Recall $|f(x)| \leq F(x)$ $(\forall f \in \mathcal{F}, x \in \mathcal{X})$. (Dudley).

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} |(P_n - P)f|\right] \leq C\sqrt{\frac{\mathbb{E}[F(X)^2]}{n}} \int_0^1 \sqrt{\log \sup_{Q} N(\delta \|F\|_{L^2(Q)}, \mathcal{F}, \|\cdot\|_{L^2(Q)})}\, d\delta$$

- Integral may diverge

$$\delta_0 + \int_{\delta_0}^1 \sqrt{\log \cdots\cdots}\; d\delta.$$

- In symmetrization, $\mathbb{E}[\sup (P_n - P)f] \leq \mathbb{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_1^n \varepsilon_i f(X_i)\right]$

$\leq$ Dudley integral

Gaussian complexity.

$$R_n(\mathcal{F}) \leq C \cdot \mathbb{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_1^n g_i f(X_i)\right] \quad \text{for } g_i \overset{iid}{\sim} N(0,1).$$

Application. $\Theta \subseteq \mathbb{B}_2(0, R)$ in $\mathbb{R}^d$

$$\mathcal{F} = \{f_\theta : \theta \in \Theta\} \qquad |f_{\theta_1}(x) - f_{\theta_2}(x)| \leq M(x) \cdot \|\theta_1 - \theta_2\|_2.$$

$$F(x) = R \cdot M(x) + |f_{\theta_0}(x)| \quad \text{for some } \theta_0 \in \Theta.$$

$$|f_\theta(x)| \leq F(x).$$

Take $\theta_1, \theta_2, \cdots, \theta_N$ be min-ε-covering of Θ

$$N \leq \left(1 + \frac{2R}{\varepsilon}\right)^d.$$

$\forall \theta, \ \exists j \quad \text{s.t.} \ \|\theta - \theta_j\|_2 \leq \varepsilon.$

$$\|f_\theta - f_{\theta_j}\|^2_{L^2(Q)}$$

$$= \int |f_\theta(x) - f_{\theta_j}(x)|^2 \, dQ(x)$$

$$\leq \|\theta_j - \theta\|_2^2 \cdot \int M^2(x) \, dQ(x).$$

$\varepsilon = R\delta.$

$$N(\delta \cdot \|F\|_{L^2(Q)} ; \mathcal{F}, \|\cdot\|_{L^2(Q)})$$

$$\leq N(R\delta \cdot \|M\|_{L^2(Q)} ; \mathcal{F}, \|\cdot\|_{L^2(Q)})$$

$$\leq N(R\delta ; \Theta, \|\cdot\|_2)$$

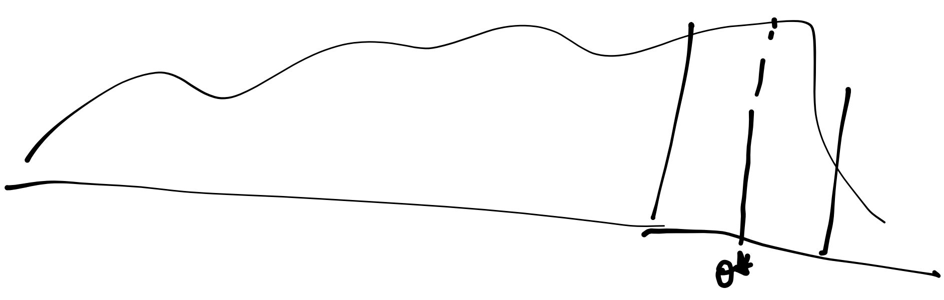
$$\leq \left(1 + \frac{2}{\delta}\right)^d.$$

$$\mathbb{E}\left[\sup_{\theta \in \Theta} |P_n f_\theta - P f_\theta|\right] \leq C \cdot \sqrt{\frac{\mathbb{E}[F(x)^2]}{n}} \int_0^1 \sqrt{d \log(1/\delta)} \, d\delta$$

$$\leq C \sqrt{\frac{d}{n}} \cdot \|F\|_{L^2(P)}.$$

---

e.g. $\quad f_\theta(x) = \mathbb{1}_{\{|x - \theta| \leq 1\}}$



$F(x) = 1.$ $N(\delta, \mathcal{F}, \|\cdot\|_{L^2(Q)}) \leq \left(\frac{c}{\varepsilon}\right)^{\delta}.$

Proof: (Probabilistic arguments).

Let $\theta_1, \theta_2, \cdots, \theta_N$ be max ε-packing under $L^2(Q)$.

