STATS 310A: Theory of Probability I	Autumn 2016/17
Lecture 18: November 30	
Lecturer: Persi Diaconis	*Scribes: Kenneth Tay*

18.1 Weak Convergence

Definition 18.1 *If F_n, F are distribution functions on $\mathbb{R}$, then F_n* ***converges weakly*** *to F if $F_n(x) \to F(x)$ for all continuity points x of F.*

If F_n converges weakly to F, we write $F_n \Rightarrow F$.

Definition 18.2 *Let $(\Omega, \mathcal{F}, P)$ be a probability space, X_n, X random variables on this space. We say that X_n* ***converges to X in probability*** *if for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\{|X_n - X| > \varepsilon\} = 0.$$

We write $X_n \xrightarrow{P} X$.

Definition 18.3 *We say that X_n* ***converges to X almost surely*** *if*

$$P\left\{\lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1.$$

Equivalently, X_n converges almost surely to X if for any $\varepsilon > 0$,

$$P\{|X_n - X| > \varepsilon \text{ i.o.}\} = 0.$$

Proposition 18.4 *If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{P} X$.*

However, the converse is **not** true! 2 examples:

- Let $(\Omega, \mathcal{F}, P) = ((0,1], \text{Borel sets}, \lambda)$, and let $X_1(\omega) = \delta_{(0,\frac{1}{2}]}$, $X_2(\omega) = \delta_{(\frac{1}{2},1]}$, $X_3(\omega) = \delta_{(0,\frac{1}{4}]}$, $X_4(\omega) = \delta_{(\frac{1}{4},\frac{2}{4}]}$, etc.

  $X_n \xrightarrow{P} 0$, but X_n does not converge almost surely to 0.

- Define a density

$$P(x) = \begin{cases} \dfrac{c}{|j|(\log(2+|j|))^2} & j \neq 0, \\ 0 & j = 0, \end{cases}$$

  where c is the constant that makes P a density. If $X_1, \dots$ are iid from P, then $P\{S_n/n > \varepsilon\} \to 0$, i.e. $S_n/n \xrightarrow{P} 0$, but because the mean does not exist, S_n/n does not converge to 0 almost surely.

Theorem 18.5 (Slutsky) *Suppose X_n, Y_n, Z are random variables such that $X_n \Rightarrow Z$ and $X_n - Y_n \Rightarrow 0$. Then $Y_n \Rightarrow Z$.*

Proof: Fix a continuity point x of Z. Given $\varepsilon > 0$, pick continuity points y' and y'' of Z such that $y' < x - \varepsilon < x < x + \varepsilon < y''$.

Assume that $\omega \in \{Y_n \leq x\}$. If $\omega \notin \{X_n \leq y''\}$, then

$$|X_n(\omega) - Y_n(\omega)| > y'' - x > (x + \varepsilon) - x = \varepsilon.$$

Thus, $\{Y_n \leq x\} \subseteq \{X_n \leq y''\} \cup \{|X_n - Y_n| > \varepsilon\}$. Similarly, we can obtain $\{X_n \leq y''\} \subseteq \{Y_n \leq x\} \cup \{|X_n - Y_n| > \varepsilon\}$. Thus

$$P\{X_n \leq y''\} - P\{|X_n - Y_n| > \varepsilon\} \leq P\{Y_n \leq x\} \leq P\{X_n \leq y''\} + P\{|X_n - Y_n| > \varepsilon\}.$$

Taking limits ($n \to \infty$),

$$P\{Z \leq y'\} \leq \liminf P\{Y_n \leq x\} \leq \limsup P\{Y_n \leq x\} \leq P\{Z \leq y''\}.$$

Since x is a continuity point of Z, then we can choose y' and y'' close to x. This means that

$$\liminf P\{Y_n \leq x\} = \limsup P\{Y_n \leq x\} = P\{Z \leq x\},$$

as required.

