36710 - 36752

ADVANCED PROBABILITY OVERVIEW

FALL 2020

LECTURE 22 : MON , NOV 16 , 2020

LAST TIME: BERRY-ESSEEN CLT

ASSUME A TRIANGULAR ARRAY SETTING $X_{n,k} \sim (\mu_{n,k}, \sigma^2_{n,k})$

INDEPENDENT WITHIN EACH ROW OF THE ARRAY.

LET

$$Y_k = \frac{X_{n,k} - \mu_{n,k}}{\left(\sum_{k=1}^{r_n} \sigma^2_{n,k}\right)^{1/2}} \quad \text{AND} \quad W_{r_n} = \sum_{k=1}^{r_n} Y_k \sim (0,1)$$

THEN

SMALL UNIVERSAL CONSTANT $< 1/2$

$$\sup_{x \in \mathbb{R}} \left| P_n(W_{r_n} \le x) - P_n(\underset{N(0,1)}{Z} \le x) \right| \le C \frac{\sum_{k=1}^{r_n} \mathbb{E}\left[|X_{n,k} - \mu_{n,k}|^3\right]}{\left(\sum_{k=1}^{r_n} \sigma^2_{n,k}\right)^{3/2}}$$

IF $X_{n,k} = X_k \overset{iid}{\sim} (0, \sigma^2)$ S.T. $\mathbb{E}[|X|^3] = \mu_3$

ADDITIONAL 3rd MOMENT ASSUMPTION

THE BERRY-ESSEEN BOUND IS

$$C \frac{n\mu_3}{(n\sigma^2)^{3/2}} = C \frac{\mu_3}{\sqrt{n}\,\sigma^3} \to 0 \quad \text{AS } n \to \infty \text{ LIKE } \frac{1}{\sqrt{n}}$$

EXAMPLE: $X_1, X_2, \ldots \overset{iid}{\sim} \text{Bernoulli}(p)$ $\quad \sum_{i=1}^{n} X_i \sim \text{Bin}(n, p)$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \leftarrow \frac{\bar{X}_n - p}{\sqrt{n p(1-p)}} \xrightarrow{D} N(0,1) \quad \text{FOR EACH } p \quad \text{POINTWISE RESULT}$$

(X)

LF CONDITION: $\frac{1}{p(1-p)} \mathbb{E}\left[(X-p)^2 \, 1\{|X-p| > \varepsilon\sqrt{np(1-p)}\}\right] \to 0$

WHAT IF WE WANT CONVERGENCE UNIFORMLY IN p?

LF CONDITION $(X) \le \frac{1}{p(1-p)} \max\{p^2, (1-p)^2\} \, \mathbb{P}\left(|X-p| > \underline{\varepsilon\sqrt{np(1-p)}}\right)$

SO SINCE UNIFORMITY IN p MEANS WE CAN ALLOW p TO CHANGE WITH n, WE CONSIDER SEQUENCE OF p's $\{p_n\}$ AND OBTAIN FROM THE LF CONDITION THAT

$$\sqrt{n}\,\frac{\bar{X}_n - p_n}{\sqrt{p_n(1-p_n)}} \xrightarrow{D} N(0,1)$$

AS LONG AS $\underline{n\, p_n(1-p_n) \to \infty}$

TO GET A CONVERGENCE RATE, USE BERRY-ESSEEN BOUND:

LET ε_n BE S.T. $p_n \in [\varepsilon_n, 1-\varepsilon_n]$ AND $\varepsilon_n \to 0$

THEN

WRITE $p = p_n$

$$\mathbb{E}\left[|X-p|^3\right] = p(1-p)\underbrace{\left[(1-p)^2 + p^2\right]}_{<1} < p(1-p)$$

SO

$$\frac{\mathbb{E}\left[|X-p|^3\right]}{(p(1-p))^{2/3}} \le \frac{1}{\sqrt{p(1-p)}} \le \frac{1}{\sqrt{\varepsilon_n(1-\varepsilon_n)}}$$

AND THE BARRY-ESSEEN BOUND:

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n - np_n}{\sqrt{n\, p_n(1-p_n)}} \le x\right) - \mathbb{P}(Z \le x)\right| \le C\,\frac{1}{\sqrt{n\,\varepsilon_n(1-\varepsilon_n)}}$$

WHICH VANISHES AS LONG AS $n \varepsilon_n (1-\varepsilon_n) \to \infty$.

