Chapter 2.

Elementary asymptotics

Vdv Chap. 2 and 3.

Setup $\{x_1, x_2, \ldots, x_n, \cdots\}$

$$X_i : (\Omega, \mathcal{F}, P) \longmapsto (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$$

Def. [a.s. & in prob]

A.S.: $P(\lim_{n\to\infty} \|X_n - X\| = 0) = 1$

$$\Updownarrow$$

for P-almost every $w \in \Omega$,

$$\lim_{n\to\infty} \|X_n(w) - X(w)\| = 0$$

Written as $X_n \xrightarrow{a.s.} X$

In prob: $\forall \varepsilon > 0$,

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

Written as $X_n \xrightarrow{P} X$

Thm. (i) If $X_n \xrightarrow{a.s.} X$ and

$$X_n \xrightarrow{a.s.} \tilde{X}.$$

Then $X \overset{a.s.}{=} \tilde{X}$

(ii) If $X_n \xrightarrow{P} X$ and

$$X_n \xrightarrow{P} \tilde{X}.$$

then $X \overset{a.s.}{=} \tilde{X}$.

(iii)
$$\begin{array}{c} X_n \xrightarrow{a.s.} X \\ \Downarrow \\ X_n \xrightarrow{P} X \end{array}$$

Def. [weak convergence, convergence in dist.]

An $\mathbb{R}^d$-valued RV X_n is said weakly converge to (some RV) X

if $\forall$ bounded continuous func.

$$f: \mathbb{R}^d \longmapsto \mathbb{R}$$

we have

$$\mathbb{E}[f(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[f(X)]$$

We write $X_n \Rightarrow X$

$(X_n \xrightarrow{d} X, \ X_n \rightsquigarrow X)$ Jon

Thm: [Portmanteau]

Consider $[X_n]$ to a seq. of RVs, and X to be one RV.

The following are equivalent.

(i) $\forall$ b.c. func f,

$$\mathbb{E}[f(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[f(X)]$$

(ii) $\forall$ continuity points $t \in \mathbb{R}^d$ of $P(X \le \cdot)$

$$P(X_n \le t) \xrightarrow{n\to\infty} P(X \le t)$$

(iii)

⋮ Lecture note

(VII)

(IX) Levy's continuity theorem:

$\forall t \in \mathbb{R}^d$,

$$\mathbb{E}[\exp(i \cdot t^\top X_n)] \xrightarrow{n\to\infty} \mathbb{E}[\exp(i \cdot t^\top X)]$$

(X) Cramer-Wald device

$\forall t \in \mathbb{R}^d$,

$$t^\top X_n \Rightarrow t^\top X$$

[Proof lefts to us]

Thm. (identifiasility)

If $X_n \Rightarrow X$ & $X_n \Rightarrow \tilde{X}$,

then $X \overset{d}{=} \tilde{X}$.

CMT

Thm (Continuous mapping Thm.)

Consider $g: \mathbb{R}^d \to \mathbb{R}^m$

s.t. g is continuous (no need differentible like abs)

only on support of X.

Then

(i) $X_n \Rightarrow X$ imply $g(X_n) \Rightarrow g(X)$

(ii) $X_n \xrightarrow{P} X$ implies $g(X_n) \xrightarrow{P} g(X)$

(iii) $X_n \xrightarrow{a.s.} X$ implies $g(X_n) \xrightarrow{a.s.} g(X)$

(iv) If $X_n \Rightarrow X$ and "$\exists$ a couple of $\{X_n\}$, $\{Y_n\}$ s.t.

$\|X_n - Y_n\| \xrightarrow{P} 0$"

then $Y_n \Rightarrow X$

(v) If $X_n \Rightarrow X$ and $Y_n \xrightarrow{P} c$

Then $(X_n, Y_n) \Rightarrow (X, c)$

Proof: Left to us.

Thm. (Slutsky Thm.)

$\{X_n\}$ is $\mathbb{R}^d$-valued seq of RVs,

and $X_n \Rightarrow X$.

