

STAT 7200

Introduction to Advanced Probability

Lecture 24

Taylor R. Brown

- 1 Theory of Convergence II
 - Weak Convergence
 - Characteristic Functions

Equivalent Definitions of Weak Convergence

Theorem 1 (Equivalent Definitions of Weakly Convergence)

The following statements are all equivalent definition of weak convergence:

(1) $\{\mu_n\}$ converges weakly to μ . (Original definition)

(2) $\mu_n(A) \rightarrow \mu(A)$ for all measurable set A such that $\mu(\partial A) = 0$. (∂A is defined as the boundary of set A)

(3) $\mu_n((-\infty, x]) \rightarrow \mu((-\infty, x])$ for all $x \in \mathbb{R}$ such that $\mu(\{x\}) = 0$. That is, the convergence of CDFs. (Note, $\{x\}$ is the boundary of set $(-\infty, x]$.)

(4) (Skorohod's Theorem) there are random variable Y, Y_1, Y_2, \dots defined on the same probability triple, with $\mathcal{L}(Y) = \mu$ and $\mathcal{L}(Y_n) = \mu_n$ such that $Y_n \rightarrow Y$ with probability 1 (This theorem connects the strongest type of convergence: convergence almost surely, with weak convergence.)

(5) $\int_{\mathbb{R}} f d\mu_n \rightarrow \int_{\mathbb{R}} f d\mu$ for all bounded Borel-measurable functions $f : \mathbb{R} \rightarrow \mathbb{R}$. such that $\mu(D_f) = 0$, where D_f is the set of discontinuous points of f . (The continuous condition of definition 1) is relaxed.)

Proof: (3) \Rightarrow (4)

- (3) $\mu_n((-\infty, x]) \rightarrow \mu((-\infty, x])$ for all $x \in \mathbb{R}$ such that $\mu(\{x\}) = 0$.
(4) there are random variable Y, Y_1, Y_2, \dots defined on the same probability triple, with $\mathcal{L}(Y) = \mu$ and $\mathcal{L}(Y_n) = \mu_n$ such that $Y_n \rightarrow Y$ with probability 1
- Proof: We have shown that, let $F_n(x) = \mu_n((-\infty, x])$, $F(x) = \mu((-\infty, x])$, and let (Ω, \mathcal{F}, P) be the uniform measure over $\Omega = [0, 1]$, and $Y_n(\omega) = \inf\{x : F_n(x) \geq \omega\}$, $Y(\omega) = \inf\{x : F(x) \geq \omega\}$.
 - Then Y_n has CDF $F_n(x)$ and Y has CDF $F(x)$. and $Y_n(\omega) \rightarrow Y(\omega)$ if $Y(\omega)$ is continuous at ω .

Example

- If $\mu_n \sim N(0, \frac{1}{n})$, then $\{\mu_n\}$ converges weakly to $\mu = \delta_0$,
- It is usually easier to show the weak convergences using CDF function (definition (3)). In this case, note that the CDF of μ , $F_\mu(x) = 0$ for $x < 0$, but $F_\mu(x) = 1$ for $x \geq 0$.
- Now for $x < 0$, let Z be a random variables that follow $N(0, 1)$ distribution, then as $n \rightarrow \infty$,

$$F_{\mu_n}(x) = P(Z \leq \sqrt{nx}) \rightarrow 0 = F_\mu(x)$$

- For $x > 0$,

$$F_{\mu_n}(x) = 1 - \mu_n((x, \infty)) = 1 - P(Z > \sqrt{nx}) \rightarrow 1 = F_\mu(x)$$

- For $x = 0$, $F_{\mu_n}(0) = \frac{1}{2} \not\rightarrow 1 = F_\mu(0)$ as $n \rightarrow \infty$. However, since $\mu(\{0\}) = 1 \neq 0$, we can still conclude that $\{\mu_n\}$ converges weakly to $\mu = \delta_0$.

On Skorohod's Representation Theorem

- The definition (4) is also known as Skorohod's Representation Theorem. This theorem essentially states that, if probability measure $\{\mu_n\}$ converge weakly to μ , then we can then find random variables Y, Y_1, Y_2, \dots such that $\mathcal{L}(Y_n) = \mu_n$, $\mathcal{L}(Y) = \mu$, and $\{Y_n\}$ converges to Y almost surely.
- However, if we start with random variables X_1, X_2, \dots , and $\{\mathcal{L}(X_n)\}$ converges weakly to $\mathcal{L}(X)$, this does not allow us to deduce any conclusion regarding whether $\{X_n\}$ would converge to X in probability or almost surely. In fact, do not even know whether X, X_1, X_2, \dots are defined on the same probability triple.
- However, by applying Skorohod's Representation Theorem, we will be able to find other random variables that have the same distribution as $\{X_n\}$ and X , but are defined on the same sample space and enjoy the strong condition of convergence almost surely.
- This theorem can be used to link results involving weakly converges to the results involving convergence almost surely. Example on the next slide...

