Mean Field Asymptotics in Statistical Learning.

Mar 10.

Lecture 14. Gordon's inequality.

① Gordon's inequality.

Theorem [Gordon't inequality] Let S, T be two finite set.

Let $(X_{st})_{s\in S, t\in T}$, $(Y_{st})_{s\in S, t\in T}$ be two Gaussian processes.

$\mathbb{E}[X_{st}] = \mathbb{E}[Y_{st}] = 0$, $\quad (\tilde{X}_{st})_{st} \quad (\tilde{Y}_{st}) \quad \mathbb{E}\tilde{X}_{st} = \mathbb{E}\tilde{Y}_{st} = Q(s,t)$

$$\begin{cases} \mathbb{E}[(X_{st_1} - X_{st_2})^2] \le \mathbb{E}[(Y_{st_1} - Y_{st_2})^2], & \forall\, t_1, t_2 \in T,\ s \in S. \\ \mathbb{E}[(X_{s_1t_1} - X_{s_2t_2})^2] \ge \mathbb{E}[(Y_{s_1t_2} - Y_{s_2t_2})^2], & \forall\, s_1 \ne s_2 \in S,\ t_1, t_2 \in T. \end{cases}$$

Then ① $\forall$ determistic function $Q(s,t)$

$$\mathbb{E}\Big[\min_{s\in S}\max_{t\in T} X_{st} + Q(s,t)\Big] \le \mathbb{E}\Big[\min_{s\in S}\max_{t\in T} Y_{st} + Q(s,t)\Big].$$

② If we further assume $\mathbb{E}[X_{st}^2] = \mathbb{E}[Y_{st}^2]$. then

$\forall\, \tau \in \mathbb{R}$, function $Q(s,t)$. we have

$$\mathbb{P}\Big(\min_{s\in S}\max_{t\in T}(X_{st} + Q(s,t)) \ge \tau\Big)$$
$$\le \mathbb{P}\Big(\min_{s\in S}\max_{t\in T}(Y_{st} + Q(s,t)) \ge \tau\Big).$$

Remark 1: $|S| = 1$, this reduces to Slepian / Sukdov-Fernique.

Remark 2: S, T are compact sets of Euclidean space

$Q(s,t)$ are continous.

Proof of ②: Take $f(x) = \prod_{s\in S}\left[1 - \prod_{t\in T} \mathbb{1}\{X_{st} + Q(s,t) \leq \tau\}\right]$

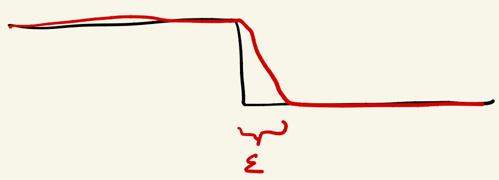
$$\Rightarrow \mathbb{E}[f(X)] = \mathbb{P}\left(\min_{s\in S}\max_{t\in T}(X_{st} + Q(s,t)) \geq \tau\right).$$

Want to apply Kahane's inequality.

Problem: $\mathbb{1}\{x \leq \tau\}$ is not smooth.

The $\psi \in C^\infty(\mathbb{R})$ with $\psi(t) = 1$, $t < 0$, $\psi(t) = 0$, $t \geq 1$, $\psi'(t) \geq 0$.

Take $f_\varepsilon(x) = \prod_{s\in S}\left[1 - \prod_{t\in T}\psi((X_{st} - Q(s,t) - \tau)/\varepsilon)\right]$



Define $n = |S|\cdot|T|$,

$A = \{((s_1,t_1),(s_2,t_2)) : s_1 \neq s_2 \in S,\ t_1, t_2 \in T\}$.

$B = \{((s,t_1),(s,t_2)) : s \in S,\ t_1, t_2 \in T\}$.

$\Rightarrow \forall (i,j) \in A, \quad \partial_i\partial_j f(x) \geq 0$

$\forall (i,j) \in B, \quad \partial_i\partial_j f(x) \leq 0$.