(i) If $\{Y_n\}$ is $\mathbb{R}^d$-valued seq of RVs s.t.

$\boxed{Y_n \xrightarrow{P} c}$, then

$$X_n + Y_n \Rightarrow X + c$$

(ii) If $\{Y_n\}$ is $\mathbb{R}^d$-valued seq. ~~st~~. $Y_n \xrightarrow{P} c$,

$$Y_n \cdot X_n \Rightarrow c \cdot X$$

(iii) If $\{Y_n\}$ is - - - - - - s.t. $Y_n \xrightarrow{P} c \ (c \neq 0)$,

then
$$\frac{X_n}{Y_n} \Rightarrow X/c$$

Proof: CMT.

LLN & CLT

If $X_1, \dots, X_n \overset{iid}{\sim} P$ of support in $\mathbb{R}$

fixed

then ① if $\mathbb{E}[X_1] < \infty$

(WLLN) $\bar{X}_n \xrightarrow{P} \mathbb{E}_P[X]$

(SLLN) $\bar{X}_n \xrightarrow{a.s.} \mathbb{E}_P[X]$

② if $\mathbb{E}[X_i^2] < \infty$, then

(CLT) $\sqrt{n}(\bar{X}_n - \mathbb{E}_P[X]) \Rightarrow N(0, Var_P[X])$

Example [Asymptotic Normality of t-statistics]

ASN

" $X_1, X_2, \dots, X_n \overset{iid}{\sim} P$ s.t. $\mathbb{E}_P[X_i^2] < \infty$ "

$$\left[\frac{\bar{X}_n - \mathbb{E}_P(X)}{S_n} \right] \leftarrow \hat{t}_n$$

where $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

Goal: $\sqrt{n}\,\hat{t}_n \Rightarrow N(0, 1)$

Proof: CLT + Slutsky

① $\sqrt{n} \frac{\bar{X}_n - \mathbb{E}_P[X]}{\sigma} \Rightarrow N(0, 1)$ which is the CLT.

$[\sigma^2 := Var_P[X]]$

② If $\sigma / S_n \xrightarrow{P} 1$, we are done.

Or equivalently [← think about why]

$$S_n^2 \xrightarrow{P} \sigma^2 \quad \text{(CMT)}$$

③ NTS: $S_n^2 \xrightarrow{P} \sigma^2$

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} [X_i - \bar{X}_n]^2$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} X_i^2 - n \bar{X}_n^2 \right]$$

$$= \frac{n}{n-1} \left[\boxed{\frac{1}{n} \sum_{i=1}^{n} X_i^2} - \boxed{\bar{X}_n^2} \right]$$

$$\frac{n}{n-1} \downarrow \qquad \frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow[\text{WLLN}]{P} \qquad \bar{X}_n^2 \xrightarrow{\text{WLLN + CMT}}$$

$$S_n^2 \xrightarrow{P} 1 \cdot \underbrace{\left[\mathbb{E}_p[X^2] - [\mathbb{E}_p(X)]^2 \right]}_{= \mathrm{Var}_p[X]}$$

We have shown

$$S_n^2 \xrightarrow{P} \mathrm{Var}_p[X]$$

Q: How fast is S_n^2 converging to $\mathrm{Var}_p[X]$???

A: In short, under $\mathbb{E}[X^4] < \infty$,

S_n^2 is converging to var in $\frac{1}{\sqrt{n}}$ rate.

Def [Big-O and small-o]

Consider $\{x_n\}_{n=1}^{\infty}$ and $\{r_n\}_{n=1}^{\infty}$ to be 2 $\mathbb{R}$-valued seq. of fixed numbers.

(i) $x_n = O(r_n)$ iff

$$\limsup_{n\to\infty} \left|\frac{x_n}{r_n}\right| < \infty$$

(ii) $x_n = o(r_n)$ iff

$$\limsup_{n\to\infty} \left|\frac{x_n}{r_n}\right| = 0$$ (translate using ε-δ)

Def [Big-O_p and small-o_p]

Consider $\{X_n\}$ and $\{R_n\}$ to be 2 seq. of $\mathbb{R}$-valued RVs.

