ULLN

$$\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \underbrace{f(X_i) - \mathbb{E}f(X)}_{(\mathbb{P}_n-\mathbb{P})f} = o_p(1).$$

$$\underbrace{\mathbb{E}_\varepsilon \frac{1}{n}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right|}$$

$\varepsilon = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases}$ Rad.

For now: $\mathcal{F}$ finite. $\mathcal{F}=\{f_1,\cdots,f_N\}$.

$$\mathbb{E}\max_{f_j\in\{f_1,\cdots,f_N\}} X_{f_j} \asymp \sqrt{\log N}$$ union bound.

"maximal inequality"

$X_{f_1},\cdots,X_{f_N}$ indep.

$\{X_f, f\in\mathcal{F}\}$.

$\mathcal{F}$: $\varepsilon$-net $\{f_1,\cdots,f_{N(\varepsilon)}\}$.

$\forall f\in\mathcal{F}$. $\pi_\varepsilon(f) = \underset{j\in\{1,\cdots,N\}}{\arg\min} \|f_j - f\|$.

$$\mathbb{E}_\varepsilon \frac{1}{n}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right|$$

$$\le \mathbb{E}_\varepsilon \frac{1}{n}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\varepsilon_i \underbrace{(f(X_i)-\pi_\varepsilon(f)(X_i))}_{\le\varepsilon.}\right| + \mathbb{E}_\varepsilon\frac{1}{n}\underbrace{\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\varepsilon_i \pi_\varepsilon(f)(X_i)\right|}_{\text{maximal inequality}}$$

$\mathcal{F}$: metric entropy condition

$\to 0$.

Unif. LLN $\Longrightarrow$ Unif. CLT

Motivation

(1) M-estimation / ERM (emp. risk minimization).

$$\theta_0 := \underset{\theta \in \Theta}{\arg\min}\ \mathbb{E}\, m(X;\theta) \equiv \mathbb{P}\, m(X;\theta).$$

$\theta \in \mathbb{R}$.

$X_1, \cdots, X_n \overset{iid}{\sim} \mathbb{P}_{\theta_0}$. $\theta_0$ unknown

$$\hat\theta_n = \underset{\theta \in \Theta}{\arg\min}\ \frac{1}{n}\sum_{i=1}^{n} m(X_i;\theta) \equiv \mathbb{P}_n\, m(X;\theta)$$

(1) $\hat\theta_n \xrightarrow{P} \theta_0$ (Shuguang: ULLN). $\sup_{\theta \in \Theta} |\hat\theta_n - \theta_0| \xrightarrow{P} 0$?

(2) $|\hat\theta_n - \theta_0| = n^{-?}$ (rate-of-convergence)

(3) $\sqrt{n}\,(\hat\theta_n - \theta_0) \overset{?}{\Rightarrow} N(0, \sigma^2)$.

$m$ convex $\frac{1}{n}\sum_{i=1}^{n} m'(X_i;\theta) = 0$.

$\mathcal{F}: \{m'(\cdot;\theta), \theta \in \Theta\}$.

$m$: $-$log-likelihood.

$L_2$-risk. $(Y - f_\theta(X_i))^2$

$L_1$-risk. "minimal deviation"

(2). Semi-parametric.

MLE. $\eta = (\theta, f)$.

$\theta \in \mathbb{R}$. $f \in \mathcal{F}$ infinite-dimensional Sobolev space.

$\hat\theta_{MLE}(\hat f)$ plug-in.

$$\sup_{f \in \mathcal{F}} \sqrt{n}\,(\hat\theta_{MLE}(\hat f) - \theta_0) \overset{?}{\Rightarrow} N(0, \sigma^2).$$

$(X, R)$ $\quad R=1, \; X = X^{true}$

$\qquad\qquad R=0, \; X = \text{"NA"}$

$\mathbb{E}X = \theta$. $\quad R$ missing me.ha.

$\hat{f}$: sample split.

