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.

$$y = Ax_0 + w, \qquad x_0 \in \mathbb{R}^d, \quad A \in \mathbb{R}^{n\times d}, \quad w \in \mathbb{R}^n.$$

$$w_i \sim_{iid} N(0, \sigma^2), \qquad \max_i |x_{0,i}| \le B$$

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

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

$A \in \mathbb{R}^{n\times d}$, $A_{ij} \sim_{iid} \mathbb{P}_A$.

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

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: SK model ($\mathbb{Z}_2$ synchro).

$$\Omega = \{\pm 1\}^n. \qquad M(\sigma) = -\lambda \langle \sigma, x_0 \rangle^2$$

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

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

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

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

$A \in \mathbb{R}^{n\times n}$, $A_{ij} \sim_{iid} \mathbb{P}_A$, $\mathbb{E}[A_{ij}] = 0$, $\mathbb{E}[A_{ij}^2] = 1$

$\sup_{ij} \mathbb{E}[|A_{ij}|^3] \le K < \infty$

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

$\lim_{n\to\infty} \left(f_0(A/\sqrt{n}) - f_0(G/\sqrt{n}) \right) = 0.$

Remark:

⊗ We don't need $\mathbb{E}[\varphi(\beta, G/\sqrt{n})]$ converge in order for Thm hold.

⊗ When $h=0$, one can show $\mathbb{E}[\varphi(\beta, G/\sqrt{n})] \to$ fixed quantity.

⊗ One can show $\varphi(\beta, A/\sqrt{n}) \overset{d}{\approx} \varphi(\beta, G/\sqrt{n})$.

$\varphi(\beta, G/\sqrt{n})$ concentrate $\longrightarrow$ $\varphi(\beta, A/\sqrt{n})$ concentrate.

⊗ Condition: $\mathbb{E}[|A_{ij}|^p] \leq K < \infty$.

weeker than sub-Gaussian.

⊗ In many MF model. universality exists.

Empirical dist of eigenvalue.

⊗ With additional conditions, one can translate the univeality of free entropy density to univeality of observables.

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

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

$$\varphi(\beta, h; A) = \frac{1}{n} \log \left\{ \sum_{\sigma} \exp(-\beta H(\cdots)) \right\}.$$

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

b) $\lim_{n\to\infty} \mathbb{P}(|\varphi(\beta, h; A) - \bar{\varphi}(\beta, h)| \geq \varepsilon) = 0$.

$h \in (-\delta + h_0, h_0 + \delta)$.

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

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

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

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

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

Right side $\xrightarrow{n\to\infty}$ $\frac{1}{\delta}(\bar{\varphi}(\beta, h_0+\delta) - \bar{\varphi}(\beta, h_0))$ $\xrightarrow{\delta\to 0}$ $\partial_h \bar{\varphi}(\beta, h_0)$.

② The Lindeberg approach. [Chatterjee, 2005]

Thm (Generalized Lindeberg principle).

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

U_i, V_i are mutually independent.

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

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

$$b_i = \left| \mathbb{E}[U_i^2] - \mathbb{E}[U_i^2] \right|.$$

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

Suppose $f \in C^3(\mathbb{R}^n)$ with $\sup_i \sup_u |\partial_i^r f(u)| \le L_r(f)$.

Then $\left| \mathbb{E}[f(U)] - \mathbb{E}[f(V)] \right|$

$$\le \underbrace{\sum_{i=1}^n a_i L_1(f) + \frac{1}{2} b_i L_2(f)}_{0} + \frac{1}{6} n L_3(f) M_3$$

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

⊗ Taylor expansion.

Proof: Define $\overline{W}_i = (U_1, U_2, \cdots U_i, V_{i+1}, \cdots, V_n)$.

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

$$\overline{W}_{i-1} = (U_1, \cdots U_{i-1}, V_i, V_{i+1}, \cdots V_n)$$

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

$$f(\overline{W}_i) = f(\overline{W}_i^0) + U_i \partial_i f(\overline{W}_i^0) + \frac{1}{2} U_i^2 \partial_i^2 f(\overline{W}_i^0) + \frac{1}{6} \partial_i^3 f(\xi_i) U_i^3$$

$$f(\overline{W}_{i-1}) = f(\overline{W}_i^0) + V_i \partial_i f(\overline{W}_i^0) + \frac{1}{2} U_i^2 \partial_i^2 f(\overline{W}_i^0) + \frac{1}{6} \partial_i^3 f(\xi_i') V_i^3$$

③ Application to SK model.

