Weak convergence of Empirical Distribution

Space of right-continuous functions

Definition 1. Let $D[0,1]$ be a set of right-continuous functions:

$$D[0,1] = \{f : [0,1] \to \mathbb{R} :\ \forall t \in [0,1)\ f(t+) = f(t),\ \forall t \in (0,1]\ \exists\ f(t-)\}.$$

Space $D[0,1]$ equipped with the uniform metric is not separable since $\rho(\delta_x, \delta_y) = 1$ for every x, where $\delta_x(u) = I_{u \geq x}$. Thus we need another metric on $D[0,1]$ instead of uniform metric.

Definition 2. Let $f, g \in D[0,1]$ and Λ be the set of strictly increasing, continuous mappings λ of $[0,1]$ onto itself, $\lambda(0) = 0$, $\lambda(1) = 1$. Then

$$\rho(f,g) = \inf\{\varepsilon :\ \exists \lambda \in \Lambda :\ \forall x \in [0,1]\ |f(\lambda(x)) - g(x)| \leq \varepsilon,\ |\lambda(x) - x| \leq \varepsilon\}.$$

The topology defined by metric ρ is called the *Skorohod topology.*

Firstly, let's prove that ρ is a metric. Really,

1. Obviously, $\rho \geq 0$ and $\rho(f,g) = 0$ iff $f = g$.
2. Since $\lambda \in \Lambda$ there exists $\lambda^{-1} \in \Lambda$ and $\sup|\lambda(x) - x| = \sup|\lambda^{-1}(x)|$. Therefore, $\rho(g,f) = \rho(f,g)$.
3. For every $\lambda_1, \lambda_2 \in \Lambda$ we have

   $$\sup_x |\lambda_2(\lambda_1(x)) - x| \leq \sup_x |\lambda_2(\lambda_1(x)) - \lambda_1(x)| + \sup_x |\lambda_1(x) - x| = \sup_x |\lambda_2(x) - x| + \sup_x |\lambda_1(x) - x|.$$

   Thus for every $f, g, h, \lambda_1, \lambda_2$ such that $\sup_x |f(\lambda_2(x)) - g(x)| \leq \varepsilon_1,\ |\lambda_2(x) - x| \leq \varepsilon_1$, $\sup_x |g(\lambda_1(x)) - h(x)| \leq \varepsilon_1,\ |\lambda_1(x) - x| \leq \varepsilon_1$ we have $\sup_x |\lambda_2(\lambda_1(x)) - x| \leq \varepsilon_1 + \varepsilon_2$ and

   $$\sup_x |f(\lambda_2(\lambda_1(x))) - h(x)| \leq \sup_x |f(\lambda_2(\lambda_1(x))) - g(\lambda_1(x))| + \sup_x |g(\lambda_1(x)) - h(x)| \leq \varepsilon_1 + \varepsilon_2.$$

   Therefore, $\rho(f,h) \leq \rho(f,g) + \rho(g,h)$.

Example 1. Let $f = \delta_u$, $g = \delta_v$, $u < v \in [0,1]$. Let's prove that $\rho(\delta_u, \delta_v) \leq v - u$. For $u > 0$ consider

$$\lambda(x) = \begin{cases} v\frac{x}{u}, & x < u, \\ 1 - (1-v)\frac{(1-x)}{(1-u)}, & x \geq u. \end{cases}$$

Then $g(\lambda(x)) = f(x)$, $\sup|\lambda(x) - x| = u - v$. Therefore, $\rho(\delta_u, \delta_v) \leq v - u$. On the other hand, if $|\lambda(x) - x| < u - v$, then

$$g(\lambda(u)) \leq g(v - 0) = 0 = f(u) - 1 \leq f(u) - (u - v).$$

Thus $\rho(\delta_u, \delta_v) = v - u$.

Space of right-continuous functions

Space $(D[0,1], \rho)$ is a separable space but it's not complete.

Example 2. Let $f_n(x) = I_{[1/2, 1/2+1/n)}$. Then $\rho(f_n, f_m) < \left|\frac{1}{n} - \frac{1}{m}\right|$ and $\{f_n\}$ is fundamental in $(D[0,1], \rho)$. But it isn't convergent.

Thus another metric is considered:

1. Let

   $$||\lambda|| = \sup\left|\ln\frac{\lambda(x) - \lambda(y)}{x - y}\right|,\ \lambda \in \Lambda_0,$$

where $\Lambda_0 = \{\lambda \in \Lambda : ||\lambda|| < \infty\}$.

2. Let
$$\rho_0(f,g) = \inf\{\varepsilon : \ \exists \lambda \in \Lambda_0 : \ \forall x \in [0,1] \ |f(\lambda(x)) - g(x)| \le \varepsilon, \ ||\lambda|| \le \varepsilon\}.$$
Since $||\lambda|| = ||\lambda^{-1}||$ and $||\lambda_2(\lambda_1)|| \le ||\lambda_2|| + ||\lambda_1||$ it's a metric on $D[0,1]$. Metrics ρ_0 and ρ are topological equivalent, but the space $(D[0,1], \rho_0)$ is complete (we don't prove this fact).

Weak convergence in $D[0,1]$

Let's remind the definition and main properties of weak convergence.

Definition 3. We say that a sequence $\{X_n\}$ of stochastic processes in D[0,1] weakly converges to the process X, and we write $X_n \xrightarrow{d} X$, if
$$\mathbf{E} f(X_n) \to \mathbf{E} f(X)$$
for every continuous (with respect to ρ_0) and bounded mapping $f : D[0,1] \to \mathbb{R}$.

