Basic Concepts in Weak Convergence

Chia-Min Wei

November 18, 2024

1 Introduction

Let $\{\mathrm{P}_n\}_{n=1}^{\infty}$ be a sequence of probability measures on a probability space. What do we mean by P_n converges to P? In this note, we introduce basic concepts of weak convergence which are used throughout the literature of empirical processes.

2 Measures on a Metric Space

We start from introducing some properties of a probability measure defined on a metric space. Let S denote a metric space and $\mathcal{S}$ denote the Borel σ-algebra.

We say that a probability measure P defined on a topological space equipped with the Borel σ-algebra $(T, B(T))$ is *regular* if for any $\epsilon > 0$ and $A \in B(T)$, there exists a closed set F and an open set G such that

$$F \subset A \subset G, \quad \mathrm{P}(G - F) < \epsilon.$$

On the other hand, we say that P is *t*ight if there exists a compact set K such that

$$\mathrm{P}(K) > 1 - \epsilon.$$

Theorem 1: Any probability measure defined on a metric space $(S, \mathcal{S})$ is regular.

Proof. For any closed set F, consider the sequence of open set

$$G_\epsilon = \{x \in S : d(x, F) < \epsilon\}.$$

Since F is closed, $G_\epsilon \downarrow F$ as $\epsilon \to 0$. Therefore, $\mathrm{P}(G_\epsilon - F) \to 0$ by the continuity of probability measures. Since $\mathcal{S}$ is generated by closed sets in S, by checking all sets $A \subset S$ that satisfy the asserted property is a sigma-field, the proof is done. □

There is an important implication of Theorem 1: to check that two probability measures on a metric space coincide, it suffices to check whether they coincide on open sets (closed sets).

> **Theorem 2:** Probability measures P and Q coincide if and only if for any bounded and continuous function $f : (S, \mathcal{S}) \to \mathbb{R}$,
>
> $$\int_S f \, d\,\mathrm{P} = \int_S f \, d\,\mathrm{Q}\,.$$

Proof. Let us apply the conclusion we just obtained. Let F be a closed set, and $\mathbf{1}_F$ its indicator function. We can approximate $\mathbf{1}_F$ with a continuous and bounded functions defined by

$$f_\epsilon(x) = (1 - d(x, F)/\epsilon)^+.$$

f_ϵ converges pointwise to $\mathbf{1}_F$ since F is closed. By Bounded Convergence Theorem,

$$\int f_\epsilon(x)\,\mathrm{P}(dx) \longrightarrow \int \mathbf{1}_F\,\mathrm{P}(dx) = \mathrm{P}(F), \quad \int f_\epsilon(x)\,\mathrm{Q}(dx) \longrightarrow \mathrm{Q}(F).$$

But $\int f_\epsilon(x)\,\mathrm{P}(dx) = \int f_\epsilon(x)\,\mathrm{Q}(dx)$ for all ϵ, and thus we must have $\mathrm{P}(F) = \mathrm{Q}(F)$. □

Now we introduce another important concept: tightness. We say that a probability measure P define on a topological space is *tight* if for any $\epsilon > 0$, there exists a compact set K such that

$$\mathrm{P}(K) > 1 - \epsilon.$$

Before we dive into a result regarding the tightness of a probability measure on a metric space, let us review some concepts in topology.

We say that a topological space is *second countable* if there exists a countable basis for the topology; *Lindelöff* if any open cover of a subset $A \subset S$ admits a countable subcover; *separable* if there exists a countable and dense subset D. ("Dense" means that any open subset of S includes an element of D.)

Proposition 1: A metric space S is second countable if and only if it is Lindelöff if and only if it is separable.

We say that a metric space (S, d) is totally bounded if for any $\epsilon > 0$, there exists a finite number of open balls whose center lies in S and their union contains S.

Theorem 3: If a metric space $(S, \mathcal{S})$ is complete and separable, then any probability measure on $(S, \mathcal{S})$ is tight.

Proof. Since S is separable, there exists a countable and dense subset $\{x_i\}_{i=1}^{\infty}$. Note that for any k, the collection of $1/k$ open balls $\{B(x_i, 1/k)\}_{i=1}^{\infty}$ covers S. Let $\epsilon > 0$. For each, choose $n_k \in \mathbb{N}$ such that

$$\mathrm{P}\left(\bigcup_{i \le n_k} B(x_i, 1/k)\right) > 1 - \epsilon/2^k.$$

Now the set

$$\bigcap_{k \ge 1} \bigcup_{i \le n_k} B(x_i, 1/k)$$

is totally bounded. Write $A_k = \bigcup_{i \le n_k} B(x_i, 1/k)$. Observe that

$$\mathrm{P}\left(\bigcap_{k \ge 1} A_k\right) = \mathrm{P}\left(\left(\bigcup_{k \ge 1} A_k^c\right)^c\right) \ge 1 - \sum_{k \ge 1} \mathrm{P}(A_k^c) > 1 - \epsilon.$$

Since S is complete, the closure K of $\bigcap_{k \ge 1} \bigcup_{i \le n_k} B(x_i, 1/k)$ is compact. And clearly $\mathrm{P}(K) > 1 - \epsilon$. □

Definition 1 (Separating Class): Let $(S, \mathcal{S})$ be a measured space. A subclass $\mathcal{A}$ of $\mathcal{S}$ is called a separating class if any two probability measures coincide on $\mathcal{S}$ if and only if they coincide on $\mathcal{A}$.

