Lec 1.

Review. $X_1, \cdots, X_n \overset{iid}{\sim} P$

$$Pf = \int f(x)\, dP(x)$$

$$\mathbb{P}_n f := \frac{1}{n}\sum_{i=1}^{n} f(X_i) \qquad f \in \mathcal{F}$$

$m_\theta : \mathcal{X} \to \mathbb{R}$
$\forall \theta \in \Theta$

(1) Consistency of M-estimators,

$$\theta_0 = \arg\max_\theta \underline{P}\, m_\theta \qquad P m_\theta := M(\theta)$$

$$\hat\theta_n = \arg\max_\theta \mathbb{P}_n m_\theta \qquad \mathbb{P}_n m_\theta := M_n(\theta)$$

Goal: $d(\hat\theta_n, \theta_0) \to 0$ in prob of $\mathbb{P}$

$\forall \delta > 0$, $\quad d(\hat\theta_n, \theta_0) \geq \delta$

$$\Rightarrow M(\hat\theta_n) \leq \sup_{\theta : d(\theta,\theta_0) \geq \delta} M(\theta)$$

$$\Rightarrow M(\hat\theta_n) - M(\theta_0) + M_n(\theta_0) - M_n(\hat\theta_n) \leq \sup_{\theta : d(\theta,\theta_0) \geq \delta} M(\theta) - M(\theta_0)$$

If we have

$$M(\theta_0) > \sup_{\theta : d(\theta,\theta_0) \geq \delta} M(\theta)$$

then

$$\varphi(\delta) \leq -M(\hat\theta_n) + M(\theta_0) - M_n(\theta_0) + M_n(\hat\theta_n)$$

$$= (M_n - M)(\hat\theta_n) - \underbrace{(M_n - M)(\theta_0)}_{\mathbb{P}_n m_\theta}$$

$$= (M_n - M)(\hat\theta_n - \theta_0)$$

$$\leq \sup_\theta |(M_n - M)\theta|$$

$$= \sup_{f \in \mathcal{F}} |(\mathbb{P}_n - P) f|$$

$$M_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m_\theta(X_i)$$

$$M(\theta) = \int m_\theta(x)\, dP(x)$$

We can show

$$\langle 1\rangle \quad \sup_{f\in\mathcal{F}} |(\mathbb{P}_n - P)f| := \|\mathbb{P}_n - P\|_{\mathcal{F}} \xrightarrow{a.s.} 0$$

$$\langle 2\rangle \quad \mathbb{E}\, \|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0$$

For $\langle 1\rangle$, it is about G-C.

(a) Finite bracketing number of $\mathcal{F}$ $\Rightarrow$ G.C.

(b) Entropy is well controlled $\Rightarrow$ G.C.

Thm 1-1: Let $\mathcal{F}$ be a class of m-functions with envelope F such that $PF < \infty$. Let $\mathcal{F}_M = \{ f 1\{F \le M\} : f \in \mathcal{F} \}$, $\forall M > 0$.

Then $\forall \epsilon > 0$, $M > 0$

$$\underbrace{\|\mathbb{P}_n - P\|_{\mathcal{F}} \xrightarrow{a.s.} 0}_{G.C.} \iff \frac{1}{n} \log N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) \xrightarrow{P} 0$$

The proof uses symmetrization.

$$\mathbb{E}\, \|\mathbb{P}_n - P\|_{\mathcal{F}} \le 2\, \mathbb{E}_X \mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^n \varepsilon_i f(x_i) \right|$$

$$\le 2\, \mathbb{E}_X \mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}_M} \frac{1}{n} \left| \sum_i \varepsilon_i f(x_i) \right| + 2P\underbrace{[F 1\{F > M\}]}_{\to 0}$$

Let $\mathcal{G}_M$ be an ε-net of $\mathcal{F}_M$ w.r.t. $L_1(\mathbb{P}_n)$, i.e. $\forall f \in \mathcal{F}_M$, $\exists g \in \mathcal{G}_M$ s.t.

$$\|f - g\|_{L_1(\mathbb{P}_n)} = \frac{1}{n}\sum_i |f(x_i) - g(x_i)| \le \varepsilon.$$

It suffices to bound

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

$$\lesssim \sqrt{\log|\mathcal{G}_M|}\ \frac{1}{\sqrt{n}} \sup_{g\in\mathcal{G}_M} \|g\|_n$$

(c) Chaining. $\mathbb{E}\sup_{t\in T}|X_t|$

Thm A. (T,d) be a metric space, and $\{X_t\}_{t\in T}$ is separable.

