Lecture 7: Ledoux-Talagrand

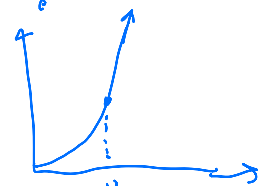
Pset 1 Due Today
Pset 2 Released tonight

Recap.

- Expand the set $\mathcal{G}$ to $\mathcal{M}$
  - Bound modulus for $\mathcal{M}$
  - Show $p_n^* \in \mathcal{M}$ w.h.p.
  - → Bound $\|\hat{\mu}_n - \mu\|_2$
- Particular example
  - $\mathcal{G} = \mathcal{G}_{mom,k}(\sigma)$ = bounded k^{th} moments
    - Issue: needed $n \gtrsim d^{k/2}$ for $p_n^* \in \mathcal{G}$
  - $\mathcal{M} = \mathcal{G}_{TV}(\mathcal{G}, \varepsilon)$, $\rho = \mathcal{O}(\varepsilon^{1-\frac{1}{k}})$

- Show $p_n^* \in \mathcal{M}$ by bounding truncated moments

$$\tilde{\psi}_k(x) = \begin{cases} x^k & : x \leq x_0 \\ x_0^k + k x_0^{k-1}(x - x_0) & : x > x_c \end{cases}$$



L-Lipschitz
$(L = k x_0^{k-1})$

⟹ invoke Ledoux-Talagrand + symmetrization

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

zero-mean

Today:

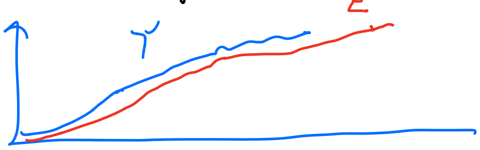
- Prove Ledoux-Talagrand
- Finish bounding ◯ using Rosenthal's inequality + Khintchine's inequality
  - ⟹ finite-sample concentration for truncated moments

Warm-up. Stochastic dominance Y, Z r.v.'s on $\mathbb{R}$

Def^{n}: Z (first-order) stochastically dominates Y if

$$\mathbb{E}[f(Y)] \leq \mathbb{E}[f(Z)] \text{ for all increasing functions } f.$$

Intuition. Push everything in Y to the right.

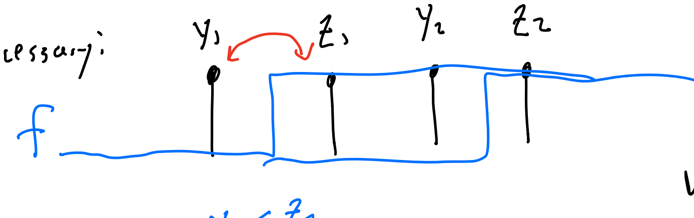


Lemma. $Y = \frac{1}{2}$ prob. of y_1 or y_2 $(y_1 \le y_2)$

$Z = \frac{1}{2}$ prob. of z_1 or z_2 $(z_1 \le z_2)$

Then Z 1st-order s.d.'s Y iff $z_2 \ge y_2$ and $z_1 \ge y_1$.

Proof. Necessary:



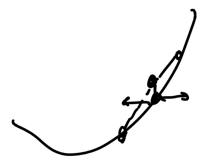
$y_1 \le z_1$

Sufficient: $\mathbb{E}[f(Z)] = \frac{1}{2} f(z_1) + \frac{1}{2} f(z_2) \ge \frac{1}{2} f(y_1) + \frac{1}{2} f(y_2) = \mathbb{E}[f(Y)]$. □

Definition Z second-order stochastically dominates Y if

$\mathbb{E}[g(Y)] \le \mathbb{E}[g(Z)]$ for all convex, increasing g.

- Second-order dominance: weaker than first-order dominance
- intuition: push to right, then spread out



Lemma. $Y = \frac{1}{2}$ prob. of y_1 or y_2 $(y_1 \le y_2)$

$Z = \frac{1}{2}$ prob. of z_1 or z_2 $(z_1 \le z_2)$

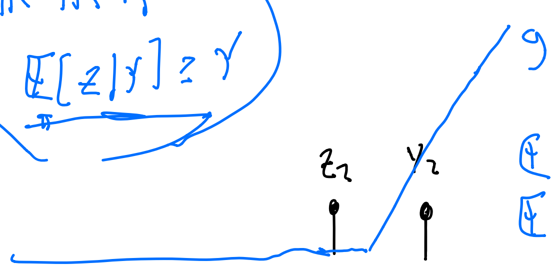
Then Z 2nd-order s.d.'s Y iff $\frac{z_1 + z_2}{2} \ge \frac{y_1 + y_2}{2}$, $\underbrace{\max(z_1, z_2)}_{z_2} \ge \underbrace{\max(y_1, y_2)}_{y_2}$.