Apply Kahane's Lemma (see Lecture 13).

$$\Rightarrow \mathbb{E}[f_\varepsilon(X)] \leq \mathbb{E}[f_\varepsilon(Y)]$$

Take $\varepsilon \to 0 \Rightarrow \mathbb{P}\left(\min_{s\in S}\max_{t\in T} X_{st} + Q(s,t) \geq \tau\right)$

$$\leq \mathbb{P}\left(\min_{s\in S}\max_{t\in T} Y_{st} + Q(s,t) \geq \tau\right). \qquad \square$$

② Example (Smallest singular value Gaussian R.M).

Thm: Let $A \in \mathbb{R}^{m\times n}$ $m > n$, $A_{ij} \sim$ iid $N(0,1)$.

Denote $\sigma_{\min}(A) = \min_{u\in S^{n-1}} \max_{v\in S^{m-1}} \langle v, Au\rangle$.

$\uparrow$ least singular value of A

Then $\mathbb{E}[\sigma_{\min}(A)] \geq \sqrt{m} - \sqrt{n}$.

Remarks: $\mathbb{E}[\sigma_{\max}(A)] \leq \sqrt{m} + \sqrt{n}$ $\quad 0<r<1$

$X \in \mathbb{R}^{m\times n}$ $\quad X = A/\sqrt{m}$ $\quad \frac{n}{m} \to r.$ $n\to\infty$.

$$1-\sqrt{r} \leq \mathbb{E}[\sigma_{\min}(X)] \leq \mathbb{E}[\sigma_{\max}(X)] \leq 1+\sqrt{r}.$$

$$1-\sqrt{r} = \lim_{\substack{n\to\infty\\ \frac{n}{m}\to r}} \sigma_{\min}(X) \leq \lim_{\substack{n\to\infty\\ \frac{n}{m}\to r}} \sigma_{\max}(X) = 1+\sqrt{r}.$$

When $m \gg n$, $\|X^TX - I_n\|_{op} = \max\{|\sigma_{\max}^2(X) - 1|, |\sigma_{\min}^2(X) - 1|\}$

$$= O(\sqrt{r}) \simeq O\left(\sqrt{\tfrac{n}{m}}\right).$$

Proof: $S = S^{n-1}$, $T = S^{m-1}$, $Y_{uv} = \langle v, Au\rangle$

$\mathbb{E}[Y_{uv}] = 0$

$\mathbb{E}[(Y_{uv} - Y_{zw})^2] = 2 - 2\langle v, z\rangle\langle u, w\rangle$

Define $(X_{uv})_{u\in S, v\in T}$ $\quad X_{uv} = \langle g, u\rangle + \langle h, v\rangle$

$g \sim N(0, I_n)$ $\quad h \sim N(0, I_m)$ $\quad g \perp h$.

$\mathbb{E}[X_{uv}] = 0$

$\mathbb{E}[(X_{uv} - X_{zw})^2] = 4 - 2\langle u, w\rangle - 2\langle v, z\rangle$.

$$\mathbb{E}[(X_{uv} - X_{zw})^2] - \mathbb{E}[(Y_{uv} - Y_{zw})^2]$$

$$= 2(1 - \langle u, w\rangle)(1 - \langle v, z\rangle).$$

"$\leq$" is enough

1) When $u = w \in S^{n-1}$

$$\mathbb{E}[(X_{uv} - X_{wz})^2] = \mathbb{E}[(Y_{uv} - Y_{uz})^2]$$

2) When $u \neq w$

$\underline{\qquad\qquad\qquad}$ $\geq \mathbb{E}[\,\cdots\,]$

$$\Rightarrow \mathbb{E}[\sigma_{\min}(X)] \geq \mathbb{E}\left[\min_{u\in S^{n-1}} \max_{v\in S^{m-1}} \langle g, u\rangle + \langle h, v\rangle\right]$$

$$= \mathbb{E}[\|h\|_2] - \mathbb{E}[\|g\|_2] \qquad h \sim N(0, I_m)$$
$$\quad \gtrsim \sqrt{m} \qquad\quad \approx \sqrt{n} \qquad\qquad g \sim N(0, I_n)$$

