Lecture No 5 March 22, 2022 (Tue)

8.1 Convergence in law and weak convergence of measures — continuation

We stated the following theorem.

[Theorem 8.1] Let $\mu_n, n = 1, 2, \dots$ and μ be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then, the following five conditions are mutually equivalent.

(1) $\mu_n \Longrightarrow \mu$ (weak convergence)

(2) For ${}^\forall G$: open set of $\mathbb{R}$, $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$

(3) For ${}^\forall C$: closed set of $\mathbb{R}$, $\limsup_{n\to\infty} \mu_n(C) \leq \mu(C)$

(4) For ${}^\forall A \in \mathcal{B}(\mathbb{R})$ s.t. $\mu(\partial A) = 0$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$, where $\partial A = \bar{A} \setminus A^\circ$ is the boundary of A.

(5) Let F_n, F be distribution functions of μ_n, μ, respectively, i.e. $F(x) = \mu((-\infty, x])$ etc. Then, at ${}^\forall$continuity point x of F, $\lim_{n\to\infty} F_n(x) = F(x)$ holds (i.e. for ${}^\forall x$ s.t. $F(x) = F(x-)$).

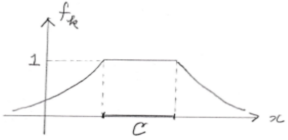
[Proof] We showed

$$(2) \iff (3),\ (2)+(3) \implies (4),\ (4) \implies (5).$$

What are left is to show:

$$(1) \implies (3) \text{ and } (5) \implies (1).$$

• (1) $\Longrightarrow$ (3)



∵ Let C be a closed set and set

$$f_k(x) := \frac{1}{\{1 + \mathrm{dist}\,(x, C)\}^k}, \quad x \in \mathbb{R},$$

where $\mathrm{dist}\,(x, C) = \inf\{|x - y|; y \in C\}$. Then, we have

$$\mu(C) = \int_{\mathbb{R}} 1_C(x)\,\mu(dx) = \lim_{k\to\infty} \int_{\mathbb{R}} f_k(x)\,\mu(dx)$$
$$= \lim_{k\to\infty} \lim_{n\to\infty} \int_{\mathbb{R}} f_k(x)\,\mu_n(dx) \geq \limsup_{n\to\infty} \mu_n(C)$$

and (3) is shown. Here, for the second equality, noting $0 \leq f_k(x) \leq 1$ and $\lim_{k\to\infty} f_k(x) = 1_C(x)$ (which follows from the closedness of C), we apply Lebesgue's convergence theorem. The third equality follows by $f_k \in C_b(\mathbb{R})$ and (1). The last inequality follows by $f_k(x) \geq 1_C(x)$. □

• Finally the "proof of (5) $\Longrightarrow$ (1)" is left.

(∵) ◦ Cut-off of the tail of F: For ${}^{\forall}\varepsilon > 0$, ${}^{\exists}a < b$ s.t a, b are both continuity points of F and $F(a) \le \varepsilon$, $1 - F(b) \le \varepsilon$ hold.

◦ From (5), for n sufficiently large, we have

$$F_n(a) \le 2\varepsilon, \quad 1 - F_n(b) \le 2\varepsilon.$$

◦ Next, for any $\delta > 0$ and $f \in C_b(\mathbb{R})$, take a simple function g such that it jumps only at continuity points of F and
$\|f - g\|_{L^\infty([a,b])} < \delta$ holds.
Then, set $g \equiv 0$ on $(-\infty, a], [b, \infty)$.

