Mean Field Asymptotics in Statistical Learning. Feb 4rd.

Lecture 6: Concentration phenomena in mean field asymptotics

Non-asymptotic regime $\mathbb{P}\left(\|\hat{\theta} - \theta\|_2^2 / d \geq C \cdot \sqrt{\frac{d \log d/\delta}{n}} \right) \leq \delta.$

Asymptotic regime.

① $\lim_{n\to\infty} \mathbb{P}\left(\left| \|\hat{\theta} - \theta\|_2^2 / n - \mathbb{E}[\|\hat{\theta} - \theta\|_2^2 / n] \right| \geq \varepsilon \right) = 0$

② $\lim_{n\to\infty} \mathbb{E}[\|\hat{\theta} - \theta\|_2^2] / n =$ some formula.

① Gaussian concentration inequality.

Prop: Let $f: \mathbb{R}^d \to \mathbb{R}$ be an L-Lipschitz function.

$$|f(x_1) - f(x_2)| \leq L \cdot \|x_1 - x_2\|_2, \quad \forall x_1, x_2 \in \mathbb{R}^d$$

Let $G = (G_1, \cdots, G_d)$ standard Gaussian r.v.

$G_i \sim_{i.i.d} N(0,1)$.

Then we have

$$\mathbb{P}\left(|f(G) - \mathbb{E} f(G)| \geq t \right) \leq 2 \cdot \exp\left\{ -\frac{t^2}{2L^2} \right\}.$$

Example: $f(G) = \frac{1}{n} \sum_{i=1}^n G_i$. $\|\nabla f(G)\|_2 = \|\frac{1}{n} \mathbf{1}_n\|_2 = \frac{1}{\sqrt{n}}$.

f is $\frac{1}{\sqrt{n}}$-Lipschitz. $G_i \sim_{i.i.d} N(0,1)$

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n G_i \right| \geq t \right) \leq 2 \cdot \exp\left\{ -\frac{nt^2}{2} \right\}.$$

Proof of Prop.: ① Gaussian Log-Sobolev inequality + Herbest.
[Thm 3.25] Ramon von Handel.

② Gaussian isoperimetric inequality.

② $\mathbb{Z}_2$ sync problem.

$\theta = (\theta_1, \theta_2, \cdots, \theta_n)^T \in \mathbb{R}^n$, $\quad \theta_i \overset{\text{i.i.d.}}{\sim} \text{Unif}(\mathbb{Z}_2 = \{\pm 1\})$

$Y = \frac{\lambda}{n}\theta\theta^T + W \in \mathbb{R}^{n\times n}$. $\quad W \sim GOE(n)$.

$\|W\|_{op}^2 \approx \text{const}$

$\|\frac{\theta\theta^T}{n}\|_{op} \approx \frac{\|\theta\|_2^2}{n} \approx 1$

$W_{ij} \sim_{\text{i.i.d.}} N(0, \frac{1}{n})$ $1 \le i < j \le n$, $\quad W_{ii} \sim_{\text{i.i.d.}} N(0, \frac{2}{n})$ $\quad 1 \le i \le n$, $\quad W_{ij} = W_{ji}$

Observe $Y \in \mathbb{R}^{n\times n}$, estimate θ.

Spectral estimator: $\quad \hat{\theta}(Y) = \hat{\theta}_{spec}(Y) = \underset{\sigma \in S^{n-1}(\sqrt{n})}{\text{argmax}} \langle \sigma, Y\sigma \rangle / n$.

a) Goal: To show $\underset{\sigma \in S^{n-1}(\sqrt{n})}{\max} \langle \sigma, Y\sigma \rangle / n$ $\approx$ its expectation. w.h.p.

Prop 1: Let $\mathcal{R}_n \subseteq \{\theta \in \mathbb{R}^n : \|\theta\|_2^2/n \le 1\}$.

Let $Y = A + W \in \mathbb{R}^{n\times n}$. $\quad$ A deterministic matrix

$W \sim GOE(n)$

Then $\delta > 0$, we have

$$\mathbb{P}_W\left(\left| \sup_{\sigma \in \mathcal{R}_n} \langle \sigma, Y\sigma \rangle / n - \mathbb{E}[\cdots] \right| \le \sqrt{\frac{4 \lg(2/\delta)}{n}} \right) \ge 1 - \delta.$$

Proof: Let $G \in \mathbb{R}^{n\times n}$ $\quad G_{ij} \sim_{\text{i.i.d}} N(0,1)$, $\quad 1 \le i, j \le n$. (Not symmetric)

Denote $\hat{W} = (G + G^T)/\sqrt{2n}$. $\quad$ Then we have $\hat{W} \overset{d}{=} W$.

