $\mathbb{R}$, there is a simpler proof using Helly's selection theorem. In general case, we need to prove in other way because the proof using Helly's selection theorem uses cumulative distribution function; i.e., X needs to be $\mathbb{R}$.

The proof that 1 implies 2 in Prokhorov's theorem is more straightforward. The proof that 2 implies 1 requires more effort to prove for us. We do not discuss the details of the proof. Instead, we explain a key lemma for the proof that 2 implies 1.

▶ **Lemma 23.** *If X is a compact metric space, then $\mathcal{P}_r$ is compact.*

The proof relies on results from vector space theory such as Alaoglu's theorem and the Riesz representation theorem. Although these theorems need to be stated in set-based vector space in Isabelle/HOL for our use, most of Isabelle/HOL's vector space library is based on type classes. The set-based vector space library by Lee [16] includes only basic definitions. Thiemann and Yamada also formalized a set-based vector space [26]. However, their work treats only finite-dimensional spaces. Since we are interested in the Lévy-Prokhorov metric rather than vector space theory, we leave the development of the set-based vector space library for future work. Thus, we formalize positive linear functionals used in proofs and their properties without mentioning vector spaces. For Alaoglu's theorem, we prove a special case of the theorem.

Proof. The idea of the proof is to make a homeomorphism between $\mathcal{P}_r$ and a compact space. Let Φ be

$$\Phi = \left(\mathbb{R}^{\mathrm{C}(X)}\right) \cap \{\varphi \mid \varphi \text{ is a positive linear functional} \wedge \varphi(1) \leq r\}. \tag{1}$$

Remember that an element φ of $\mathbb{R}^{\mathrm{C}(X)}$ is a function $\varphi : \mathrm{C}(X) \to \mathbb{R}$. We denote $\varphi(f)$ by φ_f. Then, the linearity of $\varphi \in \Phi$ means that for all $\alpha, \beta \in \mathbb{R}$ and $f, g \in \mathrm{C}(X)$, $\varphi_{\alpha f + \beta g} = \alpha\varphi_f + \beta\varphi_g$. The positiveness means that for all $f \in \mathrm{C}(X)$ such that $f \geq 0$, $\varphi_f \geq 0$.

We assume that Φ is equipped with the subspace topology of the product topology $\mathbb{R}^{\mathrm{C}(X)}$ (subspace topology of the *weak* topology*). We define the function T from $\mathcal{P}_r$ to Φ by $T(\mu)_f = \int f \mathrm{d}\mu$. It is easy to check that $T(\mu) \in \Phi$, and T is a sequential homeomorphic map. For instance, the linearity of the integral implies the linearity of $T(\mu)$. The function T is bijective by the Riesz representation theorem (Corollary 31). As Gaans stated, Φ is metrizable[3]. Thus, T is a homeomorphism[4]. Furthermore, Φ is compact by the special case of Alaoglu's theorem (Theorem 28). Hence, $\mathcal{P}_r$ is compact. ◀

▶ Remark 24. In the lecture notes, Gaans stated that the sequential compactness of a closed subset of Φ follows from its compactness. This statement is true because Φ is metrizable. However, they did not mention that in their proof.

Prokhorov's theorem is applied to prove the completeness of the Lévy-Prokhorov metric.

▶ **Corollary 25.** *If X is separable and complete, then $(\mathcal{P}(X), d_{\mathcal{P}(X)})$ is complete.*

When we prove the existence of a limit of a Cauchy sequence $\{\mu_n\}_{n\in\mathbb{N}} \subseteq \mathcal{P}(X)$, we use Prokhorov's theorem as $\Gamma = \{\mu_n\}_{n\in\mathbb{N}}$. Hence, we need to show that $\{\mu_n\}_{n\in\mathbb{N}} \subseteq \Gamma_r$ for some $r \geq 0$. This follows from the fact that $\{\mu_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence.

As a consequence of Corollary 11, Proposition 12, and Corollary 25, we have the following.

▶ **Corollary 26.** *If X is a Polish space, then so is $\mathcal{P}(X)$.*

4.3 Alaoglu's Theorem

Alaoglu's theorem (sometimes called the Banach-Alaoglu theorem) is an important result in functional analysis. The theorem states that the closed unit ball of the dual space of a normed vector space is compact. Let Y be a vector space over $\mathbb{R}$ and Y^* the dual space of Y. The *weak* topology* is a topology on Y^*, which is the coarsest topology that makes every $(\lambda f.\ f(y)) : Y^* \to \mathbb{R}$ continuous. The original statement of the Alaoglu's theorem is the following.

