Mean Field Asymptotics in Statistical Learning.

Mar 8.

Lecture 13: Gaussian comparison inequalities.

① Motivation.

② Slepian's inequality and Sudakov-Fernique inequality.

③ Examples: Wishart matrix and spiked GOE matrix.

③ Proof of Slepian's inequality.

④ Proof of Sudakov-Fernique if time permitted.

① Motivating example: the LASSO problem

Signal: $x_0 \in \mathbb{R}^d$, Noise: $w \in \mathbb{R}^n$, Sensing matrix: $A \in \mathbb{R}^{n \times d}$.

Response: $y = Ax_0 + w \in \mathbb{R}^n$, $A_{ij} \sim_{iid} N(0, \frac{1}{n})$, $w_i \sim_{iid} N(0, \sigma^2)$,

important.

LASSO estimator: $\hat{x} = \arg\min_x \frac{1}{2n} \|y - Ax\|_2^2 + \frac{\lambda}{n}\|x\|_1$

Goal: derive $m_* \equiv \lim_{d\to\infty} \mathbb{E}\left[\frac{1}{d}\sum_{i=1}^d \psi(\hat{x}_i, x_{0,i})\right]$. when $n/d \to \delta$.

Free energy approach ($\beta = \infty$): the perturbed LASSO objective.

$$f(h) \equiv \lim_{d\to\infty} \mathbb{E}\left[\min_x \left\{\frac{1}{2n}\|y - Ax\|_2^2 + \frac{\lambda}{n}\|x\|_1 + h\frac{1}{d}\sum_{i=1}^d \psi(x_i, x_{0,i})\right\}\right].$$

$$\Rightarrow \quad f'(0) = m_*.$$

(need to argue why limit and derivative can be exchanged).

Question: how to derive $f(h)$?

Intuition:
$$F = \min_x \left\{\frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1\right\}.$$
$$= \min_x \max_v \langle \underset{\substack{\uparrow\\ \text{Gaussians}}}{A}\underbrace{(x_0 - x) + w}_{\text{fixed}}, v\rangle - \frac{1}{2}\|v\|_2^2 + \lambda\|x\|_1$$

$$\Phi(x, v) \equiv \langle A(x_0 - x) + w, v\rangle - \frac{1}{2}\|v\|_2^2 + \lambda\|x\|_1$$

Gaussian process on $\mathbb{R}^d \times \mathbb{R}^n$ indexed by x, v.

Need some machinery to analyze the extremum of Gaussian processes.

Gaussian comparison inequalities.

② Slepian's inequality and its applications.

Definition (Gaussian process). A stochastic process $(X_t)_{t\in T}$ is called Gaussian process, if $\forall$ finite subset $T_0 \subset T$, the random vector has normal distribution.

Remark : a Gaussian process can be completely characterized by its mean function $\mu(t) = \mathbb{E}[X_t]$ and covariance function $\Sigma(t,s) = \mathbb{E}[X_t X_s]$. We will often consider continous Gaussian process where X_t is a continous function of t a.s., and T is a separable space. This leads to $\mathbb{E}[\sup_{t\in T} X_t] = \sup_{\substack{T_0 \subseteq T \\ \text{finite}}} \mathbb{E}[\sup_{t\in T_0} X_t]$.

Example : $X_t = \langle t, g\rangle$, $t \in S^{n-1}$, $g \sim N(0, I_d)$.

Theorem [Slepian's inequality and Sudakov-Fernique's inequality].

Let $T \subseteq \mathbb{R}^d$. Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two Gaussian processes with $\mathbb{E}[X_t] = \mathbb{E}[Y_t]$,

and $\mathbb{E}[(X_t - X_s)^2] \le \mathbb{E}[(Y_t - Y_s)^2]$. Then .

① $\mathbb{E}[\sup_{t\in T} X_t] \le \mathbb{E}[\sup_{t\in T} Y_t]$. [Sudakov-Fernique] expectation bound

② If additionally we have $\mathbb{E}[X_t^2] = \mathbb{E}[Y_t^2]$

$\Rightarrow \forall \tau \in \mathbb{R}$, $\mathbb{P}(\sup_{t\in T} X_t \ge \tau) \le \underbrace{\mathbb{P}(\sup_{t\in T} Y_t \ge \tau)}_{\text{may give a good concentration bound}}$. [Slepian].

