9. Weak Convergence

DASN, NTU https://personal.ntu.edu.sg/wptay/

9.1 Weak Convergence

Let (Ω, d) be a metric space with probability measure $\mathbb{P} : \mathcal{B} \mapsto [0,1]$, where $\mathcal{B}$ is the Borel σ-algebra. Since $\mathcal{B}$ defines $\mathbb{P}$, one may want to define the convergence of a sequence of probability distributions $\mathbb{P}_n \to \mathbb{P}$ as $\mathbb{P}_n(B) \to \mathbb{P}(B)$ for all $B \in \mathcal{B}$. However, this definition is too strong as can be seen from this example: Let δ_x be the probability measure with $\delta_x(\{x\}) = 1$. Then we desire that $\delta_{1/n} \to \delta_0$ but $\delta_{1/n}((0,1)) = 1$ and $\delta_0((0,1)) = 0$.

Let $C_b(\Omega)$ be the set of continuous and bounded functions $f : \Omega \mapsto \mathbb{R}$.

Lemma 9.1. *$\mathbb{P}$ on $(\Omega, \mathcal{B})$ is uniquely determined by $\int_\Omega f \,\mathrm{d}\mathbb{P},\ f \in C_b(\Omega)$.*

Proof. We show that $\int_\Omega f \,\mathrm{d}\mathbb{P}_1 = \int_\Omega f \,\mathrm{d}\mathbb{P}_2,\ \forall f \in C_b(\Omega) \implies \mathbb{P}_1 = \mathbb{P}_2$. For any open set $U \subset \Omega$, we have

$$d(\omega, U^c) = \begin{cases} 0, & \text{if } \omega \in U^c, \\ > 0, & \text{if } \omega \in U. \end{cases}$$

Let $f_m(\omega) = \min\left\{1, md(\omega, U^c)\right\} \in C_b(\Omega)$. We have

$$\lim_{m\to\infty} f_m(\omega) = \mathbf{1}_U(\omega).$$

From MCT, we obtain

$$\lim_{m\to\infty} \int_\Omega f_m \,\mathrm{d}\mathbb{P}_1 = \int_\Omega \lim_{m\to\infty} f_m \,\mathrm{d}\mathbb{P}_1 = \int_\Omega \mathbf{1}_U(\omega) \,\mathrm{d}\mathbb{P}_1(\omega) = \mathbb{P}_1(U).$$

Since

$$\int_\Omega f_m \,\mathrm{d}\mathbb{P}_1 = \int_\Omega f_m \,\mathrm{d}\mathbb{P}_2,$$

$\mathbb{P}_1(U) = \mathbb{P}_2(U)$ for all open U. Since the open sets generate $\mathcal{B}$ and $\{B \in \mathcal{B} : \mathbb{P}_1(B) = \mathbb{P}_2(B)\}$ is a λ-system, the $\pi - \lambda$ theorem completes the proof. ☐

Definition 9.1 (Weak convergence or Convergence in distribution)**.** *We say that $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$ if*

$$\int_\Omega f \,\mathrm{d}\mathbb{P}_n \to \int_\Omega f \,\mathrm{d}\mathbb{P},\ \forall f \in C_b(\Omega).$$

Similarly, we say that $X_n \xrightarrow{\mathrm{d}} X$ *if* $\mathbb{P}_{X_n} \xrightarrow{\mathrm{d}} \mathbb{P}_X$, *i.e.,*

$$\mathbb{E}f(X_n) = \int_\Omega f \, \mathrm{d}\mathbb{P}_{X_n} \to \int_\Omega f \, \mathrm{d}\mathbb{P}_X = \mathbb{E}f(X), \ \forall f \in C_b(\Omega).$$

In the sequel, we restrict $\Omega = \mathbb{R}$. All results can be extended to $\mathbb{R}^k$ with trivial modifications. We denote $F(t) = \mathbb{P}((\infty, t])$ as the cdf of $\mathbb{P}$ and $F_n(t)$ as the cdf of $\mathbb{P}_n$.

Theorem 9.1. $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P} \iff F_n(t) \to F(t)$ *for all continuity points* t *of* $F(\cdot)$.

Proof. Let Φ be the set of continuity points of $F(t)$.