[$\{X_t, t\in T\}$ is separable if $\exists$ a countable subset $\tilde{T}$ of T and a null set N s.t. $\forall\ \omega\notin N$, and $t\in T$, $\exists$ a sequence $\{t_n\}_{n\geq 1}$ of $\tilde{T}$ such that $d(t_n,t)\to 0$ and $\lim_{n\to\infty} X_{t_n}(\omega) = X_t(\omega)$]

Suppose $\forall\ s,t\in T$ and $u\geq 0$,

(C1) $$\mathbb{P}(|X_t - X_s|\geq u) \leq 2\exp\left\{-\frac{u^2}{2d^2(s,t)}\right\}$$

Then $\forall\ t_0\in T$,

$$\mathbb{E}\sup_{t\in T}|X_t - X_0| \leq C\int_0^{\tilde{D}/4}\sqrt{\log N(\epsilon,T,d)}\,d\epsilon$$

where $\tilde{D}$ is the diameter of (T,d).

<2> Apply Chaining to bound $\mathbb{E}\|\mathbb{P}_n - P\|_{\mathcal{F}}$

$$\mathbb{G}_n := \sqrt{n}(\mathbb{P}_n - P)$$

We want to control

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} = \mathbb{E}\sup_{f\in\mathcal{F}} \frac{1}{\sqrt{n}}|(\mathbb{P}_n - P)f|$$

$$\overset{\text{sym.}}{\leq} 2\,\mathbb{E}_X \mathbb{E}_\epsilon \sup_{f\in\mathcal{F}} \frac{1}{\sqrt{n}} \Big|\underbrace{\sum_{i=1}^n \epsilon_i f(X_i)}_{X_t,\ t=f}\Big|$$

$$(T,d) = (\mathcal{F}, \overset{\|\cdot\|_n}{L_2(\mathbb{P}_n)})$$

To verify (C1), $\forall f,g\in\mathcal{F}$ and $u\geq 0$

$$\mathbb{P}_\epsilon\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i f(X_i) - \frac{1}{\sqrt{n}}\sum_i \epsilon_i g(X_i)\right| \geq u\right) \leq 2\exp\left\{-\frac{u^2}{2\|f-g\|_n^2}\right\}$$

where $\|f\|_n^2 = \frac{1}{n}\sum_i f^2(X_i)$. By Thm A (add $f\equiv 0$ to $\mathcal{F}$)

$$\mathbb{E}_\epsilon \sup_{f\in\mathcal{F}\cup\{0\}} \frac{1}{\sqrt{n}}\left|\sum_{i=1}^n \epsilon_i f(X_i)\right| \lesssim \int_0^{\tilde{D}} \sqrt{\log N(\epsilon, \mathcal{F}\cup\{0\}, \|\cdot\|_n)}\, d\epsilon$$

where $\tilde{D} = \sup_{f,g\in\mathcal{F}} \|f-g\|_n \leq 2\sup_{f\in\mathcal{F}} \|f\|_n := \sigma_{n,2}$

(1) $$\mathbb{E}_\epsilon \sup_{f\in\mathcal{F}\cup\{0\}} \frac{1}{\sqrt{n}}\left|\sum_{i=1}^n \epsilon_i f(X_i)\right| \lesssim \int_0^{\sigma_{n,2}/\|F\|_n = \theta_n \in [0,1]} \sqrt{\log N(\epsilon\|F\|_n, \mathcal{F}\cup\{0\}, \|\cdot\|_n)}\, d\epsilon\, \|F\|_n$$

