Weak convergence of probability measures

These additional notes contain a short overview of the most important results on weak convergence of probability measures. Many more details and results as well as proofs can be found in the (German) lecture notes “Wahrscheinlichkeitstheorie”.

1. Weak convergence of probability measures on metric spaces

In the sequel, (S, d) is a metric space with Borel σ-field $\mathcal{S} = \mathcal{B}(S)$. Let μ and μ_n, $n \in \mathbb{N}$, be probability measures on $(S, \mathcal{S})$. How can one define convergence of $(\mu_n)_{n\in\mathbb{N}}$ to μ?

Possible ideas could be

a) $\lim\limits_{n\to\infty} \mu_n(A) = \mu(A) \quad$ for all $A \in \mathcal{S}$.

b) $\|\mu_n - \mu\| := \sup\limits_{A\in\mathcal{S}} |\mu_n(A) - \mu(A)| \longrightarrow 0$ for $n \to \infty$ (this is the so-called *convergence in variation*).

In general, both notions are too restrictive for our purposes:

- For $\mu_n := \delta_{\frac{1}{n}}$ and $\mu := \delta_0$, neither a) nor b) is satisfied, because $A := \{0\}$ has $\mu_n(A) \equiv 0$, but $\mu(A) = 1$.
- If μ_n is a standardised binomial distribution with parameters n, p and μ is the standard normal distribution $\mathcal{N}(0, 1)$, there is a countable set $A \in \mathcal{B}(\mathbb{R})$ with $\mu_n(A) \equiv 1$, but of course $\mu(A) = 0$.

(1.1) **Definition.** We say that $(\mu_n)_{n\in\mathbb{N}}$ *converges weakly* to μ and write $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$ if

$$\lim_{n\to\infty} \int h \, d\mu_n = \int h \, d\mu \qquad \text{for all } h \in C_b(S),$$

where $C_b(S)$ denotes the space of all bounded continuous functions $h : S \to \mathbb{R}$.

(1.2) **Theorem (Portmanteau Theorem):** *The following statements are equivalent:*

1) $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$.

2) $\lim\limits_{n\to\infty} \int h \, d\mu_n = \int h \, d\mu \quad$ *for all uniformly continuous* $h \in C_b(S)$.

3) $\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$ *for all closed* $F \subseteq S$.

4) $\liminf_{n\to\infty} \mu_n(G) \ge \mu(G)$ *for all open* $G \subseteq S$.

5) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for all* μ*-boundaryless* $A \in \mathcal{S}$, *i.e.* $A \in \mathcal{S}$ *with* $\mu(\bar{A} \setminus A^\circ) = 0$, *where* $\bar{A}$ *is the closure and* A° *the interior of* A.

If one thinks of μ_n, μ as the distributions of S-valued random variables X_n, X, one often uses instead of weak convergence of μ_n to μ the terminology that the X_n converge to X in distribution. More precisely, suppose that we have for any $n \in \mathbb{N}$ a probability space $(\Omega_n, \mathcal{F}_n, P_n)$ and a measurable mapping $X_n : \Omega_n \to S$, i.e. an S-valued random variable, and also a probability space $(\Omega, \mathcal{F}, P)$ and a measurable mapping $X : \Omega \to S$. (This is always possible if we only specify the distributions μ_n, μ, but not the mappings: We can take $\Omega_n := \Omega := S$, $\mathcal{F}_n := \mathcal{F} := \mathcal{S}$, $P_n := \mu_n$, $P := \mu$ and $X_n := X := \mathrm{Id} : S \to S$.) The key point here is that all the X_n and X have the *same range* S. The distributions $\mu_n := P_n \circ X_n^{-1}$ of X_n under P_n and $\mu := P \circ X^{-1}$, the distribution of X under P, are then probability measures on $(S, \mathcal{S})$. We then say that $(X_n)_{n\in\mathbb{N}}$ *converges in distribution* to X and write $X_n \underset{n\to\infty}{\Longrightarrow} X$ if $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$. Other notations sometimes used are $X_n \xrightarrow[n\to\infty]{d} X$, $X_n \xrightarrow[n\to\infty]{\mathcal{L}} X$, $\mathcal{L}(X_n) \underset{n\to\infty}{\Longrightarrow} \mathcal{L}(X)$, and very explicitly $\mathcal{L}(X_n|P_n) \underset{n\to\infty}{\Longrightarrow} \mathcal{L}(X|P)$.

