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 with $\mathbb{E}[X_{st}] = \mathbb{E}[Y_{st}] = 0$, and

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

Then ① For any deterministic function $Q(s,t)$

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

② If we further have $\mathbb{E}[X_{st}^2] = \mathbb{E}[Y_{st}^2]$, then $\forall \tau \in \mathbb{R}$, function $Q(s,t)$, we have

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

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

Remark 2: This thm also holds when S, T are compact subsets of Euclidean spaces, and $Q(s,t)$ is a continous function.

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

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

Want to apply Kahane's inequality.

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

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

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

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

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

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

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

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

Apply Kahane's Lemma (see Lecture 13).

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

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

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

② Example (Smallest sigular value of Gaussian R.M.)

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

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

the smallest sigular value of matrix A.

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

Remark: If $X\in\mathbb{R}^{m\times n}$, $X_{ij}\sim$ iid $N(0,\frac{1}{m})$, $\quad m>n$,

$n/m\equiv r.$

we have shown that

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

Actually, one can show that a.s.

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

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

Proof: $S = S^{m-1}$, $T = S^{n-1}$, $Y_{uv} = \langle u, Av \rangle$.

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

$$\mathbb{E}[(Y_{uv} - Y_{wz})^2] = \mathbb{E}[\langle A^T, vu^T - zw^T \rangle^2] = \|vu^T - zw^T\|_F^2$$

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

The next step requires some creativity.

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

where $g \sim N(0, I_m)$ and $h \sim N(0, I_n)$ $g \perp h$

Then $\mathbb{E}[\ \ _{uv}] = 0$,

$$\mathbb{E}[(X_{uv} - X_{wz})^2] = \mathbb{E}[\langle g, u-w \rangle^2] + \mathbb{E}[\langle h, v-z \rangle^2]$$

$$= \|u-w\|_2^2 + \|v-z\|_2^2 = 4 - 2\langle u, w \rangle - 2\langle v, z \rangle.$$

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

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

To apply Gordon, "$\leq$" is enough

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

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

2) When $u \neq w$.

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

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

$$= \underbrace{\mathbb{E}[\|g\|_2]}_{\approx \sqrt{m}} - \underbrace{\mathbb{E}[\|h\|_2]}_{\approx \sqrt{n}}$$

$$\geq \sqrt{m} - \sqrt{n}.$$

$\boxed{\text{since } \mathbb{E}[\|g\|_2^2] - \sqrt{n} \text{ is increasing in } n}$

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

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

$$\begin{aligned} L &\equiv \min_{x \in \mathbb{R}^d} \tfrac{1}{2} \| y - Ax \|_2^2 + \lambda \|x\|_1 \\ &= \min_{x \in \mathbb{R}^d} \max_{v \in \mathbb{R}^n} \langle v, A(x_0 - x) \rangle + \langle v, w \rangle - \tfrac{1}{2}\|v\|_2^2 + \lambda \|x\|_1 \\ &= \min_{u \in \mathbb{R}^d} \max_{v \in \mathbb{R}^n} \langle v, \underset{\uparrow \text{ Need to be rescaled.}}{A} u \rangle + \underbrace{\langle v, w \rangle - \tfrac{1}{2}\|v\|_2^2 + \lambda \| x_0 - u \|_1}_{\psi(u,v).} \end{aligned}$$

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

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

(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 $S_w \times S_u$ (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, $\forall \mu \in \mathbb{R}$, $t > 0$,

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

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

$Y_{uv} = \|u\|_2 \cdot \langle g, v \rangle + \|v\|_2 \cdot \langle h, u \rangle.$

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

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

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

$$\begin{aligned}
\Delta &\equiv \{\mathbb{E}[(Y_{u_1v_1} - Y_{u_2v_2})^2] - \mathbb{E}[(X_{u_1v_1} - X_{u_2v_2})^2]\}/2 \\
&= \mathbb{E}[X_{u_1v_1} X_{u_2v_2}] - \mathbb{E}[Y_{u_1v_1} Y_{u_2v_2}] \\
&= \langle u_1, u_2 \rangle \langle v_1, v_2 \rangle + \|u_1\|_2 \|u_2\|_2 \|v_1\|_2 \|v_2\|_2 \\
&\quad - \langle u_1, u_2 \rangle \|v_1\|_2 \|v_2\|_2 - \langle v_1, v_2 \rangle \|u_1\|_2 \|u_2\|_2 \\
&= (\|u_1\|_2 \|u_2\|_2 - \langle u_1, u_2 \rangle)(\|v_1\|_2 \|v_2\|_2 - \langle v_1, v_2 \rangle).
\end{aligned}$$

"≤ 0" is enough to apply Gordon.

1) When $u_1 = u_2$, $\quad \Delta = 0,$

2) When $u_1 \neq u_2$, $\quad \Delta \geq 0.$

By Gordon:

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

$$\begin{aligned}
\mathbb{P}(\Phi(G) \geq \tau) &= \mathbb{P}(\min_u \max_v (\langle v, Gu \rangle + \psi(u,v)) \geq \tau \mid z < 0) \\
&\leq \mathbb{P}(\min_u \max_v (\langle v, Gu \rangle + z\|u\|_2 \|v\|_2 + \psi(u,v)) \geq \tau \mid z < 0) \\
&= 2 \cdot \mathbb{P}(\min_u \max_v (\langle v, Gu \rangle + z\|u\|_2 \|v\|_2 + \psi(u,v)) \geq \tau, z < 0) \\
&\leq 2 \cdot \mathbb{P}(\min_u \max_v (X_{uv} + \psi(u,v)) \geq \tau) \\
&\leq 2 \cdot \mathbb{P}(\phi(g,h) \geq \tau).
\end{aligned}$$

(b) When S_u, S_v are convex and $\psi(u,v)$ is convex-concave.

Strong duality holds.

$$\min_{u\in S_u}\max_{v\in S_v} \langle v, Gu\rangle + \psi(u,v) = \max_{v\in S_v}\min_{u\in S_u} \langle v, Gu\rangle + \psi(u,v)$$

$$= -\underbrace{\min_{v\in S_v}\max_{u\in S_u}\Big[\langle u, \underbrace{(-G)}_{\uparrow\ \text{Gaussian RM.}} v\rangle + \psi(u,v)\Big]}_{\text{Apply Gordon to this}}$$

$\square$.

④ Example : LASSO problem.

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

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

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

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

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

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

$$\overset{d②}{=} \min_{u\in\mathbb{R}^d} \tfrac{1}{2}\Big(\Big(\frac{\|u\|_2^2}{n}+\sigma^2\Big)^{\frac{1}{2}}\|\hat{g}\|_2 + \langle h, u\rangle/\sqrt{n}\Big)_+^2 + \lambda\|x_0-u\|_1$$

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

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

④ $\frac{\|u\|_2^2}{n} = \xi$
$\frac{\langle h, u\rangle}{n} = e$

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

assume exchangable

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

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

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

⑥ uniform Law of large numbers

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

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

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

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

(Can be verified)

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

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

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

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

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

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

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

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.