As we mentioned, the class of closed sets is a separating class for the Borel σ-algebra. Indeed, by Dynkin's π-λ Lemma, any π-system that generates the Borel σ-algebra is a separating class.

Example 2.1: Let $\mathbb{R}^{\infty}$ be the space of sequences of real numbers. Recall that the product topology of $\mathbb{R}^\infty$ is the one generated by the basis:

$$\mathcal{B} = \{O_1 \times O_2 ... O_n \times \mathbb{R} \times \mathbb{R} ... : O_i' s \text{are open}, n < \infty\}.$$

Hence, $\mathbb{R}^\infty$ is indeed a metric space. The product topology is separable. The countable collection of points

$$\mathbf{Q} = \{(q_1, ..., q_n, 0, ..., 0, ...) : q_i \in \mathbb{Q}, n < \infty\}$$

is dense in $\mathbb{R}^\infty$. Define a metric on $\mathbb{R}$ by $b(x_i, y_i) = d(x_i, y_i) \wedge 1$. We can define a metric on $\mathbb{R}^\infty$

$$\rho(x, y) = \sum_{i=1}^{\infty} 2^{-i} b(x_i, y_i).$$

Then indeed this metric induces the product topology. Therefore, $\mathbb{R}^\infty$ is a metric space. Also, with this metric,

$$x \longrightarrow y \iff x_i \longrightarrow y_i \text{ for all } i.$$

Hence, $\mathbb{R}^\infty$ is complete. We conclude that $\mathbb{R}^\infty$ is a separable and complete metric space. By Theorem 3, any probability measure on $\mathbb{R}^\infty$ is tight.

Since $\mathbb{R}^\infty$ is separable, it is also Lindelöff. This means that the σ-algebra generated by $\mathcal{B}$ is indeed the Borel σ-algebra. It is also clear that $\mathcal{B}$ is a π-system. Hence, the basis $\mathcal{B}$ is a separating class.

Example 2.2: Let $C = C[0, 1]$ be the set of continuous functions f on $[0, 1]$. Define the norm of f as $\|f\| = \sup_{x \in [0,1]} |f(x)|$, and give it the uniform metric,

$$\rho(f, g) = \|f - g\|.$$

We show that C is separable. Let D_k be the set of polygonal functions that are linear over each subinterval $[(i-1)/k, i/k]$ and have rational values at the endpoints. Since each D_k is countable, the set $D = \bigcup_{k \geq 1} D_k$ is also countable. To show that D is dense, for given f and ϵ, choose k so large so that the partition of $[0, 1]$ is so fine, that within each subinterval, $|f(x) - f(y)| < \epsilon$ for any two points x and y in that subinterval. By choosing the values of the endpoints to be rational numbers very close to the original value of f on the endpoints, we can construct a $g \in D$ such that $\rho(f, g) < \epsilon$. Recall that C is also complete. Therefore, any probability on the Borel

σ-algebra of C is tight.

Write $\mathcal{C}$ as the Borel σ-algebra of C. Define the projection of functions on $t_1, ..., t_k \in [0, 1]$ as

$$\pi_{t_1,...,t_k}(f) = (f(t_1), ..., f(t_k)).$$

$\pi : C \to \mathbb{R}^k$ is a continuous function, and thus also measurable. In C, we say that a set A is finite-dimensional if there exists $t_1, ..., t_k$ and $H \subset \mathbb{R}^k$ such that $A = \pi^{-1}_{t_1,...,t_k}(H)$. Namely,

$$A = \{f \in C : (f(t_1), f(t_2), ..., f(t_k)) \in H\}.$$

Now for any set $\pi^{-1}_{t_1,...,t_k}(H)$ and $s_1, ..., s_l \in [0, 1]$, the set can be written as $\pi^{-1}_{t_1,...,t_k,s_1,...,s_l}(H')$ for some $H' \subset \mathbb{R}^{k+l}$. Hence, for any

$$\begin{aligned} \pi^{-1}_{t_1,...,t_k}(H_1) \cap \pi^{-1}_{s_1,...,s_l}(H_2) &= \pi^{-1}_{t_1,...,t_k,s_1,...,s_l}(H'_1) \cap \pi^{-1}_{t_1,...,t_k,s_1,...,s_l}(H'_2) \\ &= \pi^{-1}_{t_1,....,t_k,s_1,...s_l}(H'_1 \cap H'_2). \end{aligned}$$

This proves that the collection of finite-dimensional sets is a π-system. Call such collection C_F. Now each closed ball in C can be written as a countable intersection of sets in C_F

$$\overline{B}(f, \epsilon) = \bigcap_{r\in\mathbb{Q}} \{g : |g(r) - f(r)| \leq \epsilon\}.$$

Hence, $\sigma(C_F)$ contains all closed balls, and thus all open balls. Since C is separable and thus Lindelöff, $\sigma(C_F)$ contains all open sets. Since C_F is a π-system and $\sigma(C_F) = \mathcal{C}$, C_F is a separating class.

3 Weak Convergence of Probability Measures

Definition 2 (Weak Convergence): We say that a sequence of probability measure $\{P_n\}$ defined on $(S, \mathcal{S})$ converges weakly to a probability measure P, denoted as $P_n \Rightarrow P$, if for any bounded and continuous real function f we have

$$\int_S f \, d\,\mathrm{P}_n \longrightarrow \int_S f \, d\,\mathrm{P}\,.$$

Definition 3: Let X_n's and X be random variables with realized values on $(S, \mathcal{S}')$. Let $\mu_n(\mu)$ be the measure on $\mathcal{S}'$ induced by $X_n(X)$. We say that X_n converge to X weakly if μ_n converge to μ weakly.