Def: A class $\mathcal{F}$ of m-functions with envelope F satisfies the uniform entropy bound if and only if $J(1,\mathcal{F},F) < \infty$ where, $\forall \delta > 0$,

$$J(\delta,\mathcal{F},F) = \int_0^\delta \sup_Q \sqrt{\log N(\epsilon\|F\|_{Q,2}, \mathcal{F}\cup\{0\}, L_2(Q))}\, d\epsilon$$

where the supremum is taken over all finitely discrete prob. measures on $\mathcal{X}$, and $\|F\|_{Q,2} = \int F^2(x)\, dQ(x) > 0$.

$$\text{LHS of (1)} \overset{\text{def'n}}{\leq} J(\theta_n, \mathcal{F}, F)\, \|F\|_{n,2}$$

$$\Rightarrow \mathbb{E}\|\mathbb{P}_n - P\|_{\mathcal{F}} \lesssim \mathbb{E}_X\left[J(\theta_n,\mathcal{F},F)\,\|F\|_{n,2}\right] \leq J(1,\mathcal{F},F)\,\mathbb{E}\left[\|F\|_{n,2}\right]$$

Since
$$\mathbb{E}\|F\|_{n,2} \le \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, F(X_i)} \le \|F\|_{L_2(P)}$$
then
$$\mathbb{E}\,\|\mathbb{P}_n - P\|_{\mathcal{F}} \lesssim J(1, \mathcal{F}, F)\, \|F\|_{L_2(P)}$$

$\Rightarrow$ **Thm B**: If $\mathcal{F}$ is a class of m-functions with envelope F, then
$$\mathbb{E}\,\|G_n\|_{\mathcal{F}} \lesssim \mathbb{E}\left[J(\theta_n, \mathcal{F}, F)\, \|F\|_{n,2}\right]$$
$$\lesssim J(1, \mathcal{F}, F)\, \|F\|_{L_2(P)},$$
with $\theta_n = \sup_{f\in\mathcal{F}} \|f\|_{n,2} / \|F\|_{n,2}$

Remark:

Thm B': Suppose $0 < \|F\|_{L_2(P)} < \infty$ and let $\sigma^2 > 0$ be $\sup_{f\in\mathcal{F}} \|f\|^2_{L_2(P)} \le \sigma^2 \le \|F\|^2_{L_2(P)}$. Let $\delta = \frac{\sigma}{\|F\|_{P,2}}$. One has
$$\mathbb{E}\,\|G_n\|_{\mathcal{F}} \lesssim J(\delta, \mathcal{F}, F)\, \|F\|_{L_2(P)} + \frac{B}{\delta^2 \sqrt{n}} J^2(\delta, \mathcal{F}, F)$$
where $B = \sqrt{\mathbb{E} \max_{1\le i\le n} F^2(X_i)}$,

VW, 2011 and Chernozhukov et al, 2014.

Example 1. $\mathcal{F}$ is a class of B-uniformly bounded functions such that
$$N(\epsilon, \mathcal{F}, \|\cdot\|_n) \le C\, v\, (16e)^v \left(\frac{B}{\epsilon}\right)^{2v}, \ \forall \epsilon > 0$$

By Thm
$$\int_0^B \sqrt{\log(\epsilon, \mathcal{F}\cup\{0\}, \|\cdot\|)}\, d\epsilon \le \int_0^B \sqrt{\log\left(1 + c\left(\frac{B}{\epsilon}\right)^{2v}\right)}\, d\epsilon$$
$$\le C' B \sqrt{v}$$
$$\mathbb{E}\,\|G_n\|_{\mathcal{F}} \lesssim B\sqrt{v} \Rightarrow \mathbb{E}\,\|\mathbb{P}_n - P\|_{\mathcal{F}} \lesssim B\sqrt{\frac{v}{n}}$$

Example: $\mathcal{F} = \{ m_\theta(\cdot) : \theta \in B_2(1) \subset \mathbb{R}^d \}$ and $\mathcal{F} = -\mathcal{F}$.