IF $\varepsilon_n = \frac{1}{\sqrt{n}}$ OR $n^{-\alpha}$ $\alpha \in (0,1)$

MULTIVARIATE CLT

CONSIDER A TRIANGULAR ARRAYS OF $\overset{\text{CENTERED}}{\downarrow}$ RANDOM VECTORS IN $\mathbb{R}^d$

$$Y_{n,1}, Y_{n,2}, \dots Y_{n,r_n}$$

THAT ARE ASSUMED INDEPENDENT AND S.T.

$$\sum_{k=1}^{r_n} \mathbb{E}\left[\|Y_{n,k}\|^2 \mathbb{1}\{\|Y_{n,k}\| > \varepsilon\} \right] \to 0 \quad \forall \varepsilon > 0$$

AND

$$\sum_{k=1}^{r_n} \operatorname{COV}[Y_{n,k}] \to \underset{d \times d}{\Sigma}$$

THEN

$$\sum_{k=1}^{r_n} Y_{n,k} \xrightarrow{D} N(0, \Sigma)$$

EXAMPLE: LINEAR REGRESSION WITH DETERMINISTIC COVARIATES

$$\underset{n\times 1}{Y} = X\beta + \varepsilon$$

X: $n \times d$ DETERMINISTIC MATRIX

β: $d \times 1$ VECTOR OF UNKNOWN REGRESSION COEFF.

ε: $n \times 1$ VECTOR CONSISTING OF iid $(0, \sigma^2)$

WE OBSERVE (Y, X) AND ESTIMATE β WITH

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

ASSUME THAT X IS OF FULL RANK d.

$$\mathbb{E}[\hat{\beta}] = \underbrace{(X^T X)^{-1} X^T X}_{I}\beta + \underbrace{\mathbb{E}[(X^T X)^{-1} X^T \varepsilon]}_{=0} = \beta$$

ALSO $\quad \mathrm{cov}[\hat{\beta}] = \sigma^2 (X^T X)^{-1} \cancel{X^T X} \cancel{(X^T X)^{-1}} = \sigma^2 (X^T X)^{-1}$

TO GET A CLT FOR $\hat{\beta}$:

$$(X^T X)^{1/2}(\hat{\beta} - \beta) = \underbrace{(X^T X)^{-1/2} X^T}_{A_n} \varepsilon = A_n \varepsilon = \sum_{i=1}^{n} A_{n,i}\, \varepsilon_i$$

$\downarrow$

$A_{n,i}$: ith COLUMN OF A_n

SO LF CONDITION IS

$$\underbrace{\sum_{i=1}^{n} \|A_{n,i}\|^2\, \mathbb{E}\left[\varepsilon_i^2\, \mathbb{1}\{\|A_{n,i}\|\, |\varepsilon_i| > \eta\}\right]}_{(*)} \to 0$$

FOR EACH $\eta > 0$

THIS IS SATISFIED IF

$\max_i \|A_{n,i}\| \to 0$ $\quad$ BECAUSE $(*)$ IS BOUNDED BY

$$d \max_i \mathbb{E}\left[\varepsilon_i^2\, \mathbb{1}\{\|A_{n,i}\|\, |\varepsilon_i| > \eta\}\right]$$

BECAUSE $\quad \sum_{i=1}^{n} \|A_{n,i}\|^2 = \mathrm{Tr}(A_n A_n^T) = \mathrm{tr}(I_d) = d$

FOR EXAMPLE, IF $d = 2$ $\qquad Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$i = 1, \dots, n \qquad x_i \in \mathbb{R}$

$\hookrightarrow$ DETERMINISTIC

$(X^T X)^{-1/2} X^T$

THEN

$$A_n = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & \bar{X}_n \\ \bar{X}_n & \bar{X}_n^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ x_1 & x_2 & \dots & x_{n-1} & x_n \end{bmatrix}$$

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$

SO LF CONDITION HOLDS IF $\bar{X}_n^2$ IS BOUNDED

AND $\quad \max_i |x_i| = o(\sqrt{n})$

SO BY CLT, UNDER CONDITIONS ON A_n,

$$(X^T X)^{1/2} (\hat\beta - \beta) \xrightarrow{D} N(0, \sigma^2 I_d)$$

IF $\sigma^{-2} \frac{\overset{d\times n}{X^T}\overset{n\times d}{X}}{n} \to \underset{d\times d}{\Sigma}$ AS $n \to \infty$

THEN, BY SLUTSKY'S THEOREM,

$$\sqrt{n}(\hat\beta - \beta) \xrightarrow{D} N(0, \underset{\sigma^{-2}\left(\frac{X^T X}{n}\right)^{-1}}{\underset{\approx}{\Sigma}})$$

<u>REMARK</u>. WE WANT TO SHOW JUST CONSISTENCY OF $\hat\beta$. THEN THE CALCULATIONS ARE SIMPLER. ASSUMING $\sigma^{2} \frac{X^T X}{n} \to \Sigma$

THEN

$$\hat\beta - \beta = \left(\frac{X^T X}{n}\right)^{-1} \underline{\frac{X^T \varepsilon}{n}} \xrightarrow{P} 0$$

BECAUSE $\frac{X^T\varepsilon}{n} \xrightarrow{P} 0$ AND SLUTSKY'S THEOREM

$\Downarrow$

BECAUSE $\operatorname{COV}\left(\frac{X^T\varepsilon}{n}\right) = \frac{\sigma^2}{n} \underbrace{\frac{X^T X}{n}}_{\to \Sigma} \to 0$