$$\square$$

$$\gtrsim \sqrt{m} - \sqrt{n}$$

$$\boxed{\begin{array}{l} f(m) \equiv \mathbb{E}_{h \sim N(0, I_m)}[\|h\|_2] - \sqrt{m} \\ \\ f(m) \text{ is increasing in } m. \end{array}}$$

$$\mathbb{E}[\|h\|_2] \leq \sqrt{\mathbb{E}[\|h\|_2^2]}$$

concentration

$$m = \mathbb{E}[\|h\|_2^2] \approx \|h\|_2^2$$

③ Convex Gaussian minimax theorem (CGMT). [Stojnic, 2013] [Thrampoulidis, Oymak, Hassibi, 2015].

Many problems can be written as

$$\Phi(G) \equiv \min_{u \in S_u} \max_{v \in S_v} \langle u, Gv \rangle + \psi(u,v).$$

where $S_u \subseteq \mathbb{R}^m$, $S_v \subseteq \mathbb{R}^n$, $G \in \mathbb{R}^{m \times n}$, $G_{ij} \sim_{iid} N(0,1)$.

$\psi: \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$.

Example: LASSO problem

$y = Ax_0 + w$, $A_{ij} \sim N(0, \frac{1}{n})$.

$$L \equiv \min_{x \in \mathbb{R}^d} \frac{1}{2} \| y - Ax \|_2^2 + \lambda \|x\|_1$$

$$= \min_{x \in \mathbb{R}^d} \max_{v \in \mathbb{R}^n} \langle v, A(x_0 - x) \rangle + \langle v, w \rangle - \frac{1}{2} \|v\|_2^2 + \lambda \|x\|_1$$

$$= \min_{u \in \mathbb{R}^d} \max_{v \in \mathbb{R}^n} \langle v, \underset{\substack{\uparrow \\ \text{Need to be rescaled.}}}{A} u \rangle + \underbrace{\langle v, w \rangle - \frac{1}{2} \|v\|_2^2 + \lambda \| x_0 - u \|_1}_{\psi(u,v).}$$

The Gaussian process to be compared with.

$$\phi(g,h) \equiv \min_{u \in S_u} \max_{v \in S_v} \|u\|_2 \langle g, v \rangle + \|v\|_2 \langle h, u \rangle + \psi(u,v).$$

$g \sim N(0, I_n)$, $h \sim N(0, I_m)$. Simpler to analyze.

Thm (CGMT). Let $S_u \subseteq \mathbb{R}^m$, $S_v \subseteq \mathbb{R}^n$ be compact sets, ψ be continous on $S_u \times S_v$, and $G \in \mathbb{R}^{m \times n}$, $g \in \mathbb{R}^n$, $h \in \mathbb{R}^m$, with $G_{ij}, g_j, h_i \sim$ i.i.d $N(0,1)$.

Define $$\Phi(G) = \min_{u \in S_u} \max_{v \in S_v} \langle u, Gv \rangle + \psi(u,v) \quad \text{(PO)}$$

$$\phi(g,h) = \min_{u \in S_u} \max_{v \in S_v} \|u\|_2 \langle g, v \rangle + \|v\|_2 \langle h, u \rangle + \psi(u,v) \quad \text{(AO)}$$

(a) $\forall \tau \in \mathbb{R}$

$$\mathbb{P}(\Phi(G) \le \tau) \le \underline{2}\, \mathbb{P}(\phi(g,h) \le \tau)$$

(b) Further assume that S_u, S_v are convex set and ψ is convex-concave on $S_u \times S_v$ (Strong duality hold)

Then for all $\tau \in \mathbb{R}$.

$$\mathbb{P}(\Phi(G) \geq \tau) \leq 2\,\mathbb{P}(\phi(g,h) \geq \tau).$$

In particular, $\forall \mu \in \mathbb{R}$, $t > 0$,

δ $\quad t = \sqrt{\frac{\log(c/\delta)}{n}}$

$$\mathbb{P}(|\Phi(G) - \mu| > t) \leq 2 \cdot \mathbb{P}(|\phi(g,h) - \mu| \geq t).$$

$Var(X_{uv}) \neq Var(Y_{uv})$

Proof:

a) $X_{uv} = \langle u, Gv \rangle + z\|u\|_2\|v\|_2$ $\quad z \sim N(0,1)$.