Suppose $|m_{\theta_1}(x) - m_{\theta_2}(x)| \leq L \, \|\theta_1 - \theta_2\|$, $\forall \theta_1, \theta_2$ and x

$\Rightarrow \mathbb{E} \| \mathbb{P}_n - P \|_{\mathcal{F}} \lesssim L \sqrt{\frac{d}{n}}$

<3> Maximal inequalities with bracketing.

Def: $J_{[\,]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{\log N_{[\,]}(\epsilon, \mathcal{F} \cup \{0\}, L_2(P))} \, d\epsilon$

Thm C. $\mathcal{F}$ is a class of m-functions with envelope F.

$$\mathbb{E} \| \mathbb{G}_n \|_{\mathcal{F}} \lesssim J_{[\,]}(\|F\|_{L_2(P)}, \mathcal{F} \cup \{0\}, L_2(P))$$

Thm C' If $\|f\|^2_{L_2(P)} \leq \delta^2$, and $\|f\|_\infty \leq M$ $\forall f \in \mathcal{F}$,

then

$$\mathbb{E} \| \mathbb{G}_n \|_{\mathcal{F}} \lesssim J_{[\,]}(\delta, \mathcal{F}, L_2(P)) \left(1 + \frac{M \, J_{[\,]}(\delta, \mathcal{F}, L_2(P))}{\delta^2 \sqrt{n}} \right)$$

Lec 2: VC - dimension

This is a combinatorial restriction of a class $\mathcal{F}$ of m-functions, on $\mathcal{X}$. We would show

$$\sup_Q N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq K \left(\frac{1}{\epsilon} \right)^V, \quad 0 < \epsilon < 1$$

for some $V > 0$,

where K is a universal constant. As a result

$$\int_0^\delta \sqrt{\log(1/\epsilon) V} \, d\epsilon \asymp \sqrt{V} \, \delta \log(1/\delta)$$

Vapnik and Červonenkis in the 1970's

Boolean function $f : \mathcal{X} \to \{0, 1\}$, with its class

$$\mathcal{F} = \{ \mathbb{1}_C(\cdot) : C \in \mathcal{C} \} \qquad \text{eg: } C = (-\infty, a], \; a \in \mathbb{R}$$

$\mathcal{C} = \{(-\infty, t] : t \in \mathbb{R}\}$.

⟨1⟩. VC classes to sets.

Def 1: Let $\{x_1, \cdots, x_n\}$ be a subset of $\mathcal{X}$ and $\mathcal{C}$ be a collection of subsets of $\mathcal{X}$. We say $\mathcal{C}$ picks out a subset $A \subseteq \{x_1, \cdots, x_n\}$ if $\exists\, C \in \mathcal{C}$ s.t. $A = C \cap \{x_1, \cdots, x_n\}$.

The collection $\mathcal{C}$ is said to shatter $\{x_1, \cdots, x_n\}$ if it picks out all subsets of $\{x_1, \cdots, x_n\}$.

Def 2. The VC dimension, $V(\mathcal{C})$, is the largest integer n such that $\exists \{x_1, \cdots, x_n\} \in \mathcal{X}$ can be shattered by $\mathcal{C}$.

Def 3. The VC index is defined as

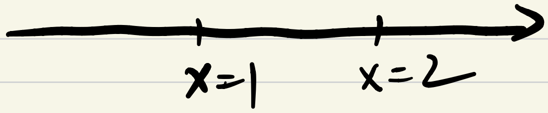
$$\Delta_n(\mathcal{C}; x_1, \cdots, x_n) = \left| \{ C \cap \{x_1, \cdots, x_n\} : C \in \mathcal{C} \} \right|$$

and

$$V(\mathcal{C}) = \sup \{ n : \max_{x_1, \cdots, x_n \in \mathcal{X}} \Delta_n(\mathcal{C}; x_1, \cdots, x_n) = 2^n \}.$$

$VC(\mathcal{C})$ is a complexity measure. We say $VC(\mathcal{C}) \le V$ if no $(V+1)$ points can be shattered by $\mathcal{C}$.