○ Under the above preparation, we can estimate as

$$\begin{aligned}
&\left|\int_{\mathbb{R}} f(x)\,\mu_n(dx) - \int_{\mathbb{R}} f(x)\,\mu(dx)\right| \\
&\quad\le \left|\int_{\mathbb{R}} f(x)\,\mu_n(dx) - \int_{\mathbb{R}} g(x)\,\mu_n(dx)\right| \\
&\qquad + \left|\int_{\mathbb{R}} g(x)\,\mu_n(dx) - \int_{\mathbb{R}} g(x)\,\mu(dx)\right| \\
&\qquad + \left|\int_{\mathbb{R}} f(x)\,\mu(dx) - \int_{\mathbb{R}} g(x)\,\mu(dx)\right|
\end{aligned}$$

By the condition (5), (2nd term)$\to 0$. Furthermore, estimating integrals separately on $[a, b]$ and other region, we obtain for large enough n,

$$\begin{aligned}
&(\text{1st term}) \le \delta + 4\varepsilon \|f\|_{L^\infty(\mathbb{R})} \\
&(\text{3rd term}) \le \delta + 2\varepsilon \|f\|_{L^\infty(\mathbb{R})}.
\end{aligned}$$

Thus, (1) is shown. □

[Remark]

- "$\mu_n \Longrightarrow \mu$ (weak convergence)" is metrizable, that is, one can construct a metric on the space $\mathcal{P}(\mathbb{R}) := \{$Borel probability measures on $\mathbb{R}\}$ in such a manner that the convergence determined this metric coincides the weak convergence. For such metrics, we have
- Lévy's distance: For distribution functions F, G,

$$\rho(F, G) = \inf\left\{\varepsilon > 0;\ F(x-\varepsilon) - \varepsilon \le G(x) \le F(x+\varepsilon) + \varepsilon,\ {}^{\forall}x \in \mathbb{R}\right\}$$

- Or, take $\{f_n\}_{n=1}^{\infty} \subset C_b(\mathbb{R})$: dense, and set

$$\tilde{\rho}(\mu, \nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|\int_{\mathbb{R}} f_n d\mu - \int_{\mathbb{R}} f_n d\nu|}{1 + |\int_{\mathbb{R}} f_n d\mu - \int_{\mathbb{R}} f_n d\nu|}, \quad \mu, \nu \in \mathcal{P}(\mathbb{R}).$$

P: Show that both ρ and $\tilde{\rho}$ are metrics on $\mathcal{P}(\mathbb{R})$ and determine the same convergence as the weak convergence.

- Prokhorov's theorem
  - $\{\mu_\alpha\}_\alpha \subset \mathcal{P}(\mathbb{R})$ is weakly compact (or relatively compact) $\underset{\text{def}}{\Longleftrightarrow}$ For ${}^\forall$subsequence (μ_{α_n}) of $(\mu_\alpha)_\alpha$, one can find further subsequence $(\mu_{\alpha_{n_k}})$ and $\mu \in \mathcal{P}(\mathbb{R})$ s.t. $\mu_{\alpha_{n_k}} \Longrightarrow \mu$ holds.
  - (Prokhorov's theorem) $\{\mu_\alpha\}$: weakly compact $\Longleftrightarrow$ $\{\mu_\alpha\}$: tight i.e. For ${}^\forall\varepsilon > 0$, ${}^\exists M > 0$ s.t. $\inf_\alpha \mu_\alpha(|x| \leq M) \geq 1 - \varepsilon$.

Prokhorov (Wikipedia)

[Example] $\{\mu_n := \delta_n\}_n$ (δ–measure concentrated at the point n) is not tight. The mass at n moves out to $+\infty$. For the corresponding distribution functions, we see

$$F_n(x) = 1_{[n,\infty)}(x) \underset{n\to\infty}{\longrightarrow} 0, \ {}^\forall x \in \mathbb{R}$$

so that the limit ($\equiv 0$) is not a distribution function.

We prepare a lemma to prove Prokhorov's theorem:

[Helly's selection theorem] For $^{\forall}$sequence of distribution functions $\{F_n\}$, $^{\exists}F_{n_k}$: subsequence, $^{\exists}F$: increasing and right-continuous s.t. $\lim_{k\to\infty} F_{n_k}(x) = F(x)$ holds for $^{\forall}$continuity point x of F. □

[Outline of proof]

• First, by diagonal argument, one can find a subsequence F_{n_k} which converges for $^{\forall}x \in \mathbb{Q}$. Denote its limit by $\tilde{F}(y), y \in \mathbb{Q}$.
• We define F on $\mathbb{R}$ by extending $\tilde{F}(y)$ such that F is right-continuous: $F(x) = \inf\{\tilde{F}(y); y \in \mathbb{Q}, y > x\}$.
• We may only show the convergence at $^{\forall}$continuity point of F. (Leave as an exercise). □

P: Show the last point.