'$\Rightarrow$': Fix $\epsilon > 0$ and $t \in \Phi$. Let $\varphi_1(x)$ and $\varphi_2(x)$ be the continuous bounded functions shown in Fig. 1, i.e.,

$$\varphi_i(x) = \begin{cases} 1, & \text{if } x \le t - \epsilon, \\ \text{linear}, & \text{if } t - \epsilon < x \le t, \\ 0, & \text{if } x > t. \end{cases}$$

The function φ_2 is defined similarly.

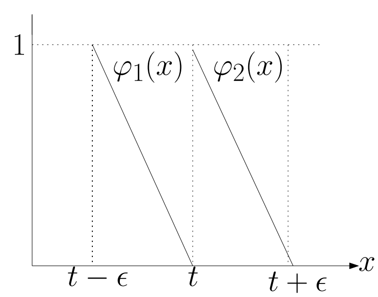

Fig. 1: Approximation of $\mathbf{1}_{(-\infty,t]}(x)$.

Then we can approximate $\mathbf{1}_{(-\infty,t]}(x)$ as

$$\mathbf{1}_{(-\infty,t-\epsilon]}(x) \le \varphi_1(x) \le \mathbf{1}_{(-\infty,t]}(x) \le \varphi_2(x) \le \mathbf{1}_{(-\infty,t+\epsilon]}(x).$$

By taking the expectation of the above inequality, we obtain

$$\begin{aligned} F(t-\epsilon) &\le \int \varphi_1 \, \mathrm{d}\mathbb{P} \\ &= \lim_{n\to\infty} \int \varphi_1 \, \mathrm{d}\mathbb{P}_n \\ &\le \liminf_{n\to\infty} F_n(t) \\ &\le \limsup_{n\to\infty} F_n(t) \\ &\le \lim_{n\to\infty} \int \varphi_2 \, \mathrm{d}\mathbb{P}_n \end{aligned}$$

$$
\begin{aligned}
&= \int \varphi_2 \, \mathrm{d}\mathbb{P} \\
&\leq F(t+\epsilon).
\end{aligned}
$$

By taking $\epsilon \to 0$, since $t \in \Phi$, we have

$$
\lim_{n\to\infty} F_n(t) = F(t).
$$

Remark 9.1. *Similar proof steps as above hold if $C_b(\mathbb{R})$ is replaced with $C_b^k(\mathbb{R})$, the set of k-differentiable bounded functions for $k \geq 1$.*

'$\Leftarrow$': We first show that $\mathbb{R}\backslash\Phi$ (the set of discontinuity points) is countable. Suppose $x \in \mathbb{R}\backslash\Phi$. Let

$$
\begin{aligned}
a(x) &= \sup\{F(y) : y < x\}, \\
b(x) &= \inf\{F(y) : y > x\}.
\end{aligned}
$$

Because the set of rational numbers $\mathbb{Q}$ is dense in $\mathbb{R}$, $\exists\, r_x \in (a(x), b(x)),\ \text{s.t. } r_x \in \mathbb{Q}$. Since the intervals $(a(x), b(x))$ are disjoint for all $x \in \mathbb{R} \setminus \Phi$, the mapping $x \mapsto r_x$ is one-to-one. Therefore, $\mathbb{R} \setminus \Phi$ is countable and the claim is proved. This implies that Φ is dense.

Note that $\mathbb{P}((-a, a]^c) = F(-a) + 1 - F(a)$. We then have $\forall\, \epsilon > 0$,

$$
\exists \pm M(\epsilon) \in \Phi, \ \text{s.t.} \ \ \mathbb{P}((-M(\epsilon), M(\epsilon)]^c) \leq \epsilon.
$$

Furthermore, we have

$$
\begin{aligned}
F_n(M(\epsilon)) &\to F(M(\epsilon)), \\
F_n(-M(\epsilon)) &\to F(-M(\epsilon)).
\end{aligned}
$$

Therefore, $\forall\, n$ sufficiently large, $\mathbb{P}_n\left((-M(\epsilon), M(\epsilon)]^c\right) \leq 2\epsilon$. Choose $-M(\epsilon) = x_1^k \leq x_2^k \leq \ldots \leq x_k^k = M(\epsilon),\ x_i^k \in \Phi$ such that $\lim_{k\to\infty} \max |x_{i+1}^k - x_i^k| = 0$. For $f \in C_b(\mathbb{R})$, let

$$
f_k(x) = \sum_{1<i\leq k} f(x_i^k) \mathbf{1}_{\left(x_{i-1}^k, x_i^k\right]}(x) \in C_b(\mathbb{R}).
$$