$\theta_0$ solves $\mathbb{E}\psi(X;\theta) \equiv 0 \quad \psi = m'$

$\hat{\theta}_n$ solves $\frac{1}{n}\sum_{i=1}^{n} \psi(X_i;\theta) \equiv 0$.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(X_i;\hat{\theta}_n) - \sqrt{n}\,\mathbb{E}\psi(X;\theta) \equiv 0.$$

$$= \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\psi(X_i;\hat{\theta}_n) - \mathbb{E}\psi(X;\hat{\theta}_n)\right) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\psi(X_i;\theta_0) - \mathbb{E}\psi(X;\theta_0)\right)}$$

$$+ \sqrt{n}\left(\mathbb{E}\psi(X;\hat{\theta}_n) - \mathbb{E}\psi(X;\theta_0)\right) \qquad \sqrt{n}\,C\,(\hat{\theta}_n - \theta_0)$$

$$+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\psi(X_i;\theta_0) - \mathbb{E}\psi(X;\theta_0)\right) \rightsquigarrow N(0, \sigma^2).$$

$$= \sqrt{n}\,(\mathbb{P}_n - \mathbb{P})\left(\psi(X;\hat{\theta}_n) - \psi(X;\theta_0)\right) = o_p(1) \qquad \checkmark.$$

$+ \;\text{—}\; + \;\text{—}\;.$

$$\sqrt{n}\,(\mathbb{P}_n - \mathbb{P})(\cdot) = \mathbb{G}_n(\cdot).$$

$$\mathbb{G}_n\left(\psi(X;\hat{\theta}_n) - \psi(X;\theta_0)\right)$$

$$\mathbb{G}_n\big(\underbrace{f(\hat{\theta}_n) - f(\theta_0)}_{\text{"}f\text{"}}\big)$$

$$\sup_{\theta_n \in \Theta} \mathbb{G}_n f(\theta_n) - f(\theta_0) = o_p(1).$$

$\theta_n \in \{\Theta : \|\theta_n - \theta_0\| \le \delta\}$.  $\mathbb{G}_n f(\theta_0)$

In general: $\mathcal{F}$ is Donsker if $\{\mathbb{G}_n f : f \in \mathcal{F}\} \Rightarrow \mathbb{G}_P$.
in $\ell^\infty(\mathcal{F})$.

Thm: (Dudley)

Define $$J(\delta, \mathcal{F}, L_2) = \int_0^\delta \sup_{Q} \sqrt{\log N(\epsilon \|F\|_{2,Q}, \mathcal{F}, L_2(Q))}\, d\epsilon.$$

$F$: envelope of $\mathcal{F}$.  $\|F\|^2_{L_2(P)} < \infty$. unif. entropy integral
$\mathbb{P} F^2$

If $J(\delta, \mathcal{F}, L_2) \to 0$ as $\delta \to 0$, then $\mathcal{F}$ is Donsker

Pf: Lemma. $\{\mathbb{G}_n f : f \in \mathcal{F}\} \Rightarrow \{\mathbb{G}_P f . f \in \mathcal{F}\}$ ← tight as $n \to \infty$.
in $\ell^\infty(\mathcal{F})$  Gaussian process

iff. (i) $\{\mathbb{G}_n f_1, \cdots, \mathbb{G}_n f_k\} \Rightarrow \{\mathbb{G}_P f_1, \cdots \mathbb{G}_P f_k\}$ $\forall k \in \mathbb{N}$.
$f_1, \cdots, f_k \in \mathcal{F}$.

(ii) asymp. equicontinuity.

$$\forall \epsilon > 0, \lim_{\delta \to 0} \lim_{n \to \infty} \mathbb{P}\left( \sup_{f, g \in \mathcal{F}, \|f - g\| \le \delta} |\mathbb{G}_n f - \mathbb{G}_n g| > \epsilon \right) \to 0.$$

(Bodhisattva Sen. for proof)

Back to Dudley:

Define $\mathcal{G}_\delta := \{ f-g;\ f\in\mathcal{F},\ g\in\mathcal{F},\ \|f-g\|\le\delta \}$.

$\mathcal{G}_\delta$: envelope $2F$.

need to show: $\forall\ \epsilon>0$,

$$\lim_{\delta\to 0}\ \lim_{n\to\infty}\ \mathbb{P}\left(\sup_{h\in\mathcal{G}_\delta} |\mathbb{G}_n h| \ge \epsilon\right)$$

Markov.

$$\le \frac{1}{\epsilon}\ \underbrace{\lim_{\delta\to 0}\ \lim_{n\to\infty}\ \mathbb{E}\left(\sup_{h\in\mathcal{G}_\delta} |\mathbb{G}_n h|\right)} \to 0 \qquad \checkmark.$$

$$\mathbb{E}\left(\sup_{h\in\mathcal{G}_\delta} |\mathbb{G}_n h|\right) \qquad \frac{1}{\sqrt{n}}.$$