$Y_{uv} = \|u\|\langle g, v\rangle + \|v\|\langle h, u\rangle$

$$\boxed{\mathbb{E}[\min_u \max_v (X_{uv} + \psi(u,v))] \leq \mathbb{E}[\min_u \max_v (Y_{uv} + \psi(u,v))]}$$

$\mathbb{E}[X_{uv}] = \mathbb{E}[Y_{uv}] = 0$.

$\mathbb{E}[X_{uv}^2] = \|uv\|_F^2 + \|u\|_2^2\|v\|_2^2 = 2\|u\|_2^2\|v\|_2^2$

$\mathbb{E}[Y_{uv}^2] = 2\|u\|_2^2\|v\|_2^2$

$$\Delta = \{\mathbb{E}[(Y_{u_1v_1} - Y_{u_2v_2})^2] - \mathbb{E}[(X_{u_1v_1} - X_{u_2v_2})^2]\}/2$$

$$= (\|u_1\|_2\|u_2\|_2 - \langle u_1, u_2\rangle)(\|v_1\|_2 \cdot \|v_2\|_2 - \langle v_1, v_2\rangle).$$

By Gordon

$$\mathbb{P}(\min_u \max_v (X_{uv} + \psi(u,v)) \geq \tau) \leq \underbrace{\mathbb{P}(\min_u \max_v (Y_{uv} + \psi(u,v)) \geq \tau)}_{\mathbb{P}(\phi(g,h) \geq \tau)}$$

$$\mathbb{P}(\Phi(G) \geq \tau) = \mathbb{P}(\min_u \max_v (\langle v, Gu\rangle + \psi(u,v)) \geq \tau \mid z > 0)$$

$$\leq \mathbb{P}(\underbrace{\min_u \max_v (\langle v, Gu\rangle + z\|u\|_2\|v\|_2 + \psi(u,v)) \geq \tau}_{\{\min_u \max_v X_{uv} + \psi(u,v) \geq \tau\} \equiv \mathcal{E}} \mid z > 0)$$

$$= \mathbb{P}(\mathcal{E} \mid z > 0) = \frac{\mathbb{P}(\mathcal{E}, z > 0)}{\mathbb{P}(z > 0)} = 2\,\mathbb{P}(\mathcal{E}, z > 0)$$

$$\leq 2\,\mathbb{P}(\mathcal{E}) \leq 2\,\mathbb{P}(\phi(g,h) \geq \tau).$$

b). S_u, S_v convex, $\psi(u,v)$ convex-concave.

$$\min_u \max_v \langle u, Xv\rangle + \psi(u,v) \equiv \max_v \min_u \langle u, Xv\rangle + \psi(u,v)$$

$$= -\min_v \max_u \langle v, -X^T u\rangle + \boxed{[-\psi(u,v)]}$$

$$\mathbb{P}\left(\min_v \max_u \langle v, -X^T u\rangle + [-\psi(u,v)] \geq \tau \right)$$

$$\leq 2\, \mathbb{P}\left(\min_v \max_u \left(\|u\| < - - - \right) \geq \tau \right) \qquad \square$$

④ Example: LASSO problem.

$A_{ij} \sim N(0, \frac{1}{n})$.