We start from some simple examples to illustrate the ideas behind the definition.

Example 3.1: On an arbitrary metric space S, let $\delta_x(A) = \mathbf{1}_A(x)$ be the probability measure that assigns unit mass on the point x. If $x_n \to x$ and f is continuous, then $f(x_n) \to f(x)$, and thus $\delta_{x_n} \Rightarrow \delta_x$. On the other hand if $x_n \not\to x$, there exists $\epsilon > 0$ such that $d(x_n, x) > \epsilon$ for infinitely many n's. Simply choose the bounded and continuous function $f(y) = (1 - d(y, x)/\epsilon)^+$. Then $f(x) = 1$ but $f(x_n) = 0$ for infinitely many n's. This shows that $\delta_{x_n} \not\Rightarrow \delta_x$. Therefore, $\delta_{x_n} \Rightarrow \delta_x$ if and only if $x_n \to x$.

Example 3.2: Let $S = [0, 1]$ with the usual metric. Consider a sequence $\{A_n\}$ of sets $A_n = \{x_{kn}\}_{k=1}^{r_n}$ for each n. Suppose $\{A_n\}$ is asymptotically uniform in the sense that for any subinterval $J \subset [0, 1]$,

$$\frac{1}{r_n} \#\{k : x_{nk} \in J\} \longrightarrow |J|.$$

Define P_n to be uniform on A_n and P be the Lebesgue measure on $[0, 1]$. Then $\mathrm{P}_n \Rightarrow \mathrm{P}$. Let f be continuous and bounded defined on $[0, 1]$. f is Lebesgue integrable and also Reimann integrable. For any $\epsilon > 0$, there exists fine enough partition $\{J_1, ..., J_m\}$ such that the upper Riemann sum and the lower Riemann sum are within ϵ of the integral.

$$\sum_{i=1}^{m} \underline{v}_i |J_i| + \epsilon \geq \int_0^1 f \, d\,\mathrm{P}, \quad \sum_{i=1}^{m} \overline{v}_i |J_i| - \epsilon \leq \int_0^1 \, d\,\mathrm{P}\,.$$

Asymptotic on n,

$$\begin{aligned}
\int_0^1 f \, d\,\mathrm{P}_n &= \sum_{k=1}^{r_n} \frac{1}{r_n} f(x_{nk}) \\
&\leq \sum_{i=1}^{m} \frac{1}{r_n} \#\{k : x_{nk} \in J_i\} \overline{v}_i \\
&\longrightarrow \sum_{i=1}^{m} |J_i| \overline{v}_i \leq \int_0^1 f \, d\,\mathrm{P} + \epsilon.
\end{aligned}$$

Similarly, one can prove $\int_0^1 f\,d\,\mathrm{P}_n$ is asymptotically larger or equal to $\int_0^1 f\,d\,\mathrm{P}$. This proves that $\int_0^1 f\,d\,\mathrm{P}_n \to \int_0^1 f\,d\,\mathrm{P}$. Hence, $\mathrm{P}_n \Rightarrow \mathrm{P}$.

Definition 4 (P-continuity Set): We call a set $A \subset S$ a P-continiuty set, if $\mathrm{P}(\partial A) = 0$, where ∂A denotes the boundary of A. ($\partial A = \overline{A} - \mathrm{int}(A)$).

The following theorem provides useful conditions equivalent to weak convergence.

Theorem 4 (Portmanteau Theorem): Suppose $\{P_n\}$ and P are probability measures defined on $(S, \mathcal{S})$. These conditions are all equivalent to $\mathrm{P}_n \Rightarrow \mathrm{P}$:

(i) For any continuous and bounded real f, $\int_S f\,d\,\mathrm{P}_n \to \int_S f\,d\,\mathrm{P}$.

(ii) For any uniformly continuous and bounded real f, $\int_S f\,d\,\mathrm{P}_n \to \int_S f\,d\,\mathrm{P}$.

(iii) $\limsup_{n\to\infty} \mathrm{P}_n(F) \le \mathrm{P}(F)$ for all closed F.

(iv) $\liminf_{n\to\infty} \mathrm{P}_n(G) \ge \mathrm{P}(G)$ for all open G.

(v) $\mathrm{P}_n(A) \to \mathrm{P}(A)$ for all P-continiuty sets A.

Recall that in Example 3.1, we see that $\delta_{x_n} \Rightarrow \delta_x \iff x_n \to x$. If we choose $A = \{x\}$, then apparently $\delta_{x_n}(A) = 0 \not\to 1 = \delta_x(A)$. This does not contradict with Theorem 4 because $\{x\}$ is not a P-continuity set.

Proof.

- (i) $\implies$ (ii): Trivial
- (ii) $\implies$ (iii): Let F be a closed set in S. Set for all $\epsilon > 0$, $f_\epsilon(x) = (1 - d(x,F)/\epsilon)^+$ and $F_\epsilon = \{x : d(x,F) < \epsilon\}$. Since F is closed, $F_\epsilon \downarrow F$ as $\epsilon \to 0$. Also, $\int f_\epsilon\,d\,\mathrm{P}_n \ge \mathrm{P}_n(F)$ for all n. Fix $\delta > 0$. There exists a small enough ϵ such that $\mathrm{P}(F_\epsilon) \le \mathrm{P}(F) + \delta$. Note that

$$\mathrm{P}_n(F) \le \int f_\epsilon\,d\,\mathrm{P}_n \to \int f_\epsilon\,d\,\mathrm{P} \le \mathrm{P}(F_\epsilon) \le \mathrm{P}(F) + \delta$$
$$\implies \limsup_{n\to\infty} \mathrm{P}_n(F) \le \mathrm{P}(F) + \delta.$$

- (iii) $\implies$ (iv): It follows easily from complement arguments.

