STAT 7200

Introduction to Advanced Probability
Lecture 19

Taylor R. Brown

- Weak Convergence

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

(3) $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ for all $x \in \mathrm{R}$ 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 $Y, Y_1, Y_2, \cdots$ defined on the same probability triple, with $\mathcal{L}(Y) = \mu$ and $\mathcal{L}(Y_n) = \mu_n$ such that $Y_n \to Y$ with probability 1 (This theorem connects the strongest type of convergence: convergence almost surely, with the week convergence.)

(5) $\int_{\mathrm{R}} f d\mu_n \to \int_{\mathrm{R}} f d\mu$ for all bounded Borel-measurable functions $f : \mathrm{R} \to \mathrm{R}$. 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.)

Structure of Proof

- Our proof will follow the following structure:
- We have proved: (5) $\Rightarrow$ (1), (5) $\Rightarrow$ (2) and (2) $\Rightarrow$ (3)

Proof: (1) $\Rightarrow$ (3)

- (1) $\{\mu_n\}$ converges weakly to μ: $\int_{\mathrm{R}} f d\mu_n \to \int_{\mathrm{R}} f d\mu$ for all bounded continuous functions f.

  (3) $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ for all $x \in \mathrm{R}$ such that $\mu(\{x\}) = 0$.

- **Strategy:** We can not apply (1) directly by setting $f = 1_{(-\infty,x]}$ since $1_{(-\infty,x]}$, although bounded, is discontinuous at x. We may resolve this issue by constructing continuous approximation of $1_{(-\infty,x]}$.
- **Proof:** For any $\varepsilon > 0$ (which is used to control how good the approximation is), define $f(t) = 1$ for $t \leq x$ and 0 for $t \geq x + \varepsilon$, but let $f(t)$ be a linear function on $(x, x + \varepsilon)$.

- As f is now continuous and $1_{(-\infty,x]} \leq f \leq 1_{(-\infty,x+\varepsilon]}$:

$$\limsup_n \mu_n((-\infty, x]) \leq \limsup_n \int f d\mu_n = \int f d\mu \leq \mu((-\infty, x + \varepsilon])$$

- Let $\varepsilon \to 0$. By the continuity of probability, we have $\limsup_n \mu_n((-\infty, x]) \leq \mu((-\infty, x])$

Proof: (1) $\Rightarrow$ (3): continued

- **Proof: continued** Similarly, define $f(t) = 1$ for $t \leq x - \varepsilon$ and 0 for $t \geq x$, but let $f(t)$ be a linear function on $(x - \varepsilon, x)$. Then f is linear and $1_{(-\infty, x-\varepsilon]} \leq f \leq 1_{(-\infty, x]}$. And:

$$\liminf_n \mu_n((-\infty, x]) \geq \liminf_n \int f d\mu_n = \int f d\mu \geq \mu((-\infty, x - \varepsilon])$$

- Let $\varepsilon \to 0$, $\liminf_n \mu_n((-\infty, x]) \geq \mu((-\infty, x)) = \mu((-\infty, x])$. The last equality holds since $\mu(\{x\}) = 0$.
- In summary:

$$\liminf_n \mu_n((-\infty, x]) \geq \mu((-\infty, x]) \geq \limsup_n \mu_n((-\infty, x])$$

- we then must have:

$$\lim_n \mu_n((-\infty, x]) = \mu((-\infty, x])$$

Proof: (4) $\Rightarrow$ (5)

- (4) there are random variable $Y, Y_1, Y_2, \cdots$ defined on the same probability triple, with $\mathcal{L}(Y) = \mu$ and $\mathcal{L}(Y_n) = \mu_n$ such that $Y_n \to Y$ with probability 1.

  (5) $\int_{\mathrm{R}} f d\mu_n \to \int_{\mathrm{R}} f d\mu$ for all bounded Borel-measurable functions $f : \mathrm{R} \to \mathrm{R}$. such that $\mu(D_f) = 0$, where D_f is the set of discontinuous points of f.

- **Proof:** Pick an appropriate f. First, we want to show that $P\left(f(Y_n) \to f(Y)\right) = 1$. Note that
  - $0 \leq P(Y_n(\omega) \to Y(\omega), D_f) \leq P(D_f) = 0$
  - $1 = P(Y_n \to Y) = P(Y_n \to Y, D_f) + P(Y_n \to Y, D_f^c) = P(Y_n \to Y, D_f^c)$
  - $\{\omega : f(Y_n) \to f(Y)\} \supseteq \{\omega : Y_n(\omega) \to Y(\omega)\} \cap \{\omega : Y(\omega) \in D_f^c\}$

  so $f(Y_n) \to f(Y)$ wp1 by (4) and monotonicity of P.

- Because f is bounded, $f(Y)$ is integrable, so $\mathrm{E}[f(Y_n)] \to \mathrm{E}[f(Y)]$ by the dominated convergence theorem.

Proof: (3) $\Rightarrow$ (4)

- (3) $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ for all $x \in \mathrm{R}$ such that $\mu(\{x\}) = 0$.