▶ **Theorem 27** (Alaoglu's Theorem)**.** *Let Y be a normed vector space and $B^* = \{\varphi \in Y^* \mid \|\varphi\| \leq r\}$. Then, B^* is compact in Y^* with respect to the weak* topology.*

We do not prove the above form of the theorem due to the lack of set-based vector space library in Isabelle/HOL. Instead, we prove a special case of Alaoglu's theorem for our use.

---

[3] Since X is compact, $\mathrm{C}(X)$ along with the topology of uniform convergence is separable (Theorem 2.4.3 [24]). Let $\{g_n\}_{n\in\mathbb{N}}$ be a dense subset of $\mathrm{C}(X)$. Then, the metric on Φ is, for instance, given by

$$d(\varphi, \psi) = \sum_{n=0}^{\infty} \frac{1}{2^{n+1}} \min(1, |\varphi(g_n) - \psi(g_n)|).$$

[4] A function f from a first-countable space is continuous iff it is sequentially continuous.

▶ **Theorem 28.** *If a topological space X is compact, then Φ defined by (1) in the proof of Lemma 23 is compact.*

▶ Remark 29. While the Alaoglu's theorem says that $\{\varphi \in C(X)^* \mid \|\varphi\| \leq r\}$ is compact, Theorem 28 states that $\Phi = \{\varphi \in C(X)^* \mid \|\varphi\| \leq r \wedge \varphi \text{ is positive}\}$ is compact. Note that $\|\varphi\| = \varphi(1)$ when $\varphi \in C(X)^*$ is positive.

Proof Outline. We formalize the theorem following the proof in the lecture notes by Heil [6]. The proof is simple. We first observe that $\prod_{f \in \mathrm{C}(X)} [-r\|f\|, r\|f\|]$ is compact in $\mathbb{R}^{\mathrm{C}(X)}$ by Tychonoff's theorem. Note that every $f \in \mathrm{C}(X)$ is bounded because X is compact. We then show that $\Phi \subseteq \prod_{f \in \mathrm{C}(X)} [-r\|f\|, r\|f\|]$ and Φ is closed. The fact that Φ is closed is shown by the characterization of closed sets by limit (Lemma 1). ◀

4.4 The Riesz Representation Theorem

The Riesz representation theorem (sometimes called the Riesz-Markov representation theorem or Riesz-Markov-Kakutani representation theorem) states that a real-valued (or complex-valued) positive linear functional is represented by the Lebesgue integration with respect to a unique measure. We prove the Riesz representation theorem following the book by Rudin [22].

▶ **Theorem 30** (The Riesz representation theorem)**.** *Let X be a locally compact Hausdorff space and φ a real-valued positive linear functional on $C_c(X)$, where $C_c(X)$ is the set of all continuous functions on X whose closed support is compact. Then, there exists a σ-algebra $\mathcal{M}$ in X and a unique measure μ on $(X, \mathcal{M})$ such that:*

- $\varphi(f) = \int f \mathrm{d}\mu$ *for all* $f \in C_c(X)$,
- $\Sigma_X \subseteq \mathcal{M}$,
- $\mu(K) < \infty$ *for all compact sets* K,
- $\mu(A) = \inf\{\mu(U) \mid A \subseteq U,\ U \text{ is open}\}$ *for all* $A \in \mathcal{M}$,
- $\mu(A) = \sup\{\mu(K) \mid K \subseteq A,\ K \text{ is compact}\}$ *for all open sets A and for all* $A \in \mathcal{M}$ *such that* $\mu(A) < \infty$, *and*
- μ *is a complete measure, i.e., if* $E \in \mathcal{M}$, $A \subseteq E$, *and* $\mu(E) = 0$, *then* $A \in \mathcal{M}$.

In the book, the proof of the Riesz representation theorem is divided into ten steps and uses two lemmas. Their proofs consist of around nine pages, whereas we spent more than 2,100 lines for their proofs. The proof requires Urysohn's lemma on locally compact Hausdorff space. Although Isabelle/HOL's library has several forms of Urysohn's lemmas and lemmas related to locally compact spaces, the library does not include Urysohn's lemma on locally compact Hausdorff space. Hence, we formalized the lemma by ourselves.

