

Summary¹ of Convergence for Probability Measures.

Notations:

- (1) $(\mathbf{X}, \mathcal{X})$, a measurable space;
- (2) $\mathcal{P}(\mathbf{X})$, the collection of probability measures on $(\mathbf{X}, \mathcal{X})$;
- (3) $\nu, \mu, \mu_1, \mu_2, \dots \in \mathcal{P}(\mathbf{X})$;
- (4) $\mu[f] = \int_{\mathbf{X}} f d\mu$, $f : \mathbf{X} \rightarrow \mathbb{R}$.

General Definitions.

- **Separating class:** sub-set $\mathcal{C}$ of $\mathcal{X}$ such that, for every two probability measures μ, ν on $(\mathbf{X}, \mathcal{X})$, the equality $\mu(A) = \nu(A)$, $A \in \mathcal{C}$ implies $\mu(A) = \nu(A)$, $A \in \mathcal{X}$.
- **Set-wise (strong) convergence:** $\lim_{n \rightarrow \infty} \mu_n \stackrel{s}{=} \mu$: $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$ for all $A \in \mathcal{X}$;
- **Total variation distance:** $d_{TV}(\mu, \nu) = 2 \sup_{A \in \mathcal{X}} |\mu(A) - \nu(A)|$;
- **Convergence in total variation:** $\lim_{n \rightarrow \infty} \mu_n \stackrel{TV}{=} \mu$: $\lim_{n \rightarrow \infty} d_{TV}(\mu_n, \mu) = 0$.

The special case: $\mathbf{X}$ is a metric space, with metric ρ , and $\mathcal{X} = \mathcal{B}(\mathbf{X})$.

DEFINITIONS.

- **weak convergence** $\lim_{n \rightarrow \infty} \mu_n \stackrel{w}{=} \mu$: $\lim_{n \rightarrow \infty} \mu_n[f] = \mu[f]$ for all bounded continuous f .
- **Convergence-determining class:** sub-set $\tilde{\mathcal{C}}$ of $\mathcal{X}$ such that, for all probability measures μ_n, μ on $(\mathbf{X}, \mathcal{X})$, $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$, $A \in \tilde{\mathcal{C}}$, implies $\lim_{n \rightarrow \infty} \mu_n \stackrel{w}{=} \mu$.
- **the Lévy-Prokhorov metric:**

$$d_{LP}(\mu, \nu) = \inf\{\varepsilon > 0 : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon, A \in \mathcal{B}(\mathbf{X})\},$$

where $A^\varepsilon = \{x \in \mathbf{X} : \inf_{y \in A} \rho(x, y) \leq \varepsilon\}$.

- **tightness:** the collection $\{\mu_n, n \geq 1\}$ is tight if, for every $\varepsilon > 0$, there exists a compact $K_\varepsilon \subset \mathbf{X}$ so that, for all n , $\mu_n(K_\varepsilon) \geq 1 - \varepsilon$.
- **relative compactness:** the collection $\{\mu_n, n \geq 1\}$ is relatively compact if every sub-sequence contains a weakly converging sub-sub-sequence.

RESULTS.

- (1) Convergence-determining class is separating (by uniqueness of weak limit); separating class might not be convergence-determining (for example, equality of finite dimensional distributions for cadlag processes is enough for equality of corresponding measures but not enough for weak convergence).
- (2) Weak convergence defines topology on $\mathcal{P}(\mathbf{X})$.
- (3) **Portmanteau theorem:** the following conditions are equivalent
 - $\lim_{n \rightarrow \infty} \mu_n \stackrel{w}{=} \mu$;
 - $\lim_{n \rightarrow \infty} \mu_n[f] = \mu[f]$ EITHER for all bounded, *uniformly* continuous f OR for all bounded *Lipschitz* continuous² f ;
 - $\limsup_{n \rightarrow \infty} \mu_n(G) \leq \mu(G)$ for all *closed* sets $G \subset \mathbf{X}$;
 - $\liminf_{n \rightarrow \infty} \mu_n(G) \geq \mu(G)$ for all *open* sets $G \subset \mathbf{X}$;
 - $\lim_{n \rightarrow \infty} \mu_n(G) = \mu(G)$ for all measurable G with boundary ∂G such that $\mu(\partial G) = 0$.
- (4) **Continuous mapping theorem:** if $\mathbf{Y}$ is another metric space and $h : \mathbf{X} \rightarrow \mathbf{Y}$ is a continuous function, then weak convergence $\lim_{n \rightarrow \infty} \mu_n \stackrel{w}{=} \mu$ implies weak convergence $\lim_{n \rightarrow \infty} \mu_n \circ h^{-1} \stackrel{w}{=} \mu \circ h^{-1}$, where, for a measure $\nu \in \mathcal{P}(\mathbf{X})$, the measure $\nu \circ h^{-1} \in \mathcal{P}(\mathbf{Y})$ is defined by $\nu \circ h^{-1}(B) = \nu(x \in \mathbf{X} : h(x) \in B)$, $B \in \mathcal{B}(\mathbf{Y})$.
- (5) **Prohorov's theorem:** tightness implies relative compactness.