■

Note: We actually used Slutsky's Theorem in one of our card guessing examples (Lecture 16). Let

$$X_i = \begin{cases} 1 & \text{with probability } \frac{1}{i}, \\ 0 & \text{otherwise.} \end{cases}$$

If $S_n = \sum_{i=1}^{n} X_i$, then $\mathbb{E}S_n = \mu_n = \sum_{i=1}^{n} \frac{1}{i} \sim \log n$ and Var $S_n = \sigma_n^2 = \sum_{i=1}^{n} \frac{1}{i}\left(1 - \frac{1}{i}\right) \sim \log n$.

The Central Limit Theorem says that $P\left\{\frac{S_n - \mu_n}{\sigma_n} \leq x\right\} \to \Phi(x)$. We then made the leap to say that $P\left\{\frac{S_n - \log n}{\sqrt{\log n}} \leq x\right\} \to \Phi(x)$. We need Slutsky's Theorem to make that step rigorous.

18.2 Characteristic Functions & Fourier Analysis

Let μ be a probability on $\mathbb{R}^n$.

Definition 18.6 *The* ***characteristic function*** *or* ***Fourier transform*** *of μ at $t \in \mathbb{R}^n$ is*

$$\phi(t) := \mathbb{E}(e^{it \cdot X}) = \int_{\mathbb{R}^n} \cos(t \cdot x) + i \sin(t \cdot x) \mu(dx).$$

Here are some facts about characteristic functions:

1. ϕ exists for all t. (This is not the case for moment generating functions).

2. $\phi(0) = 1$.

3. If X and Y are independent, then

$$\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t).$$

Equivalently, if $X \sim \mu$ and $Y \sim \nu$ independent, then

$$\phi_{\mu * \nu}(t) = \phi_\mu(t) \cdot \phi_\nu(t).$$

4. $\phi_\mu(t) = \phi_\nu(t)$ for all t iff $\mu = \nu$.

5. (Continuity Theorem) If F_k, F are distribution functions on $\mathbb{R}^n$, then $F_k \Rightarrow F$ iff $\phi_k(t) = \phi(t)$ for all t.

18.2.1 Example: Card guessing

Let

$$X_i = \begin{cases} i & \text{with probability } \dfrac{1}{i}, \\ 0 & \text{with probability } 1 - \dfrac{1}{i}. \end{cases}$$

Let $S_n = \sum_{i=1}^{n} X_i$. Recall that in Lecture 16, we showed that the Lindeberg condition does not hold, and we noted that the CLT fails in this case. We will use characteristic functions to figure out the distribution of S_n.

First, let's compute the characteristic function for each X_j:

$$\begin{aligned} \phi_j(t) &= \mathbb{E}\left[e^{itX_j}\right] \\ &= \left(1 - \frac{1}{j}\right) + \frac{e^{itj}}{j} \\ &= 1 - \frac{1 - e^{itj}}{j}. \end{aligned}$$

Hence we can compute the characteristic function of S_n:

$$\begin{aligned} \phi_{S_n}(t) &= \prod_{j=1}^{n} \left[1 - \frac{1 - e^{itj}}{j}\right] \\ &= \exp\left\{\sum_{i=1}^{n} \log\left[1 - \frac{1 - e^{itj}}{j}\right]\right\} \\ &\approx \exp\left\{-\sum_{i=1}^{n} \frac{1 - e^{itj}}{j}\right\}. \end{aligned}$$

Next, we look at S_n/n instead so that the exponent becomes a Riemann sum:

$$\begin{aligned}\phi_{S_n/n}(t) &= \mathbb{E}\left[e^{it\frac{S_n}{n}}\right] \\ &= \phi_{S_n}\left(\frac{t}{n}\right) \\ &\approx \exp\left\{-\frac{1}{n}\sum_{i=1}^{n}\frac{1-e^{itj/n}}{j/n}\right\} \\ &\approx \exp\left\{-\int_0^1 \frac{1-e^{itx}}{x}dx\right\} =: F_\infty(t).\end{aligned}$$

Thus $P\left\{\frac{S_n}{n} \le x\right\} \to F_\infty(x)$.