③

Example 1 (Norm of Gaussian random matrices).

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

Then $\mathbb{E}[\|A\|_{op}] \le \sqrt{m} + \sqrt{n}$.

(So when $m \propto n$, $\mathbb{E}[\|A/\sqrt{n}\|_{op}] \le \sqrt{\frac{m}{n}} + 1 = O(1)$).

Proof : Variational representation of operator norm

$$\|A\|_{op} = \max_{\substack{u\in S^{n-1} \\ v\in S^{m-1}}} \langle u, A^T v\rangle = \max_{(u,v)\in T} \underbrace{X_{uv}}_{\text{Gaussian process}}$$

where $T = S^{n-1} \times S^{m-1}$, $X_{uv} = \langle u, A^T v\rangle$.

$\mathbb{E}[X_{uv}] = 0$, $\mathbb{E}[X_{uv}^2] = \mathbb{E}[\langle A, vu^T\rangle^2] = \|uv^T\|_F^2 = \|u\|_2^2 \|v\|_2^2 = 1$.

$$\mathbb{E}[(X_{uv} - X_{wz})^2] = \mathbb{E}[\langle A, vu^T - zw^T\rangle^2] = \|vu^T - zw^T\|_F^2$$
$$= 2 - 2\langle v, z\rangle\langle u, w\rangle.$$

The next step requires some creativity.

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

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

Then $\mathbb{E}[Y_{uv}] = 0$, $\mathbb{E}[Y_{uv}^2] = \mathbb{E}[\langle g,u\rangle^2] + \mathbb{E}[\langle h,v\rangle^2] = 2$.

$$\mathbb{E}[(Y_{uv} - Y_{wz})^2] = \mathbb{E}[\langle g, u-w\rangle^2] + \mathbb{E}[\langle h, v-z\rangle^2]$$
$$= \|u-w\|_2^2 + \|v-z\|_2^2 = 4 - 2\langle u,w\rangle - 2\langle v,z\rangle.$$

$$\Rightarrow \mathbb{E}[(Y_{uv}-Y_{wz})^2] - \mathbb{E}[(X_{uv}-X_{wz})^2] = 2 - 2\langle u,w\rangle - 2\langle v,z\rangle + 2\langle u,w\rangle\langle v,z\rangle$$
$$= 2\,\underbrace{(1-\langle u,w\rangle)}_{\geq 0}\,\underbrace{(1-\langle v,z\rangle)}_{\geq 0} \geq 0 .$$

By Sudakov-Fernique inequality, we have

$$\mathbb{E}[\|A\|_{op}] \leq \mathbb{E}\Big[\sup_{\substack{u\in S^{n-1}\\ v\in S^{m-1}}} \langle g,u\rangle + \langle h,v\rangle\Big]$$

quite tight.

$$= \mathbb{E}[\|g\|_2] + \mathbb{E}[\|h\|_2] \leq \mathbb{E}[\|g\|_2^2]^{\frac{1}{2}} + \mathbb{E}[\|h\|_2^2]^{\frac{1}{2}}$$

$$= \sqrt{n} + \sqrt{m}.$$ $\square$.

✩ The key is to smartly find a Gaussian process Y_{uv}.

Example 2 (The spiked GOE matrix).

Proposition: $\theta \in S^{n-1}$, $\lambda \geq 0$, $W \sim GOE(n)$. $(W_{ii} \sim N(0, \frac{2}{n}),\ W_{ij} \sim N(0, \frac{1}{n}))$

$Y = \lambda\theta\theta^T + W \in \mathbb{R}^{n\times n}$. Denote $\lambda_{max}(Y) = \sup_{\sigma\in S^{n-1}} \langle\sigma, Y\sigma\rangle$.