  (4) there are random variables $Y, Y_1, Y_2, \ldots$ defined on the same probability triple, with $\mathcal{L}(Y) = \mu$ and $\mathcal{L}(Y_n) = \mu_n$ such that $Y_n \to Y$ with probability 1

- **Strategy:** We will construct random variables with CDFs $F_n(x) = \mu_n((-\infty, x])$, $F(x) = \mu((-\infty, x])$, then we will show the convergence of these random variables using the fact that the corresponding CDFs converge.

Proof: (3) $\Rightarrow$ (4): Probability Integral Transform

- **Proof:** We can construct random variable with given CDF using probability integral transform theorem.
- This theorem states that, for random variable U that follows uniform distribution, given any CDF $F(x)$, define quantile function $Q(p) = \inf\{x : F(x) \geq p\}$, then the random variable $Q(U)$ follows distribution with CDF $F(x)$.

- By definition $F(q) \geq p \Rightarrow Q(p) \leq q$

$$\mathrm{P}[Q(U) \leq x] = \mathrm{P}[F(x) \geq U] = F(x)$$

- Other useful results include:

  a) $F(q) < p \Rightarrow Q(p) \geq q$
  b) When the CDF is continuous and strictly increasing, the quantile function is the inverse of CDF.
  c) The quantile function $Q(p)$ is a non-decreasing function, same as the CDF.

Proof: (3) $\Rightarrow$ (4): continued

- **Proof: continued** Let $F_n(x) = \mu_n((-\infty, x])$, $F(x) = \mu((-\infty, x])$, and let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability triple with the uniform measure over $\Omega = [0,1]$, and $Y_n(\omega) = \inf\{y : F_n(y) \geq \omega\}$, $Y(\omega) = \inf\{y : F(y) \geq \omega\}$. Then Y_n has CDF $F_n(x)$ and Y has CDF $F(x)$.
  - Now we will show that $Y_n(\omega) \to Y(\omega)$ if $Y(\omega)$ is continuous at ω.
  - Firstly, define $Y(\omega) = y$. Then $y - \varepsilon < Y(\omega) < y + \varepsilon$ implies:

$$F(y - \varepsilon) < \omega < F(y + \varepsilon)$$

    if $Y(\omega)$ is continuous at ω.
  - The reason: if $F(y + \varepsilon) = \omega$, for any $\delta > 0$, $F(y + \varepsilon) < \omega + \delta$, then $Y(\omega + \delta) \geq y + \varepsilon = Y(\omega) + \varepsilon$. This indicates that there is a jump of at least size ε of $Y(\omega)$ at ω, which is a contradiction to the continuity of Y at ω.
  - Thus, $F(y - \varepsilon) < \omega < F(y + \varepsilon)$

Proof: (3) $\Rightarrow$ (4): continued

- **Proof: continued** In the previous slide, if Y is continuous at ω, and $Y(\omega) = y$, then $F(y-\varepsilon) < \omega < F(y+\varepsilon)$ for all $\varepsilon > 0$.
- Now for a particular ε, we can always find $0 < \varepsilon' < \varepsilon$, such that $\mu(y-\varepsilon') = \mu(y+\varepsilon') = 0$. ($\mathrm{P}(x) > 0$ only for at most countably many $\{x\}$). Then $F_n(y-\varepsilon') \to F(y-\varepsilon')$ and $F_n(y+\varepsilon') \to F(y+\varepsilon')$. Thus, for large enough n, we have:

$$F_n(y-\varepsilon') < \omega < F_n(y+\varepsilon')$$

- Since $\omega < F_n(y+\varepsilon')$, $\omega \leq F_n(y+\varepsilon')$, then:
$$Y_n(\omega) \leq y + \varepsilon' = Y(\omega) + \varepsilon'.$$

- Since $F_n(y-\varepsilon') < \omega$, then $y - \varepsilon' < Y_n(\omega)$, which implies a weaker inequality:
$$Y_n(\omega) \geq y - \varepsilon' = Y(\omega) - \varepsilon'.$$

- In summary, $|Y_n(\omega) - Y(\omega)| \leq \varepsilon' < \varepsilon$ for large enough n. Thus, $Y_n(\omega) \to Y(\omega)$ when Y is continuous at ω

Proof: (3) $\Rightarrow$ (4): continued

- **Proof: continued** Finally, we need to establish the fact that Y is continuous with probability 1. Or equivalently, D_Y, the set of the discontinuous points of Y, has probability 0. This statement can be justified based on : a) Y is defined on a uniform measure over $[0, 1]$, b)D_Y is at most countable.
- To show D_f is at most countable, let us first create a partition of $\mathrm{R} = \bigcup_{z \in \mathrm{Z}} (z, z+1]$, and define $\Omega_z = \{\omega : z < Y(\omega) \leq z+1\} = Y^{-1}\left((z, z+1]\right)$. Since Y is a non-decreasing function, each Ω_z should be an interval on $[0, 1]$ and $\{\Omega_z\}$ forms a partition of $[0, 1]$.
- Then Let $D_f^z =$ Discontinuous points in Ω_z. Clearly $D_f = \bigcup_z D_f^z$.
- Next define $D_f^z \supseteq D_f^z(m) = \{$ jumps that have size $\geq m^{-1}\}$.
- Clearly $|D_f^z| \leq m$ and $D_f = \bigcup_z \bigcup_m D_f^z(m)$. Therefore the number of discontinuous points is at most countable.