"symm"

$$\le \mathbb{E}_X \mathbb{E}_\varepsilon \sup_{h\in\mathcal{G}_\delta} \frac{1}{\sqrt{n}}\left|\sum_{i=1}^n h(X_i)\varepsilon_i\right|$$

$$\underbrace{\mathbb{E}_\varepsilon \sup_{h\in\mathcal{G}_\delta} \frac{1}{\sqrt{n}}\left|\sum_{i=1}^n h(X_i)\varepsilon_i\right|} \qquad X_h.$$

$$\overset{?}{\le}\ \underbrace{\mathbb{E}_\varepsilon \sup_{h\in\mathcal{G}_\delta} \frac{1}{\sqrt{n}}\sum_{i=1}^n \left|(h(X_i)-\Pi h(X_i))\varepsilon_i\right| + \frac{1}{\sqrt{n}}\sum_{i=1}^n \left|\Pi h(X_i)\varepsilon_i\right|}$$

Chaining.

Assume: $\mathcal{G}_\delta$ finite. Fix $h_0\in\mathcal{G}_\delta$.

$T_1$ $\epsilon_1$-net. $\mathbb{E}\sup_{h\in\mathcal{G}_\delta}|X_h|$

$\Pi_1$

$$\le \mathbb{E}\sup_{h\in\mathcal{G}_\delta} |X_h - X_{\Pi_1(h)}| + |X_{\Pi_1(h)}|$$

$T_2$ $\epsilon_2$-net $\quad \le \mathbb{E} \sup_{h \in \mathcal{G}_\delta} |X_{\pi_1(h)}| + \sup_{h \in \mathcal{G}_\delta} |\underline{X_{\pi_1(h)}} - \underline{X_{\pi_2(h)}}|$

$$+ \mathbb{E} \sup_{h \in \mathcal{G}_\delta} |X_{\pi_2(h)} - X_h|$$

$\vdots$

$T_N$ $\epsilon_N$-net $\quad \le \sum_{k=1}^{N} \mathbb{E} \sup_{h \in \mathcal{G}_\delta} |X_{\pi_k(h)} - X_{\pi_{k-1}(h)}| \qquad \pi_N(h) = h.$

$h \quad \pi_1(h)$

$\pi_2(h)$

$\ldots$

$$\lesssim \sum_{k=1}^{N} \sqrt{\log |T_{k-1}| |T_k|} \cdot \{ \underbrace{|X_{\pi_k(h)} - X_h|}_{\le \epsilon_k} + \underbrace{|X_{\pi_{k-1}(h)} - X_h|}_{\le \epsilon_{k-1}} \}$$

$\pi_N(h) = h.$ $\quad \lesssim 2\sqrt{2} \sum_{k=1}^{N} \epsilon_{k-1} \sqrt{\log \underline{N(\epsilon_k, \mathcal{G}_\delta, L_2)}\, N(\epsilon_{k-1}, \mathcal{G}_\delta, L_2)}.$

$\epsilon_{N+1}$

$$\lesssim \sum_{k=1}^{N} \frac{\epsilon_{k-1}}{\epsilon_k} \epsilon_k \sqrt{\log N(\epsilon_k, \mathcal{G}_\delta, L_2)} \lesssim \rho \sum_{k=1}^{N} \epsilon_k \sqrt{\log N(\epsilon_k, \mathcal{G}_\delta, L_2)}$$

$\frac{\epsilon_{k-1}}{\epsilon_k} = \rho \qquad \lesssim \sum_{k=1}^{N} \int_{\epsilon_{k+1}}^{\epsilon_k} \sqrt{\log N(\epsilon, \mathcal{G}_\delta, L_2)}\, d\epsilon.$

$\rho = 0.5$ $\qquad = \int_{\epsilon_N}^{\epsilon_1} \sqrt{\log N(\epsilon, \mathcal{G}_\delta, L_2)}\, d\epsilon$

$\epsilon_{N+1}$

$\epsilon_0$ : Diam $(\mathcal{G}_\delta)$. $\quad \epsilon_1 = \frac{\text{Diam}(\mathcal{G}_\delta)}{2}.$

$$\le \int_{0}^{\frac{\text{Diam}(\mathcal{G}_\delta)}{2}} \sqrt{\log N(\epsilon, \mathcal{G}_\delta, L_2)}\, d\epsilon$$

$\epsilon$

$$\leq \int_{0}^{\frac{\mathrm{diam}_n(\delta_S)}{2}} \sup_{Q} \sqrt{\log N(\epsilon, \delta_S, L_2(Q))}\, d\epsilon .$$