$g(x) = x$

$\pi \in \Pi(Y, Z)$

$\mathbb{E}_\pi[Z \mid Y] \ge Y$

Proof.



$\mathbb{E}[g(Y)] > 0$

$\mathbb{E}[g(Z)] = 0$

necessary

Sufficiency.

$$\frac{y_2 - z_1}{z_2 - z_1} g(z_2) + \frac{z_2 - y_2}{z_2 - z_1} g(z_1) \geq g(y_2)$$

$$\frac{z_2 - y_2}{z_2 - z_1} g(z_2) + \frac{y_2 - z_1}{z_2 - z_1} g(z_1) \geq g(z_1 + z_2 - y_2)$$

$$g(z_2) + g(z_1) \geq g(y_2) + g(z_1 + z_2 - y_2) \geq g(y_2) + g(y_1) \quad \blacksquare$$

Thm (Ledoux-Talagrand)

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

$\varepsilon_1, \dots, \varepsilon_n \sim \text{Uniform}(\{\pm 1\})$

T collection of n-tuples $(t_1, \dots, t_n)$ ← $\langle x_i - \mu, v \rangle$

Then, for all convex, increasing g,

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

Interpretation. φ Lipschitz ($L = 1$ WLOG)

$\sup_{t \in T} \sum_i \varepsilon_i \varphi(t_i)$ s.d.'d by $\sup_{t \in T} \sum_i \varepsilon_i t_i$

Without sup: $\sum_i \varepsilon_i \varphi(t_i) = N(0, \sum_i \varphi(t_i)^2)$ (with $|\varphi(t_i)| \leq |t_i|$) and $N(0, \sum_i t_i^2)$

$\varepsilon_i \varphi(t_i)$: 0-mean random variable $\Rightarrow \sum_i$

Wish sup. Supremum of "sub-Gaussian" quantities
→ all correlated

Ledoux-Talagrand: handling correlations as long as φ is Lipschitz

Proof of Ledoux-Talagrand.

$\varphi(0)=0$

a_+, a_-: largest value of a

Warm-up result: $(n=2)$

✓ **Claim.** $\mathbb{E}_\varepsilon\left[g\left(\sup_{(a,b)\in T} a + \varepsilon\varphi(b)\right)\right] \le \mathbb{E}_\varepsilon\left[g\left(\sup_{(a,b)\in T} a + \varepsilon b\right)\right]$

only (a_+, b_+) (a_-, b_-) matter

no ε

a vs. $\varphi(a)$

sup over all T

≥ sup over $\{(a_+,b_+), (a_-,b_-)\}$

Only need to worry about two points:

(a_+, b_+): maximizer of $a+\varphi(b)$

(a_-, b_-): maximizer of $a-\varphi(b)$

$(L=1)$

$Y_1 = a_- - \varphi(b_-)$

$Y_2 = a_+ + \varphi(b_+)$

$\frac{1}{2}$ mixture of δ_{Y_1} and δ_{Y_2}

↓ sub.'d

$Z_1 = \max(a_- - b_-, a_+ - b_+)$

$Z_2 = \max(a_- + b_-, a_+ + b_+)$

$\frac{1}{2}$ mixture of δ_{Z_1} and δ_{Z_2}

✓ $\frac{Y_1+Y_2}{2} \le \frac{Z_1+Z_2}{2}$ and $\max(Y_1, Y_2) \le \max(Z_1, Z_2)$ ✓

somehow make this into absolute values

$$\frac{a_- + a_+ + (\varphi(b_+) - \varphi(b_-))}{2} \le \frac{\max(a_+ - b_+,\ a_- - b_-) + \max(a_- + b_-,\ a_+ + b_+)}{2}$$