Example 1: $\mathcal{X} = \mathbb{R}$ $\quad \mathcal{C} = \{(-\infty, c] : c \in \mathbb{R}\}$



$A = \{ \{\}, \{1\}, \{2\}, \{1,2\} \}$

$VC(\mathcal{C}) = 1$

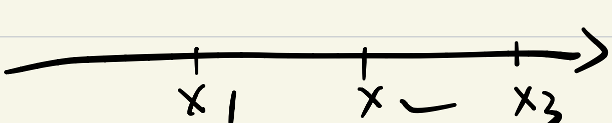
① $A = \{1\}$ $\quad$? $\exists\, C = (-\infty, c]$ s.t. $(-\infty, c] \cap \{1,2\} = \{1\}$

② $A = \{1,2\}$ $\quad \exists\, C = (-\infty, c]$ s.t. $(-\infty, c] \cap \{1,2\} = \{1,2\}$

③ $A = \{2\}$

Example 1': $\quad \mathcal{C} = \{(a, b], a < b \in \mathbb{R}\}$

$VC(\mathcal{C}) = 2$



Example 2: $\mathcal{X} = \mathbb{R}^d$ $\mathcal{C} = \{ (-\infty, c_i]_{i=1}^d , c_i \in \mathbb{R} \}$

$VC(\mathcal{C}) = d$

$\mathcal{C} = \{ [a_i, b_i]_{i=1}^d , a_i, b_i \in \mathbb{R} \}$

$VC(\mathcal{C}) = 2d$.

$\mathcal{C} = \{ a^T x + b \le 0 : a \in \mathbb{R}^d, b \in \mathbb{R} \}$,

$VC(\mathcal{C}) = d+1$

Lemma (Sauer's)

$\forall n \geq 1$.

$$\Delta_n(\mathcal{C}; x_1, \cdots, x_n) \le \sum_{j=0}^{V(\mathcal{C})} \binom{n}{j}$$

If $n \le V(\mathcal{C})$, then RHS $\le 2^n$.

If $n > V(\mathcal{C})$, then RHS $\le \left(\frac{ne}{V(\mathcal{C})}\right)^{V(\mathcal{C})}$

---

(2) VC classes of Boolean functions

For any $x_1, \cdots, x_n \in \mathcal{X}$.

$$\mathcal{F}(x_1, \cdots, x_n) := \{ (f(x_1), \cdots, f(x_n)) : f \in \mathcal{F} \}$$

which is a subset of $\{0,1\}^n$.

Def 4: We say $\{x_1, \cdots, x_n\}$ is shattered by $\mathcal{F}$ if

$\Delta_n(\mathcal{F}; x_1, \cdots, x_n) = |\mathcal{F}(x_1, \cdots, x_n)| = 2^n$

The $VC(\mathcal{F})$ is defined as the largest n for which $\exists\, x_1, \cdots, x_n$ can be shattered.

Remark: $\forall f \in \mathcal{F}$, $C_f = \{ x \in \mathcal{X} : f(x) = 1 \} \Rightarrow \mathcal{C} = \{ C_f : f \in \mathcal{F} \}$

$|\mathcal{F}(x_1, \cdots, x_n)| = \Delta_n(\mathcal{F}; x_1, \cdots, x_n) = \Delta_n(\mathcal{C}; x_1, \cdots, x_n)$.

<3>. Covering number bound for VC classes of sets.