The most special case: $\mathbf{X}$ is a complete separable metric³ space.

RESULTS.

- (1) $\lim_{n \rightarrow \infty} \mu_n \stackrel{w}{=} \mu$ if and only if $\lim_{n \rightarrow \infty} d_{LP}(\mu_n, \mu) = 0$.

¹Sergey Lototsky, USC; version of May 15, 2026.

² $|f(x) - f(y)| \leq L_f \rho(x, y)$, $x, y \in \mathbf{X}$

³That is, Polish metric, which is not exactly the same as simply Polish...

(2) **Prohorov's theorem:** tightness is equivalent⁴ to relative compactness.

Further developments.

Let $\mathfrak{F}$ be a (sufficiently rich) collection of functions $f : \mathbf{X} \rightarrow \mathbb{R}$. Then we can define the corresponding $\mathfrak{F}$ -convergence by

$$\lim_{n \rightarrow \infty} \mu_n \stackrel{\mathfrak{F}}{=} \mu \iff \lim_{n \rightarrow \infty} \mu_n[f] = \mu[f] \quad \forall f \in \mathfrak{F}.$$

In particular, for a metric space $\mathbf{X}$, the **vague** convergence is the $\mathfrak{F}$ -convergence with $\mathfrak{F}$ equal the set of compactly supported functions (or, sometimes, the closure of this set with respect to uniform convergence).

Similarly, given another (sufficiently rich) collection $\mathfrak{G}$ of function, we can define the *pseudometric*⁵ on $\mathcal{P}(\mathbf{X})$ by

$$d_{\mathfrak{G}}(\mu, \nu) = \sup_{f \in \mathfrak{G}} |\mu[f] - \nu[f]|,$$

as well as the convergence in this pseudometric. For example,

- (1) If $\mathfrak{G}$ is the set of all bounded measurable functions, then $d_{\mathfrak{G}} = d_{\text{TV}}$.
- (2) If $\mathbf{X} = \mathbb{R}$ and $\mathfrak{G}$ is the collection of indicator functions of the sets $(-\infty, x]$ for all $x \in \mathbb{R}$, then $d_{\mathfrak{G}}$ is the **Kolmogorov** metric.
- (3) If $\mathbf{X}$ is a metric space and $\mathfrak{G}$ is the collection of continuous functions satisfying $|f| \leq 1$, then the $d_{\mathfrak{G}}$ is the **Radon** metric.

Further modifications of this construction lead to the **Wasserstein** distance $\mathcal{W}_p$, $p \geq 1$.

For a comprehensive survey of this topic, see the paper

Alison L. Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.

Random Probability Measures.

DEFINITION. A **random probability measure** μ on $(\mathbf{X}, \mathcal{X})$ is a *kernel* from the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbf{X}, \mathcal{X})$. In other words, $\mu = \mu(\omega, A)$, $\omega \in \Omega$, $A \in \mathcal{X}$, is a function such that $\omega \mapsto \mu(\omega, A)$ is a random variable for every A and $A \mapsto \mu(\omega, A)$ is a probability measure on $(\mathbf{X}, \mathcal{X})$ for every ω . In particular, $0 \leq \mu(\omega, A) \leq 1$.

Many examples from applications involve random *point measures* of the form

$$\mu(A) = \frac{1}{N} \sum_{k=1}^N 1(\xi_k \in A),$$

where ξ_k , $k = 1, \dots, N$ are $\mathbf{X}$ -valued random variables. Non-probability measures can also appear, for example, in the study of random processes with jumps.

A sequence $\{\mu_n, n \geq 1\}$ of random probability measures can converge to the limit μ (random or deterministic) in a variety of ways, *combining different modes of convergence for measures and for random variables*. Here are some examples:

- weak convergence in probability or with probability one (or in distribution, or in L_p): for every bounded continuous $f : \mathbf{X} \rightarrow \mathbb{R}$, $\lim_{n \rightarrow \infty} \mu_n[f] = \mu[f]$ in probability or with probability one (or in distribution, or in L_p);
- convergence in expectation: for every bounded continuous f , $\lim_{n \rightarrow \infty} \mathbb{E} \mu_n[f] = \mathbb{E} \mu[f]$

Vague convergence can be considered instead of weak⁶. For details, see

Olav Kallenberg. *Random Measures, Theory and Applications*. Probability Theory and Stochastic Modelling. Vol. 77. Springer, 2017,

in particular, Section 1.3 and Chapter 4.

⁴If $\mathbf{X}$ is not complete and separable, then a single probability measure $\mu \in \mathcal{P}(\mathbf{X})$ might not be tight.

⁵all properties of distance are obvious except $d_{\mathfrak{G}}(\mu, \nu) = 0$ might not imply $\mu = \nu$.

⁶In particular, this is why no topology is fixed on $\mathcal{P}(\mathbf{X})$ in the basic definition of a random measure.