Taking the expectation over $\mathbb{P}_n$, we obtain

$$
\begin{aligned}
\int f_k \, \mathrm{d}\mathbb{P}_n &= \sum_{1<i\leq k} f(x_i^k) \left(F_n(x_i^k) - F_n(x_{i-1}^k)\right) \\
&\xrightarrow{n\to\infty} \sum_{1<i\leq k} f(x_i^k) \left(F(x_i^k) - F(x_{i-1}^k)\right) \\
&= \int f_k \, \mathrm{d}\mathbb{P}. \qquad (1)
\end{aligned}
$$

Let

$$
\eta_k(M(\epsilon)) = \sup_{|x|\leq M(\epsilon)} |f_k(x) - f(x)|\,.
$$

Since $f \in C_b(\mathbb{R})$ is continuous, it is uniformly continuous on $[-M(\epsilon), M(\epsilon)]$ and we have $\lim_{k\to\infty} \eta_k(M(\epsilon)) = 0$. We have

$$\begin{aligned}\left|\int f \,\mathrm{d}\mathbb{P} - \int f_k \,\mathrm{d}\mathbb{P}\right| &\le 2\|f\|_\infty \mathbb{P}\left((-M(\epsilon), M(\epsilon)]^c\right) + \eta_k(M(\epsilon)) \\ &\le 2\|f\|_\infty \epsilon + \eta_k(M(\epsilon)), \end{aligned} \tag{2}$$

and for n large,

$$\begin{aligned}\left|\int f \,\mathrm{d}\mathbb{P}_n - \int f_k \,\mathrm{d}\mathbb{P}_n\right| &\le 2\|f\|_\infty \mathbb{P}_n\left((-M(\epsilon), M(\epsilon)]^c\right) + \eta_k(M(\epsilon)) \\ &\le 2\|f\|_\infty \epsilon + \eta_k(M(\epsilon)). \end{aligned} \tag{3}$$

From (1) to (3), we therefore obtain

$$\begin{aligned}&\left|\int f \,\mathrm{d}\mathbb{P}_n - \int f \,\mathrm{d}\mathbb{P}\right| \\ \le &\left|\int f \,\mathrm{d}\mathbb{P}_n - \int f_k \,\mathrm{d}\mathbb{P}_n\right| + \left|\int f_k \,\mathrm{d}\mathbb{P}_n - \int f_k \,\mathrm{d}\mathbb{P}\right| + \left|\int f_k \,\mathrm{d}\mathbb{P} - \int f \,\mathrm{d}\mathbb{P}\right| \to 0,\end{aligned}$$

as $n \to \infty$, then $k \to \infty$ and finally $\epsilon \to 0$. □

Lemma 9.2. *If* $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$, *$g$ is a continuous mapping, then*

$$\mathbb{P}_n \circ g^{-1} \xrightarrow{\mathrm{d}} \mathbb{P}_n \circ g^{-1},$$

or in other words,

$$X_n \xrightarrow{\mathrm{d}} X \implies g(X_n) \xrightarrow{\mathrm{d}} g(X).$$

Proof. Since $f \in C_b(\mathbb{R}) \implies f \circ g \in C_b(\mathbb{R})$, the result follows. □

Lemma 9.3. *If* $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$, *then* $\exists$ *r.v.s* $Y_n \sim$ *cdf* F_n, $Y \sim$ *cdf* F, s. t. $Y_n \to Y$ *a.s.*

Proof. From Lemma 8.2, we know that, for the probability space $((0,1), \mathcal{B}(0,1), \lambda)$, where λ is Lebesgue measure, we have r.v.s

$$\begin{aligned}F_n^{-1}(\omega) &= \inf\left\{y : F_n(y) \ge \omega\right\} \sim \text{cdf } F_n, \\ F^{-1}(\omega) &= \inf\left\{y : F(y) \ge \omega\right\} \sim \text{cdf } F.\end{aligned}$$

Let Ω_0 be the set of continuity points of F^{-1}. A similar argument as that in the proof of Theorem 9.1 shows that Ω_0^c is countable, hence $\lambda(\Omega_0) = 1$. In the following, we show that $F_n^{-1}(\omega) \to F^{-1}(\omega)$ for all $\omega \in \Omega_0$.