- (iii), (iv) $\implies$ (v): Since A is a P-continuity set, $\mathrm{P}(\overline{A}) = \mathrm{P}(A) = \mathrm{P}(\mathrm{int}A)$. We then have

$$\limsup_{n\to\infty} \mathrm{P}_n(\overline{A}) \leq \mathrm{P}(\overline{A}) = \mathrm{P}(A)$$
$$\liminf_{n\to\infty} \mathrm{P}_n(\mathrm{int}A) \geq \mathrm{P}(\mathrm{int}A) = \mathrm{P}(A).$$

  This then implies

$$\limsup_{n\to\infty} \mathrm{P}_n(A) = \liminf_{n\to\infty} \mathrm{P}_n(A) = \mathrm{P}(A).$$

- (v) $\implies$ (i): By linearity of integrals, we can assume f is bounded between 0 and 1. Observe that

$$\int_S f\, d\,\mathrm{P} = \int_0^1 \mathrm{P}(f > t)\, dt, \quad \int_S f\, d\,\mathrm{P}_n = \int_0^1 \mathrm{P}_n(f > t)\, dt.$$

  Since f is continuous, $\partial\{s : f(s) > t\} \subset \{s : f(s) = t\}$. But $\mathrm{P}(s : f(s) = t)$ can be strictly positive only for countably many t's. By (v), $\mathrm{P}_n(f > t) \to \mathrm{P}(f > t)$ for almost every t. Hence, by BCT,

$$\int_0^1 \mathrm{P}_n(f > t)\, dt \to \int_0^1 \mathrm{P}(f > t)\, dt.$$

□

It will be nice if we only need to check whether P_n converges to P on a certain class of sets in $\mathcal{S}$ to unsure that $\mathrm{P}_n \Rightarrow \mathrm{P}$.

Theorem 5: Suppose (i) that $\mathcal{A}_P$ is a π-system and (ii) that each open set is a countable union of $\mathcal{A}_P$ sets. If $\mathrm{P}_n(A) \to \mathrm{P}(A)$ for every A in $\mathcal{A}_P$, then $\mathrm{P}_n \Rightarrow \mathrm{P}$.

Theorem 6: Suppose (i) that $\mathcal{A}_P$ is a π-system and (ii) that S is separable, and for every $x \in S$ and $\epsilon > 0$, there exists $A \in \mathcal{A}_P$ such that

$$x \in \mathrm{int}(A) \subset A \subset B(x, \epsilon).$$

If $\mathrm{P}_n(A) \to \mathrm{P}(A)$ for all $A \in \mathcal{A}_P$, then $\mathrm{P}_n \Rightarrow \mathrm{P}$.

Definition 5 (Convergence-Determining Class): We call a subclass $\mathcal{A}$ of $\mathcal{S}$ a *convergence-determining class* if, *for any* $\{\mathrm{P}_n\}$ and P, $\mathrm{P}_n(A) \to \mathrm{P}(A)$ for all P-continuity A in $\mathcal{A}$ implies $\mathrm{P}_n \Rightarrow \mathrm{P}$.

To ensure that a collection of $\mathcal{A}$ is convergence-determining, we must make sure that the class of P-continiuty sets $\mathcal{A}_P$ in $\mathcal{A}$ satisfies the conditions of Theorem 6 **for any** P. Fix any $x \in S$ and $\epsilon > 0$. Let $\mathcal{A}_{x,\epsilon}$ denote the collection of sets in $\mathcal{A}$ such that

$$x \in \mathrm{int}(A) \subset A \subset B(x, \epsilon),$$

and let $\partial\mathcal{A}_{x,\epsilon}$ denote the collection of their boundaries.

Theorem 7: Suppose that (i) that $\mathcal{A}$ is a π-system and (ii) that S is separable and for each $x \in S$ and ϵ, $\partial\mathcal{A}_{x,\epsilon}$ either contains $\varnothing$ or contains uncountably many disjoint sets. Then $\mathcal{A}$ is a convergence-determining class.

Proof. Let $\{\mathrm{P}_n\}$ and P be given arbitrary. Let $\mathcal{A}_P$ denote the collection of P-continuity sets in $\mathcal{A}$. Apparently, $\mathcal{A}_P$ is a π-system. Now fix $x \in S$ and $\epsilon > 0$. Since $\partial\mathcal{A}_{x,\epsilon}$ must contain a set E with $\mathrm{P}(E) = 0$, this means that there is a P-continiuty set in $\mathcal{A}_{x,\epsilon}$. Hence, $\mathcal{A}_P$ satisfies the conditions in Theorem 6. This shows that $\mathcal{A}$ is a convergence-determining class. □

Example 3.3: The collection of $\mathcal{A}$ finite intersections of open balls form a convergence-determining class. Because

$$\partial B(x, r) \subset \{y : d(x, y) = r\},$$

and thus either $\varnothing$ is in $\partial\mathcal{A}_{x,\epsilon}$ or there are uncountably many disjoint sets in $\partial\mathcal{A}_{x,\epsilon}$.

