Mean Field Asymptotics in Statistical Learning.

Mar 31.

Lecture 18. Lindeberg approach and universality.

[Carmona, Hu, 2004].
[Chatterjee, 2005]
[Korada, Montanari, 2010]

① The Universality phenomenon.

Example 1: The LASSO problem.

$y = A x_0 + w$, $\quad x_0 \in \mathbb{R}^d$, $\quad A \in \mathbb{R}^{n\times d}$, $\quad w \in \mathbb{R}^n$.

$w_i \sim_{iid} N(0, \sigma^2)$.

$$L(A) \equiv \min_{x\in[-B,B]} \left\{ \frac{1}{2n} \| y - A x \|_2^2 + \frac{\lambda}{n} \| x \|_1 \right\}$$

Thm: Let $G \in \mathbb{R}^{n\times d}$ with $G_{ij} \sim_{iid} N(0,1)$.
$A \in \mathbb{R}^{n\times d}$ with $A_{ij} \sim_{iid} \mathbb{P}_A$

$\mathbb{E}_A[A_{ij}] = 0$, $\quad \mathbb{E}_A[A_{ij}^2] = 1$, $\quad \sup_{ij} \mathbb{E}_A[A_{ij}^6] \le K < \infty$.

Then $\lim_{\substack{n\to\infty \\ \frac{n}{d}\to\delta}} \left(\mathbb{E}[L(A/\sqrt{n})] - \mathbb{E}[L(G/\sqrt{n})] \right) = 0$.

Example 2: The SK model ($\mathbb{Z}_2$ synchronization).

$\Omega = \{\pm 1\}^n$

$$H(\sigma, A_n) = -\frac{1}{\sqrt{2}} \sum_{ij=1}^{n} A_{ij} \sigma_i \sigma_j + h M(\sigma).$$

$M(x) = \frac{1}{n}\langle x, x_0 \rangle^2$.
M deterministic

Free entropy density: $\varphi(\beta, A_n) \equiv \frac{1}{n} \log \left\{ \sum_{\sigma\in\Omega} \exp\{-\beta H(\sigma, A_n)\} \right\}$.

Ground state energy: $f_0(A_n) \equiv \min_{\sigma\in\Omega} H(\sigma, A_n)$.

Theorem: Let $G \in \mathbb{R}^{n\times n}$ with $G_{ij} \sim_{iid} N(0,1)$.
Let $A \in \mathbb{R}^{n\times n}$ with $A_{ij} \sim_{iid} \mathbb{P}_A$.

$\mathbb{E}[A_{ij}] = 0$, $\quad \mathbb{E}[A_{ij}^2] = 1$, $\quad \sup_{ij} \mathbb{E}[|A_{ij}|^3] \le K < \infty$

Then $\lim_{n\to\infty} \left(\mathbb{E}[\varphi(\beta, A/\sqrt{n})] - \mathbb{E}[\varphi(\beta, G/\sqrt{n})] \right) = 0$.
$\lim_{n\to\infty} \left(\mathbb{E}[f_0(A/\sqrt{n})] - \mathbb{E}[f_0(G/\sqrt{n})] \right) = 0$

Remark:
- ⊗ We don't even need to know $\lim_{n\to\infty} \mathbb{E}[f(\beta, G/\sqrt{n})]$ exists before we know the universality. (for some M, this may not exists)
- ⊗ When $h = 0$, one can show that the limit exists.
- ⊗ One can show that $\varphi(\beta, A/\sqrt{n}) \overset{d}{\approx} \varphi(\beta, G/\sqrt{n})$, both concentrate.

⊗ This principle can be extended to other mean-field models, e.g. empirical distribution of eigenvalues, etc.

⊗ With additional conditions, one can show the universality of limiting observables.