[Proof of Prokhorov's theorem]
(We only show [$\Longleftarrow$], which is important for application.)

• Missing is that the limit F obtained by Helly's selection theorem is a distribution function (or that the limit μ of $\mu_{\alpha_{n_k}}$ is a probability measure), but the tightness guarantees this.

• Indeed, take $^\forall$subsequence (μ_{α_n}) of (μ_α), then by Helly's selection theorem, $^\exists$subsequence $(F_{\alpha_{n_k}})$ of the corresponding distribution functions (F_{α_n}) and $^\exists F$ s.t.

$$\lim_{k\to\infty} F_{\alpha_{n_k}}(x) = F(x)$$

holds for $^\forall$continuity point x of F. However, by tightness, $^\forall\varepsilon > 0$, $^\exists M > 0$ (we can assume $\pm M$ are both continuity points of F by making M slightly larger if necessary) s.t.

$$\mu_{\alpha_{n_k}}([-M, M]) \left(= F_{\alpha_{n_k}}(M) - F_{\alpha_{n_k}}(-M)\right) \geq 1 - \varepsilon$$

holds. By letting $k \to \infty$, we obtain $F(M) - F(-M) \geq 1 - \varepsilon$ and this implies that F is a distribution function of a certain probability measure $\mu \in \mathcal{P}(\mathbb{R})$. □

8.2 Characteristic function

As a preparation for the proof of central limit theorem (CLT), we introduce a characteristic function of $\mu \in \mathcal{P}(\mathbb{R})$ (i.e. probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$) as a Fourier transform with variable ξ replaced by $-\xi$, or Fourier inverse transform.

[Definition 8.2] **Characteristic function** of $\mu \in \mathcal{P}(\mathbb{R})$ is defined by

$$\varphi(\xi) \equiv \varphi_\mu(\xi) := \int_{\mathbb{R}} e^{i\xi x}\,\mu(dx), \quad \xi \in \mathbb{R},$$

where $i = \sqrt{-1}$. Characteristic function of a real-valued r.v. X is the characteristic function of P_X which is a distribution of X, i.e.,

$$\varphi_X(\xi) = E\left[e^{i\xi X}\right], \quad \xi \in \mathbb{R} \qquad \square$$

[Example] (1) The characteristic function of Gaussian distribution $\mu_{m,v}(dx) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x-m)^2}{2v}} dx$ with mean $m \in \mathbb{R}$ and variance $v > 0$ is given by

$$\varphi_{\mu_{m,v}}(\xi) = \exp\left\{ im\xi - \frac{1}{2} v\xi^2 \right\}$$

(2) For $\lambda > 0$, let μ_λ be a Poisson distribution with parameter λ, i.e. $\mu_\lambda(\ell) = \frac{\lambda^\ell}{\ell!} e^{-\lambda}$, $\ell \in \mathbb{Z}_+$. Then,

$$\varphi_{\mu_\lambda}(\xi) = \exp\left\{ \lambda(e^{i\xi} - 1) \right\} \qquad \square$$

⊙

$$\begin{aligned} \varphi_{\mu_\lambda}(\xi) &= \sum_{\ell=0}^{\infty} e^{i\xi\ell} \frac{\lambda^\ell}{\ell!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{\ell=0}^{\infty} \frac{1}{\ell!} (\lambda e^{i\xi})^\ell = e^{-\lambda} e^{\lambda e^{i\xi}} \qquad \square \end{aligned}$$

• Uniqueness theorem (i.e. characteristic functions uniquely characterize distributions)

[Theorem 8.2] Let $\mu, \tilde{\mu} \in \mathcal{P}(\mathbb{R})$. Then,

$$\varphi_\mu(\xi) = \varphi_{\tilde{\mu}}(\xi),\ ^\forall \xi \in \mathbb{R} \iff \mu = \tilde{\mu} \qquad \square$$