For any continuity point y of F such that $y < F^{-1}(\omega)$, we have $F(y) < \omega$ (cf. Week 8), hence for all n sufficiently large, $F_n(y) < \omega$ since $F_n(y) \to F(y)$ from Theorem 9.1. Thus,

$F_n^{-1}(\omega) \geq y$ and $\liminf_{n\to\infty} F_n^{-1}(\omega) \geq F^{-1}(\omega)$ by taking $y \to F^{-1}(\omega)$ (note that set of continuity points of F is dense).

For any continuity point y of F such that $y > F^{-1}(\omega)$, $F(y) > \omega$ since $\omega \in \Omega_0$ (cf. Week 8), hence for all n sufficiently large, $F_n(y) > \omega$. Thus, $F_n^{-1}(\omega) \leq y$ and $\limsup_{n\to\infty} F_n^{-1}(\omega) \leq F^{-1}(\omega)$. $\square$

Lemma 9.4. *If for any subsequence $(n(k))_{k\geq 1}$, $\exists$ subsubsequence $(n(k(r)))_{r\geq 1}$ s. t. $\mathbb{P}_{n(k(r))} \xrightarrow{\mathrm{d}} \mathbb{P}$, then*

$$\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}.$$

Proof. Suppose otherwise. Then $\exists f \in C_b(\mathbb{R}), \epsilon > 0$, sequence $(n(k))_{k\geq 1}$, such that

$$\left| \int f \, \mathrm{d}\mathbb{P}_{n(k)} - \int f \, \mathrm{d}\mathbb{P} \right| > \epsilon,$$

for all $k \geq 1$. Then, any subsequence of distributions indexed by $(n(k(r)))_{r\geq 1}$ cannot converge, a contradiction. $\square$

Lemma 9.5. $X_n \xrightarrow{\mathrm{p}} X \implies X_n \xrightarrow{\mathrm{d}} X$.

Proof. From Corollary 6.1, for any sequence $(n(k))_{k\geq 1}$, $\exists$ subsequence $(n(k(r)))_{r\geq 1}$, s. t. $X_{n(k(r))} \to X$ a.s. Let $f \in C_b(\mathbb{R})$. By DCT, we obtain

$$\lim_{r\to\infty} \mathbb{E}\big[f(X_{n(k(r))})\big] = \mathbb{E}\Big[\lim_{r\to\infty} f(X_{n(k(r))})\Big] = \mathbb{E}[f(X)],$$

Therefore, $X_{n(k(r))} \xrightarrow{\mathrm{d}} X$. From Lemma 9.4, we have $X_n \xrightarrow{\mathrm{d}} X$. $\square$

In general, we can prove $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$ in two steps:

(i) Show that for all sequences $(n(k))_{k\geq 1}$, $\exists\ (n(k(r)))_{r\geq 1}$, s. t. $\mathbb{P}_{n(k(r))}$ converges. This is done using the concept of uniform tightness, which we discuss below.

(ii) Show that all subsequential distribution limits above are the same $\mathbb{P}$. This is done through the use of characteristic functions, which we discuss next week.

9.2 Uniform Tightness

Definition 9.2 (Uniform tightness). *$(\mathbb{P}_n)_{n\geq 1}$ is uniformly tight if $\forall\, \epsilon > 0$, $\exists$ compact $K \subset \mathbb{R}$ such that*

$$\mathbb{P}_n(K) \geq 1 - \epsilon, \ \forall\, n \geq 1.$$

Theorem 9.2 (Helly's selection theorem)**.** *If $(\mathbb{P})_{n\geq 1}$ is uniformly tight, then $\exists\ (n(k))_{k\geq 1}$ such that*

$$\mathbb{P}_{n(k)} \xrightarrow{\mathrm{d}} \mathbb{P}, \text{ for some } \mathbb{P}.$$

Lemma 9.6 (Cantor's Diagonalization)**.** *Suppose A is a countable set, and $f_n : A \mapsto \mathbb{R}$. Then, $\exists\ (n(k))_{k\geq 1}$,* s. t. *$f_{n(k)}(a)$ converges (or goes to $\pm\infty$), $\forall\, a \in A$.*

Proof. Let $A = \{a_1, a_2, \ldots\}$. From Lemma 1.5, we have

$$\begin{aligned}
&\exists\ (n_1(k))_{k\geq 1}\,, \text{ s. t. } \ f_{n_1(k)}(a_1) \text{ converges},\\
&\exists\ (n_2(k))_{k\geq 1} \subset (n_1(k))_{k\geq 1}\,, \text{ s. t. } \ f_{n_2(k)}(a_2) \text{ converges},\\
&\ldots
\end{aligned}$$

