Lecture 6: Finite-sample Analysis via Expanding the Set

Reminder:
- HW1 due Tuesday
- Lecture feedback form

Recap:

- $TV(p, p_n) = 1$ ☹
- Relax TV to $\widetilde{TV}$
- Project under $\widetilde{TV}$ to resilient dist^{n}s

works:

- $\widetilde{TV}(p, p_n) = O(\sqrt{\tfrac{d}{n}})$
- modulus bounded: "mean cross lemma"

$$\mathcal{G}_{TV}(\rho, \epsilon) = \{\text{all } (\rho,\epsilon)\text{-resilient dist}^n\text{s}\}$$

This time:

- Keep TV (even though $TV(p, p_n) = 1$)
- Expand set $\mathcal{G}$ to $\mathcal{M}$
- $\tilde{p}_n$: empirical dist^n of $\tilde{p}$
- p_n^*: empirical dist^n of p^*

$TV(p_n^*, \tilde{p}_n)$ is small

far from p^*

$\in \mathcal{M}$ ← analyze modulus for $\mathcal{M}$

$\Rightarrow$ application: bounded k^{th} moments

$k = 2$ (bounded covariance)

p^* "~~oblivious~~" $\longrightarrow \tilde{p}$

samples | "adaptive" | samples

$X_1', \dots, X_n'$ (need to couple randomness) $\longrightarrow X_1, \dots, X_n$

p_n^* $\qquad$ $\tilde{p}_n$

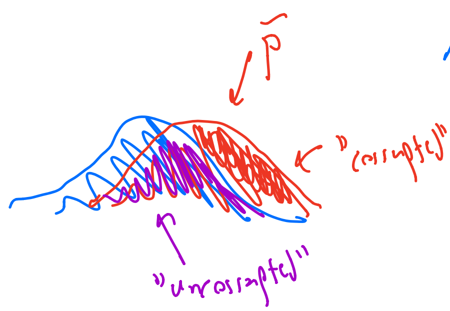
$TV(p_n^*, \tilde{p}_n) = 1$

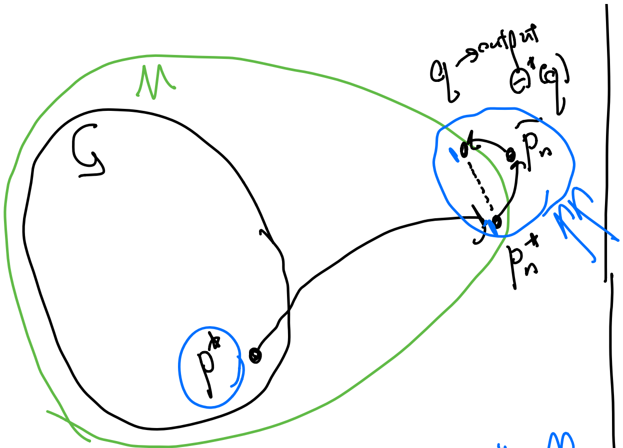
p_n^*: sample $X_1^*, \dots, X_n^* \sim p^*$

$\Updownarrow$ $TV(p_n^*, \tilde{p}_n) = \epsilon$

$\tilde{p}_n$: pick ~~ϵn~~ Binom(n, ϵ) of the X_i and allow an adversary to replace them arbitrarily (adaptive)

/ with draws from some distribution q (oblivious)





MD functional: project onto $\mathcal{M}$ under TV

Lemma. The error $L(p_n^*, \theta^*(q))$ of MD functional is at most

$$\mathfrak{m}(\mathcal{M}, \underbrace{2\,TV(p_n^*, \bar{p}_n)}_{\Delta\text{-ineq.}})$$

assuming that $p_n^* \in \mathcal{M}$.

3 things:

- $p_n^* \in \mathcal{M}$ with high probability
- $L(p_n^*, \theta) \approx L(p^*, \theta)$
- $\mathfrak{m}(\mathcal{M}, 4\varepsilon)$ small

Note: $TV(p_n^*, \bar{p}_n) = \frac{1}{n}\mathrm{Binom}(n, \varepsilon)$
$\le 2\varepsilon$ w.p. $1 - \exp(-\frac{\varepsilon n}{3})$.

