Mean Field Asymptotics in Statistical Learning.

Feb 3rd.

Lecture 5. $\mathbb{Z}_2$ synchronization and the free energy approach.

① Asymptotic risk in $\mathbb{Z}_2$ synchronization.

a) Model setup.

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

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

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

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

Expected risk :
$$R(\hat{\Theta}) = \mathbb{E}_{\substack{\theta \sim \mathrm{Unif}(\mathbb{Z}_2^n) \\ Y \sim \mathbb{P}(Y|\theta)}} \| \hat{\Theta}(Y) - \theta\theta^T \|_F^2 / n^2$$

☆ Connection to stochastic block model.

People form two groups (more generally, k groups).

People from same group form an edge w/ prob p independently $\quad (p > q) \quad p = \frac{a}{n}$

People from different group form an edge w/ prob q independently. $\quad q = \frac{b}{n}$.

Observe the adjacency matrix A. Infer the two groups.

$\theta \sim \mathrm{Unif}(\{\pm 1\}^n)$. Give θ, generate a graph $G = (V, E)$, $\quad V = \{1, 2, \cdots, n\}$,

$V_+ \equiv \{ i \in V : \theta_i = +1 \}$, $\quad V_- \equiv \{ i \in V : \theta_i = -1 \}$.

$$A_{ij} \sim \begin{cases} \mathrm{Ber}(p) & i, j \text{ in the same group} \\ \mathrm{Ber}(q) & i, j \text{ in different group.} \end{cases} \qquad A_{ii} = 0.$$

$$\mathbb{E}[A \mid \theta] = \tfrac{1}{2}(p+q) \mathbf{1}\mathbf{1}^T + \tfrac{1}{2}(p-q)\theta\theta^T - pI.$$

$Y \equiv A - \frac{1}{2}(p+q)\mathbf{1}\mathbf{1}^T + pI$. then $\mathbb{E}[Y | \theta] = \frac{1}{2}(p-q)\theta\theta^T = \frac{a-b}{2n}\theta\theta^T$

$W = Y - \frac{a-b}{2n}\theta\theta^T$ (mean 0 noise).

$\mathrm{Var}(W_{ij}) = \frac{1}{2}p(1-p) + \frac{1}{2}q(1-q) \approx \frac{a+b}{2n}$ $\qquad Y = \frac{a-b}{2n}\theta\theta^T + W$

Effective SNR: $\lambda \equiv \dfrac{a-b}{\sqrt{2(a+b)}}$

b) Estimators in $\mathbb{Z}_2$ sync

MLE : $\hat{\theta}_{ML}(Y) \equiv \underset{\sigma \in \{\pm 1\}^n}{\operatorname{argmax}} \langle \sigma, Y\sigma \rangle$. Computationally NP hard.

$\{\pm 1\}^n \subseteq S^{n-1}(\sqrt{n})$

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

leading eigenvector.

$\hat{\Theta}_{spec}(Y) = \hat{\theta}_{spec} \hat{\theta}_{spec}^T \in \mathbb{R}^{n \times n}$.

SDP estimator : $\hat{\Theta}_{SDP}(Y) = \underset{X}{\operatorname{argmax}} \langle Y, X \rangle$