Example 3.4: Consider the collection of rectangles in $\mathbb{R}^k$, sets of the form $\{x : b < x \leq a\}$. The collection satisfies Theorem 7, and hence is a convergence-determining class.

Example 3.5: In $\mathbb{R}^n$, the class $\mathcal{A}$ of sets

$$Q_x = \{y : y \leq x\}$$

is also a convergence-determining class. Suppose $\mathrm{P}_n(Q_x) \to \mathrm{P}(Q_x)$ for each Q_x with $\mathrm{P}(\partial Q_x) = 0$. For each $1 \leq i \leq k$, define $E_i = \{t : \mathrm{P}\{x : x_i = t\} > 0\}$. E_i is at most countable. Hence, there are uncountably many rectangles (in the form $(a, b]$) such that each vertex $x = (x_1, ..., x_k)$ satisfy $x_i \notin E_i$. Let $\mathcal{A}_P$ be the collection of such rectangles. Such collection satisfies the condition in Theorem 6. For any $A \in \mathcal{A}_P$, for each vertex x of A, $\mathrm{P}(\partial Q_x) = 0$. A can be written as inclusion and exclusions of the Q_x's. It then follows that $\mathrm{P}_n(A) \to \mathrm{P}(A)$ by the inclusion-exclusion formula. And thus $\mathrm{P}_n \Rightarrow \mathrm{P}$.

There is another way to state that $\mathcal{A}$ is a convergence-determining class. For any probability measure P, define $F(x) = \mathrm{P}\{y : y \leq x\}$. Then $\mathrm{P}_n \Rightarrow \mathrm{P}$ if and only if $F_n(x) \to F(x)$ for all x at which F is continuous.

Hence, for $\mathbb{R}^n$-valued random variables X_n, saying that X_n converges weakly to X is equivalent to saying that X_n converges to X in distribution.

Suppose that $h : (S, \mathcal{S}) \to (S', \mathcal{S}')$ is a measurable function that maps S into S'. For any probability measure P on $(S, \mathcal{S})$, h then induces a measure on S', $\mathrm{P}\, h^{-1}$, defined by

$$\mathrm{P}\, h^{-1}(A) = \mathrm{P}(h^{-1}(A)).$$

Theorem 8 (Continuous Mapping Theorem): Let $h : (S, \mathcal{S}) \to (S', \mathcal{S}')$ be a continuous function, and suppose $\mathrm{P}_n \Rightarrow \mathrm{P}$ on $(S, \mathcal{S})$. Then $\mathrm{P}_n\, h^{-1} \Rightarrow\ h^{-1}$ on $(S, \mathcal{S})$.

Proof. Let f be a continuous function from $(S', \mathcal{S}')$ into $(\mathbb{R}, \mathcal{R})$. Since f and h are continuous, $f \circ h : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{R})$ is also continuous. Hence, by change of variable,

$$\int_{S'} f\, d\mathrm{P}_n\, h^{-1} = \int_S f \circ h\, d\mathrm{P}_n \longrightarrow \int_S f \circ h\, d\mathrm{P} = \int_{S'} f\, d\mathrm{P}\, h^{-1}.$$

□

Corollary: Let X_n's and X be $\mathbb{R}^n$-valued, and suppose X_n converges to X in distribution. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be continuous. Then $f(X_n)$ converges to $f(X)$ in distribution.

4 Prohorov's Theorem

In the previous section, we discussed how to check if a sequence of probability measure $\{\mu_n\}$ converges weakly to a probability measure μ by introducing the concept of convergence-determining class (Definition 5). But how do we know if $\{\mu_n\}$ converges weakly in the first place? We first introduce the notion of relative compactness.

Definition 6 (Relatively Compact): Let $\mathcal{P}$ be a family of probability measures defined on $(S, \mathcal{S})$. Then we say that $\mathcal{P}$ is *relatively compact* if for any sequence in $\mathcal{P}$, there exists a subsequence that converges weakly to some **probability measure**.

Let us recall a result from probability theory:

Theorem 9 (Helly's Selection Theorem): Let $\{\mu_n\}_{n=1}^{\infty}$ be a sequence of sub-probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then there exists a subsequence $\{\mu_{n_k}\}_{k=1}^{\infty}$ and a sub-probability measure μ such that

$$\mu_n(a, b] \longrightarrow \mu(a, b]$$

for all $a, b \in [-\infty, \infty]$ such that $\mu(\partial(a, b]) = 0$. We say that μ_{n_k} converges *vaguely* to μ.

Remark: If a sequence of sub-probability measure $\{\mu_n\}$ converges vaguely to a μ, then such μ is unique. Say μ_n converges vaguely to ν_1 and ν_2. Then ν_1 and μ_2 agrees on the π-system:

$$\{(a, b] : \nu_1(\partial(a, b]) = 0 = \nu_2(\partial(a, b]) = 0\},$$

which generates $\mathcal{B}(\mathbb{R})$.