(i) $X_n = O_p(R_n)$ iff $\forall \varepsilon > 0$, $\exists$ some $M = M_\varepsilon > 0$ s.t.

$$\liminf \mathbb{P}[|X_n| \leq M_\varepsilon \cdot |R_n|] > 1 - \varepsilon$$

(ii) $X_n = o_p(R_n)$ iff $\forall M > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n| \leq M |R_n|) = 1$$

[Good exercise]

$$X_n = o_p(1) \iff X_n \xrightarrow{P} 0$$

<u>Remark</u>: we can extend the above def. to $\mathbb{R}^d$.

Thm [Prokhorov] Consider $\{X_n\}$ to be a seq. of $\mathbb{R}^d$-valued RVs.

(i) If $X_n \Rightarrow$ (some) X, then

$$X_n = O_p(1)$$

(ii) If $X_n = O_p(1)$, then $\exists$ a subseq. $\{n_j\}$ of $\{n\}$

$$X_{n_j} \Rightarrow \text{(some) } X$$

[If $X_n = O_p(1)$, we say X_n is uniformly tight]

Thm. [Operations on $O_p(\cdot)$ and $o_p(\cdot)$]

(i) $O_p(1) \cdot O_p(1) = O_p(1)$

⋮

Example [sample mean]

$X_1, \dots, X_n \overset{iid}{\sim} P$

$\mathbb{E}_P[X^2] < \infty$

$\Longrightarrow \sqrt{n}\,(\bar{X}_n - \mathbb{E}_P[X]) \Rightarrow N(0, \sigma^2)$

$\Longrightarrow \sqrt{n}\,(\bar{X}_n - \mathbb{E}_P[X]) \overset{\text{Prokhorov}}{=} O_p(1)$

$\Longrightarrow \bar{X}_n - \mathbb{E}_P[X] = O_p[1/\sqrt{n}]$

$= o_p(1)$

Ex. [sample variance]

$X_1, \dots, X_n \overset{iid}{\sim} P$

$\mathbb{E}_P[X^4] < \infty$

$Y_n := \frac{1}{n}\sum_{i=1}^{n} X_i^2$

$S_n^2 = \frac{n}{n-1}\left(Y_n - \bar{X}_n^2\right)$

$\Longrightarrow Y_n = \mathbb{E}_P[X^2] + O_p(1/\sqrt{n})$

$\bar{X}_n = \mathbb{E}_P[X] + O_p(1/\sqrt{n})$

$\Longrightarrow \bar{X}_n^2 = [\mathbb{E}_P(X)]^2 + O_p(1/\sqrt{n})$

$\Longrightarrow S_n^2 = (1 + O(\frac{1}{n}))\,(\mathbb{E}_P(X^2) - [\mathbb{E}_P(X)]^2 + O_p(1/\sqrt{n}))$

$\Longrightarrow S_n^2 = \mathrm{Var}_P(X) + \underline{O_p(1/\sqrt{n})}$ don't necessary best.

O_p: Stochastically bounded. Best: minimax rate of convergence

Thm. [$\mathbb{R}^d$- CLT]

$X_1, \dots, X_n \overset{iid}{\sim} P$ supported over $\mathbb{R}^d$

$\mathbb{E}_P\left[\|X\|^2\right] < \infty$

$$\Longrightarrow \sqrt{n}(\bar{X}_n - \mu) \Rightarrow \mathcal{N}(0, \Sigma)$$

$\mu = \mathbb{E}X,\ \Sigma = \mathrm{Cov}[X]$

$\quad = \mathbb{E}[(X-\mu)\cdot(X-\mu)^T] \in \mathbb{R}^{d\times d}$

Proof. Cramer-Wald device $+\ \mathbb{R}^1$-CLT

We have $\forall t \in \mathbb{R}^d$,

$$\sqrt{n}\,\big(\underbrace{t^T\bar{X}_n}_{\frac{1}{n}\sum_{i=1}^{n}(t^T X_i)} - t^T\mu\big) \Rightarrow \mathcal{N}(0, t^T\Sigma t)$$

Thm. [Lindeberg-Feller Thm.]