$L(p_n^*, \theta) \approx L(p^*, \theta)$ ✓

⇒ "uniform convergence"

$L(p, \theta) = \|\mu(p) - \theta\|_2$

⇒ $\|\mu(p^*) - \mu(p_n^*)\|_2 \to 0$

$\mathfrak{m}(\mathcal{M}, 4\varepsilon)$: pick a good $\mathcal{M}$
(generally $\mathcal{G}_{TV}(\rho, \varepsilon)$)

$p_n^* \in \mathcal{M}$ w.h.p. ⇒ challenge

Aside. Sometimes convenient to project onto $\mathcal{M}$ but show that $p_n^* \in \mathcal{G}'$, where $\mathcal{G} \subset \mathcal{G}' \subset \mathcal{M}$

⇒ generalized modulus

$$\mathfrak{m}(\mathcal{M}, \mathcal{G}', 2\varepsilon) = \sup_{\substack{p \in \mathcal{G}' \\ q \in \mathcal{M} \\ TV(p,q) \le 2\varepsilon}} L(p, \theta^*(q))$$

$p_n^* \in \mathcal{G}'$

$q \in \mathcal{M}$

3 things:

$$\begin{cases} \bullet\ p_n^* \in \mathcal{M} \text{ with high probability} \\ \bullet\ L(p_n^*, \theta) \approx L(p^*, \theta) \\ \bullet\ \mathfrak{m}(\mathcal{M}, 2\varepsilon) \text{ small} \end{cases}$$

Note: $\mathrm{TV}(p_n^*, \tilde{p}_n) = \frac{1}{n}\mathrm{Binom}(n, \varepsilon)$

$\lesssim 2\varepsilon$ w.p. $1 - \exp\left(-\frac{\varepsilon n}{3}\right)$.

Application. Bounded k^{th} moments

"Orlicz norm"

$$\|p\|_\psi = \inf\left\{\sigma \text{ s.t. } \mathbb{E}\left[\psi\left(\frac{|\langle x-\mu, v\rangle|}{\sigma}\right)\right] \leq 1 \quad \forall \|v\|_2 = 1\right\}$$

- $\psi(x) = x^2 \Rightarrow \mathbb{E}[\langle x-\mu, v\rangle^2] \leq \sigma^2 \quad \forall \|v\|_2 = 1 \Leftrightarrow$ bounded covariance

$\psi(x) = e^{x^2} - 1 \Rightarrow$ sub-Gaussian

$\psi(x) = x^k \Rightarrow \mathbb{E}[|\langle x-\mu, v\rangle|^k] \leq \sigma^k \quad \forall \|v\|_2 = 1$ "bounded k^{th} moments"

$$\boxed{\triangleq \mathcal{G}_{\mathrm{mom},k}(\sigma)}$$

General result for Orlicz norms: constant factor??

$$\|p\|_\psi \leq \sigma \Rightarrow \left(\sigma\varepsilon\psi^{-1}\left(\tfrac{1}{\varepsilon}\right), \varepsilon\right)\text{-resilient}$$

$$\psi(x) = x^2 \Rightarrow \varepsilon\psi^{-1}\left(\tfrac{1}{\varepsilon}\right) = \varepsilon \cdot \sqrt{\tfrac{1}{\varepsilon}} = \sqrt{\varepsilon}$$

$$\psi(x) = e^{x^2} - 1 \Rightarrow \approx \varepsilon\sqrt{\log(1/\varepsilon)}$$

$$\psi(x) = x^k \Rightarrow \varepsilon \cdot \left(\tfrac{1}{\varepsilon}\right)^{\frac{1}{k}} = \varepsilon^{1-\frac{1}{k}}$$

e.g. $k = 4 \Rightarrow \varepsilon^{3/4}$ (better $\varepsilon^{1/2}$)

Bounded moments $\Rightarrow$ resilience

Naive approach: • project $\tilde{p}_n$ onto $\mathcal{G}_{mom,k}(\sigma)$

i.e. $\mathcal{M} = \mathcal{G} = \mathcal{G}_{mom,k}(\sigma)$.

Bounded k^{th} moments: $\underbrace{\mathbb{E}[|\langle x-\mu, v\rangle|^k]}$ $\le \sigma^k$

$x_1, \dots, x_n \sim p^*$

x_1 appears in p_n^* with weight $\frac{1}{n}$.

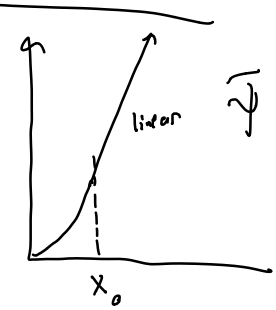
$$\frac{1}{n}\langle x_1-\mu, v\rangle^k \to \frac{1}{n}\underbrace{\|x_1-\mu\|_2}_{\approx\sqrt{d}}{}^k = \frac{1}{n}\cdot\sqrt{d}^k = \frac{d^{k/2}}{n}$$

$n \gtrsim d^{k/2}$

Goal: Get good estimates for $n \approx d$.