Then
$$\limsup_{n\to\infty} \mathbb{E}[\|Y\|_{op}] \leq \begin{cases} 2, & \lambda < 1 \\ \lambda + \frac{1}{\lambda}, & \lambda \geq 1. \end{cases}$$

Proof: $$\lambda_{max}(Y) = \sup_{t\in S^{n-1}} \underbrace{\langle t, Wt\rangle + \lambda\langle t,\theta\rangle^2}_{X_t}$$

$$Y_t = \frac{2}{\sqrt{n}}\langle g, t\rangle + \lambda\langle t,\theta\rangle^2. \qquad g \sim N(0, I_n)$$

$$\mathbb{E}[(X_t - X_s)^2] = \mathbb{E}[\langle W, tt^T - ss^T\rangle^2] \qquad G_{ij} \sim N(0,1)$$
$$= \frac{2}{n}\mathbb{E}[\langle G, tt^T - ss^T\rangle^2]$$
$$= \frac{2}{n}\|tt^T - ss^T\|_F^2 = \frac{4}{n}(1 - \langle t,s\rangle^2)$$

$$\mathbb{E}[(Y_t - Y_s)^2] = \frac{4}{n} \mathbb{E}[\langle g, t-s\rangle^2] = \frac{4}{n} \|t-s\|_2^2 = \frac{4}{n}(2 - 2\langle t, s\rangle)$$

$$\mathbb{E}[(X_t - X_s)^2] - \mathbb{E}[(Y_t - Y_s)^2] = \frac{4}{n}(1 - \langle t, s\rangle^2) \geq 0.$$

$$\Rightarrow \quad \mathbb{E}[\|Y\|_{op}] \leq \mathbb{E}[\sup_{t \in S^{n-1}} Y_t].$$

$$\sup_{t \in S^{n-1}} Y_t \leq \inf_{\zeta > \lambda} \sup_t \left[\frac{2}{\sqrt{n}}\langle g, t\rangle + \lambda\langle t, \theta\rangle^2 - \zeta(\|t\|_2^2 - 1)\right]$$

$$= \inf_{\zeta > \lambda} \left[\frac{1}{n}\langle g, (\zeta I - \lambda\theta\theta^T)^{-1} g\rangle + \zeta\right]$$

$$\boxed{\mathbb{E}[\inf_\zeta Z(\zeta)] \leq \inf_\zeta \mathbb{E}[Z(\zeta)].}$$

$$\mathbb{E}[\|Y\|_{op}] \leq \mathbb{E}\left[\sup_{t \in T} Y_t\right] \leq \inf_{\zeta > \lambda} \mathbb{E}\left[\frac{1}{n}\langle g, (\zeta I - \lambda\theta\theta^T)^{-1} g\rangle + \zeta\right]$$

$$= \inf_{\zeta > \lambda} \frac{1}{n} \mathrm{tr}[(\zeta I - \lambda\theta\theta^T)^{-1}] + \zeta.$$

$$\boxed{\text{One way : solve } \frac{1}{n}\mathrm{tr}[(\zeta I - \lambda\theta\theta^T)^{-2}] = 1 .}$$

Some cruder bound:

① When $\lambda = 0$. take $\zeta = 1$, $\Rightarrow \mathbb{E}[\|W\|_{op}] \leq 2$.

② When $\lambda < 1$, take $\zeta = 1$, $\Rightarrow$

$$\mathbb{E}[\|Y\|_{op}] \leq \frac{n-1}{n} + \frac{1}{n}\frac{1}{1-\lambda} + 1 = 2 + \frac{\lambda}{n(1-\lambda)}$$

$$\Rightarrow \limsup_{n\to\infty} \mathbb{E}[\|Y\|_{op}] \leq 2$$

③ When $\lambda \geq 1$, take $\zeta = \lambda + \frac{1}{\sqrt{n}}$ $\Rightarrow$

$$\mathbb{E}[\|Y\|_{op}] \leq \frac{n-1}{n}\frac{1}{\lambda + \frac{1}{\sqrt{n}}} + \frac{\sqrt{n}}{n} + \lambda + \frac{1}{\sqrt{n}}.$$

$$\Rightarrow \limsup_{n\to\infty} \mathbb{E}[\|Y\|_{op}] \leq \lambda + \frac{1}{\lambda}.$$

$\square$

④ Proof of Slepian's inequality.

☆ **Lemma**: [Kahane, 1986].

Let $X = (X_i)_{i \in [n]}$ and $Y = (Y_i)_{i \in [n]}$ be two Gaussian vectors, with $\mathbb{E}X = \mathbb{E}Y$.