2. Tightness and Prohorov's theorem

In this section, (S, d) is a metric space with Borel σ-field $\mathcal{S} = \mathcal{B}(S)$, and $\mathcal{M}_1(S)$ is the set of all probability measures on $(S, \mathcal{S})$. Convergence of a sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}_1(S)$ to μ then means that $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$, i.e. $\lim\limits_{n\to\infty} \int h\, d\mu_n = \int h\, d\mu$ for all $h \in C_b(S)$.

(2.1) **Remark.** The *topology* on $\mathcal{M}_1(S)$ associated to weak convergence has as a neighbourhood basis the sets of the form

$$U_{\varepsilon,h_1,\dots,h_n}(\mu) := \left\{\nu \in \mathcal{M}_1(S) \,\middle|\, \left|\int h_i\, d\nu - \int h_i\, d\mu\right| < \varepsilon, \quad i = 1, \dots, n\right\}$$

with $\varepsilon > 0$, $n \in \mathbb{N}$, $h_i \in C_b(S)$.

(Recall that a neighbourhood basis $\mathcal{U}$ of a point x is a system of neighbourhoods of x such that each neighbourhood of x contains some $U \in \mathcal{U}$.) $\diamond$

(2.2) **Definition.** A set $\mathcal{M} \subseteq \mathcal{M}_1(S)$ is called *relatively sequentially compact* if each sequence in $\mathcal{M}$ contains a weakly convergent subsequence.

The goal of this section is a *characterisation* of relatively sequentially compact subsets of $\mathcal{M}_1(S)$.

(2.3) **Remark.** On $\mathcal{M}_1(S)$, we have the above topology corresponding to weak convergence and hence also a notion of compactness: $\mathcal{M} \subseteq \mathcal{M}_1(S)$ is called *compact* if every covering of $\mathcal{M}$ by open sets contains a finite covering of $\mathcal{M}$. In general, one has for topological spaces neither "compact $\Longrightarrow$ sequentially compact" (!) nor "sequentially compact $\Longrightarrow$ compact". However, the two notions are equivalent for metric spaces. So it is of interest whether the topology of weak convergence is metrisable, i.e. if there exists a metric ϱ on $\mathcal{M}_1(S)$ such that $\varrho(\mu_n, \mu) \longrightarrow 0$ for $n \to \infty$ if and only if $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$. This is possible if S is separable. $\diamond$

(2.4) **Example.** Take $S = \mathbb{R}$ and $\mu_n = \delta_{x_n}$ for a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$. If we have $\lim\limits_{n\to\infty} x_n = +\infty$, one cannot hope in general to find a weakly convergent subsequence of (μ_n); for $x_n = n$, we have for example $\int h\, d\mu_n = h(n)$, and this need not converge *simultaneously for all* $h \in C_b(\mathbb{R})$ to some limit; one can for instance look at $h(x) = \sin x$. But if the sequence (x_n) is bounded, it has a convergent subsequence $x_{n_k} \longrightarrow x$ for $k \to \infty$, and then we obviously have $\mu_{n_k} \underset{k\to\infty}{\Longrightarrow} \delta_x$.

The above example shows that if we want to obtain relative sequential compactness, we need to impose some kind of *boundedness condition*; the mass of μ_n should not be allowed to wander away as $n \to \infty$.

(2.5) **Definition.** $\mathcal{M} \subseteq \mathcal{M}_1(S)$ is called *tight* if for every $\varepsilon > 0$, there exists a compact set $K \subseteq S$ with $\mu(K) \geq 1 - \varepsilon$ for all $\mu \in \mathcal{M}$.

The main result of this section is now

(2.6) **Theorem (Prohorov):** *Consider* $\mathcal{M} \subseteq \mathcal{M}_1(S)$.