Idea: Replace moments w/ "truncated moments"

$$\tilde{\psi}(x) = \begin{cases} x^k & : x \le x_0 \\ x_0^k + (x-x_0)\cdot kx_0^{k-1} & : x > x_0 \end{cases}$$



$\Rightarrow \tilde{\psi}$ is L-Lipschitz for $L = kx_0^{k-1}$

$\|p_n^*\|_{\tilde{\psi}}$ is small w.h.p. $\Rightarrow$ resilience

$\to$ Ledoux-Talagrand contraction inequality

$\Rightarrow$ Concentration bounds for Lipschitz

Goal: Bound $\|p_n^*\|_{\tilde{\psi}}$

$\tilde{\psi}$ is L-Lipschitz

$X_1, \dots, X_n \sim p^*$

Show $\sup_{\|v\|_2 \le 1} \frac{1}{n} \sum_{i=1}^{n} \tilde{\psi}\left(\frac{|\langle x_i - \mu, v\rangle|}{\sigma}\right)$ small with high probability

---

Theorem (Ledoux-Talagrand)

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be L-Lipschitz, s.t. $\varphi(0) = 0$

Then for any convex, increasing function g, and Rademacher variables $\varepsilon_1, \dots, \varepsilon_n \sim \{\pm 1\}$, we have

$$\mathbb{E}_{\varepsilon_{1:n}}\left[g\left(\sup_{t \in T} \sum_{i=1}^{n} \varepsilon_i \varphi(t_i)\right)\right] \le \mathbb{E}_{\varepsilon_{1:n}}\left[g\left(L \cdot \sup_{t \in T} \sum_{i=1}^{n} \varepsilon_i t_i\right)\right]$$

for any $T \subseteq \mathbb{R}^n$.

$t_i = \frac{|\langle x_i - \mu, v\rangle|}{\sigma}$

$\varphi(t) = \tilde{\psi}(t)$

$g(z) = z$

$$\mathbb{E}_{\varepsilon_{1:n}}\left[\sup_{\|v\|_2 \le 1} \sum_{i=1}^{n} \varepsilon_i \tilde{\psi}\left(\frac{|\langle x_i - \mu, v\rangle|}{\sigma}\right)\right] \le \frac{L}{\sigma} \cdot \mathbb{E}_{\varepsilon_{1:n}}\left[\sup_{\|v\|_2 \le 1} \sum_{i=1}^{n} \varepsilon_i |\langle x_i - \mu, v\rangle|\right.$$

Define $\mu_{\tilde{\psi}}(v) = \mathbb{E}_{x \sim p^*}\left[\tilde{\psi}\left(\frac{|\langle x - \mu, v\rangle|}{\sigma}\right)\right]$ $\Rightarrow \le 1$.

Note: $\mu_{\tilde{\psi}}(v) \le 1$ since $\|p^*\|_{\psi} \le \|p^*\|_{\tilde{\psi}} \le \sigma$

$$\underbrace{\sup_{\|v\|_2 \le 1}}_{\text{uniform convergence}} \left| \underbrace{\frac{1}{n}\sum_{i=1}^{n} \tilde{\varphi}\left(\frac{|\langle x_i-\mu, v\rangle|}{\sigma}\right)}_{\text{finite-sample}} - \underbrace{\mu_{\tilde{\varphi}}(v)}_{\text{infinite-sample}} \right|$$

$$\mathbb{E}_{X_1,\dots,X_n}\left[\sup_{\|v\|_2\le 1}\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{\varphi}\left(\frac{|\langle x_i-\mu,v\rangle|}{\sigma}\right)-\mu_{\tilde{\varphi}}(v)\right|^K\right]$$

$$\le \mathbb{E}_{X_1,\dots,X_n,X_1',\dots,X_n'}\left[\sup_{\|v\|_2\le 1}\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{\varphi}\left(\frac{|\langle x_i-\mu,v\rangle|}{\sigma}\right)-\tilde{\varphi}\left(\frac{|\langle x_i'-\mu,v\rangle|}{\sigma}\right)\right|^K\right]$$

$\mathbb{E}[\cdot]$

$$= \mathbb{E}_{X,X',\varepsilon}\left[\sup_{\|v\|_2\le 1}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left(\tilde{\varphi}\left(\frac{|\langle x_i-\mu,v\rangle|}{\sigma}\right)-\tilde{\varphi}\left(\frac{|\langle x_i'-\mu,v\rangle|}{\sigma}\right)\right)\right|^K\right]$$

Δ-ineq. $\le \mathbb{E}[\sup|\xi_1| + |\xi_2|]$

$$\le 2^K\,\mathbb{E}_{X,\varepsilon}\left[\sup_{\|v\|_2\le 1}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\tilde{\varphi}\left(\frac{|\langle x_i-\mu,v\rangle|}{\sigma}\right)\right|^K\right]$$

$g(z) = z^K$

Ledoux-Talagrand

$$\le 2\,\mathbb{E}_{X,\varepsilon}\left[\sup_{\|v\|_2\le 1}\left|\frac{L}{\sigma}\cdot\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\langle x_i-\mu,v\rangle\right|^K\right]$$

$$= \left(\frac{2L}{\sigma}\right)^K\mathbb{E}_{X,\varepsilon}\left[\sup_{\|v\|_2\le 1}\left|\left\langle\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(x_i-\mu), v\right\rangle\right|^K\right]$$

$$= \left(\frac{2L}{\sigma}\right)^K\mathbb{E}_{X,\varepsilon}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(x_i-\mu)\right\|_2^K\right]$$