$\le |b_+ - b_-|$

$a_+ + a_- + b_- - b_+$ $\quad$ $a_+ + a_- + b_+ - b_-$

$a_+ + a_- + |b_+ - b_-|$

$$\max(a_- - \varphi(b_-),\ a_+ + \varphi(b_+)) \le \max(a_- - b_-,\ a_+ - b_+,\ a_- + b_-,\ a_+ + b_+)$$

$$= \max(a_- + |b_-|,\ a_+ + |b_+|)$$

$\varphi(0) = 0 \Rightarrow |\varphi(b)| \le |b|$

$$\le \max(a_- + |b_-|,\ a_+ + |b_+|)$$

Claim. $\mathbb{E}_\varepsilon\left[g\left(\sup_{(a,b)\in T} a + \varepsilon\varphi(b)\right)\right] \le \mathbb{E}_\varepsilon\left[g\left(\sup_{(a,b)\in T} a + \varepsilon b\right)\right]$

Applying the claim

$$\mathbb{E}_{\varepsilon_{1:n-1}}\mathbb{E}_{\varepsilon_n}\left[g\left(\sup_{t\in T}\sum_{i=1}^{n-1}\varepsilon_i\varphi(t_i) + \varepsilon_n\varphi(t_n)\right)\middle|\ \varepsilon_{1:n-1}\right]$$

a $\quad$ $\varepsilon \cdot \varphi(b)$

$$\leq \mathbb{E}_{\varepsilon_{1:n-1}}\left[\mathbb{E}_{\varepsilon_n}\left[g\left(\sup_{t\in T}\sum_{i=1}^{n-1}\varepsilon_i\varphi(t_i) + \varepsilon_n t_n\right)\right]\right]$$

$$= \mathbb{E}_{\varepsilon_{1:n-2},\,\varepsilon_n}\left[\mathbb{E}_{\varepsilon_{n-1}}\left[g\left(\underbrace{\sup_{t\in T}\sum_{i=1}^{n-2}\varepsilon_i\varphi(t_i) + \varepsilon_n t_n}_{a} + \underbrace{\varepsilon_{n-1}\varphi(t_{n-1})}_{\varepsilon\cdot\varphi(b)}\right)\right]\right]$$

$\downarrow$ $\varepsilon_{n-1} t_{n-1}$

$\vdots$

$$= \mathbb{E}\left[g\left(\sup_{t\in T}\sum_{i=1}^{n}\varepsilon_i t_i\right)\right]. \checkmark$$

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

$$\Rightarrow \text{Bend} \quad \mathbb{E}_{X_{1:n},\,\varepsilon_{1:n}}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(X_i-\mu)\right\|_2^K\right]$$

Note, without ε_i, $\mathbb{E}_{X_{1:n}}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right\|_2^K\right]$

$$= \mathbb{E}_{X_{1:n}}\left[\|\hat{\mu}_n-\mu\|_2^K\right] \leftarrow$$

→ Annoying b/c ℓ^2-norm

Key. "de-coupling inequality" Khintchine's inequality

$\exists$ constants A_k, B_k s.t.

$$A_k \|z\|_2 \le \mathbb{E}_{\varepsilon'}\left[\left|\sum_i \varepsilon_i' z_i\right|^k\right]^{1/k} \le B_k \|z\|_2$$

$$\mathbb{E}_{\varepsilon'}\left[|\langle \varepsilon', z\rangle|^k\right]^{1/k}$$

$A_k = \Theta(1)$

$B_k = \Theta(\sqrt{k})$

$$\mathbb{E}_{X_{1:n}, \varepsilon_{1:n}}\left[\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i (X_i - \mu)\right\|_2^k\right]^{1/k}$$

$$\le O(1) \cdot \mathbb{E}_{X, \varepsilon, \varepsilon'}\left[\left|\frac{1}{n}\sum_{i=1}^n \varepsilon_i \underbrace{\langle X_i - \mu, \varepsilon'\rangle}_{z_i(\varepsilon')}\right|^k\right]^{1/k}$$

Fix ε': zero-mean

$$\mathbb{E}_{\varepsilon'}\left[\mathbb{E}_{X,\varepsilon}\left[\left|\frac{1}{n}\sum_{i=1}^n z_i(\varepsilon')\right|^k\right]^{1/k}\right.$$

→ applying Rosenthal's inequality

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^n z_i(\varepsilon')\right|^k\right]^{1/k} \le O\left(\frac{k}{n}\right) \mathbb{E}\left[\sum_i |z_i(\varepsilon')|^k\right]^{1/k}$$

$1/2$

Rosenthal

$$+ \mathcal{O}\left(\frac{\sqrt{k}}{n}\right) \cdot \left(\sum_i \mathbb{E}\left[z_i(\xi')^2\right]\right)$$

....