Denote

$$A \equiv \{(j,k) \in [n]^2 : \mathbb{E}[X_j X_k] < \mathbb{E}[Y_j Y_k]\}.$$

$$B \equiv \{(j,k) \in [n]^2 : \mathbb{E}[X_j X_k] > \mathbb{E}[Y_j Y_k]\}.$$

Let $f(x) = f(x_1, \cdots, x_N) \in C^2(\mathbb{R}^n)$ satisfies ($\mathbb{E}[\partial_i \partial_j f(X)]$ can be well defined for Gaussian X).

$$\begin{cases} \partial_j \partial_k f(x) \geq 0, & (j,k) \in A, \\ \partial_j \partial_k f(x) \leq 0, & (j,k) \in B. \end{cases}$$

Then we have $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$.

Proof idea: interpolation method.

Proof: W.L.O.G, we assume that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$.

Define $Z(\theta) = X\cos\theta + Y\sin\theta$ for $0 \leq \theta \leq \frac{\pi}{2}$.

and $\varphi(\theta) = \mathbb{E}[f(Z(\theta))]$.

$\Rightarrow \varphi(0) = \mathbb{E}[f(X)]$, $\quad \varphi(1) = \mathbb{E}[f(Y)]$.

Hope to show $\varphi'(\theta) \geq 0$, $\forall\ \theta \in [0, \pi/2]$.

$$\varphi'(\theta) = \mathbb{E}[\langle \nabla f(Z(\theta)), Z'(\theta) \rangle]$$

$$= \sum_{i=1}^{N} \left[-\sin\theta\, \mathbb{E}[X_i \partial_i f(Z(\theta))] + \cos\theta \cdot \mathbb{E}[Y_i \partial_i f(Z(\theta))] \right].$$

Intergation by part:

$$\mathbb{E}[X_i \partial_i f(X\cos\theta + Y\sin\theta)] = \cos\theta \sum_{j=1}^{n} \mathbb{E}[X_i X_j] \cdot \mathbb{E}[\partial_i \partial_j f(Z(\theta))].$$

$$\mathbb{E}[Y_i \partial_i f(X\cos\theta + Y\sin\theta)] = \sin\theta \sum_{j=1}^{n} \mathbb{E}[Y_i Y_j] \cdot \mathbb{E}[\partial_i \partial_j f(Z(\theta))].$$

$$= \cos\theta \cdot \sin\theta \cdot \sum_{i,j=1}^{n} \mathbb{E}[\partial_i \partial_j f(Z(\theta))] \left\{ \mathbb{E}[Y_i Y_j] - \mathbb{E}[X_i X_j] \right\} \geq 0$$

By assumption □.

Remark: If $\mathbb{E}[X] = \mathbb{E}[Y] = \mu \neq 0$, just define $\tilde{f}(x) = f(x - \mu)$.

Corollary: Let $X, Y \in \mathbb{R}^n$ be n-dimensional Gaussian random vectors, with $\mathbb{E}[X] = \mathbb{E}[Y]$. Assume that $\mathbb{E}[X_i^2] = \mathbb{E}[Y_i^2]$, $\forall i \in [n]$, and $\mathbb{E}[X_i X_j] \geq \mathbb{E}[Y_i Y_j]$ $\forall i,j \in [n]$.

Then

(1) For any real $\eta_1, \dots, \eta_n$,

$$\mathbb{P}(X_i < \eta_i,\ \forall i) \geq \mathbb{P}(Y_i < \eta_i,\ \forall i),$$

(2) $$\mathbb{P}(\max_{i \in [n]} X_i \geq \eta) \leq \mathbb{P}(\max_{i \in [n]} Y_i \geq \eta).$$

(3) $$\mathbb{E}[\max_{i \in [n]} X_i] \leq \mathbb{E}[\max_{i \in [n]} Y_i].$$

Proof: Clearly (1) $\Rightarrow$ (2) $\Rightarrow$ (3).

To prove (1), conceptually we want to take

$$f(x_1, \dots, x_n) = \prod_{i=1}^n \underbrace{1\{x_i < \eta_i\}}_{\text{Not smooth.}}$$

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

Take $f_\varepsilon(x_1, \dots x_N) = \prod_{i=1}^n \psi((x_i - \eta_i)/\varepsilon)$. $\Rightarrow$ $\partial_i \partial_j f_\varepsilon \geq 0$.