Proof. Let F_n denote the cumulative distribution function corresponding to μ_n. Enumerate the set of all rational numbers $\{q_i\}_{i=1}^{\infty}$. By Bolzano Weierstrass Theorem, there exists a subsequence of $\{F_n\}$, $\{F_{1k}\}$ such that $\{F_{1k}(q_1)\}$ converges to some point $a_1 \in [0, 1]$. Further from this subsequence, there exists a subsequence $\{F_{2k}\}$ such that $\{F_{2k}(q_2)\}$ converges to some $a_2 \in [0, 1]$. Iteratively, we have for all q_j, sequences $\{F_{jk}\}$ such that $\{F_{jk}(q_j)\}$ converges to some point a_j, and that $\{F_{jk}\}$ is a subsequence of $\{F_{(j-1)k}\}$. Now consider the sequence $\{G_k\}_{k=1}^{\infty} = \{F_{kk}\}_{k=1}^{\infty}$. Then $G_k(q_j) \to a_j$ for all $q_j \in \mathbb{Q}$. Now define for all $x \in \mathbb{R}$,

$$G(x) = \inf\{a_j : j \text{ such that } q_j \geq x\}.$$

Then $G(x)$ is nondecreasing and right continuous. Moreover, $G(q_j) = a_j$ for all $j \in \mathbb{N}$. Hence, $G_k(q) \to G(q)$ for all $q \in \mathbb{Q}$. Our proof is done if we can show that $G_k(x) \to G(x)$ for all x such that G is continuous. Let x be a point at which G is continuous. Let $\epsilon > 0$. Then there exists $q, q' \in \mathbb{Q}$ so close to x such that

$$G(x) - \epsilon \leq G(q) \leq G(x) \leq G(q') \leq G(x) + \epsilon.$$

For all k,

$$G_k(q) \leq G_k(x) \leq G_k(q').$$

Taking $k \to \infty$, we have $G_k(q) \to G(q)$ and $G_k(q') \to G(q')$, and thus

$$G(x) - \epsilon \leq \liminf G_k(x) \leq \limsup G_k(x) \leq G(x) + \epsilon.$$

Since ϵ is arbitrary, we have

$$\lim_{k \to \infty} G_k(x) = G(x).$$

□

Therefore, any sequence of probability measures $\{\mu_n\}$ are guaranteed to have a subsequence that converges to a sub-probability measure. However, it is not guaranteed that such measure is a probability measure.

Example 4.1: Consider the sequence of probability measure $\{\delta_n\}_{n=1}^{\infty}$ defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where δ_n is the probability that assigns probability 1 to the point n. $\{\delta_n\}$ converges vaguely to μ that assigns probability 0 to any set. In this case, measure is *escaping* to infinity.

Example 4.2: Let $\{\mu_n\}_{n=1}^{\infty}$ be the sequence of probability measure that has uniform distribution on $[-n, n]$. Then μ_n converges vaguely to μ that assigns probability 0 to any set. In this case, measure simply *evaporates*.

Hence, we need conditions on $\{\mu_n\}_{n=1}^{\infty}$ that guarantees that measures do not escape or evaporate. Moreover, for probability measures defined on $(C, \mathcal{C})$, we don't even have Theorem 9 to ensure vague convergence.

Example 4.3: Consider $\{\delta_n\}_{n=1}^{\infty}$ defined by δ_n assigning probability 1 to the continuous function z_n that increases linearly on $[0, 1/n]$ and decreases linearly on $[1/n, 2/n]$, and stays at 0 to the right of $2/n$. Let δ_0 be the probability measure that assigns probability 1 to the constant 0 function. Note that for any $t_1, ..., t_k \in [0, 1]$,

$$\delta_n(A) \longrightarrow \delta_0(A),$$

where $A = \{f \in C : (t_1, ..., t_k) \in H\}$ for some $H \in \mathbb{R}^k$. However, since $d(z_n, 0) = 1$ for all n, $z_n \not\to 0$, and hence $\delta_n \not\Rightarrow \delta_0$. This shows that the collection C_F of finite dimensional sets is a separating class, but **not** a convergence determining class.

However, if we do know that $\{\mathrm{P}_n\}$ *is relatively compact, and* $\mathrm{P}_n(A) \to \mathrm{P}(A)$ *for all* $A \in C_F$*, then we are guaranteed that* $\mathrm{P}_n \Rightarrow \mathrm{P}$. For any subsequence of $\{P_n\}$, say $\{P'_n\}$, there exists $\{P'_{nk}\}$ that converges to some probability measure P'. But then P' and P must agree on the separating class C_F, and so $\mathrm{P} = \mathrm{P}'$.

Now suppose we know that a sequence of probability measures $\{\mathrm{P}_n\}_{n=1}^{\infty}$ on $(C, \mathcal{C})$ is relatively compact, and that for all $t_1, ..., t_k \in [0, 1]$, there exists some probability measure $\mu_{t_1,...,t_k}$ on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ such that

$$\mathrm{P}_n\, \pi^{-1}_{t_1,...,t_k} \Rightarrow \mu_{t_1,...,t_k}.$$

We can then conclude that there exists a probability measure P *on* $(C, \mathcal{C})$ *such that its finite dimensional distribution*

$$\mathrm{P}\, \pi^{-1}_{t_1,...,t_k} = \mu_{t_1,...,t_k}.$$

Definition 7 (Tightness): We say that a family $\mathcal{P}$ of probability measures defined on $(S, \mathcal{S})$ is *tight* if for every ϵ there exists a compact set $K \subset S$ such that

$$\mathrm{P}(K) > 1 - \epsilon$$

for all $\mathrm{P} \in \mathcal{P}$.

Theorem 10 (Prohorov's Theorem): Let $\mathcal{P}$ be a family of probability measures defined on a metric space $(S, \mathcal{S})$. If $\mathcal{P}$ is tight, then it is relatively compact. If $(S, \mathcal{S})$ is separable and complete, the converse also holds.

Proof. Suppose $(S, \mathcal{S})$ is separable and complete and that $\mathcal{P}$ is relatively compact.