$+ h \sum_{i=1}^{d} \varphi(u_i, x_{0,i})$

$$L \equiv \min_{u \in \mathbb{R}^d} \max_{v \in \mathbb{R}^n} \langle v, A u\rangle + \langle v, w\rangle - \tfrac{1}{2}\|v\|_2^2 + \lambda \|x_0 - u\|_1.$$

$$= \min_{u \in \mathbb{R}^d} \max_{v \in \mathbb{R}^n} \langle v, G u\rangle/\sqrt{n} + \underbrace{\langle v, w\rangle - \tfrac{1}{2}\|v\|_2^2 + \lambda \|x_0 - u\|_1}_{\psi(u,v)}$$

CGMT

$$\approx \min_{u \in \mathbb{R}^d} \max_{v \in \mathbb{R}^n} \|u\|_2 \langle g, v\rangle/\sqrt{n} + \|v\|_2 \langle h, u\rangle/\sqrt{n}$$
$$+ \langle v, w\rangle - \tfrac{1}{2}\|v\|_2^2 + \lambda \|x_0 - u\|_1$$

①
$$\max_v \left[\langle a, v\rangle + b\|v\|_2 - \tfrac{1}{2}\|v\|_2^2 \right] \qquad v = t \cdot \frac{a}{\|a\|_2}$$
$$= \max_{t \geq 0} \left[(\|a\|_2 + b)t - \tfrac{1}{2}t^2 \right] = \tfrac{1}{2}(\|a\| + b)_+^2$$

$$\overset{①}{=} \min_{u \in \mathbb{R}^d} \tfrac{1}{2}\left(\big\| \|u\|_2\, g/\sqrt{n} + w \big\|_2 + \langle h, u\rangle/\sqrt{n} \right)_+^2 + \lambda \|x_0 - u\|_1$$

② $g \sim N(0, I_n)$
$w \sim N(0, \sigma^2 I_n)$
$\tilde{g} \sim N(0, I_n)$

$$\overset{d\,②}{=} \min_{u \in \mathbb{R}^d} \tfrac{1}{2}\left(\left(\frac{\|u\|_2^2}{n} + \sigma^2 \right)^{\frac{1}{2}} \|\tilde{g}\|_2 + \langle h, u\rangle/\sqrt{n} \right)_+^2 + \lambda \|x_0 - u\|_1$$

③ w.h.p
$1-\varepsilon \leq \|\tilde{g}\|/\sqrt{n} \leq 1+\varepsilon$

$$\overset{③}{\approx} n \times \min_{u \in \mathbb{R}^d} \left\{ \tfrac{1}{2}\left(\left(\frac{\|u\|_2^2}{n} + \sigma^2 \right)^{\frac{1}{2}} + \frac{\langle h, u\rangle}{n} \right)_+^2 + \frac{\lambda}{n}\|x_0 - u\|_1 \right\}$$

④ $\frac{\|u\|_2^2}{n} = \zeta$

$\frac{\langle h, u\rangle}{n} = e$

$$\overset{④}{=} n \times \boxed{\min_{u \in \mathbb{R}^d}} \boxed{\min_{\zeta \geq 0, e} \max_{\mu, \nu}} \left\{ \tfrac{1}{2}\left((\zeta + \sigma^2)^{\frac{1}{2}} + e \right)_+^2 + \frac{\lambda}{n}\|x_0 - u\|_1 \right.$$
$$\left. + \frac{\mu}{2}\left(\frac{\|u\|_2^2}{n} - \zeta \right) - \nu\left(\frac{\langle h, u\rangle}{n} - e \right) \right\}$$

assume exchangable

⑤
$$\arg\min_{u \in \mathbb{R}^d} \left[\lambda \|x_0 - u\|_1 + \frac{\mu}{2}\|u\|_2^2 - \nu\langle h, u\rangle \right]$$
$$= \arg\min_{x \in \mathbb{R}^d} \left[\|x - x_0 - \tfrac{\nu}{\mu}h\|_2^2 + \tfrac{2\lambda}{\mu}\|x\|_1 \right] - x_0$$
$$= \eta\left(x_0 + \tfrac{\nu}{\mu}h\, ;\, \tfrac{\lambda}{\mu} \right) - x_0$$