1) *If* $\mathcal{M}$ *is tight, then* $\mathcal{M}$ *is relatively sequentially compact.*

2) *Suppose that* S *is complete and separable. If* $\mathcal{M}$ *is relatively sequentially compact, then* $\mathcal{M}$ *is also tight.*

3. Weak convergence on $S = C[0,1]$

In this section, we take as $S = C[0,1]$ the space of continuous functions $x : [0,1] \to \mathbb{R}$ with the sup-norm $\|x\| := \sup_{0 \le t \le 1} |x(t)|$ and the corresponding metric $d(x,y) := \|x - y\|$. Then S is a Banach space and separable, because $[0,1]$ is compact; see Dieudonné (1969), "Foundations of Modern Analysis", 7.4.4. The Borel σ-field $\mathcal{S} = \mathcal{B}(S) = \sigma(C_b(S))$ is also generated by the system $\mathcal{Z}$ of all *cylinder sets*

$$Z = \{x \in S \mid x(t_j) \in A_j, \ j = 1, \ldots, n\}$$

with $n \in \mathbb{N}$, $0 \le t_1 < t_2 < \cdots < t_n \le 1$ and $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$, i.e.

$$\mathcal{B}(S) = \sigma(\mathcal{Z}). \tag{3.1}$$

Moreover, $\mathcal{Z}$ is clearly closed under taking intersections..

A probability measure μ on $C[0,1]$, or more precisely on $(S, \mathcal{B}(S))$, corresponds to a *real-valued stochastic process* $X = (X_t)_{0 \le t \le 1}$ with *continuous trajectories*: If we have such an X defined on $(\Omega, \mathcal{F}, P)$, we obtain μ on $C[0,1]$ as the image of P under X, i.e. as the distribution of X under P; conversely, if we have μ on $C[0,1]$, we can take $\Omega := S = C[0,1]$, $\mathcal{F} := \mathcal{S} = \mathcal{B}(C[0,1])$, $P := \mu$, and the coordinate process X with $X_t(\omega) := \omega(t)$, $0 \le t \le 1$. Moreover, μ is uniquely determined by its *finite-dimensional marginal distributions*

$$\mu^{(J)} := \mu \circ \pi_J^{-1}, \qquad J \subseteq [0,1] \text{ finite, on } \mathbb{R}^J,$$

where $\pi_J : S \to \mathbb{R}^{|J|}, x \mapsto \pi_J(x) := \left(x(t_j)\right)_{t_j \in J}$ are the canonical projections.

Because all the $\pi_J : S \to \mathbb{R}^{|J|}$ are continuous, $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$ on $C[0,1]$ implies the weak convergence of all finite-dimensional marginal distributions, i.e. $\mu_n^{(J)} \underset{n\to\infty}{\Longrightarrow} \mu^{(J)}$ on $\mathbb{R}^{|J|}$ for all finite $J \subseteq [0,1]$. Indeed, if $g \in C_b(\mathbb{R}^{|J|})$, we have $g \circ \pi_J \in C_b(S)$, and the transformation theorem yields

$$\int_{\mathbb{R}^{|J|}} g \, d\mu_n^{(J)} = \int_{\mathbb{R}^{|J|}} g \, d(\mu_n \circ \pi_J^{-1}) = \int_S (g \circ \pi_J) \, d\mu_n$$
$$\longrightarrow \int_S (g \circ \pi_J) \, d\mu = \int_{\mathbb{R}^{|J|}} g \, d(\mu \circ \pi_J^{-1}) = \int_{\mathbb{R}^{|J|}} g \, d\mu^{(J)}.$$

However, the converse is not true, as shown by the following simple counterexample.

(3.2) **Example.** Take $x \equiv 0$, $\mu = \delta_x$ and $\mu_n = \delta_{x_n}$ with the x_n piecewise linear, $x_n(0) = 0$, $x_n(\frac{1}{n}) = 1$ and $x_n(t) = 0$ for $t \geq \frac{2}{n}$ [$\longrightarrow$ picture!]. Then the x_n clearly converge to x pointwise, but not uniformly.

For $J = \{t_1, \ldots, t_m\} \subseteq [0, 1]$, we have

$$\mu_n \circ \pi_J^{-1} \underset{n\to\infty}{\Longrightarrow} \mu \circ \pi_J^{-1}$$

due to pointwise convergence, because

$$\int_{I\!R^{|J|}} g \, d(\mu_n \circ \pi_J^{-1}) = \int_S (g \circ \pi_J) \, d\mu_n = \Big(g(x_n(t_j))\Big)_{j=1,\ldots,m}.$$

But $(\mu_n)_{n\in I\!N}$ does not converge weakly to μ, because the x_n do not converge to x uniformly; for example, the function $h(x) := \min(\|x\|, 1)$ is in $C_b(C[0, 1])$, and $\int h \, d\mu_n = h(x_n) \equiv 1$, but $\int h \, d\mu = h(x) = h(0) = 0$.