We use the following corollary in the proof of Prokhorov's theorem.

▶ **Corollary 31.** *Let X be a compact metric space and φ be a real-valued positive linear functional on $\mathrm{C}(X)$. Then, there exists a unique measure μ on X such that for all $f \in \mathrm{C}(X)$,*

$$\varphi(f) = \int f \mathrm{d}\mu.$$

5 Measurable Spaces of Finite Measures

In this section, we discuss the measurable space of all finite measures. Measurable spaces on a set of measures are used in stochastic processes and semantics of probabilistic programs. In stochastic processes, measures are usually indexed by time or states. A stochastic process is interpreted as a measurable function from its index set to the space of measures. In the semantics of probabilistic programs, the Giry monad G (or sub-Giry monad) gives a standard semantics of probabilistic programs where $G(M)$ is the measurable space of all probability measures on M defined independently from metric or topology.

We will show that this type of measurable space of all finite measures is generated from the topology of weak convergence when the underlying topological space is a Polish space.

▶ **Definition 32.** *Let M be a measurable space. The space of finite measures on M is denoted by $(\mathcal{P}(M), \Sigma_{\mathcal{P}(M)})$, where $\Sigma_{\mathcal{P}(M)}$ is the least σ-algebra that makes $(\lambda\mu.\ \mu(A))$ measurable for all $A \in \Sigma_M$.*

Note that this definition does not use any metric or topology. In Isabelle/HOL's library, the space of all sub-probability measures $\mathcal{P}_{\text{sprob}}(M)$ and the space of all probability measures $\mathcal{P}_{\text{prob}}(M)$ are already formalized by Eberl et al. [5] (*subprob-algebra M* and *prob-algebra M*, respectively). We have formalized the space of all finite measures in the same way as *subprob-algebra*. Subsequently, we have shown that $\mathcal{P}_{\text{sprob}}(M)$ and $\mathcal{P}_{\text{prob}}(M)$ are subspaces of $\mathcal{P}(M)$.

The following lemma follows immediately from the Portmanteau theorem[5].

▶ **Lemma 33** (Corollary 17.21 [14])**.** *For open $U \subseteq X$, $(\lambda\mu.\ \mu(U)) : (\mathcal{P}(X), \mathcal{O}_{d_{\mathcal{P}(X)}}) \to \mathbb{R}$ is lower semi-continuous. For closed $C \subseteq X$, $(\lambda\mu.\ \mu(C)) : (\mathcal{P}(X), \mathcal{O}_{d_{\mathcal{P}(X)}}) \to \mathbb{R}$ is upper semi-continuous.*

▶ **Corollary 34.** $\Sigma_{\mathcal{P}(X)} \subseteq \Sigma_{d_{\mathcal{P}(X)}}$.

Proof. From the definition of $\Sigma_{\mathcal{P}(X)}$, it is sufficient to show that for all $A \in \Sigma_X$, $(\lambda\mu.\ \mu(A))$ is a measurable function from $(\mathcal{P}(X), \Sigma_{d_{\mathcal{P}(X)}})$ to $\mathbb{R}$. It is easy to check the measurability because by Lemma 33, $(\lambda\mu.\ \mu(U)) : (\mathcal{P}(X), d_{\mathcal{P}(X)}) \to \mathbb{R}$ is lower semi-continuous for all open sets $U \subseteq X$, hence measurable. ◀

The inverse inclusion holds when X is separable and complete.

▶ **Theorem 35.** *If a metric space X is separable and complete, then $\Sigma_{\mathcal{P}(X)} = \Sigma_{d_{\mathcal{P}(X)}}$.*

▶ **Corollary 36.** *If X is a Polish space, then $\Sigma_{\mathcal{P}(X)} = \Sigma_{(\mathcal{P}(X), \mathcal{O}_{\mathrm{WC}_X})}$.*

We constructed the proof of Theorem 35 by ourselves because we could not find any proof for the statement. We provide an informal proof here.