Define $\hat{Y} = A + \hat{W}$.

$$f(G) = \sup_{\sigma \in \mathcal{R}_n} \langle \sigma, \hat{Y}\sigma \rangle / n. = \sup_{\sigma \in \mathcal{R}_n} \langle \sigma, \frac{(G+G^T)}{\sqrt{2n}}\sigma \rangle / n.$$

(Heuristic: $\nabla_G f(G) = 2\,\frac{\sigma_* \sigma_*^T/\sqrt{2n}}{n}$ $\qquad \sigma_* = \text{arg sup}$

$\|\nabla_G f(G)\|_F = \sqrt{\frac{2}{n}} \left\|\frac{\sigma_*\sigma_*^T}{n}\right\|_F = \sqrt{\frac{2}{n}} \|\sigma_*\|_2^2/n = \sqrt{\frac{2}{n}}$.)

$f(G_1) - f(G_2)$ $\qquad G_1, G_2 \in \mathbb{R}^{n\times n}$.

$$= \sup_{\sigma \in \mathcal{R}_n} \langle \sigma, \hat{Y}_1 \sigma \rangle / n + \inf_{\sigma \in \mathcal{R}_n} - \langle \sigma, \hat{Y}_2 \sigma \rangle / n.$$

$$\le \langle \sigma_*, \hat{Y}_1 \sigma_* \rangle / n - \langle \sigma_*, \hat{Y}_2 \sigma_* \rangle / n$$

σ_* is $\underset{\sigma \in \mathcal{R}_n}{\text{argsup}} \langle \sigma, \hat{Y}_1 \sigma \rangle$

$$= \langle G_1 - G_2, \sigma_* \sigma_*^T \rangle \sqrt{\frac{2}{n}} / n$$

$$\le \|G_1 - G_2\|_{op} \underbrace{\|\sigma_*\|_2^2/n}_{\le 1} \cdot \sqrt{\frac{2}{n}}$$

$$\le \|G_1 - G_2\|_F \cdot \sqrt{\frac{2}{n}}.$$

$$\mathbb{P}\left(\left| f(G) - \mathbb{E}[f(G)] \right| \geq t \right) \leq 2 \cdot \exp\left\{ - \frac{nt^2}{4} \right\}.$$

$$2 \exp\left\{ - \frac{nt^2}{4} \right\} = \delta. \qquad t = \sqrt{\frac{4 \log(2/\delta)}{n}}$$ □

b). Show the concentration $\langle \hat{\theta}, \theta \rangle^2 / n^2$. $\|\hat{\theta}\|_2^2 / n$.

$$\hat{\theta} = \arg\sup_{\sigma \in S^{n-1}(\sqrt{n})} \langle \sigma, Y \sigma \rangle / n.$$

$$H_h(\sigma) = \langle \sigma, Y\sigma \rangle / n + h \|\sigma\|_2^2 / n$$

0

BBP transition.

$$U_n(\lambda) = \sup_{\sigma \in S^{n-1}(\sqrt{n})} \langle \sigma, Y \sigma \rangle / n$$

$$\lim_{n \to \infty} \mathbb{E}\, U_n(\lambda) = \begin{cases} 2 & , \lambda \leq 1 \\ \lambda + \frac{1}{\lambda} & , \lambda > 1. \end{cases}$$

$$\frac{d}{d\lambda} U_n(\lambda) = \frac{d}{d\lambda} \sup_{\sigma \in S \ldots} \left[\frac{\lambda}{n^2} \langle \sigma, \theta \rangle^2 + \langle \sigma, W\sigma \rangle / n \right]$$

$$= \frac{1}{n^2} \langle \hat{\theta}, \theta \rangle^2.$$

Heuristic.

$$\lim_{n \to \infty} \mathbb{P}\left(\left| U_n(\lambda) - \mathbb{E}[U_n(\lambda)] \right| \geq \varepsilon \right) = 0$$

$$\lim_{n \to \infty} \frac{1}{n} \langle \sigma, \hat{\theta} \rangle^2 \simeq \partial \lambda \lim_{n \to \infty} \mathbb{E}[U_n(\lambda)] = \partial \lambda \begin{cases} 2 & , \lambda \leq 1 \\ \lambda + \frac{1}{\lambda} & , \lambda > 1 \end{cases}$$

$$= \begin{cases} 0 & , \lambda \leq 1 \\ 1 - \frac{1}{\lambda^2} & , \lambda > 1. \end{cases}$$