BERRY-ESSEEN BOUNDS FOR THE MULTIVARIATE CLT

$$X_1, X_2, \ldots \overset{iid}{\sim} (\mu, \Sigma)$$

LET $\mathcal{A}$ BE A COLLECTION OF SUBSETS OF $\mathbb{R}^d$. A HIGH-DIM CLT STATEMENT LOOK LIKE

$$\underline{e_n(\mathcal{A})} = \sup_{A \in \mathcal{A}} \left| P_n\left(\sqrt{n}(\bar X_n - \mu) \in A\right) - P_n(Z \in A) \right| \le \text{SOME BOUND!}$$

$Z \sim N(0, \Sigma)$

WHEN $d=1$ THEN WE TOOK $\mathcal{A} = \{(-\infty, x],\ x \in \mathbb{R}\}$

PUNCHLINE: WHEN $d>1$ THE BOUND WE OBTAIN DEPENDS ON $\mathcal{A}$!!

BENTKUS (2003) $\mu = 0$ $\Sigma = I_d$ $\mathcal{A}$ = WELL BEHAVED

$$e_n(\mathcal{A}) \le C(d, \mathcal{A}) \frac{\mathbb{E}\left[\|X_1\|^3\right]}{\sqrt{n}}$$

WHERE $C(d, \mathcal{A})$ IS GAUSSIAN ISOPERIMETRIC CONSTANT FOR $\mathcal{A}$

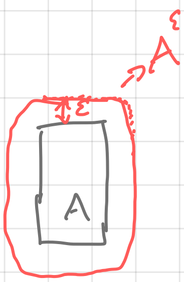
NOTICE $\mathbb{E}\left[\|X_1\|^3\right] = \mathbb{E}\left[\left(\|X_1\|^2\right)^{3/2}\right]$

$$\underset{\text{BY JENSEN}}{\ge} \left(\mathbb{E}\left[\|X_1\|^2\right]\right)^{3/2} = d^{3/2}$$

WHAT ABOUT $C(d, \mathcal{A})$? THIS CONSTANT IS SUCH THAT

$$\Pr\big(\underset{\sim N(0, I_d)}{Z} \in A^{\varepsilon} \setminus A\big) \le C(d, \mathcal{A})\,\varepsilon \qquad \forall A \in \mathcal{A}$$

$$\Pr\big(Z \in A \setminus A^{-\varepsilon}\big) \le C(d, \mathcal{A})\,\varepsilon \qquad \forall \varepsilon > 0$$



WHERE FOR ANY SET A AND $\varepsilon > 0$

$$A^{\varepsilon} = \{x \in \mathbb{R}^d : d(x, A) \le \varepsilon\}$$

$$A^{-\varepsilon} = \{x \in A : \underset{\substack{\Downarrow \\ y:\ \|x-y\| < \varepsilon}}{B(x, \varepsilon)} \subset A\}$$

IF $\mathcal{A}$ IS THE CLASS OF EUCLIDEAN BALLS, THEN $C(d, \mathcal{A}) =$ CONSTANT!

$\mathcal{A}$ IS THE CLASS OF CONVEX SETS THEN $C(d, \mathcal{A}) \asymp d^{1/4}$

$\Downarrow$

LEADS TO A BOUND ON $e_n(\mathcal{A})$ OF ORDER $\sqrt{\frac{d^7}{n}}$

IT TURNS OUT THAT IF $\mathcal{A}$ IS THE CLASS OF HYPER-RECTANGLES

THEN ASSUMING BOUNDED RANDOM VECTORS YOU CAN GET A

BOUND OF ORDER $\dfrac{\log^{\alpha} d \sqrt{\log n}}{\sqrt{n}}$ SOME $\alpha > 0$

$\downarrow$ VANISH EVEN IF $d >> n$.

CHERNOZUKOV & CO-AUTHORS (2013)

BEST RATE BY CUCHIBOTLA AND RINALDO (2020)

CONDITIONAL EXPECTATION

Def LET $(\Omega, \mathcal{F}, P)$ BE A PROBABILITY SPACE AND $\mathcal{C} \subseteq \mathcal{F}$ A SUB-σ-FIELD. LET X BE A RV THAT IS $\mathcal{F}/\mathcal{B}^1$ MEAS. S.T. $\mathbb{E}|X| < \infty$. LET $\mathbb{E}[X | \mathcal{C}]$ STANDS FOR ANY FUNCTION $h : \Omega \to \mathbb{R}$ THAT IS $\mathcal{C}/\mathcal{B}^1$ MEAS.

S.T.

$$\int_C h \, dP = \int_C X \, dP \qquad \text{FOR ALL } C \in \mathcal{C}.$$

WE CALL SUCH A FUNCTION h A VERSION OF THE CONDITIONAL EXPECTATION OF X GIVEN $\mathcal{C}$.