Statement 1: For any open sets $\{G_n\}$ such that $G_n \uparrow S$ and $\epsilon > 0$, there exists N such that for all $n \geq N$, $\mathrm{P}(G_n) \geq 1 - \epsilon$ for all $P \in \mathcal{P}$.

proof of claim. Suppose this is not true. Then for each n, we have some $\mathrm{P}_n \in \mathcal{P}$ such that $\mathrm{P}_n(G_n) \leq 1 - \epsilon$. Since $\mathcal{P}$ is relatively compact, there exists a subsequence $\{\mathrm{P}_{n_i}\}$ of $\{\mathrm{P}_n\}$ that weakly converges to some probability measure Q. Fixing any n, for all $n_i > n$,

$$\mathrm{P}_{n_i}(G_n) \leq \mathrm{P}_{n_i}(G_{n_i}) \leq 1 - \epsilon.$$

By Theorem 4,

$$\mathrm{Q}(G_n) \leq \liminf_i \mathrm{P}_{n_i}(G_n) \leq 1 - \epsilon.$$

And since $G_n \uparrow S$, we reach $\mathrm{Q}(S) \leq 1 - \epsilon$. A contradiction. □

Fix $\epsilon > 0$. Now for each k let $\{A_{ki}\}_{i=1}^{\infty}$ be a sequence of open balls with radius $1/k$ that covers S. Such sequence can be found since S is separable. By the claim above, for each k, there exists n_k such that $\mathrm{P}(\bigcup_{i \leq n_k} A_{ki}) > 1 - \epsilon/2^k$ for all $P \in \mathcal{P}$. The set

$$A = \bigcap_{k \geq 1} \bigcup_{i \leq n_k} A_{k_i}$$

is a totally bounded set. Since S is complete, the closure K of A is compact. Moreover, $\mathrm{P}(K) \geq 1 - \epsilon$ for all $\mathrm{P} \in \mathcal{P}$.

Now we prove the opposite direction. Suppose $\mathcal{P}$ is tight on a metric space $(S, \mathcal{S})$. Let $\{\mathrm{P}_n\}$ be a sequence of $\mathcal{P}$. We want to find a subsequence $\{\mathrm{P}_{n_i}\}$ and construct a probability measure Q such that $\mathrm{P}_{n_i} \Rightarrow \mathrm{Q}$.

Finding the subsequence: Choose compact sets K_u in such a way that $\mathrm{P}(K_u) \geq 1 - 1/u$ for all $\mathrm{P} \in \mathcal{P}$. The set $\bigcup_u K_u$ is separable. And hence there exists a countable collection $\mathcal{A}$ of open sets that satisfies the following property:

For any open G and $x \in \bigcup_u K_u$, there exists $A \in \mathcal{A}$ such that $x \in A \subset \overline{A} \subset G$.

Define $\mathcal{H}$ to be the set that consists of

$$\varnothing \text{ and finite unions of the form } \overline{A} \cap K_u \text{ where } A \in \mathcal{A}.$$

Note that $\mathcal{H}$ is countable. Therefore, using the diagonal method, we can find a subsequence $\{\mathrm{P}_{n_i}\}$ such that $\{\mathrm{P}_{n_i}(H)\}$ converges for all $H \in \mathcal{H}$. Define

$$\alpha(H) \coloneqq \lim_i \mathrm{P}_{n_i}(H).$$

Our goal is to construct a probability measure P such that

$$\mathrm{P}(G) = \sup_{H \subset G} \alpha(H)$$

for any open set G. If we succeed in doing so, then for any open set G, observe that

$$\liminf_i \mathrm{P}_{n_i}(G) \geq \alpha(H)$$

for all $H \subset G$, and so

$$\liminf_i \mathrm{P}_{n_i}(G) \geq \sup_{H \subset G} \alpha(H) = \mathrm{P}(G).$$

By Theorem 4, we can then conclude that $\mathrm{P}_{n_i} \Rightarrow \mathrm{P}$.

Construction of P: Note that $\mathcal{H}$ is closed under finite unions. Also, $\alpha(H)$ satisfies:

- $\alpha(H_1) \leq \alpha(H_2)$ if $H_1 \subset H_2$.
- $\alpha(H_1 \cup H_2) = \alpha(H_1) + \alpha(H_2)$ for all H_1, H_2.
- $\alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2)$.
- $\alpha(\varnothing) = 0$.

For any open sets G, define

$$\beta(G) = \sup_{H \subset G} \alpha(H).$$

Finally, for any $M \in \mathcal{S}$, define

$$\gamma(M) = \inf_{M \subset G} \beta(G).$$

We want to prove two things. First, γ is an outer measure. Suppose we succeed in doing so. Recall that the set

$$\mathcal{M} = \{M \subset S : \gamma(A) = \gamma(M \cap A) + \gamma(M^c \cap A) \text{ for all } A \subset S\}$$

is a σ-field, and that γ is a measure when restricted on $\mathcal{M}$. The second thing we want to prove is that all closed sets are in $\mathcal{M}$. If that holds, we can then conclude that $\mathcal{S} \subset \mathcal{M}$. This means that the restriction of γ to $\mathcal{S}$ is a measure. Let us call it P. $\mathrm{P}(G) = \gamma(G) = \beta(G)$ for all open G. And so

$$\mathrm{P}(S) = \beta(S) = \sup_{H \subset S} \alpha(H) \geq \sup_u \alpha(K_u) \geq \sup_u (1 - u^{-1}) = 1.$$

(Note that K_u's are in $\mathcal{H}$.) Therefore, P is indeed a probability measure.