1)
$$\Delta_n^+(\lambda, \delta) = \frac{U_n(\lambda + \delta) - U_n(\lambda)}{\delta}.$$

$$\Delta_n^-(\lambda, \delta) = \frac{U_n(\lambda) - U_n(\lambda - \delta)}{\delta}.$$

$$\Delta_n^-(\lambda, \delta) \leq \langle \hat{\theta}, \theta \rangle^2 / n^2 \leq \Delta_n^+(\lambda, \delta).$$

$$U_n(\lambda + \delta) = \sup_{\sigma \in \ldots} \langle \sigma, W\sigma \rangle / n + \frac{\lambda + \delta}{n^2} \langle \sigma, \theta \rangle^2.$$

$\sigma_* = \hat{\theta}$
$$\geq \langle \sigma_*, W \sigma_* \rangle / n + \frac{\lambda}{n^2} \langle \sigma_*, \theta \rangle + \frac{\delta}{n^2} \langle \sigma_*, \theta \rangle^2$$

$$= U_n(\lambda) + \frac{\delta}{n^2} \langle \hat{\theta}, \theta \rangle^2.$$

2)
$$\Delta^+(\lambda, \delta) = \frac{u(\lambda + \delta) - u(\lambda)}{\delta} \qquad u(\lambda) = \begin{cases} 2 & , \lambda < 1 \\ \lambda + \frac{1}{\lambda} & , \lambda \geq 1. \end{cases}$$

$$\Delta^-(\lambda, \delta) = \frac{u(\lambda) - u(\lambda - \delta)}{\delta}$$

$$\lim_{n \to \infty} \mathbb{P}\left(\left| \Delta_n^\pm - \Delta^\pm \right| \geq \varepsilon \right) = 0$$

$\Rightarrow \quad \forall \varepsilon > 0, \quad$ and $\quad \delta > 0.$

$$\lim_{n\to\infty} \mathbb{P}\left(\Delta^-(\lambda,\delta) - \varepsilon \le \langle \hat{\theta}, \theta \rangle^2 / n^2 \le \Delta^+(\lambda,\delta) + \varepsilon \right) = 0.$$

$$\lim_{\delta\to 0} \Delta^{\pm}(\lambda,\delta) = \Delta(\lambda) = \begin{cases} 0, & \lambda < 1 \\ 1 - \frac{1}{\lambda^2}, & \lambda > 1. \end{cases}$$

$$\Rightarrow \lim_{n\to\infty} \mathbb{P}\left(\left| \langle \hat{\theta}, \theta \rangle^2 / n^2 - \Delta(\lambda) \right| \ge \varepsilon \right) = 0$$

---

③ Concentration of Lasso training loss.

Consider the LASSO proble.

$A \in \mathbb{R}^{n\times d}$. $\quad A_{ij} \sim N(0, \frac{1}{n})$. $\quad x_0 \in \mathbb{R}^d$. with $\|x_0\|_2^2/d \le M$.

$y = Ax_0 + \varepsilon$. $\quad$ with $\varepsilon = (\varepsilon_i)_{i\in[n]} \sim_{i.i.d.} N(0, \tau^2)$.

$$\hat{x}_1 = \arg\min_x \frac{1}{\sqrt{n}} \|y - Ax\|_2 + \frac{\lambda}{d}\|x\|_1 \quad \leftarrow \text{sqrt Lasso.}$$

$+h\ f(\hat{x}_1)$

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

$$\|u\|_2^2 = \sup_{v\in\mathbb{R}^d} 2\langle u, v\rangle - \frac{1}{2}\|v\|_2^2. \quad \leftarrow \|v\|_2 \le K\cdot\sqrt{d}.$$

Prop: Let $\Omega \subseteq \mathbb{R}^d$ compact region. $\quad \Omega = \{x \in \mathbb{R}^d : \|x\|_2^2/d \le D\}$.

$$\sup_{x\in\Omega} \{\|x\|_2^2/d\} \le D.$$

Define $\quad f_\Omega(A, \varepsilon) = \min_{x\in\Omega} \frac{1}{\sqrt{n}} \|y - Ax\|_2 + \frac{\lambda}{d}\|x\|_1$.

