36710 – 36752

ADVANCED PROBABILITY OVERVIEW

FALL 2020

LECTURE 19: WED, NOV 4, 2020

LAST TIME: THE CONTINUOUS MAPPING THEOREM. IF $\{X_n\}$ AND X TAKE VALUES ON METRIC SPACE $\mathcal{X}$ AND $g: \mathcal{X} \to \mathcal{Y}$ S.T. $P_n(X \in C_g) = 1$ WHERE $C_g = \{x \in \mathcal{X} : g \text{ IS CONTINUOUS AT } x\}$. THEN

$$X_n \overset{\Rightarrow}{\underset{a.s.}{\overset{P}{\rightarrow}}} X \quad \text{IMPLIES} \quad g(X_n) \overset{\Rightarrow}{\underset{a.s.}{\overset{P}{\rightarrow}}} X$$

SLUTSKY'S THEOREM IF $X_n \xrightarrow{D} X$ AND $Y_n \xrightarrow{D} c$ ← CONSTANT

THEN
1) $X_n + Y_n \xrightarrow{D} X + c$

HOLDS TRUE IF X_n AND Y_n ARE MATRICES ⟵ 2) $X_n \cdot Y_n \xrightarrow{D} X \cdot c$

3) $\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{c}$, $c \neq 0$

Claim IF $X_n \xrightarrow{D} X$ [IN $(\mathbb{R}^d, \mathcal{B}^d)$] AND F_X (c.d.f. OF X) IS CONTINUOUS, THEN

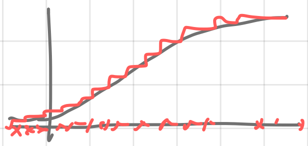
$$\sup_{x \in \mathbb{R}^d} |F_n(x) - F(x)| \to 0 \quad \text{AS } n \to \infty$$

UNIFORM CONVERGENCE

THM (GLIVENKO - CANTELLI) APPLICATION OF SLLN

LET $X_1, X_2, \ldots$ BE IID FROM A DISTRIBUTION OVER $(\mathbb{R}, \mathcal{B}^1)$ WITH c.d.f. F. FOR EACH n LET $F_n : \mathbb{R} \to [0,1]$ BE GIVEN BY:

$$x \longmapsto F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \le x\}$$



F_n: EMPIRICAL cdf (A RANDOM c.d.f.)

THEN

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0$$

REMARK: IT DOES NOT FOLLOW FROM SLLN, WHICH ONLY GIVES THAT

$$F_n(x) \xrightarrow{a.s.} \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \le x\}\right] = \underbrace{\frac{1}{n} \sum_{i=1}^{n} \Pr(X_i \le x)}_{F(x)}$$

$$= F(x)$$

FOR EACH FIXED x.

PF/ BY SLLN FOR EACH FIXED x, $F_n(x) \xrightarrow{a.s.} F(x)$ AND

$$F_n(x^-) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i < x\} \xrightarrow{a.s.} F(x^-) = \lim_{y \uparrow x} F(y)$$

$\forall \varepsilon > 0$ (SMALL) LET $-\infty = x_0 < x_1 < \ldots < x_k = \infty$ WHERE $k = k(\varepsilon)$ BE S.T. $\boxed{F(x_i^-) - F(x_{i-1})} < \varepsilon$ [POINTS AT WHICH HAS JUMP DISCONTINUITIES LARGER THAN ε ARE AMONG THE x_i'S]

FOR $x_{i-1} \le x < x_i$ WE HAVE

$$F_n(x) - F(x) \le F_n(x_i^-) - F(x_{i-1})$$

$$\le F_n(x_i^-) - F(x_i^-) + \varepsilon$$

SIMILARLY

$$F_n(x) - F(x) \ge F_n(x_{i-1}) - F(x_{i-1}) - \varepsilon$$

SO FOR ANY $x \in \mathbb{R}$

$$|F_n(x) - F(x)| \le \max_{i=1,\dots,k} \left\{ |F_n(x_i^-) - F(x_i^-)|, |F_n(x_{i-1}) - F(x_{i-1})| \right\} + \varepsilon$$

$$\le \varepsilon + \varepsilon = 2\varepsilon \qquad \text{a.s.}$$

SINCE $\varepsilon > 0$ IS ARBITRARY

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0 \qquad \blacksquare$$

LET $\mathbb{P}_n$ BE THE <u>EMPIRICAL MEASURE</u> ASSOCIATED WITH $X_1, \dots, X_n$ (→ SAMPLE) $\overset{iid}{\sim} F$

$$B \in \mathcal{B}^1 \longmapsto \mathbb{P}_n(B) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \in B\}$$