If we write $H(\sigma; A; h) = -\frac{1}{\sqrt{2}} \sum_{i,j=1}^{n} A_{ij} \sigma_i \sigma_j + h M(\sigma)$.

and $$\langle g \rangle_{\beta, h, A} \equiv \frac{\sum_\sigma g(\sigma) \exp\{-\beta H(\sigma; A; h)\}}{\sum_\sigma \exp\{-\beta H(\sigma. A; h)\}}$$

$$\varphi(\beta, h; A) \equiv \frac{1}{n} \log \left\{ \sum_{\sigma \in \{\pm 1\}^n} \exp\{-\beta H(\sigma; A; h)\} \right\}$$

Prop : Suppose a) $\bar{\varphi}(\beta, h) \triangleq \lim_{n\to\infty} \mathbb{E}[\varphi(\beta, h; A/\sqrt{n})]$ exists

b) $\lim_{n\to\infty} \mathbb{P}\left(\left| \varphi(\beta, h, A/\sqrt{n}) - \bar{\varphi}(\beta, h) \right| \geq \varepsilon \right) = 0$, $\forall h \in (h_0 - \delta, h_0 + \delta)$ for some $\delta > 0$

c) $\partial_h \bar{\varphi}(\beta, h_0)$ exists at some h_0.

Then $$\lim_{n\to\infty} \mathbb{P}\left(\left| \langle M \rangle_{\beta, h_0, A/\sqrt{n}} - \partial_h \bar{\varphi}(\beta, h_0) \right| \geq \varepsilon \right) = 0.$$

Proof idea: $\varphi(\beta, h, A/\sqrt{n})$ is convex in h. $\Rightarrow$

$$\frac{1}{\delta}\left[\varphi(\beta, h_0 - \delta; A/\sqrt{n}) - \varphi(\beta, h_0; A/\sqrt{n})\right] \leq \langle M \rangle_{\beta, h_0, A/\sqrt{n}} \leq \frac{1}{\delta}\left[\varphi(\beta, h_0 + \delta, A/\sqrt{n}) - \varphi(\beta, h_0; A/\sqrt{n})\right]$$

Left side:

$$\xrightarrow[\text{a), b)}]{n\to\infty} \frac{1}{\delta}\left(\bar{\varphi}(\beta; h_0 - \delta) - \bar{\varphi}(\beta; h_0) \right) \xrightarrow[\text{c)}]{\delta \to 0} \partial_h \bar{\varphi}(\beta; h_0)$$

Right side:

$$\xrightarrow[\text{a), b)}]{n\to\infty} \frac{1}{\delta}\left(\bar{\varphi}(\beta; h_0 + \delta) - \bar{\varphi}(\beta; h_0) \right) \xrightarrow[\text{c)}]{\delta \to 0} \partial_h \bar{\varphi}(\beta; h_0)$$

② The Lindeberg approach

[Chatterjee, 2005].

Thm (Generalized Lindeberg principle).

Let $U=(U_1, \cdots, U_n)$ and $V=(V_1, \cdots, V_n)$.

be two random vectors with mutually independent component.

For $1\le i\le n$, define

$$a_i = |\mathbb{E}[U_i]-\mathbb{E}[V_i]|$$

$$b_i = |\mathbb{E}[U_i^2]-\mathbb{E}[V_i^2]|$$

assume $\max\limits_i \{\mathbb{E}[|U_i|^3]+\mathbb{E}[|V_i|^3]\}\le M_3$.

Suppose $f\in C^3(\mathbb{R})$. with $\sup\limits_i\sup\limits_u |\partial_i^r f(u)|\le L_r(f)$, then

$$|\mathbb{E}[f(U)]-\mathbb{E}[f(V)]|\le \sum_{i=1}^n \left(a_i L_1(f)+\tfrac{1}{2}b_i L_2(f)\right)+\tfrac{1}{6}nL_3(f)M_3.$$

Proof idea: ⊗ decomposition into telescope sum, change one coordinate each time.

⊗ Taylor expansion.

Proof: Denote $\overline{W}_i=(U_1, \cdots, U_i, V_{i+1}, \cdots, V_n)$

$\overline{W}_i^\circ = (U_1, \cdots, U_{i-1}, 0, V_{i+1}, \cdots, V_n)$.