Setup: Traingular-array.

$n=1$	X_{11}			
$n=2$	X_{21}	X_{22}		
$n=3$	X_{31}	X_{32}	X_{33}	
	$\vdots$			
$n=n$	X_{n1}	X_{n2}	$\cdots$	X_{nn}

① $\{ X_{n_i} := i = 1, \dots, n \}$ are $\mathbb{R}$-valued and indep
mutually.

② $\mu_{n_i} := \mathbb{E}[X_{n_i}] < \infty$, (May not identically distributed)
$\sigma_{n_i}^2 = Var[X_{n_i}] < \infty$.

Then, define

$$\underset{\text{pretty small}}{\longleftarrow} Y_{n_i} := \frac{X_{n_i} - \mu_{n_i}}{\sqrt{\sum_{i=1}^{n} \sigma_{n_i}^2}} \longrightarrow \text{very large}$$

③ Lindeberg cond. :

$$\forall \varepsilon > 0, \sum_{i=1}^{n} \mathbb{E}\left[Y_{n_i}^2 \mathbb{1}\{ |Y_{n_i}| \geq \varepsilon \} \right] \xrightarrow{n\to\infty} 0$$

Then,

$$\sum_{i=1}^{n} Y_{n_i} \Rightarrow \mathcal{N}(0, 1)$$

Goal today: Asymptotic Normality of OLS.

Often random design can be truncted down to

Thm. [ASN for simple LR under fixed design]

Setup: $\underbrace{Y_i}_{\text{response}}^{\in \mathbb{R}} = \underbrace{\overset{\in\mathbb{R}}{\beta_0} + \beta_1^T}_{\text{coef.}} \underbrace{x_i}_{\text{covariates}}^{\mathbb{R}^d} + \underbrace{\varepsilon_i}_{\text{random noise}}^{\mathbb{R}}$

ε_i are IID with $\mathbb{E}\,\varepsilon_i^2 < \infty$.

Remark: $Y_i = Y_i^{(n)}$ (or Y_{ni}) $\quad d = d_n$

Goal: to derive a necessary & sufficient cond. for $\hat{\beta}_n := (\mathcal{X}_n^T \mathcal{X}_n)^{-1} \mathcal{X}_n^T Y$ with

$$\mathcal{X}_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \\ x_1 & x_2 & \cdots & x_n \end{pmatrix}^T \in \mathbb{R}^{n \times (d+1)}$$

to be ASN.

$$\frac{(\mathcal{X}_n^T \mathcal{X}_n)^{1/2} (\hat{\beta}_n - \beta)}{\sigma} \Longrightarrow \mathcal{N}(0, I_d)$$

$\sigma := \sigma^{(n)} = \sqrt{\operatorname{Var}(\varepsilon_i)}$

Claim: Define

$$H = (H^{(n)}) = X_n (X_n^T X_n)^{-1} X_n^T$$

$$= \begin{pmatrix} H_{11} & H_{12} & \cdots & H_{1n} \\ H_{21} & H_{22} & \cdots & H_{2n} \\ \vdots & & & \\ H_{n1} & H_{n2} & \cdots & H_{nn} \end{pmatrix}$$

with H_{jj}'s called the leverage score for the j-th sample point.

Then, the "iff" cond. for ASN of $\hat{\beta}_n$ is

$$\max_{i=1,\dots,n} H_{ii}^{(n)} \xrightarrow{n\to\infty} 0$$

Proof. Cramer-Wald
+
L-F CLT
+
DCT (Dominated convergence Thm.)

① $\hat{\beta}_n = \beta + (X_n^T X_n)^{-1} X_n^T \vec{\varepsilon}_n$

w. $\vec{\varepsilon}_n = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$.