[Outline of proof] ($\Longleftarrow$) is obvious. We may show ($\Longrightarrow$) only.
• Inversion formula: the distribution function $F = F_\mu(x)$ of $\mu \in \mathcal{P}(\mathbb{R})$ is reconstructed from the characteristic function φ_μ as

$$F(b) - F(a) = \lim_{R\to\infty} \frac{1}{2\pi} \int_{-R}^{R} \frac{e^{-i\xi b} - e^{-i\xi a}}{-i\xi} \varphi_\mu(\xi)\, d\xi$$

for continuity points $a, b \in \mathbb{R}$ of F. ([Remark] From $\frac{e^{-i\xi b} - e^{-i\xi a}}{-i\xi} = \int_a^b e^{-i\xi x} dx$, RHS has the form of Fourier inverse transform. We omit the proof of inversion formula.)

• Once inversion formula is shown, $\varphi_\mu = \varphi_{\tilde{\mu}}$ implies $F_\mu = F_{\tilde{\mu}}$ and this further implies $\mu = \tilde{\mu}$. $\square$

• Equivalence of convergence of characteristic functions and weak convergence of distributions

[Theorem 8.3] $\mu_n \in \mathcal{P}(\mathbb{R}), n = 1, 2, \ldots,$
φ_n: corresponding characteristic functions.
(1) If $\mu_n \Longrightarrow {}^{\exists}\mu \in \mathcal{P}(\mathbb{R})$, then $\varphi_n(\xi) \longrightarrow \varphi_\mu(\xi)$, ${}^{\forall}\xi \in \mathbb{R}$
(2) Conversely, if $\varphi_n(\xi) \longrightarrow \varphi(\xi)$, ${}^{\forall}\xi \in \mathbb{R}$ and the limit φ is continuous at $\xi = 0$, then ${}^{\exists}\mu \in \mathcal{P}(\mathbb{R})$ s.t. $\varphi = \varphi_\mu$ and $\mu_n \Longrightarrow \mu$ holds. □

[Proof] (1) is obvious from the definition of weak convergence by noting that the real and imaginary parts of $f(x) = e^{i\xi x}$ both belong to $C_b(\mathbb{R})$.

Next we show (2).

To show (2),

- $(\mu_n)_{n=1,2,\dots}$ is tight, and therefore weakly compact.

⚆ We show, by the continuity of φ at 0, that ${}^\forall\varepsilon > 0$, ${}^\exists M > 0$ s.t. $\overline{\lim_{n\to\infty}} \mu_n(|x| > M) \le \varepsilon$; the detail is omitted. Once this is shown, we can apply Prokhorov's theorem. □

- By the weak compactness, from ${}^\forall$subsequence $\{\mu_{n'}\}$ of $\{\mu_n\}$, one can choose ${}^\exists$further subsequence $\{\mu_{n''}\}$ s.t. $\mu_{n''} \Longrightarrow {}^\exists\mu \in \mathcal{P}(\mathbb{R})$. Then, by (1), $\varphi_{n''}(\xi) \to \varphi_\mu(\xi)$ so that $\varphi_\mu(\xi) = \varphi(\xi)$. In particular, since φ_μ is uniquely determined, by Theorem 8.2, the limit μ is uniquely determined independently of the choice of subsequences. Therefore, $\{\mu_n\}$ itself converges (without taking subsequence) and we obtain $\mu_n \Longrightarrow \mu$. This completes the proof of Theorem 8.3. □

- Characterization of characteristic functions

[Proposition 8.4] (Properties of characteristic functions)
Let $\varphi = \varphi_X$ be a characteristic function of some r.v. X. Then, the following hold.
(1) $\varphi(0) = 1$ and φ is uniformly continuous on $\mathbb{R}$.
(2) positive definite: For ${}^\forall n \in \mathbb{N}$ and ${}^\forall \xi_j \in \mathbb{R}, 1 \leq j \leq n$, $n \times n$ Hermite matrix $(\varphi(\xi_j - \xi_k))_{1 \leq j,k \leq n}$ is non-negative definite:

$$\sum_{j,k=1}^{n} \varphi(\xi_j - \xi_k) z_j \bar{z}_k \geq 0, \quad {}^\forall z_j \in \mathbb{C},$$