$\forall i \neq j$, $\varepsilon^2 \leq \|f_{\theta_i} - f_{\theta_j}\|^2_{L^2(Q)} = \mathbb{P}(f_{\theta_i}(X) \neq f_{\theta_j}(X))$ $(X \sim Q)$.

Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} Q$.

$A = \{(f_{\theta_j}(X_1), f_{\theta_j}(X_2), \cdots, f_{\theta_j}(X_n)) \in \{0,1\}^n : j \in [N]\}$.

$\forall i \neq j$, $\mathbb{P}[\quad]$

$\leq (1 - \varepsilon^2)^n \leq \exp(-\varepsilon^2 n).$

$\mathbb{P}(\exists i, j,\ s.t.\ (f_{\theta_i}(X_1), \cdots, f_{\theta_i}(X_n)) = (f_{\theta_j}(X_1), \cdots, f_{\theta_j}(X_n)))$

$\leq \binom{N}{2} \cdot \exp(-\varepsilon^2 n).$

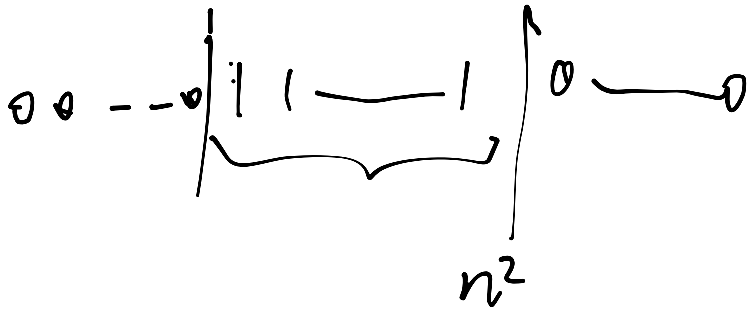
$n = \left\lceil \frac{2 \log N}{\varepsilon^2} \right\rceil$ $\mathbb{P}(\cdots) \leq \frac{1}{2}.$

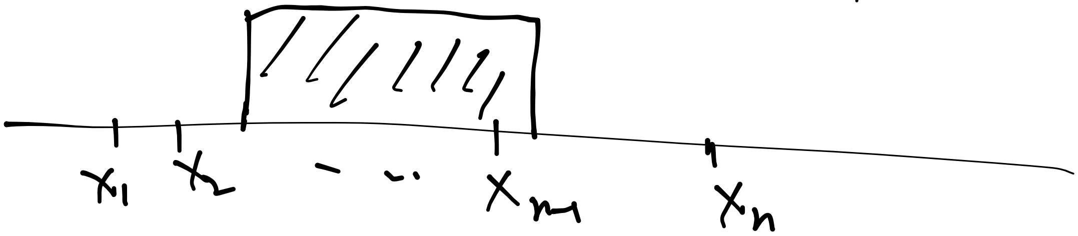
$\exists$ a deterministic seq. $x_1, x_2, \cdots, x_n$

s.t. $(f_{\theta_j}(x_1), \cdots f_{\theta_j}(x_n)) \in \{0,1\}^n$

are all distinct for $j = 1, 2, \cdots N$

(Assume w.l.o.g. $x_1 \leq x_2 \leq x_3 \cdots \leq x_n$).





$$N \le n^2 = \frac{4\log^2 N}{\varepsilon^4}.$$

$$N \le \left(\frac{c}{\varepsilon}\right)^8.$$

VC-class

VC (subgraph) dim.

Application to M-estimator.

Recall. $$F(\hat\theta_n) - F(\theta^*) = \underbrace{F(\hat\theta_n) - F_n(\hat\theta_n)}_{\le \sup_{\theta \in ?} |F_n(\theta) - F(\theta)|} + \underbrace{F_n(\hat\theta_n) - F_n(\theta^*)}_{\le 0} + \underbrace{F_n(\theta^*) - F(\theta^*)}_{LLN}$$

$\cancel{\Theta}$

e.g. $F(\theta) - F(\theta^*) \ge \|\theta - \theta^*\|_2^2$

$$\mathbb{E}\left[\sup |F_n(\theta) - F(\theta)|\right] \le \sqrt{\frac{d}{n}}$$

$$\|\theta - \theta^*\|_2 \le \left(\frac{d}{n}\right)^{1/4}.$$

$$\mathbb{E}\left[\sup_{\substack{\theta \in \Theta \\ \|\theta - \theta^*\| \le u}} \left|(P_n - P)(f_\theta - f_{\theta^*})\right|\right] \le \phi_n(u)$$

Thm (informal). $\phi_n(\delta_n) \leq \delta_n^2$ for some δ_n

then $\| \hat{\theta}_n - \theta^* \| \lesssim \delta_n$ w.h.p.