② $\hat{\beta}_n - \beta = \underbrace{(X_n^\top X_n)}^{-1} X_n^\top \vec{\varepsilon}_n$

$$(X_n^\top X_n)^{1/2}(\hat{\beta}_n - \beta) = (X_n^\top X_n)^{-1/2} X_n^\top \vec{\varepsilon}_n$$

$$\underbrace{(X_n^\top X_n)^{1/2}(\hat{\beta}_n - \beta)}_{\sigma} = \underbrace{(X_n^\top X_n)^{-1/2} X_n^\top \vec{\varepsilon}_n}_{\sigma}$$

WLOG, we can assume $\underset{\text{Var}(\varepsilon_i)}{\overset{\shortparallel}{\sigma}} = 1$.

③ Define

$a_{ni} :=$ i-th column of $(X_n^\top X_n)^{-1/2} X_n^\top$

(Please verify $\|a_{ni}\|^2 = H_{ii}$)

④ $\forall t \in \mathbb{R}^{d+1} \setminus 0_{d+1}$,

$(a_{n1}, a_{n2}, \cdots, a_{nn})$

$$t^\top \boxed{(X_n^\top X_n)^{-\frac{1}{2}} X_n^\top} \vec{\varepsilon}_n$$

$$= \sum_{i=1}^{n} \underbrace{(t^\top a_{ni}) \cdot \varepsilon_i}$$

x_i in the L-F CLT

⑤ Define

$$\sigma_{ni}^2 := Var\left((t^T a_{ni}) \varepsilon_i \right) = [t^T a_{ni}]^2 \cdot 1$$

$$\sigma_n^2 = \sum_{i=1}^n \sigma_{ni}^2 = \sum_{i=1}^n [t^T a_{ni}]^2 = \|t\|^2$$

please verify

⑥ Define

ln L-F cond.

$$Z_{ni} = \frac{(t^T a_{ni}) \varepsilon_i}{\sigma_n}$$

Fix $\varepsilon > 0$,

$$\sum_{i=1}^n \mathbb{E}\left[Z_{ni}^2 \mathbb{1}\{ |Z_{ni}| \geq \varepsilon \} \right]$$

plug in

$$= \sigma_n^{-2} \cdot \sum_{i=1}^n \left[(t^T a_{ni})^2 \cdot \mathbb{E}\left[\varepsilon_i^2 \cdot \mathbb{1}\{ |t^T a_{ni}| \cdot |\varepsilon_i| \geq \sigma_n \varepsilon \} \right] \right]$$

$$\leq \underbrace{\sigma_n^{-2}}_{\|t\|^{-2}} \cdot \max_{i=1,\dots,n} \mathbb{E}\left[\varepsilon_i^2 \mathbb{1}\{ \cdots \} \right] \cdot \underbrace{\sum_{i=1}^n (t^T a_{ni})^2}_{= \|t\|^2}$$

$$= \max_{i=1,\dots,n} \underline{\mathbb{E}\left[\varepsilon_i^2 \mathbb{1}\{ \cdots \} \right]}$$

$\boxed{}$

Goal: can drag $\lim_{n\to\infty}$ inside using DCT.

$$\boxed{} := \max_i \mathbb{E}\left[\varepsilon_i^2 \cdot \mathbb{1}\left\{\boxed{\underbrace{|t^T a_{ni}|}}^{\le \|t\|\|a_{ni}\|} |\varepsilon_i| \ge \sigma_n \cdot \varepsilon\right\}\right]$$

$$\le \max_i \mathbb{E}\left[\varepsilon_i^2 \mathbb{1}\left\{\cancel{\|t\|}\|a_{ni}\||\varepsilon_i| \ge \cancel{\sigma_n} \cdot \varepsilon\right\}\right]$$

$$= \max_i \mathbb{E}\left[\varepsilon_i^2 \mathbb{1}\left\{\|a_{ni}\||\varepsilon_i| \ge \varepsilon\right\}\right]$$

$(= \|a_{ni}\|^2)$

⑦ If $\max_{i=1,\dots,n} H_{ii} \to 0$ as $n \to \infty$.