($\bar{z}$ is a complex conjugate of $z \in \mathbb{C}$) □

[Proof] $\varphi(0) = 1$ is obvious. For $^\forall \xi_1, \xi_2 \in \mathbb{R}$,

$$
\begin{aligned}
|\varphi(\xi_1) - \varphi(\xi_2)| &= \left| E\left[e^{i\xi_1 X} - e^{i\xi_2 X} \right] \right| \\
&\le E\left[\left| e^{i(\xi_1 - \xi_2)X} - 1 \right| \right] \\
&\le E\left[(|\xi_1 - \xi_2| \cdot |X|) \wedge 2 \right] \underset{|\xi_1 - \xi_2| \to 0}{\longrightarrow} 0
\end{aligned}
$$

and this implies the uniform continuity. For the last inequality, we used $|e^{ix} - 1| \le |x| \wedge 2$.

(2) is easy from

$$
\sum_{j,k=1}^{n} \varphi(\xi_j - \xi_k) z_j \bar{z}_k = E\left[\left| \sum_{j=1}^{n} e^{i\xi_j X} z_j \right|^2 \right] \ge 0 \qquad \square
$$

Conversely, two conditions (1), (2) in Proposition 8.4 characterize for a function φ to be a characteristic function of some distribution. The continuity of φ in (1) is enough to be assumed only at $\xi = 0$. We omit the proof.

[Theorem 8.5] (Bochner's theorem) If a function $\varphi = \varphi(\xi) : \mathbb{R} \to \mathbb{C}$ satisfies three conditions

"continuous at $\xi = 0$, $\varphi(0) = 1$, positive definite",

then φ is a characteristic function of some $\mu \in \mathcal{P}(\mathbb{R})$. $\square$


Bochner (from Wikipedia)

8.3 Central limit theorem

As before, a sequence of real-valued r.v.'s are defined on a probability space $(\Omega, \mathcal{F}, P)$.

[Theorem 8.6] (Central limit theorem) Assume $\{X_n\}_{n=1}^{\infty}$: *i.i.d.*, $E[X_n^2] < \infty$ and $\mathrm{Var}(X_n) = v > 0$. Then,

$$Z_n := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - m), \quad m = E[X_1]$$

converges in law as $n \to \infty$ to a r.v. Z distributed under $N(0, v)$ (Gaussian distribution with mean 0 and variance v). In particular,

$$\lim_{n\to\infty} P(a \le Z_n \le b) = \frac{1}{\sqrt{2\pi v}} \int_a^b e^{-\frac{x^2}{2v}}\, dx$$

holds for $-\infty < a < b < \infty$. □

[Proof] • Replacing $\frac{1}{\sqrt{v}}(X_n - m)$ with X_n, we may assume $m = 0, v = 1$. Note that *i.i.d.* property is kept.

• Then the characteristic function of Z_n is computed as

$$\varphi_{Z_n}(\xi) = E\left[e^{i\frac{\xi}{\sqrt{n}}\sum_{k=1}^n X_k}\right] \underset{\perp\!\!\!\perp}{=} \prod_{k=1}^n E\left[e^{i\frac{\xi}{\sqrt{n}}X_k}\right] \underset{\text{same distri}}{=} \left(\varphi\left(\frac{\xi}{\sqrt{n}}\right)\right)^n,$$

where φ is a characteristic function of $X(= X_1)$. However, from $E[X^2] = 1$, we see $\varphi \in C^2(\mathbb{R})$ and $\varphi, \varphi', \varphi''$ are all bounded. Indeed,

$$\varphi(\xi) = E\left[e^{i\xi X}\right], \quad \varphi'(\xi) = E\left[iXe^{i\xi X}\right],$$
$$\varphi''(\xi) = E\left[-X^2 e^{i\xi X}\right]$$

Exchange of the derivative $\frac{d}{d\xi}$ and integral $E[\,\cdot\,]$ is guaranteed by noting $\left|iXe^{i\xi X}\right| \leq |X|$, $\left|-X^2 e^{i\xi X}\right| \leq X^2$. The continuity of φ', φ'' follows by Lebesgue's convergence theorem.