Theorem 1. *The following conditions are eqiuvalent:*

1. $X_n \xrightarrow{d} X$.
2. $\mathbf{P}(X_n \in A) \xrightarrow{d} \mathbf{P}(X \in A)$ *for every* $A : \mathbf{P}(X \in \partial A) = 0$*, where* ∂A *is*

Theorem 2. *If* $X_n \xrightarrow{d} X$ *and* f *is continuous (with respect to* ρ_0*) mapping* $f : D[0,1] \to \mathbb{R}$*, then* $f(X_n) \xrightarrow{d} f(X)$*. Moreover, the condition of continuity of* f *can be weakened to* $\mathbf{P}(X \in C_f) = 0$*, where* C_f *is the set of discontinuites of* f.

Definition 4. We say that X_n converges to X *in the sense of finite-dimensional distributions* and we write $X_n \xrightarrow{f.d.} X$, if $(X_n(t_1), ..., X_n(t_k)) \xrightarrow{d} (X(t_1), ..., X(t_k))$ for every $t_1, ..., t_k \in [0,1]$.

The weak convergence is stronger than the convergence in the sense of finite-dimensional distributions:

Example 3. Consider the sample space $(\Omega, \mathcal{F}, P)$, where $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $P = \lambda$ — Lebesque measure. Let
$$X_n = \begin{cases} n\omega, & \omega \in \left[0, \frac{1}{n}\right], \\ 2n - n\omega, & \omega \in \left[\frac{1}{n}, \frac{2}{n}\right], \\ 0, & \omega \ge \frac{2}{n}. \end{cases}$$
Then $(X_n(t_1), ..., X_n(t_k) \stackrel{d}{=} (0, ..., 0)$ for any $t_1, ..., t_k \in [0,1]$ and large enough n. Therefore, $X_n \xrightarrow{f.d.} X$, where $X(t) = 0$ for every $t \in [0,1]$. However,
$$1 = \mathbf{E} X_n \not\to \mathbf{E} X = 0,$$
so $X_n \stackrel{d}{\not\to} X$.

Definition 5. The set $\mathcal{P}$ of probability measures is called to be *weakly sequentially compact* if for every sequence $P_n \in \mathcal{P}$ there exists a weakly convergent subsequence P_{n_k}.

Lemm 1. *Let* $X_n \xrightarrow{f.d.} X$*. Then* $X_n \xrightarrow{d} X$ *iff* $\{\mathbf{P}_{X_n}\}$ *is a weakly sequentially compact set.*

Proof. Obviously, a weakly convergent sequence X_n is a weakly sequntially compact set. Let us prove that if $\{\mathbf{P}_{X_n}\}$ is a weakly sequentially compact set, $X_n \xrightarrow{f.d.} X$, then $X_n \xrightarrow{d} X$. Assume the converse. Then $X_n \stackrel{d}{\not\to} X$. Thus there exists $\varepsilon > 0$, $A \in \mathcal{B}(\mathbb{R}^{[0,1]}) : \mathbf{P}(X \in \partial A) = 0$ and n_k such that
$$|\mathbf{P}(X_{n_k} \in A) - \mathbf{P}(X \in A)| > \varepsilon.$$
Let $\mathbf{P}_{X_{n_{k_l}}}$ be a convergent subseqence of $\mathbf{P}_{X_{n_k}}$, so $X_{n_{k_l}} \xrightarrow{d} Y$ for some Y. Therefore, $X_{n_{k_l}} \xrightarrow{f.d.} Y$ and $X \stackrel{d}{=} Y$. Thus
$$\mathbf{P}(X_{n_{k_l}} \in A) \to \mathbf{P}(X \in A), \text{ but } \left|\mathbf{P}(X_{n_{k_l}} \in A) - \mathbf{P}(X \in A)\right| > \varepsilon.$$
The contradiction proves the lemma. ∎

Definition 6. We say that the family of probability measures $\mathcal{P}$ on $D[0,1]$ is tight if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset D[0,1]$ such that $\mathbf{P}(K_\varepsilon) \geq 1 - \varepsilon$ for every $\mathbf{P} \in \mathcal{P}$.

Theorem 3. *(Yu. V. Prohorov). Let $\mathcal{S}$ be a complete separable space. Then a set $\mathcal{P}$ of measures on $\mathcal{S}$ is weakly sequentially compact iff $\mathcal{P}$ is tight.*

Therefore, $X_n \stackrel{d}{\to} X$ iff $X_n \stackrel{f.d.}{\to} X$ and $\{P_{X_n}\}$ is tight.

The following theorem is proved in Billingsley, "Convergence of probability measures" (Theorem 15.6).

Theorem 4. *Let $X_n() \in D[0,1]$ be a sequence of stohastic processes. Suppose that there exists $\gamma > 0$, $\alpha > 1$ and nondecreasing continuous function G on $[0,1]$ such that*

$$\mathbf{P}(|X_n(t) - X_n(s)| \geq \varepsilon, |X_n(r) - X_n(t)| \geq \varepsilon) \leq \frac{1}{\varepsilon^{2\gamma}} \left(G(r) - G(s)\right)^\alpha,$$

for every $s < t < r$; then $\mathbf{P}(X_n() \in \cdot)$ is tight.