18.2.1.1 Historical aside on $\mathcal{F}_\infty$

There are 3 classes of limits:

- Stable laws (limits of sums of iid random variables),
- Infinitely divisible laws (possible limits of triangular arrays), and
- Class L (limits of sums of independent but not necessarily identically distributed random variables).

In one of their books, Enedenko-Kolmogorov "proved" that distribution functions in Class L are unimodal. Kai-Lai Chung, in the process of translating the book, found that the proof was wrong. The proof used the fact that " if f and g are unimodal, then $f * g$ are unimodal", which is not true.

Ibraogimov published a counter-example, which involved F_∞. (F_∞ has a density which is flat between 0 and 1 and decreases thereafter.) Sun found that his proof was wrong.

Ibraogimov later published a proof that distributions in Class L are unimodal. Kantor found that Ibraogimov's proof wrong, and eventually Yamazato published a proof which was correct.

F_∞ appears in number theory and in analysis of the Fast Fourier Transform (FFT). Early versions of FFT (Cooley-Tukey algorithm) needed a running time of $f(n) = n\sum_{p|n} p$. What is the average running time of this algorithm? F_∞ appears in the analysis.

18.2.2 Skorohod's Theorem

Theorem 18.7 (Skorohod) *Let F_n, F be distribution functions on $\mathbb{R}$ such that $F_n \Rightarrow F$.*

Then there exist $(\Omega, \mathcal{F}, P)$ and random variables Y_n, Y with $P\{Y_n \le x\} = F_n(x)$ and $P\{Y \le x\} = F(x)$ such that $Y_n(\omega) \to Y(\omega)$ for all ω.

Proof: Let $(\Omega, \mathcal{F}, P) = ((0,1], \text{Borel sets}, \lambda)$.

For any ω, define $Y_n(\omega) = \inf\{x : \omega \le F_n(x)\}$, $Y(\omega) = \inf\{x : \omega \le F(x)\}$. We showed in a previous lecture that $P\{Y_n \le x\} = F_n(x)$, $P\{Y \le x\} = F(x)$.

Fix $\omega \in (0,1]$. Given $\varepsilon > 0$, choose x in the interval $(Y(\omega)-\varepsilon, Y(\omega))$ with x being a continuity point of F. $x < Y(\omega)$ implies that $F(x) < \omega$. Hence, $F_n(x) < \omega$ for sufficiently large n, i.e. $Y(\omega) - \varepsilon < x < Y_n(\omega)$ for sufficiently large n.

Taking lim infs on both sides, we get $Y(\omega) - \varepsilon \leq \liminf Y_n(\omega)$. Since this holds for all ε, $Y(\omega) \leq \liminf Y_n(\omega)$.

Similarly, if $\omega < \omega'$, $\limsup Y_n(\omega) \leq Y(\omega')$. If Y is continuous at ω, then we can let ω' approach ω, so $\limsup Y_n(\omega) \leq Y(\omega)$, which implies that $\lim Y_n(\omega) = Y(\omega)$.

Now consider what happens if Y is not continuous at ω. Since Y_n and Y are increasing functions on $(0,1]$, they can have at most countably many discontinuities. Hence, we can just set $Y_n(\omega) = Y(\omega) = 0$ at these points. ■

Corollary 18.8 *If $F_n \Rightarrow F$, then for every bounded continuous function f, $\int f dF_n \to \int f dF$.*

Proof: Let Y_n, Y be random variables such that $Y_n \sim F_n$, $Y \sim F$. Using Skorohod's Theorem and the Bounded Convergence Theorem,

$$\begin{aligned}\lim \int f dF_n &= \lim \mathbb{E}[f(Y_n)] \\ &= \mathbb{E}[f(Y)] \\ &= \int f dF.\end{aligned}$$

■

Corollary 18.9 *If $F_n \Rightarrow F$, then $\phi_n(t) \to \phi(t)$ for all t.*

Proof: For any t, let $f(X) = e^{itX}$ in the previous corollary. ■