Therefore, by Taylor's formula, $^\exists\eta : |\eta| \le \frac{|\xi|}{\sqrt{n}}$ s.t.

$$\varphi\left(\frac{\xi}{\sqrt{n}}\right) = \varphi(0) + \frac{\xi}{\sqrt{n}}\varphi'(0) + \frac{\xi^2}{2n}\varphi''(\eta).$$

By $\varphi(0) = 1, \varphi'(0) = E[iX] = 0$, $\varphi''(0) = -E[X_1^2] = -1$,

$$\begin{aligned}\log \varphi_{Z_n}(\xi) &= n \log \varphi\left(\frac{\xi}{\sqrt{n}}\right) \\ &= n \log\left(1 + \frac{\xi^2}{2n}\varphi''(\eta)\right) \\ &= \frac{\xi^2}{2}\varphi''(0) + o(1) \longrightarrow\ -\frac{\xi^2}{2},\end{aligned}$$

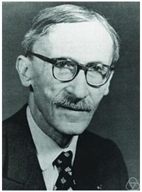
where $\log z$ is a function on $\mathbb{C}$ (enough to be defined near 1, so we may take principal value). For the 3rd line, we used $\log(1+\delta) = \delta + o(\delta)$ $(\delta(\in \mathbb{C}) \to 0)$. Thus, we have

$$\varphi_{Z_n}(\xi) \ \longrightarrow\ \exp\left\{-\frac{\xi^2}{2}\right\} = \varphi_Z(\xi),$$

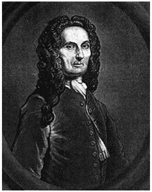
where $Z \underset{\text{law}}{=} N(0,1)$. Therefore, by Theorem 8.3-(2), we see that Z_n converges in law to Z as $n \to \infty$. □

[Remark] The proof of CLT is easy if we use characteristic functions. This proof was originally given by P. Lévy in the beginning of 1930's.

Before him, for example for X_n taking discrete values, the probability is described by a combinatorial formula and, by estimating it for example by Stirling's formula for $n!$, one can give the proof of CLT. However, the proof is more complicated; cf. de Moivre (1730), Gauss (1809), Laplace (1810).



Lévy



de Moivre

Gauss

Laplace

(from Wikipedia)

8.4 Poisson's law of small numbers

[Theorem 8.8] (Law of small numbers) Let $(X_j^{(n)})_{j=1}^n$ be *i.i.d.* taking values in $\{0, 1\}$ and $p_n := P(X_j^{(n)} = 1) \in (0, 1)$ satisfies

$$\lim_{n\to\infty} np_n = \lambda > 0 \qquad (\text{i.e., } p_n \sim \tfrac{\lambda}{n})$$

Then, $Z_n := \sum_{j=1}^n X_j^{(n)}$ converges in law as $n \to \infty$ to a r.v. Z distributed under the Poisson distribution with parameter λ, i.e., Z is a $\{0\} \cup \mathbb{N}$–valued r.v. such that

$$P(Z = \ell) = e^{-\lambda}\frac{\lambda^\ell}{\ell!}, \ell = 0, 1, 2, \ldots \qquad \square$$

Poisson (from Wikipedia)

⦾ We compute the characteristic function. For $\xi \in \mathbb{R}$,

$$\varphi_{Z_n}(\xi) \underset{i.i.d.}{=} \left(E[e^{i\xi X_1^{(n)}}]\right)^n = \left((1-p_n) + p_n e^{i\xi}\right)^n$$

$\therefore$

$$\begin{aligned}\log \varphi_{Z_n}(\xi) &= n \log\left(1 + p_n(e^{i\xi} - 1)\right) \\ &\sim np_n(e^{i\xi} - 1) \underset{n\to\infty}{\longrightarrow} \lambda(e^{i\xi} - 1)\end{aligned}$$

$\therefore\ \varphi_{Z_n}(\xi) \to e^{\lambda(e^{i\xi}-1)}$, which is a characteristic function of the Poisson distribution with parameter λ. This concludes the proof of Theorem 8.8. □