On Skorohod's Representation Theorem: Example

Proposition 2

If $X_n \geq 0$, and $\{\mathcal{L}(X_n)\}$ converges weakly to $\mathcal{L}(X)$ then
 $E(X) \leq \liminf E(X_n)$

- **Proof:** Fatou's Lemma (Theorem 9.1.1) Told us that $E[\liminf_n X_n] \leq \liminf_n E[X_n]$. $\liminf_n X_n$ is only defined if all of these random variables are defined on the same probability space.
- However, by applying Skorohod's Representation Theorem, we can find random variables Y, Y_1, Y_2, \dots such that $\mathcal{L}(Y_n) = \mathcal{L}(X_n)$, $\mathcal{L}(Y) = \mathcal{L}(X)$, and $\{Y_n\}$ converges to Y almost surely. Then:

$$E(X) = E(Y) = E(\liminf_n Y_n) \leq \liminf_n E(Y_n) = \liminf_n E(X_n)$$

Convergence in Probability Implies Weak Convergence

Theorem 3 (Convergence in Probability Implies Weak Convergence)

If $\{X_n\} \rightarrow X$ in probability, then $\{\mathcal{L}(X_n)\}$ converges weakly to $\mathcal{L}(X)$

- **Proof:** We need to show $P(X_n \leq z) \rightarrow P(X \leq z)$ for $P(X = z) = 0$ based on the fact that, for any $\varepsilon > 0$, $P(|X_n - X| \geq \varepsilon) \rightarrow 0$.
- We will show $\limsup_n P(X_n \leq z) \leq P(X \leq z) \leq \liminf_n P(X_n \leq z)$
 - $P(X_n \leq z) = P(X_n \leq z, X \leq z + \varepsilon) + P(X_n \leq z, X > z + \varepsilon) \leq P(X \leq z + \varepsilon) + P(|X - X_n| \geq \varepsilon)$
 - Then we have: $\limsup_n P(X_n \leq z) \leq P(X \leq z + \varepsilon)$, let $\varepsilon \rightarrow 0$, $\limsup_n P(X_n \leq z) \leq P(X \leq z)$
 - $P(X \leq z - \varepsilon) = P(X \leq z - \varepsilon, X_n \leq z) + P(X \leq z - \varepsilon, X_n > z) \leq P(X_n \leq z) + P(|X - X_n| \geq \varepsilon)$
 - Then we have: $\liminf_n P(X_n \leq z) \geq P(X \leq z - \varepsilon)$, let $\varepsilon \rightarrow 0$, $\liminf_n P(X_n \leq z) \geq P(X < z) = P(X \leq z)$ when $P(X = z) = 0$.

Characteristic Function: Definition and Properties

- **Characteristic Function** of a random variable X is defined as:

$$\phi_X(t) = E(e^{itX}) = E[\cos(tX)] + iE[\sin(tX)], \quad t \in \mathbb{R}$$

- Unlike moment generating function, characteristic function is always finite for any $t \in \mathbb{R}$, since:

$$|e^{itX}| = 1$$

- Similar to the moment generating function, we can use the derivative of characteristic function to find the moment:

$$\phi_X^{(k)}(0) = i^k E(X^k), \quad \text{if } E(|X^k|) < \infty$$

- **Remark:** If the moment generating function is finite in an interval around 0, then all the moments are finite. However, the knowledge on characteristic function (which is always finite) does not guarantee the existence of moment of any order.

Characteristic Function: A Uniformly Continuous Function

- **Continuous Function** A function $f(x)$ is continuous on set A if for any $\varepsilon > 0$, $x \in A$, there is a $\delta > 0$ that depends on both ε and x , so that for any $|h| < \delta$, $|f(x+h) - f(x)| < \varepsilon$.
- **Uniformly Continuous Function** A function $f(x)$ is **uniformly continuous** on A if for any $\varepsilon > 0$, there is a $\delta > 0$ that only depends on ε , so that for any $x \in A$, and for any $|h| < \delta$, $|f(x+h) - f(x)| < \varepsilon$.
- **Example (1):** $f(x) = x$ on \mathbb{R} is uniformly continuous, as $|f(x+h) - f(x)| = |h|$ does not depend on x .
- **Example (2):** $f(x) = \frac{1}{x}$ on $(0, \infty)$ is continuous but not uniformly continuous.

$$|f(x+h) - f(x)| = \left| \frac{1}{x+h} - \frac{1}{x} \right| = \left| \frac{h}{x(x+h)} \right|$$

- If $f(x) = \frac{1}{x}$ is uniformly continuous, let us choose $\varepsilon = 1$, then there must be a δ that does not depend on x , so that as long as $|h| < \delta$, $\left| \frac{h}{x(x+h)} \right| < 1$ for all $x > 0$. However, for $0 < h < 1/2$, we can always let $x = h$, and $\left| \frac{h}{x(x+h)} \right| = \left| \frac{1}{2h} \right| > 1$, a contradiction.

Characteristic Function: A Uniformly Continuous Function

- **Example (2):** $f(x) = \frac{1}{x}$ on $(0, \infty)$ is continuous but not uniformly continuous.
- **Example (3):** $f(x) = \frac{1}{x}$ on $[1, \infty)$ is uniformly continuous:
- $|f(x+h) - f(x)| = \left| \frac{1}{x+h} - \frac{1}{x} \right| = \left| \frac{h}{x(x+h)} \right| \leq |h/1 + h|$

Characteristic Function : A Uniformly Continuous Function

Proposition 4

Characteristic function is a uniformly continuous function on \mathbb{R} .

- **Proof:** Denote the distribution of random variable X as μ :

$$\begin{aligned} |\phi_X(t+h) - \phi_X(t)| &= \left| \int (e^{i(t+h)x} - e^{itx}) \mu(dx) \right| \\ &\leq \int |e^{i(t+h)x} - e^{itx}| \mu(dx) \\ &= \int |e^{itx}| |e^{ihx} - 1| \mu(dx) = \int |e^{ihx} - 1| \mu(dx), \end{aligned}$$

which does not depend on t .

- Furthermore, as $|e^{ihx} - 1| \leq 2$, by dominated convergence theorem:

$$\lim_{h \rightarrow 0} \int |e^{ihx} - 1| \mu(dx) = \int \left(\lim_{h \rightarrow 0} |e^{ihx} - 1| \right) \mu(dx) = 0.$$

So $\phi_X(t)$ is a uniformly continuous function on \mathbb{R} .