$\exists K < \infty$, s.t. $\forall \delta > 0$, we have

$$\mathbb{P}\left(\left| f_\Omega(A,\varepsilon) - \mathbb{E}[f_\Omega(A,\varepsilon)] \right| \ge K\left(\sqrt{\frac{d}{n}(M+D)} + \tau \right) \sqrt{\frac{\log(2/\delta)}{n}} \right) \le \delta.$$

Proof: Define $\bar{A} = \sqrt{n}\, A$, $\quad \bar{\varepsilon} = \varepsilon/\tau$.

Define $F(\bar{A}, \bar{\varepsilon}) = f_\Omega(\bar{A}/\sqrt{n}, \bar{\varepsilon}\cdot\tau) = f_\Omega(A, \varepsilon)$.

$$F(\bar{A}, \bar{\varepsilon}) = \min_{x\in\Omega} \frac{1}{\sqrt{n}} \|A(x_0 - x) + \varepsilon\|_2 + \frac{\lambda}{n}\|x\|_1.$$

$$\|u\|_2 = \sup_{\|v\|_2 \le 1} \langle u, v \rangle.$$

$$= \min_{x\in\Omega} \sup_{\substack{v \\ \|v\|_2\le 1}} \frac{1}{\sqrt{n}} \langle v, A(x_0 - x) + \varepsilon \rangle + \frac{\lambda}{n}\|x\|_1$$

$$= \min_{x\in\Omega} \sup_{\|v\|_2\le 1} \underbrace{\frac{1}{n}\langle v, \bar{A}(x_0 - x)\rangle + \frac{\tau}{\sqrt{n}}\langle v, \bar{\varepsilon}\rangle + \frac{\lambda}{n}\|x\|_1}_{L(\bar{A}, \bar{\varepsilon}, x, v)}$$

$$F(\bar{A}_1, \bar{\varepsilon}_1) - F(\bar{A}_2, \bar{\varepsilon}_2)$$

$$= \min_{x \in \Omega} \sup_{\|v\|_2 \le 1} L(\bar{A}_1, \bar{\varepsilon}_1, x, v) + \max_{x \in \Omega} \inf_{\|v\|_2 \le 1} \left[-L(\bar{A}_2, \bar{\varepsilon}_2, x, v) \right]$$

$x_* = \text{argmax}$

$$\le \sup_{\|v\|_2 \le 1} L(\bar{A}_1, \bar{\varepsilon}_1, x_*, v) + \inf_{\|v\|_2 \le 1} \left[-L(\bar{A}_2, \bar{\varepsilon}_2, x_*, v) \right].$$

$v_* = \text{argmax}$

$$\le L(\bar{A}_1, \bar{\varepsilon}_1, x_*, v_*) - L(\bar{A}_2, \bar{\varepsilon}_2, x_*, v_*).$$

$$= \frac{1}{n} \langle v, (\bar{A}_1 - \bar{A}_2)(x_0 - x_*) \rangle + \frac{\tau}{\sqrt{n}} \langle v_*, \bar{\varepsilon}_1 - \bar{\varepsilon}_2 \rangle.$$

$$\le \frac{1}{n} \|\bar{A}_1 - \bar{A}_2\|_{op} \underbrace{\|v_*\|_2}_{\le 1} \underbrace{\|x_0 - x_*\|_2}_{\le \sqrt{d} \cdot C} + \frac{\tau}{\sqrt{n}} \|\bar{\varepsilon}_1 - \bar{\varepsilon}_2\|_2 \underbrace{\|v_*\|_2}_{\le 1}.$$

$n \propto d.$

$$\le C \cdot \left(\|\bar{A}_1 - \bar{A}_2\|_F + \|\bar{\varepsilon}_1 - \bar{\varepsilon}_2\|_2 \right) / \sqrt{n}.$$

$$\le C_1 \sqrt{\|\bar{A}_1 - \bar{A}_2\|_F^2 + \|\bar{\varepsilon}_1 - \bar{\varepsilon}_2\|_2^2} \, / \sqrt{n}.$$

$$G = (\bar{A}, \bar{\varepsilon}) \in \mathbb{R}^{nd+n}$$

$\|G\|_2.$

F is $\frac{C_1}{\sqrt{n}}$ - Lipschit in G.