$\overset{CMT}{\Longrightarrow} \max_{i=1,\dots,n} \|a_{ni}\| \to 0$.

Lastly, since

$$\varepsilon_i^2 \mathbb{1}\{\cdot\} \le \varepsilon_i^2 \text{ and } \mathbb{E}\,\varepsilon_i^2 < \infty$$

$$\Downarrow \text{DCT}$$

$$\lim_{n\to\infty} \mathbb{E}[\cdot] = \mathbb{E}[\underbrace{\lim_{n\to\infty} \cdot}_{=0}] = 0$$

sufficient

$\square$

Chap 2.6 Delta Method [Multi-variate]

$X_1, \dots, X_n \overset{iid}{\sim} P_{\theta_0} \in \{P_\theta : \theta \in \Theta\}$

allowed to be multi- or infinite-dimensional

$\Psi : \Theta \longrightarrow \mathbb{R}^d$

$\psi_0 := \Psi(\theta_0)$

Example. If P_{θ_0} is indexing the CDF of X s.t.

$$\mathbb{E}_{P_{\theta_0}}[X] < \infty, \text{ if}$$

$$\Psi : \theta \longmapsto \mathbb{E}_{P_\theta}[x]$$

ψ_0 is just the population mean

$\psi_n :=$ an estimator of $\psi_0 \in \mathbb{R}^d$

Assume $\exists$ a seq $r_n \nearrow \infty$, s.t.

$$(*) \quad r_n(\psi_n - \psi_0) \Longrightarrow Z$$

a well-define distr on $\mathbb{R}^d$.

eg: $\psi_0 = \mathbb{E}_{P_{\theta_0}}[x]$

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$r_n = \sqrt{n}$

as long as $\mathbb{E}_{P_{\theta_0}}[x^2] < \infty$,

$$\sqrt{n}(\psi_n - \psi_0) \Rightarrow N(0, \Sigma) \quad \Sigma \to \mathrm{cov}(X)$$

Q: Suppose (*) is true, and we have

$$f: \mathbb{R}^d \longmapsto \mathbb{R}.$$

is "smooth enough". Then what is the limit of

$$f(\psi_n) - f(\psi_0) \quad ???$$

A: It depends on the smoothness cond.

Assume $\exists \nabla f(\cdot)$ at ψ_0 s.t.

$$(**) \quad \lim_{\varepsilon \to 0} \sup_{\substack{h \in \mathbb{R}^d \\ \|h\|=1}} \frac{|f(\psi_0 + \varepsilon h) - f(\psi_0) - \varepsilon \langle h, \nabla f(\psi_0)\rangle|}{\varepsilon} = 0$$

Remark "sup" is "uniform convergence"

(**) is true if

(i) f is partially differentiable surronding ψ_0;

(ii) the partial derivatives are cont. at ψ_0.

Thm. [Multi-Delta] Suppose (*), (**) are true,

Then

(i) $f(\psi_n) - f(\psi_0) - \langle \psi_n - \psi_0, \nabla f(\psi_0) \rangle$

$= o_p(r_n^{-1})$

Slutsky + CMT $\cdot r_n$

(ii) $r_n (f(\psi_n) - f(\psi_0)) \Rightarrow \langle Z, \nabla f(\psi_0) \rangle$

Pf. (i) $\Rightarrow$ (ii)

(i) (**) + (*) + Prokhorov + CMT

$$\lim_{\varepsilon \to 0} \sup_{\substack{h \in \mathbb{R}^d \\ \|h\|=1}} \frac{|f(\psi_0 + \varepsilon h) - f(\psi_0) - \varepsilon \langle h, \nabla f(\psi_0) \rangle|}{\varepsilon} = 0$$

"Top 5 fundamental tricks in Math Stats"

$$|f(\psi_n) - f(\psi_0) - \langle \psi_n - \psi_n, \nabla f(\psi_0)\rangle|$$

$$= |f(\psi_0 + \varepsilon_n \cdot h_n) - f(\psi_0) - \varepsilon_n \langle h_n, \nabla f(\psi_0)\rangle|$$