If we now think in general of μ_n and μ as the distributions of continuous stochastic processes X^n and X, respectively, the next theorem is the key result on *convergence in distribution of continuous stochastic processes*.

(3.3) **Theorem.** *For probability measures $(\mu_n)_{n\in I\!N}, \mu$ on $(C[0, 1], \mathcal{B}(C[0, 1]))$, the following are equivalent:*

1) $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$.

2) *All finite-dimensional marginal distributions of the μ_n converge weakly to the corresponding finite-dimensional marginal distributions of μ, and the sequence $(\mu_n)_{n\in I\!N}$ is tight.*

To analyse the tightness of a given set $\mathcal{M} \subseteq \mathcal{M}_1(C[0, 1])$, we need a description of the (relatively) compact subsets of $C[0, 1]$. For that purpose, define for $x \in C[0, 1]$ and $\delta > 0$ the *modulus of continuity* of x as

$$w_\delta(x) := \sup\Big\{|x(t) - x(s)| \,\Big|\, s, t \in [0, 1] \text{ with } |t - s| \leq \delta\Big\}.$$

Then we have

$$\lim_{\delta\searrow 0} w_\delta(x) = 0 \qquad \text{for every } x \in C[0, 1],$$

because each x is uniformly continuous on $[0, 1]$. Moreover, $x \mapsto w_\delta(x)$ is continuous (as a mapping from $C[0, 1]$ to $I\!R$ for fixed $\delta > 0$), because $|w_\delta(x) - w_\delta(y)| \leq 2\|x - y\|$ (as can be checked easily), and $\delta \mapsto w_\delta(x)$ is clearly increasing for every fixed x.

(3.4) **Proposition (Arzelà–Ascoli):** *A set* $A \subseteq C[0,1]$ *is relatively compact in* $C[0,1]$ *if and only if*

1) A is uniformly bounded, i.e. $\sup_{x \in A} \|x\| < \infty$,

and

2) A is uniformly (over its elements x) uniformly continuous, meaning that

$$\lim_{\delta \searrow 0} \sup_{x \in A} w_\delta(x) = 0.$$

(3.5) **Remark.** If one already has 2), one can also replace 1) by

1′) A is uniformly bounded at 0, i.e. $\sup_{x \in A} |x(0)| < \infty$. ◇

(3.6) **Proposition.** *Fix* $\mathcal{M} \subseteq \mathcal{M}_1(C[0,1])$ *and denote by* $\mu^0 := \mu(\{x(0) \in \cdot\,\}) = \mu \circ \pi_{\{0\}}^{-1}$ *for* $\mu \in \mathcal{M}$ *the distribution of "coordinate 0". Then* $\mathcal{M}$ *is tight if and only if we have both that* $\{\mu^0 \mid \mu \in \mathcal{M}\}$ *is tight on* $\mathbb{R}$ *and that*

$$\lim_{\delta \searrow 0} \sup_{\mu \in \mathcal{M}} \mu(\{x \,|\, w_\delta(x) \geq \eta\}) = 0 \qquad \text{for all } \eta > 0, \tag{3.7}$$

i.e. if $w_\delta(\,\cdot\,)$ *goes* μ*-stochastically to 0, uniformly over* $\mathcal{M}$*, as* $\delta \to 0$.

(3.8) **Remark.** If $\mathcal{M}$ is a sequence $(\mu_n)_{n \in \mathbb{N}}$, then instead of (3.7) in Proposition (3.6), it is already sufficient if we have

$$\lim_{\delta \searrow 0} \limsup_{n \to \infty} \mu_n(\{x \,|\, w_\delta(x) \geq \eta\}) = 0 \qquad \text{for all } \eta > 0. \tag{3.9}$$

◇

Remark. In Proposition (3.6), the family $\{\mu^0 \mid \mu \in \mathcal{M}\}$ is for example trivially tight if

$$\mu(\{x \,|\, x(0) = 0\}) = 1 \qquad \text{for all } \mu \in \mathcal{M}.$$

◇