Proof of Theorem 35. Since $\Sigma_{d_{\mathcal{P}(X)}}$ is generated from closed balls, it is sufficient to prove that every closed ball is a member of $\Sigma_{\mathcal{P}(X)}$. Let μ be a finite measure on X and $\varepsilon \geq 0$. Our goal is to show that $cBall_{\mathcal{P}(X)}(\mu, \varepsilon) \in \Sigma_{\mathcal{P}(X)}$. Let $\mathcal{O}_b$ be a countable base of X and $\mathcal{O}_{\mathrm{bfU}}$ the set of all finite unions of elements of $\mathcal{O}_b$. Then, $\mathcal{O}_{\mathrm{bfU}}$ is also countable.

▷ Claim 37.

$$cBall_{\mathcal{P}(X)}(\mu, \varepsilon) = \bigcap_{U \in \mathcal{O}_{\mathrm{bfU}}} \left(\bigcap_{n \in \mathbb{N}} (\lambda\nu.\ \nu(U))^{-1} \left(-\infty, \mu\left(U^{\left(\varepsilon + \frac{1}{1+n}\right)}\right) + \varepsilon + \frac{1}{1+n} \right] \cap \left(\lambda\nu.\ \nu\left(U^{\left(\varepsilon + \frac{1}{1+n}\right)}\right)\right)^{-1} \left[\mu(U) - \left(\varepsilon + \frac{1}{1+n}\right), \infty \right) \right) \quad (2)$$

[5] Remember that for a first-countable space X,
- $f : X \to \mathbb{R}$ is lower semi-continuous iff $(x_n \longrightarrow x)$ F_{seq} in X implies $f(x) \leq \mathrm{Liminf}_{F_{\text{seq}}}\{f(x_n)\}_{n\in\mathbb{N}}$.
- $f : X \to \mathbb{R}$ is upper semi-continuous iff $(x_n \longrightarrow x)$ F_{seq} in X implies $f(x) \geq \mathrm{Limsup}_{F_{\text{seq}}}\{f(x_n)\}_{n\in\mathbb{N}}$.

If the above claim is shown, $cBall(\mu, \varepsilon) \in \Sigma_{\mathcal{P}(X)}$ follows from the definition of $\Sigma_{\mathcal{P}(X)}$.

The inclusion $\subseteq$ in equation (2) is directly proven by unfolding the definition of the Lévy-Prokhorov metric. Hence, we show $\supseteq$ of (2). Let us assume that ν is a member of the right hand side of (2). Then, for all $U \in \mathcal{O}_{\mathrm{bfU}}$ and $n \in \mathbb{N}$, we have

$$\nu(U) \leq \mu\left(U^{\left(\varepsilon+\frac{1}{1+n}\right)}\right) + \varepsilon + \frac{1}{1+n}, \quad \mu(U) \leq \nu\left(U^{\left(\varepsilon+\frac{1}{1+n}\right)}\right) + \varepsilon + \frac{1}{1+n}. \tag{3}$$

We show $\nu \in cBall_{\mathcal{P}(X)}(\mu, \varepsilon)$ by proving that $d_{\mathcal{P}(X)}(\mu, \nu) < \varepsilon'$ for all $\varepsilon' > \varepsilon$. Let $\varepsilon' > \varepsilon$, then there exists $n \in \mathbb{N}$ such that $\varepsilon + \frac{1}{1+n} < \varepsilon'$. For an open set $A \subseteq X$, we have

$$\begin{aligned}
\mu(A) &= \sup\{\mu(K) \mid K \subseteq A,\ K \text{ is compact}\} \quad \text{(Corollary 19)} \\
&\leq \sup\{\mu(U) \mid U \subseteq A,\ U \in \mathcal{O}_{\mathrm{bfU}}\} \qquad (4) \\
&\leq \sup\left\{\nu\left(U^{\left(\varepsilon+\frac{1}{1+n}\right)}\right) + \varepsilon + \frac{1}{1+n} \mid U \subseteq A,\ U \in \mathcal{O}_{\mathrm{bfU}}\right\} \quad \text{(by (3))} \\
&\leq \nu\left(A^{\left(\varepsilon+\frac{1}{1+n}\right)}\right) + \varepsilon + \frac{1}{1+n}.
\end{aligned}$$

The inequality (4) above is shown as follows: Since $\mathcal{O}_b$ is a base of X, there exists $\mathcal{O}' \subseteq \mathcal{O}_b$ such that $A = \bigcup_{U \in \mathcal{O}'} U$. If $K \subseteq A$ is compact, there exists a finite subset $\mathcal{O}'_{\mathrm{fin}} \subseteq \mathcal{O}'$ such that $K \subseteq \bigcup_{U \in \mathcal{O}'_{\mathrm{fin}}} U$. By the definition of $\mathcal{O}_{\mathrm{bfU}}$, we have $\bigcup_{U \in \mathcal{O}'_{\mathrm{fin}}} U \in \mathcal{O}_{\mathrm{bfU}}$. Thus, (4) holds.