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

$\Rightarrow$ send $\varepsilon \to 0$ and apply monotone convergence thm. $\square$

Remark: For $(X_t)_{t \in T}$ Gaussian process, when T is finite, it directly gives the theorem. When T is uncountable but separable (has a countable dense subset), when X_t is a continous Gaussian process, we have

$$\mathbb{E}[\sup_{t \in T} X_t] = \sup_{\substack{T_0 \subseteq T \\ \text{finite}}} \mathbb{E}[\sup_{t \in T_0} X_t].$$

⑤ Proof of Sudakov - Fernique.

Proposition [Sudakov - Fernique's inequality].

Let $X, Y \in \mathbb{R}^n$ be two Gaussian random vectors with $\mathbb{E}[X] = \mathbb{E}[Y]$, and $\mathbb{E}[(X_i - X_j)^2] \leq \mathbb{E}[(Y_i - Y_j)^2]$. Then.

$$\mathbb{E}[\max_{i \in [n]} X_i] \leq \mathbb{E}[\max_{i \in [n]} Y_i].$$

Proof: Let $\mu = \mathbb{E}X = \mathbb{E}Y$.

Define $\tilde{X} = X - \mu$, $\tilde{Y} = Y - \mu \in \mathbb{R}^n$.

and $Z(\theta) \equiv \cos\theta \cdot \tilde{X} + \sin\theta \cdot \tilde{Y} + \mu$

Fix $\beta > 0$, define $F_\beta : \mathbb{R}^n \to \mathbb{R}$

$$F_\beta(x) \equiv \beta^{-1} \log\left(\sum_{i=1}^n e^{\beta x_i}\right).$$

For $\theta \in [0, \pi/2]$, let $\varphi(t) = \mathbb{E}[F_\beta(Z(\theta))]$.

$$\varphi'(\theta) = \cos\theta \sin\theta \sum_{i,j=1}^n \mathbb{E}\left[\frac{\partial^2 F_\beta}{\partial x_j \partial x_i}(Z(\theta))\right] \cdot \left\{\mathbb{E}[\tilde{Y}_i \tilde{Y}_j] - \mathbb{E}[\tilde{X}_i \tilde{X}_j]\right\}$$

same calculation as before

Define $p_i(x) = \partial_{x_i} F_\beta(x) = \dfrac{e^{\beta x_i}}{\sum_{j=1}^n e^{\beta x_j}}$, prob dist on $\mathbb{R}^n$.

$$\Rightarrow \frac{\partial^2 F_\beta}{\partial x_i \partial x_j}(x) = \begin{cases} \beta(p_i(x) - p_i(x)^2) & \text{if } i = j \\ -\beta p_i(x) p_j(x) & \text{if } i \neq j. \end{cases}$$

$$\sum_{i,j=1}^n \frac{\partial^2 F_\beta}{\partial x_j \partial x_i}(x) \left\{\mathbb{E}[\tilde{Y}_i \tilde{Y}_j] - \mathbb{E}[\tilde{X}_i \tilde{X}_j]\right\}$$

$$= \underbrace{\beta \sum_{i=1}^n p_i(x)\left[\mathbb{E}[\tilde{Y}_i^2] - \mathbb{E}[\tilde{X}_i^2]\right]} - \beta \sum_{i,j=1}^n p_i(x) p_j(x) \left[\mathbb{E}[\tilde{Y}_i \tilde{Y}_j] - \mathbb{E}[\tilde{X}_i \tilde{X}_j]\right]$$

$$= \frac{\beta}{2} \sum_{i,j=1}^n p_i(x) p_j(x) \left[\mathbb{E}[\tilde{Y}_i^2] - \mathbb{E}[\tilde{X}_i^2] + \mathbb{E}[\tilde{Y}_j^2] - \mathbb{E}[\tilde{X}_j^2]\right].$$

$$= \frac{\beta}{2} \sum_{i,j=1}^n p_i(x) p_j(x) \left[\mathbb{E}[(Y_i - Y_j)^2] - \mathbb{E}[(X_i - X_j)^2]\right] \geq 0.$$

$$\Rightarrow \quad \mathbb{E}[F_\beta(X)] \leq \mathbb{E}[F_\beta(Y)].$$

$$\beta \to \infty \quad \Rightarrow \quad \mathbb{E}[\max_{i \in [n]} X_i] \leq \mathbb{E}[\max_{i \in [n]} Y_i]. \qquad \square$$