Summary[1] of Convergence for Probability Measures.

Notations:

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

General Definitions.

- `Set-wise (strong) convergence`: $\lim_{n\to\infty} \mu_n \overset{s}{=} \mu$: $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for all $A \in \mathcal{X}$;
- `Total variation distance`: $\mathrm{d}_{\mathrm{TV}}(\mu, \nu) = 2 \sup_{A \in \mathcal{X}} |\mu(A) - \nu(A)|$;
- `Convergence in total variation`: $\lim_{n\to\infty} \mu_n \overset{\mathrm{TV}}{=} \mu$: $\lim_{n\to\infty} \mathrm{d}_{\mathrm{TV}}(\mu_n, \mu) = 0$.

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

Definitions.

- `weak convergence` $\lim_{n\to\infty} \mu_n \overset{w}{=} \mu$: $\lim_{n\to\infty} \mu_n[f] = \mu[f]$ for all bounded continuous f.
- the `Lévy-Prokhorov metric`:

$$\mathrm{d}_{LP}(\mu, \nu) = \inf\{\varepsilon > 0 : \mu(A) \le \nu(A^\varepsilon) + \varepsilon,\ A \in \mathcal{B}(\boldsymbol{X})\},$$

  where $A^\varepsilon = \{x \in \boldsymbol{X} : \inf_{y \in A} \rho(x, y) \le \varepsilon\}$.
- `tightness`: the collection $\{\mu_n,\ n \ge 1\}$ is tight if, for every $\varepsilon > 0$, there exists a compact $K_\varepsilon \subset \mathbf{X}$ so that, for all n, $\mu_n(K_\varepsilon) \ge 1 - \varepsilon$.
- `relative compactness`: the collection $\{\mu_n,\ n \ge 1\}$ is relatively compact if every subsequence contains a weakly converging sub-sub-sequence.

Results.

(1) Weak convergence defines topology on $\mathcal{P}(\boldsymbol{X})$.
(2) `Portmanteau theorem`: the following conditions are equivalent
  - $\lim_{n\to\infty} \mu_n \overset{w}{=} \mu$;
  - $\lim_{n\to\infty} \mu_n[f] = \mu[f]$ EITHER for all bounded, *uniformly* continuous f OR for all bounded *Lipschitz* continuous[2] f;
  - $\limsup_{n\to\infty} \mu_n(G) \le \mu(G)$ for all *closed* sets $G \subset \boldsymbol{X}$;
  - $\liminf_{n\to\infty} \mu_n(G) \ge \mu(G)$ for all *open* sets $G \subset \boldsymbol{X}$;
  - $\lim_{n\to\infty} \mu_n(G) = \mu(G)$ for all measurable G with boundary ∂G such that $\mu(\partial G) = 0$.

(3) `Continuous mapping theorem`: if $\boldsymbol{Y}$ is another metric space and $h : \boldsymbol{X} \to \boldsymbol{Y}$ is a continuous function, then weak convergence $\lim_{n\to\infty} \mu_n \overset{w}{=} \mu$ implies weak convergence $\lim_{n\to\infty} \mu_n \circ h^{-1} \overset{w}{=} \mu \circ h^{-1}$, where, for a measure $\nu \in \mathcal{P}(\boldsymbol{X})$, the measure $\nu \circ f^{-1} \in \mathcal{P}(\boldsymbol{Y})$ is defined by $\nu \circ h^{-1}(B) = \nu\big(x \in \boldsymbol{X} : h(x) \in B\big)$, $B \in \mathcal{B}(\boldsymbol{Y})$.
(4) `Prohorov's theorem`: tightness implies relative compactness.

The most special case: $\boldsymbol{X}$ is a complete separable metric[3] space.

Results.

(1) $\lim_{n\to\infty} \mu_n \overset{w}{=} \mu$ if and only if $\lim_{n\to\infty} \mathrm{d}_{LP}(\mu_n, \mu) = 0$.
(2) `Prohorov's theorem`: tightness is equivalent[4] to relative compactness.

---

[1]Sergey Lototsky, USC; version of June 11, 2023.
[2]$|f(x) - f(y)| \le L_f \rho(x, y), x, y \in \boldsymbol{X}$
[3]That is, Polish metric, which is not exactly the same as simply Polish...
[4]If $\boldsymbol{X}$ is not complete and separable, then a single probability measure $\mu \in \mathcal{P}(\boldsymbol{X})$ might not be tight.

Further developments.

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

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

In particular, for a metric space $\boldsymbol{X}$, 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 $\mathcal{P}(\boldsymbol{X})$ by

$$\mathrm{d}_{\mathfrak{G}}(\mu, \nu) = \sup_{f\in\mathfrak{G}} \big|\mu[f] - \nu[f]\big|,$$

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

(1) If $\mathfrak{G}$ is the set of all bounded measurable functions, then $\mathrm{d}_{\mathfrak{G}} = \mathrm{d}_{\mathrm{TV}}$.
(2) If $\boldsymbol{X} = \mathbb{R}$ and $\mathfrak{G}$ is the collection of indicator functions of the sets $(-\infty, x]$ for all $x \in \mathbb{R}$, then $\mathrm{d}_{\mathfrak{G}}$ is the `Kolmogorov` metric.
(3) If $\boldsymbol{X}$ is a metric space and $\mathfrak{G}$ is the collection of continuous functions satisfying $|f| \leq 1$, then the $\mathrm{d}_{\mathfrak{G}}$ 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 $(\boldsymbol{X}, \mathcal{X})$ is a *kernel* from the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\boldsymbol{X}, \mathcal{X})$. 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 $(\boldsymbol{X}, \mathcal{X})$ 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 $\xi_k,\ k = 1, \ldots, N$ are $\boldsymbol{X}$-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 $f : \boldsymbol{X} \to \mathbb{R}$, $\lim_{n\to\infty} \mu_n[f] = \mu[f]$ in probability or with probability one (or in distribution, or in L_p);
- convergence in expectation: for every bounded continuous f, $\lim_{n\to\infty} \mathbb{E}\mu_n[f] = \mathbb{E}\mu[f]$

Vague convergence can be considered instead of weak[6]. 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.

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

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