$\uparrow$

$\varepsilon_n := \|\psi_n - \psi_0\|_2$

$$h_n := \begin{cases} \frac{\psi_n - \psi_0}{\varepsilon_n}, & \text{if } \varepsilon_n \neq 0 \\ 0, & \text{if } \varepsilon_n = 0 \end{cases}$$

$\|h_n\| = 1$ as long as $\varepsilon_n \neq 0$.

$$\leq \sup_{h : \|h\| = 1} |f(\psi_0 + \varepsilon_n h) - f(\psi_0) - \varepsilon_n \cdot \langle h, \nabla f(\psi_0)\rangle|$$

$$= \varepsilon_n \cdot g(\varepsilon_n)$$

with $g : \varepsilon \longmapsto \begin{cases} \sup_{\|h\|=1} \frac{|f(\psi_0 + \varepsilon h) - f(\psi_0) - \langle \cdot, \cdot \rangle|}{\varepsilon} & \text{if } \varepsilon \neq 0 \\ 0, & \text{o.w.} \end{cases}$

By (**), g is cont. at 0.

It remains to show

$$\varepsilon_n \cdot g(\varepsilon_n) = o_p(r_n^{-1})$$

(i) Prokhorov

$$r_n(\psi_n - \psi_0) \Rightarrow Z$$

$$\Downarrow$$

$$\varepsilon_n = O_p(r_n^{-1})$$

(ii) Since g is conts. at 0 and

$$\varepsilon_n = o_p(1)$$

$$\Downarrow \text{CMT}$$

$$g(\varepsilon_n) = o_p(1)$$

(iii) $\varepsilon_n g(\varepsilon_n) = O_p(r_n^{-1})\, o_p(1)$

$$= o_p(r_n^{-1})$$

$\square$

The most general version:

$$f: \mathbb{R}^d \longmapsto \mathbb{R}^p, \quad p \geq 1.$$

Smoothness cond.

f: is called differentiable at ψ_0 if

$$(***) \quad \limsup_{\substack{\varepsilon \to 0 \\ h \in \mathbb{R}^d \\ \|h\|=1}} \frac{\|f(\psi_0 + \varepsilon h) - f(\psi_0) - \varepsilon \cdot J_f \cdot h\|}{\varepsilon} = 0$$

for a $\mathbb{R}^p \times \mathbb{R}^d$ Jacobian matrix J_f.

Thm. (Multi-Delta, $\mathbb{R}^d \to \mathbb{R}^p$)

Thm. 2.6.3 in Lec #2 proof.

Ex. [Relative risk]

$$\underset{\text{data}}{\mathcal{X}} = (\underset{\text{treatment}}{\overset{\in\{0,1\}}{T}}, \underset{\text{outcome}}{Y})$$

Y is usually dependent on T.

Data $(X_1, X_2, \ldots, X_n) \overset{iid}{\sim} \mathcal{X}$

$$\psi_0 := \begin{pmatrix} \mathbb{E}_{\theta_0}[YT] =: \psi_{0,1} \\ \\ \mathbb{E}_{\theta_0}[Y(1-T)] =: \psi_{0,2} \end{pmatrix} \in \mathbb{R}^2$$

$$\text{relative risk} := \frac{\psi_{0,1}}{\psi_{0,2}}$$

To estimate RR:

$$\underset{\text{an est. of } \psi_0}{\psi_n} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n Y_i T_i \\ \\ \frac{1}{n}\sum_{i=1}^n Y_i(1-T_i) \end{pmatrix}$$

$$\sqrt{n}(\psi_n - \psi_0) \Rightarrow \mathcal{N}(0, \underset{\text{calculate}}{\Sigma})$$

Use $f: \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \mapsto \frac{z_1}{z_2}$

Verify f satisfies (**) smooth,

$$\Downarrow$$

$$\sqrt{n}\,\big(\underbrace{f(\psi_n)}_{\text{the est. of RR}} - f(\psi_0)\big) \Rightarrow N(0, \sigma_f^2)$$