↓ RANDOM PROB. MEASURE

$\mathbb{E}[\mathbb{P}_n(B)] = \Pr(X \in B)$

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = \sup_{A \in \mathcal{A}} |\mathbb{P}_n(A) - P(A)|$$

$P(A) = \Pr(X \in A)$ PROB. DISTR. OF X

AND $\mathcal{A} = \{(-\infty, x], \ x \in \mathbb{R}\}$

<u>DKW INEQUALITY</u>: $\varepsilon > 0$

$$\Pr\left(\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \ge \varepsilon \right) \le 2 \exp\{-2n\varepsilon^2\}$$

THIS IMPLIES THE GLIVENKO CANTELLI THEOREM BY BOREL-CANTELLI

BECAUSE $\sum_n 2\exp\{-2n\varepsilon^2\} < \infty$ [SO PROB. THAT $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \ge \varepsilon$ i.o. IS ZERO].

IF ε IS OF ORDER $\sqrt{\frac{\log n}{2n}}$ THEN THEN

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| < \sqrt{\frac{\log n}{2n}} \quad \text{WITH PROB} \ge \frac{2}{n}.$$

MORE GENERALLY ONE CAN ADDRESS BOUNDING

$$\sup_{A \in \mathcal{A}} |P_n(A) - P(A)|$$

USING VC THEORY.

r_n IS POSITIVE $\uparrow$

O_p (BIG OH-P NOTATION) $\qquad X_n = o_p(r_n) \iff \frac{X_n}{r_n} \xrightarrow{P} 0$

RECALL THAT $x_n = O(r_n)$ IF $\exists\, C > 0$ S.T. $\left|\frac{x_n}{r_n}\right| < C$ FOR ALL n.

IF $\{X_n\}$ IS A SEQUENCE OF RV'S AND $\{r_n\}$ A SEQUENCE OF POSITIVE NUMBERS WE SAY THAT $X_n = O_p(r_n)$ WHEN $\forall \varepsilon > 0$, $\exists\, C = C(\varepsilon)$ S.T.

$$P_n(|X_n| \geq C r_n) \leq \varepsilon \quad \text{ALL } n.$$

IF $r_n = 1$ ALL n, $X_n = O_p(1)$ MEANS THAT X_n IS A TIGHT SEQUENCE OR A SEQUENCE BOUNDED IN PROB.

- IF $\{X_n\}$ ARE RANDOM VECTORS, THEN $X_n = O_p(r_n)$ WHEN $\|X_n\| = O_p(r_n)$

<u>EXAMPLE</u> $X_1, X_2, \ldots$ iid WITH $\mathbb{E}[X_1] = \mu$ AND $\mathrm{Var}[X_1] = \sigma^2$

THEN $\frac{S_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{P} \mu$

OR, EQUIVALENTLY, $\frac{S_n}{n} = \mu + o_p(1)$

IN FACT WE HAVE ALSO THAT

$$\Pr\left(\sqrt{n}\left|\frac{S_n}{n} - \mu\right| > c\right) \leq \frac{\sigma^2}{c^2}$$

SO, FOR EACH $\varepsilon > 0$, PICK C LARGE ENOUGH S.T.

$$P_n\left(\sqrt{n}\left|\frac{S_n}{n} - \mu\right| > C\right) < \varepsilon$$

WHICH IMPLIES THAT $\sqrt{n}\left(\frac{S_n}{n} - \mu\right) = O_p(1)$ OR

$$\frac{S_n}{n} = \mu + \underbrace{O_p\left(\frac{1}{\sqrt{n}}\right)}_{\text{ALSO } o_p(1)} \quad \text{AS } n \to \infty$$