Thm D: $\exists$ universal con't $K>0$, such that for any VC classes $\mathcal{C}$ of sets, any prob. measure Q, $r \geq 1$

$$N(\epsilon, \mathcal{C}, L_r(Q)) \leq K V (4e)^V \left(\frac{1}{\epsilon}\right)^{rV}, \quad \forall\, 0<\epsilon<1$$

and $V = VC(\mathcal{C})$

We prove a weaker version

$$\sup_Q N(\epsilon, \mathcal{C}, L_r(Q)) \leq \left(\frac{c_1}{\epsilon}\right)^{c_2 r V}$$

Proof: Fix $0<\epsilon<1$, and let $X_1, \cdots X_n \overset{iid}{\sim} Q$. Let

$$m := D(\epsilon, \mathcal{C}, L_1(Q))$$

be the ϵ-packing number, meaning $\exists\, C_1, \cdots, C_m \in \mathcal{C}$ such

$$Q\, |1_{C_i} - 1_{C_j}| > \epsilon, \quad \forall\, i \neq j$$

Let $\mathcal{F} = \{1_C : C \in \mathcal{C}\}$, in particular $f_1, \cdots f_m$ corresponding to $C_1 \cdots C_m$, which satisfy

$$\|f_i - f_j\|_{L_1(Q)} = \int |f_i(x) - f_j(x)|\, dQ(x)$$
$$= Q\,|1_{C_i} - 1_{C_j}| > \epsilon$$

As a result,

$$\mathbb{P}(f_i(X_1) = f_j(X_1)) = 1 - \mathbb{P}(f_i(X_1) \neq f_j(X_1))$$
$$= 1 - Q\,|1_{C_i} - 1_{C_j}|$$
$$\leq 1 - \epsilon \leq e^{-\epsilon}$$

Then for every $k \geq 1$,

$$\mathbb{P}(f_i(X_1) = f_j(X_1), \cdots, f_i(X_k) = f_j(X_k)) \leq e^{-k\epsilon}$$

By taking a union bound,

$$\mathbb{P}(f_i(X_1) = f_j(X_1), \cdots, f_i(X_k) = f_j(X_k), \text{ for some } 1 \leq i \neq j \leq m)$$
$$\leq \binom{m}{2} e^{-k\epsilon} \leq \frac{m^2}{2} e^{-k\epsilon}$$

Recall $\mathcal{F}(x_1, \cdots, x_k) = \{ f(x_1) \cdots f(x_k) : f \in \mathcal{F} \}$, we have

$$\mathbb{P}\left(|\mathcal{F}(x_1, \cdots, x_k)| \geq m \right) \geq \mathbb{P}\left(|\{ f_i(x_1), \cdots, f_i(x_k) : 1 \leq i \leq m \}| \geq m \right)$$

$$\geq 1 - \frac{m^2}{2} e^{-k\epsilon}$$

$$\geq \frac{1}{2} \qquad \text{by } k := \left\lceil \frac{2\log m}{\epsilon} \right\rceil$$

This implies $\exists \{z_1, \cdots, z_k\}$ s.t. $|\mathcal{F}(z_1, \cdots, z_k)| \geq m$.

$$\underset{\substack{\| \\ D(\epsilon, \mathcal{C}, L_1(Q))}}{m} \leq |\mathcal{F}(z_1, \cdots, z_k)| \leq \max_{x_1, \cdots, x_k} \Delta_n(\mathcal{C}; x_1, \cdots, x_n) \overset{\text{Sauer's}}{\leq} \sum_{j=0}^{V} \binom{k}{j}$$

Case 1: $k \leq V$

then $N(\epsilon, \mathcal{C}, L_1(Q)) \leq D(\epsilon, \mathcal{C}, L_1(Q)) \leq 2^V \leq \left(\frac{2}{\epsilon}\right)^V$

Case 2: $k > V$.

$N(\epsilon, \mathcal{C}, L_1(Q)) \leq m \leq \left(\frac{ke}{V}\right)^V \qquad k = \left\lceil \frac{2\log m}{\epsilon} \right\rceil$

$$\Rightarrow \quad m \leq \left(\frac{8e}{\epsilon}\right)^V$$

This completes the proof for $r = 1$.

For $r \geq 1$, since $Q|1_C - 1_D| = \|1_C - 1_D\|_{L_1(Q)} = \|1_C - 1_D\|_{L_r(Q)}^r$

then $N(\epsilon, \mathcal{C}, L_r(Q)) = N(\epsilon^r, \mathcal{C}, L_1(Q))$ $\square$