Similarly, we have $\nu(A) \leq \mu\left(A^{\left(\varepsilon+\frac{1}{1+n}\right)}\right) + \varepsilon + \frac{1}{1+n}$ for all open sets $A \subseteq X$. Hence,

$$\begin{aligned}
d_{\mathcal{P}(X)}(\mu, \nu) &= \inf\{\alpha > 0 \mid \forall A\text{: open}.\ \mu(A) \leq \nu(A^\alpha) + \alpha \land \nu(A) \leq \mu(A^\alpha) + \alpha\} \\
&\leq \varepsilon + \frac{1}{1+n} < \varepsilon'.
\end{aligned}$$ ◀

Corollary 36 is applied to prove that the space of finite measures is a standard Borel space, which is a measurable space generated from a Polish space. Many practical spaces (e.g. $\mathbb{R}$, $\mathbb{N}$, and countable product spaces of standard Borel spaces) are standard Borel spaces. Standard Borel spaces have good properties such as Kuratowski's theorem stating that any standard Borel space is either a countable discrete space or isomorphic to $\mathbb{R}$. In our previous work [11], we formalized the notion of standard Borel space. As a consequence of Corollary 26 and Corollary 36, we obtain the following.

▶ **Corollary 38.** *If M is a standard Borel space, then so is $\mathcal{P}(M)$.*

▶ **Corollary 39.** *If M is a standard Borel space, then $\mathcal{P}_{\mathrm{sprob}}(M)$ and $\mathcal{P}_{\mathrm{prob}}(M)$ are also standard Borel spaces.*

6 Conclusion

We formalized the Lévy-Prokhorov metric and related notions to show that the measurable space of finite measures on a standard Borel space is a standard Borel space. We also showed important mathematical theorems such as the Riesz representation theorem and Prokhorov's theorem. Our formalization consists of around 11,000 lines (4,400 lines for the Riesz representation theorem and 6,600 lines for the Lévy-Prokhorov metric) including comments and blank lines.

Formalization of the large deviation theory and transportation theory could be interesting future works. Both of these theories depend on Prokhorov's theorem for the proof of important theorems in their fields, namely, Sanov's theorem and the existence of an optimal coupling.

References

1 Jeremy Avigad, Johannes Hölzl, and Luke Serafin. A formally verified proof of the central limit theorem. *Journal of Automated Reasoning*, 59(4):389–423, 2017. doi:10.1007/s10817-017-9404-x.

2 Clemens Ballarin. Tutorial to locales and locale interpretation. *Contribuciones científicas en honor de Mirian Andrés Gómez*, pages 123–140, 2010.

3 Patrick Billingsley. *Convergence of Probability Measures*. Wiley series in probability and mathematical statistics, Tracts on probability and statistics. Wiley, New York, United States, 1968.

4 Jean-Dominique Deuschel and Daniel W. Stroock. *Large Deviations*, volume 137 of *Pure and Applied Mathematics*. Elsevier Science, 1989.

5 Manuel Eberl, Johannes Hölzl, and Tobias Nipkow. A verified compiler for probability density functions. In *European Symposium on Programming (ESOP 2015)*, volume 9032 of *LNCS*, pages 80–104. Springer, 2015. doi:10.1007/978-3-662-46669-8_4.

6 Christopher E. Heil. Alaoglu's theorem. https://heil.math.gatech.edu/6338/summer08/section9f.pdf, 2008. Lecture notes on MATH 6338 (Real Analysis II) at Gerogia Insisute of Technology, Accessed: January 5, 2024.

7 Christopher E. Heil. Nets, directed sets, and convergence. https://heil.math.gatech.edu/6338/summer08/section9b.pdf, 2008. Lecture notes on MATH 6338 (Real Analysis II) at Gerogia Insisute of Technology, Accessed: January 5, 2024.