Then $$\mathbb{E}[f(U)]-\mathbb{E}[f(V)]=\sum_{i=1}^n\left(\mathbb{E}[f(\overline{W}_i)]-\mathbb{E}[f(\overline{W}_{i-1})]\right).$$

By 3rd order Taylor expansion, we have

$$f(\overline{W}_i)=f(\overline{W}_i^\circ)+U_i\partial_i f(\overline{W}_i^\circ)+\frac{U_i^2}{2}\partial_i^2 f(\overline{W}_i^\circ)+\frac{1}{2}\int_0^{U_i}\partial_i^3 f(U_i^{i-1}, s, V_{i+1}^n)(U_i-s)^2ds.$$

$$f(\overline{W}_{i-1})=f(\overline{W}_i^\circ)+V_i\partial_i f(\overline{W}_i^\circ)+\frac{V_i^2}{2}\partial_i^2 f(\overline{W}_i^\circ)+\frac{1}{2}\int_0^{V_i}\partial_i^3 f(U_i^{i-1}, s, V_{i+1}^n)(V_i-s)^2ds.$$

$$\Rightarrow \mathbb{E}[f(U)]-\mathbb{E}[f(V)]=\sum_{i=1}^n\Big\{\mathbb{E}[(U_i-V_i)\partial_i f(\overline{W}_i^\circ)]+\tfrac{1}{2}\mathbb{E}[(U_i^2-V_i^2)\partial_i^2 f(\overline{W}_i^\circ)]$$

$$+\mathbb{E}\left[\tfrac{1}{2}\int_0^{U_i}\partial_i^3 f(U_i^{i-1}, s, V_{i+1}^n)(U_i-s)^2ds\right]+\mathbb{E}\left[\tfrac{1}{2}\int_0^{V_i}\partial_i^3 f(U_i^{i-1}, s, U_{i+1}^n)(V_i-s)^2ds\right]$$

$$\le \sum_{i=1}^n\left(a_iL_1(f)+\tfrac{1}{2}b_iL_2(f)\right)+\tfrac{1}{6}nL_3(f)\cdot M_3.$$ □

③ Application to the SK model.

$$\Omega = \{\pm 1\}^n$$

$$H(\sigma, A) = -\frac{1}{\sqrt{2}} \sum_{ij=1}^{n} A_{ij}\sigma_i\sigma_j + h\, M(\sigma).$$

Free entropy density $\varphi(\beta, A) \equiv \frac{1}{n} \log \left\{ \sum_{\sigma\in\{\pm1\}^n} \exp\{-\beta H(\sigma, A)\} \right\}.$

Ground state energy: $f_0(A_n) \equiv \min_{\sigma\in\Omega} H(\sigma, A_n)$.

> **Theorem:** Let $G \in \mathbb{R}^{n\times n}$ with $G_{ij} \sim_{iid} N(0,1)$.
> Let $A \in \mathbb{R}^{n\times n}$ with $A_{ij} \sim_{iid} \mathbb{P}_A$.
>
> $\mathbb{E}[A_{ij}] = 0$, $\mathbb{E}[A_{ij}^2] = 1$, $\sup_{ij} \mathbb{E}[A_{ij}^6] \le K < \infty$
>
> Then $\lim_{n\to\infty} \left(\mathbb{E}[\varphi(\beta, A/\sqrt{n})] - \mathbb{E}[\varphi(\beta, G/\sqrt{n})]\right) = 0.$
>
> $\lim_{n\to\infty} \left(\mathbb{E}[f_0(A/\sqrt{n})] - \mathbb{E}[f_0(G/\sqrt{n})]\right) = 0$