Therefore, for the diagonal sequence $(n_k(k))_{k\geq 1}$, $f_{n_k(k)}(a_l)$ converges $\forall\, l$. □

Proof of Theorem 9.2. Since $\mathbb{Q}$ is a dense subset of $\mathbb{R}$ and countable, by Lemma 9.6, $\exists\ (n(k))_{k\geq 1}$,

$$\text{s. t. } \ F_{n(k)}(q) \to F(q), \forall\, q \in \mathbb{Q}.$$

Now we extend the definition of F on $\mathbb{Q}$ to $\mathbb{R}$ by defining $F(x) = \inf\big\{F(q) : x < q, q \in \mathbb{Q}\big\}$. We prove that $F(x)$ is a cdf.

- It is obvious $F(x)$ is a non-decreasing function.
- It is right-continuous because

$$\begin{aligned}
\lim_{x_n \downarrow x} F(x_n) &= \lim_{n\to\infty} \inf\{F(q) : x_n < q \in \mathbb{Q}\}\\
&= \inf_{n\geq 1} \inf\{F(q) : x_n < q \in \mathbb{Q}\}\\
&= \inf\{F(q) : x < q \in \mathbb{Q}\}\\
&= F(x),
\end{aligned}$$

  where the second equality follows because $\inf\{F(q) : x_n < q \in \mathbb{Q}\}$ is a decreasing sequence for $x_n \downarrow x$.

- Since $(\mathbb{P}_n)_{n\geq 1}$ is uniformly tight, $\forall\, \epsilon > 0$, $\exists\, [-M, M]$,

$$\text{s. t. } \ 1 - F_n(M) + F_n(-M) \leq \epsilon, \ \forall\, n.$$

  Choose $r < -M < M < s$, $r, s \in \mathbb{Q}$. We have

$$1 - F(s) + F(r) = \lim_{k\to\infty} \big(1 - F_{n(k)}(s) + F_{n(k)}(r)\big) \leq \epsilon,$$

  which implies

$$\begin{aligned}
&\lim_{x\to-\infty} F(x) = 0,\\
&\lim_{x\to\infty} F(x) = 1.
\end{aligned}$$

Therefore, $F(x)$ is a cdf.

Let x be a continuity point of F and $a, b \in \mathbb{Q}$, s. t. $a < x < b$. We have

$$F(a) = \lim_{k\to\infty} F_{n(k)}(a) \le \liminf_{k\to\infty} F_{n(k)}(x) \le \limsup_{k\to\infty} F_{n(k)}(x) \le \lim_{k\to\infty} F_{n(k)}(b) = F(b).$$

Taking $a \uparrow x$ and $b \downarrow x$, we obtain

$$\lim_{k\to\infty} F_{n(k)}(x) = F(x).$$

From Theorem 9.1, we have

$$\mathbb{P}_{n(k)} \xrightarrow{\mathrm{d}} \mathbb{P}.$$

□

Lemma 9.7. *If $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$, then $(\mathbb{P}_n)_{n\ge 1}$ is uniformly tight.*

Proof. Let $\epsilon > 0$. Choose $M > 0$, s. t. $\mathbb{P}([-M, M]^c) \le \epsilon$. Let

$$f(x) = \begin{cases} 0, & \text{if } |x| \le M, \\ \text{linear}, & \text{if } M < x < 2M, \text{or} -2M < x < -M, \\ 1, & \text{if } |x| \ge 2M. \end{cases}$$

Note that $f \in C_b(\mathbb{R})$. We have

$$\limsup_{n\to\infty} \mathbb{P}_n([-2M, 2M]^c) \le \limsup_{n\to\infty} \int f(x)\,\mathrm{d}\mathbb{P}_n(x) = \int f(x)\,\mathrm{d}\mathbb{P}(x) \le \mathbb{P}([-M, M]^c) \le \epsilon.$$

Therefore, $\exists\, n_0$, s. t. $\forall\, n \ge n_0, \mathbb{P}_n([-2M, 2M]^c) \le 2\epsilon$. For each $n < n_0$, we choose M_n such that

$$\mathbb{P}_n([-M_n, M_n]^c) \le 2\epsilon.$$

Let $M^* = \max\{M_1, M_2, \ldots, M_{n-1}, 2M\}$. We then have $\mathbb{P}_n([-M^*, M^*]^c) \le 2\epsilon, \forall\, n \ge 1$. □