$$\overset{⑤}{=} \quad n \times \min_{\substack{\zeta \geq 0 \\ e}} \max_{\mu, \nu} \Big\{ \frac{1}{2}\left((\zeta+\sigma^2)^{\frac{1}{2}} + e \right)_+^2 + \frac{\lambda}{n} \| \eta(x_0 + \tfrac{\nu}{\mu} h ; \tfrac{\lambda}{\mu}) \|_1$$

$$+ \frac{\mu}{2}\left[\frac{\| \eta(x_0 + \frac{\nu}{\mu} h ; \frac{\lambda}{\mu}) - x_0 \|_2^2}{n} - \zeta \right] - \nu \left[\frac{\langle (\eta(x_0 + \frac{\nu}{\mu} h ; \frac{\lambda}{\mu}) - x_0), h \rangle}{n} - e \right]$$

⑥ uniform Law of large numbers

$d \to \infty$, $n/d \to \delta$, $\quad \| \eta(x_0 + \frac{\nu}{\mu} h ; \frac{\lambda}{\mu}) \|_1 / d \to \mathbb{E}[| \eta(x_0 + \frac{\nu}{\mu} G ; \frac{\lambda}{\mu}) |]$.

$$\overset{⑥}{\approx} \quad n \times \min_{\substack{\zeta \geq 0 \\ e}} \max_{\mu, \nu} \Big\{ \frac{1}{2}\left((\zeta+\sigma^2)^{\frac{1}{2}} + e \right)_+^2 + \lambda \delta^{-1} \mathbb{E}[| \eta(x_0 + \tfrac{\nu}{\mu} G ; \tfrac{\lambda}{\mu}) |]$$

$$+ \frac{\mu}{2}\left[\delta^{-1} \mathbb{E} \| \eta(x_0 + \tfrac{\nu}{\mu} G ; \tfrac{\lambda}{\mu}) - x_0 \|_2^2 - \zeta \right] - \nu \left[\delta^{-1} \mathbb{E}[(\eta(x_0 + \tfrac{\nu}{\mu} G ; \tfrac{\lambda}{\mu}) - x_0) G] - e \right]$$

Taking derivative w.r.t. ζ, e, μ, ν and ignore $(\ )_+$
(Can be verified)

$$\begin{cases} \nu + (\zeta + \sigma^2)^{\frac{1}{2}} + e = 0. \\ -\frac{\mu}{2} + \left((\zeta+\sigma^2)^{\frac{1}{2}} + e \right) \times \frac{1}{2} \frac{1}{\sqrt{\zeta + \sigma^2}} = 0, \\ e = \delta^{-1} \frac{\nu}{\mu} \cdot \mathbb{E}[\eta'(x_0 + \frac{\nu}{\mu} G ; \frac{\lambda}{\mu})], \\ \zeta = \delta^{-1} \mathbb{E}[(\eta(x_0 + \frac{\nu}{\mu} G ; \frac{\lambda}{\mu}) - x_0)^2]. \end{cases}$$

Take $\quad \tau^2 = \sigma^2 + \zeta$

$\qquad \alpha = \frac{\lambda}{\tau + e}$

$\Rightarrow \quad \nu = -(\tau + e)$

$\qquad \mu = \frac{\tau + e}{\tau}$

$$\Rightarrow \begin{cases} \lambda = \alpha \cdot \tau \cdot \left(1 - \delta^{-1} \mathbb{E}[\eta'(x_0 + \tau G ; \alpha \tau)] \right) \\ \tau^2 = \sigma^2 + \delta^{-1} \mathbb{E}[(\eta(x_0 + \tau G ; \alpha \tau) - x_0)^2]. \end{cases}$$

$$\frac{L}{n} \to \frac{\lambda^2}{2\alpha_*^2} + \lambda \delta^{-1} \mathbb{E}[| \eta(x_0 + \tau_* G ; \alpha_* \tau_*) |]$$

⑤ Technical steps in applying CGMT.

(A) Figure out how to transform the original problem to primary problem.

(B) Figure out the auxiliary optimization problem.

(C) Calculate the auxiliary problem.

(D) Extract information from the auxiliary problem.