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

$\Omega = \{\pm 1\}^n$. $\qquad M(\sigma) = -\lambda \langle \sigma, x_0 \rangle^2$

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

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

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

Thm: Let $G \in \mathbb{R}^{n \times n}$. $\quad G_{ij} \sim_{iid} N(0,1)$

$A \in \mathbb{R}^{n \times n}$ $\quad A_{ij} \sim iid\ \mathbb{P}_A$ $\quad \mathbb{E}[A_{ij}] = 0 \quad \mathbb{E}[A_{ij}^2] = 1$

$\sup_{ij} \mathbb{E}[|A_{ij}|^3] \leq K < \infty$

Then $\lim_{n\to\infty} \left(\mathbb{E}[\varphi(\beta, A/\sqrt{n})] - \mathbb{E}[\varphi(\beta, G/\sqrt{n})] \right) = 0$ ①

$\lim_{n\to\infty} \left(f_0(A/\sqrt{n}) - f_0(G/\sqrt{n}) \right) = 0$.

Proof: ① $\left| \mathbb{E}[\varphi(\beta, A/\sqrt{n})] - \mathbb{E}[\varphi(\beta, G/\sqrt{n})] \right| \to 0$.

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

$$\langle g \rangle = \frac{\sum g(\sigma) \exp(-\beta H)}{\sum \exp(-\beta H)}$$

$$\partial_{A_{ij}} f(G) = \partial_{B_{ij}} \varphi(\beta, G/\sqrt{n}) / \sqrt{n}$$

$$\varphi(\beta, B) = \frac{1}{n} \log \sum_{\sigma} \exp\left(-\frac{\beta}{\sqrt{2}} \sum_{ij} B_{ij} \sigma_i \sigma_j \right)$$

$$\partial_{B_{ij}} \varphi(\beta, B) = \frac{1}{n} \left(-\frac{\beta}{\sqrt{2}} \right) \langle \sigma_i \sigma_j \rangle$$

$$\partial_{A_{ij}}^2 f(G) = \partial_{B_{ij}}^2 \varphi(\beta, G/\sqrt{n}) / n$$

$$= \frac{1}{n} \left(-\frac{\beta}{\sqrt{2n}} \right)^2 \left(1 - \langle \sigma_i \sigma_j \rangle^2 \right)$$

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

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

$$\sup_{ij} \sup_{G} \left| \partial_{A_{ij}}^3 f(G) \right| \leq \frac{\beta^3}{\sqrt{2}\, n^{5/2}}$$

$$\mathbb{E}[A_{ij}] = \mathbb{E}[G_{ij}] \qquad \mathbb{E}[A_{ij}^2] = \mathbb{E}[G_{ij}^2]$$

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

$$\left|\mathbb{E}[f(A)] - \mathbb{E}[f(G)]\right| \overset{\text{Lindeberg}}{\leq} \frac{1}{2} n^2 \times (K+3) \times \frac{\beta^3}{\sqrt{2}\, n^{5/2}}$$

$$= \frac{\beta^3 (K+3)}{2\sqrt{2}\, n^{\frac{1}{2}}} = o(1). \quad n \to \infty.$$

② $$\varphi(\beta, A) = \frac{1}{n} \log \sum_{\sigma} \exp(-\beta H(\sigma; A))$$

$$\leq \frac{1}{n} \log \sum_{\sigma} \exp(-\beta \min_{\sigma'} H(\sigma', A))$$

$$\leq \frac{1}{n} \log 2^n \exp(-\beta \min_{\sigma'} H(\sigma', A))$$

$$= \log 2 - \frac{\beta}{n} \min_{\sigma'} H(\sigma', A).$$

$$\varphi(\beta, A) \geq \frac{1}{n} \log \exp(-\beta \min_{\sigma'} H(\sigma', A))$$

$$= -\frac{\beta}{n} \min_{\sigma'} H(\sigma', A)$$

$$f_0(A) = \frac{1}{n} \min_{\sigma'} H(\sigma', A).$$

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

$$|f_0(A) - f_0(G)| \leq \frac{\beta^3 (K+3)}{2\sqrt{2}\, n^{\frac{1}{2}}} + \frac{2\log 2}{\beta} \quad \forall \beta > 0.$$

$$\beta = n^{\frac{1}{6}} \qquad \leq O\left(\frac{1}{n^{\frac{1}{6}}}\right) = o(1). \qquad \square$$

Interpolation method [Carmona, Hu, 2004].

Lindeberg approach. [Korada, Montanari, 2010].

LASSO, CDMA.