Statement 2: If $F \subset G$ where F is closed and G is open, and if $F \subset H$ for some $H \in \mathcal{H}$, then

$$F \subset H_0 \subset G$$

for some $H_0 \in \mathcal{H}$.

Proof. Since F is closed and is contained in some K_u, it is compact. For each $x \in F$, there exists $A_x \subset \mathcal{A}$ such that

$$x \in A_x \subset \overline{A}_x \subset G.$$

There exists finitely many A_x's, say $\{A_i\}_{i=1}^n$ that covers F. Then we have

$$F \subset \bigcup_{i=1}^n (\overline{A}_i \cap K_u) \subset G.$$

□

Statement 3: γ is an outer measure on S.

Proof. We first prove that β is finitely subbadditive on the open sets. Let $H \subset G_1 \cup G_2$ where $H \in \mathcal{H}$ and G_1, G_2 are open. Define

$$F_1 := \{x \in H : \rho(x, G_1^c) \geq \rho(x, G_2^c)\}$$
$$F_2 := \{x \in H : \rho(x, G_2^c) \geq \rho(x, G_1^c)\}.$$

Then $F_1 \subset G_1$ and $F_2 \subset G_2$. If not, say $x \in F_1$ but not in G_1, then $x \in G_2$. Since G_2^c is closed, $\rho(x, G_1^c) = 0 < \rho(x, G_2^c)$, a contradiction. By Statement 2, $F_1 \subset H_1 \subset G_1$ and $F_2 \subset H_2 \subset G_2$ for some $H_1 \in \mathcal{H}$ and $H_2 \in \mathcal{H}$. But we know that

$$\alpha(H) \leq \alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2) \leq \beta(G_1) + \beta(G_2).$$

And so

$$\beta(G_1 \cup G_2) = \sup_{H \subset G_1 \cup G_2} \alpha(H) \leq \beta(G_1) + \beta(G_2).$$

Next, we prove that β is countably subadditive on the open sets. Let $H \subset \bigcup_{i=1}^{\infty} G_i$ where $H \in \mathcal{H}$ and G_i's are open. Since H is compact, there exists n such that $H \subset \bigcup_{i=1}^{n} G_i$. But by finite subadditivity,

$$\beta(H) \leq \sum_{i=1}^{n} \beta(G_i) \leq \sum_{i=1}^{\infty} \beta(G_i).$$

Finally, we can prove that γ is an outer measure. Clearly it is monotone. We now prove that it is countably subadditive. Let $\{M_i\}_{i=1}^{\infty}$ be subsets of S. By definition of γ, for each i, there exists open $G_i \supset M_i$ such that

$$\gamma(M_i) > \beta(G_i) + \epsilon/2^i.$$

Then we have

$$\gamma\left(\bigcup_{i=1}^{\infty} M_i\right) \leq \beta\left(\bigcup_{i=1}^{\infty} G_i\right) \leq \sum_{i=1}^{\infty} \beta(G_i) < \sum_{i=1}^{\infty} \gamma(M_i) + \frac{\epsilon}{2}.$$

Since this holds for all ϵ, we conclude that

$$\gamma\left(\bigcup_{i=1}^{\infty} M_i\right) \leq \sum_{i=1}^{\infty} \gamma(M_i).$$

□

Statement 4: The set of all closed sets is contained in the collection $\mathcal{M}$ of γ-measurable sets.

Proof. We first prove that $\beta(G) \geq \gamma(F \cap G) + \gamma(F^c \cap G)$ when F is closed and G is open. Fix $\epsilon > 0$. Observe that $F^c \cap G$ is open. Hence, there exists $H_1 \subset F^c \cap G$

such that

$$\alpha(H_1) \geq \beta(F^c \cap G) - \epsilon = \gamma(F^c \cap G) - \epsilon.$$

Since H_1 is compact, $H_1^c \cap G$ is open. Hence, there exists $H_0 \subset H_1^c \cap G$ such that

$$\alpha(H_0) \geq \beta(H_1^c \cap G) - \epsilon \geq \gamma(F \cap G) - \epsilon.$$

Since H_1 and H_0 are disjoint, and both are in G,

$$\beta(G) \geq \alpha(H_1 \cup H_0) = \alpha(H_1) + \alpha(H_0) \geq \gamma(F^c \cap G) + \gamma(F \cap G) - 2\epsilon.$$

Since ϵ is arbitrary,

$$\beta(G) \geq \gamma(F \cap G) + \gamma(F^c \cap G).$$

Finally, we prove that $\gamma(M) \geq \gamma(F \cap M) + \gamma(F^c \cap M)$ for all closed F. Fix $\epsilon > 0$. There exists an open set G such that $G \supset M$, and $\gamma(M) \geq \gamma(G) - \epsilon$.

$$\begin{aligned} \gamma(M) \geq \beta(G) - \epsilon &\geq \gamma(F \cap G) + \gamma(F^c \cap G) - \epsilon \\ &\geq \gamma(F \cap M) + \gamma(F^c \cap M) - \epsilon. \end{aligned}$$

Since ϵ is arbitrary, we have that

$$\gamma(M) \geq \gamma(F \cap M) + \gamma(F^c \cap M).$$

$\gamma(M) \leq \gamma(F \cap M) + \gamma(F^c \cap M)$ follows directly from Statement 3. □

□

References

Billingsley, P. (1999). *Convergence of probability measures*. Wiley.

Athreya, K. B., & Lahiri, S. N. (2006). *Measure theory and probability theory*. Springer.