- O_p / o_p CALCULUS

i) $X_n = o_p(1) \implies X_n = O_p(1)$

ii) $O_p(1) \pm O_p(1) = O_p(1)$

$o_p(1) \pm o_p(1) = o_p(1)$

iii) $O_p(1) \times o_p(1) = O_p(o_p(1)) = o_p(1)$

iv) $O_p(1) \times O_p(1) = O_p(1)$

$o_p(1) \times o_p(1) = o_p(1)$

v) $(1 + o_p(1))^{-1} = O_p(1)$

$X_n = O_p(r_n)$ AND $r_n \to 0$ THEN $X_n = o_p(1)$

<u>Lemma</u> LET $f: \mathbb{R}^d \to \mathbb{R}$ S.T. $f(0) = 0$. LET $\{X_n\}$ BE A SEQUENCE OF R.V.'S SUCH THAT X_n IS IN THE DOMAIN OF f EVENTUALLY AND $X_n \xrightarrow{P} 0$. THEN

i) IF $f(h) = o(\|h\|^p)$ AS $h \to 0$ ANY $p \geq 1$

THEN $f(X_n) = o_p(\|X_n\|^p)$

ii) IF $f(h) = O(\|h\|^p)$ AS $h \to 0$ ANY $p \geq 1$,

$f(X_n) = O_p(\|X_n\|^p)$

Pf/ LET $g(h) = \frac{f(h)}{\|h\|^p}$ FOR $h \neq 0$ AND $g(0) = 0$

THEN $f(X_n) = g(X_n) \|X_n\|^p$

i) SINCE g IS CONTINUOUS AT 0 AND $g(X_n) \xrightarrow{P} g(0) = 0$

BY CONTINUOUS MAPPING THEOREM.

ii) BY ASSUMPTION $\exists\, C > 0$ AND $\delta > 0$ S.T. $|g(h)| \leq C$

WHENEVER $\|h\|^p \leq \delta$. SO

$$P_n\left(|g(X_n)| > C\right) \leq P_n\left(\|X_n\|^p > \delta\right) \to 0. \text{ SO}$$

$$\frac{f(X_n)}{\|X_n\|^p} = g(X_n) = O_p(1) \quad \blacksquare$$

THIS ALLOWS US TO STATE AND PROVE THE <u>DELTA METHOD</u>

LET $\{X_n\}$ BE A SEQUENCE OF RV'S IN $\mathbb{R}^d$ AND $\{r_n\}$ A SEQUENCE

S.T. $r_n \to \infty$ WHERE

$$r_n (X_n - \theta) \xrightarrow{D} X \qquad \text{SOME } \theta.$$

LET $f: \mathbb{R}^d \to \mathbb{R}^k$.

QUESTION: DOES

$$r_n (f(X_n) - f(\theta)) \quad \text{CONVERGE IN DISTR.?}$$

ANSWER: YES, IF f IS DIFF. AT θ.

RECALL THAT f IS DIFF. AT θ IF THERE EXISTS A LINEAR MAPPING

(A MATRIX) $\underset{\substack{k \times d \\ \text{MATRIX}}}{f'(\theta)} : \mathbb{R}^d \to \mathbb{R}^k$ S.T.

$$f(\theta+h) = f(\theta) + \underbrace{f'(\theta)(h)}_{f'(\theta)\cdot h} + o(\|h\|) \quad \text{AS } h \to 0$$

OR $$\frac{\| f(\theta+h) - f(\theta) - f'(\theta)(h) \|}{\|h\|} \to 0 \quad \text{AS } h \to 0$$

$$f = \begin{bmatrix} f_1 \\ \vdots \\ f_k \end{bmatrix}$$

[IF $\frac{\partial f_i}{\partial x_j}$ EXISTS IN NEIGHBORHOOD OF θ AND CONTINUOUS THERE

THEN $$\underset{k \times d}{f'(\theta)} = \left(\frac{\partial f_i}{\partial x_j} \right)_{\substack{i=1,\dots,k \\ j=1,\dots,d}}$$

AND $$f'(\theta)(h) = \underset{k \times d}{f'(\theta)} \cdot \underset{d \times 1}{h}$$

IF $k=1$ THEN $f'(\theta)$ IS TRANSPOSE OF GRADIENT OF f AT θ.