8 Chris Heunen, Ohad Kammar, Sam Staton, and Hongseok Yang. A convenient category for higher-order probability theory. In *Proceedings of the 32nd Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS 2017. IEEE Press, 2017. doi:10.1109/lics.2017.8005137.

9 Michikazu Hirata. The Lévy-Prokhorov metric. *Archive of Formal Proofs*, June 2024. , Formal proof development. URL: https://isa-afp.org/entries/Levy_Prokhorov_Metric.html.

10 Michikazu Hirata. The Riesz representation theorem. *Archive of Formal Proofs*, June 2024. , Formal proof development. URL: https://isa-afp.org/entries/Riesz_Representation.html.

11 Michikazu Hirata, Yasuhiko Minamide, and Tetsuya Sato. Semantic foundations of higher-order probabilistic programs in Isabelle/HOL. In Adam Naumowicz and René Thiemann, editors, *14th International Conference on Interactive Theorem Proving (ITP 2023)*, volume 268 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 18:1–18:18, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ITP.2023.18.

12 Johannes Hölzl and Armin Heller. Three chapters of measure theory in Isabelle/HOL. In Marko van Eekelen, Herman Geuvers, Julien Schmaltz, and Freek Wiedijk, editors, *Interactive Theorem Proving*, pages 135–151, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

13 Johannes Hölzl, Fabian Immler, and Brian Huffman. Type classes and filters for mathematical analysis in Isabelle/HOL. In *Proceedings of the 4th International Conference on Interactive Theorem Proving*, ITP 2013, pages 279–294, Berlin, Heidelberg, 2013. Springer-Verlag. doi:10.1007/978-3-642-39634-2_21.

14 Alexander S. Kechris. *Classical Descriptive Set Theory*. Graduate Texts in Mathematics. Springer New York, 1995. doi:10.1007/978-1-4612-4190-4.

15 Kalle Kytölä. LevyProkhorovMetric.lean. https://github.com/leanprover-community/mathlib4/blob/master/Mathlib/MeasureTheory/Measure/LevyProkhorovMetric.lean. Accessed May 27th 2024.

16 Holden Lee. Vector spaces. *Archive of Formal Proofs*, August 2014. , Formal proof development. URL: https://isa-afp.org/entries/VectorSpace.html.

17 Paul Lévy. *Thèorie de l'addition des variables alèatoires*. Gauthier-Villars, Paris, 1937.

18 Keiko Narita, Kazuhisa Nakasho, and Yasunari Shidama. F. Riesz theorem. *Formalized Mathematics*, 25(**3**):179–184, 2017. doi:10.1515/forma-2017-0017.

19 Anthony Narkawicz. A formal proof of the Riesz representation theorem. *Journal of Formalized Reasoning*, 4(1):1–24, January 2011. doi:10.6092/issn.1972-5787/1952.

20 Lawrence C. Paulson. Porting the HOL Light metric space library. https://lawrencecpaulson.github.io/2023/07/12/Metric_spaces.html. Accessed: December 31. 2023.

21 Yuri Vasilyevich Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability and Its Applications*, 1(2):157–214, 1956. doi:10.1137/1101016.

22 Walter Rudin. *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, Inc., USA, 1987.

23 Mirah Shi. Nets and filters. https://www.uvm.edu/~smillere/TProjects/MShi20s.pdf, 2020. Accessed November 17th 2023.

24 Shashi Mohan Srivastava. *A Course on Borel Sets*. Springer, 1998. doi:10.1007/b98956.

25 The mathlib Community. The Lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs*, CPP 2020, pages 367–381. Association for Computing Machinery, 2020. doi:10.1145/3372885.3373824.

26 René Thiemann and Akihisa Yamada. Matrices, jordan normal forms, and spectral radius theory. *Archive of Formal Proofs*, August 2015. , Formal proof development. URL: https://isa-afp.org/entries/Jordan_Normal_Form.html.

27 Onno van Gaans. Probability measures on metric spaces. https://www.math.leidenuniv.nl/~vangaans/jancol1.pdf. Accessed: February 29. 2024.

28 Cédric. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008. doi:10.1007/978-3-540-71050-9.

29 Shaoyi Zhang. Existence and application of optimal markovian coupling with respect to non-negative lower semi-continuous functions. *Acta Mathematica Sinica*, 16(2):261–270, 2000. doi:10.1007/s101140000049.