Proof: ⊗

Denote $$\langle g \rangle_\beta = \frac{\sum_\sigma g(\sigma)\exp(-\beta H(\sigma, A))}{\sum_\sigma \exp(-\beta H(\sigma, A))}$$

$$f(A) = \varphi(\beta, A/\sqrt{n}).$$

$$\partial_{A_{ij}} f(A) = \frac{d}{dA_{ij}}\left[\varphi(\beta; A/\sqrt{n})\right] = \frac{1}{n}\frac{\beta}{\sqrt{2n}} \langle \sigma_i\sigma_j \rangle_\beta$$

$$\partial_{A_{ij}}^2 f(A) = \frac{1}{n}\left(\frac{\beta}{\sqrt{2n}}\right)^2 \left[\langle \sigma_i^2\sigma_j^2\rangle_\beta - \langle\sigma_i\sigma_j\rangle_\beta^2\right] = \frac{1}{n}\left(\frac{\beta}{\sqrt{2n}}\right)^2 (1 - \langle\sigma_i\sigma_j\rangle_\beta^2)$$

$$\partial_{A_{ij}}^3 f(A) = \frac{1}{n}\left(\frac{\beta}{\sqrt{2n}}\right)^3 \left\{-2\langle\sigma_i\sigma_j\rangle_\beta \left[\langle\sigma_i^2\sigma_j^2\rangle_\beta - \langle\sigma_i\sigma_j\rangle_\beta^2\right]\right\}$$

$$= \frac{\beta^3}{\sqrt{2}\, n^{5/2}} \langle\sigma_i\sigma_j\rangle (1 - \langle\sigma_i\sigma_j\rangle^2).$$

$$L_3(f) = \sup_{i,j} \sup_A \left|\partial_{A_{ij}}^3 f(A)\right| \le \frac{\beta^3}{\sqrt{2}\, n^{5/2}}$$

$$\mathbb{E}[A_{ij}] = \mathbb{E}[G_{ij}] = 0, \qquad \mathbb{E}[A_{ij}^2] = \mathbb{E}[G_{ij}^2] = 1.$$

$$\sup_{ij}\left[\mathbb{E}[|A_{ij}|^3] + \mathbb{E}[|G_{ij}|^3]\right] < K + 3 < \infty.$$

$$\Rightarrow \left|f(A) - f(G)\right| \le \frac{1}{2} n^2 \frac{\beta^3}{\sqrt{2}\, n^{5/2}} = \frac{\beta^3}{2\sqrt{2}\cdot\sqrt{n}} \to 0 \quad \text{as } n\to\infty.$$

⊗ $$\varphi(\beta, A) \equiv \log \sum_{\sigma\in\{\pm1\}^n} \exp\{-\beta H(\sigma, A)\} \le \log \sum_{\sigma\in\{\pm1\}^n} \exp\{-\beta \min_{\sigma'} H(\sigma', A)\}.$$

$$= \log 2 - \beta \min_\sigma H(\sigma, A) = \log 2 - \beta f_0(A)$$

$$\varphi(\beta, A) \ge \log \exp\{-\beta \min_{\sigma'} H(\sigma', A)\} = -\beta f_0(A)$$

$$\Rightarrow \quad -\frac{1}{\beta}\varphi(\beta, A) \le f_0(A) \le -\frac{1}{\beta}\varphi(\beta, A) + \frac{\log 2}{\beta} \qquad \forall \beta > 0$$

$$\Rightarrow \quad \left| f_0(A/\sqrt{n}) - f_0(G/\sqrt{n}) \right| \le \frac{\beta^2}{2\sqrt{2}\cdot\sqrt{n}} + \frac{2\log 2}{\beta} \qquad \forall \beta > 0$$

Taking $\beta = n^{\frac{1}{6}}$, $\Rightarrow \quad \left| f_0(A/\sqrt{n}) - f_0(G/\sqrt{n}) \right| = O\left(\frac{1}{n^{\frac{1}{6}}}\right) \to 0.$ □

Other reference :

[Carmona, Hu, 2004] : Universality of SK model using interpolation method

[Korada, Montanari, 2010] : Universality of CDMA, LASSO, using generalized Lindeberg approach.

Next lecture : Approximate message passing algorithm.