s.t. $X \succeq 0$

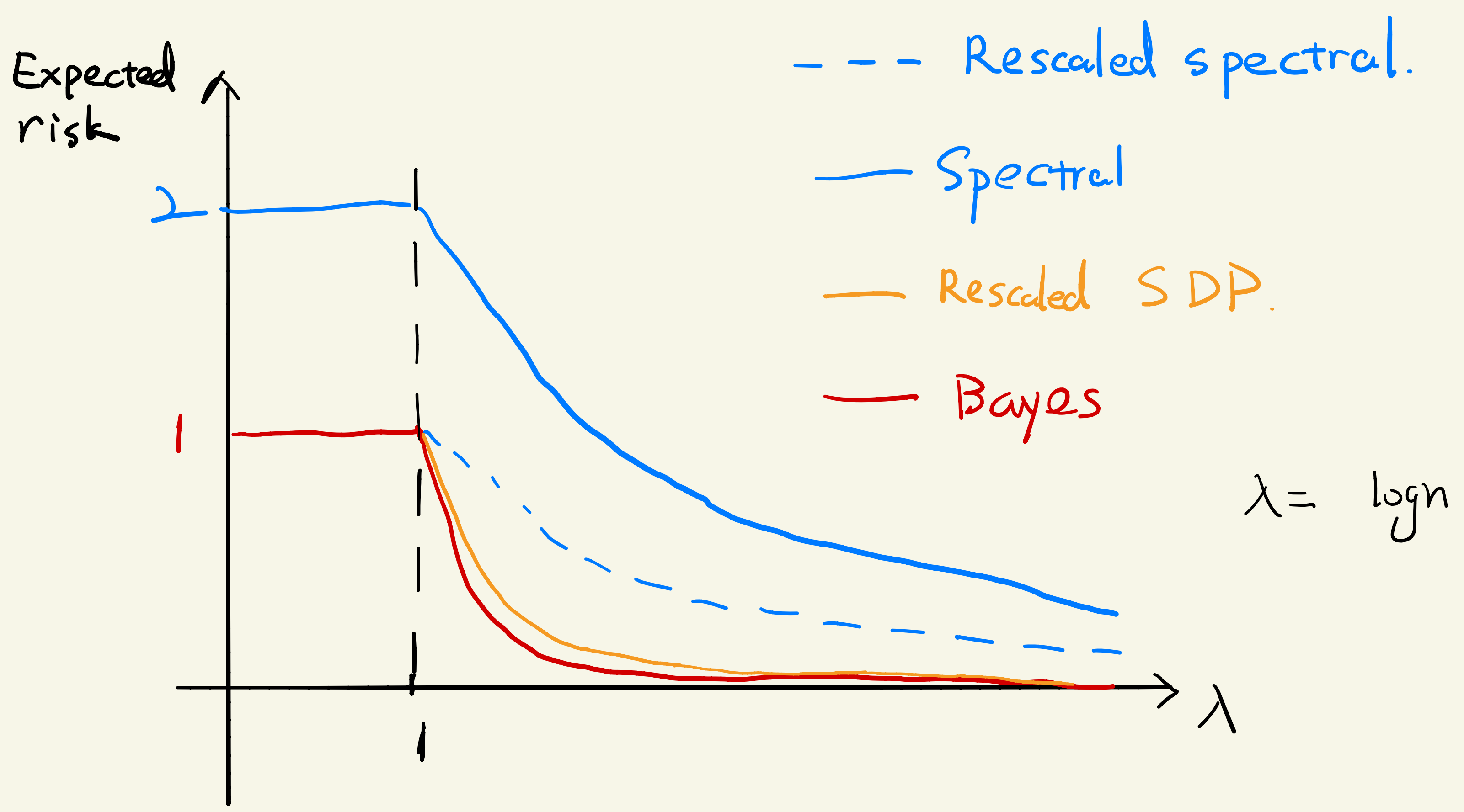
$X_{ii} = 1$

$(\operatorname{rank}(X) = 1)$

Bayes estimator : $\hat{\Theta}_{Bayes}(Y) = \mathbb{E}[\theta\theta^T \mid Y] = \sum_{\sigma \in \{\pm 1\}^n} \sigma\sigma^T p(\sigma \mid Y)$

$p(\sigma \mid Y) \propto \exp\{\lambda \langle \sigma, Y\sigma \rangle / 2\}$.

c) Expected risk $\mathbb{E} \| \hat{\Theta}(Y) - \theta\theta^T \|_F^2 / n^2$ v.s. λ.



(No statistical – computational gap in this model)

d) Asymptotic formula.

BBP phase transition.

Proposition :

(1) Spectral estimator

$$\lim_{n\to\infty} \langle \hat\theta_{spec}, \theta \rangle^2/n^2 = \begin{cases} 0, & \text{for } \lambda \le 1 \\ 1-\frac{1}{\lambda^2}, & \text{for } \lambda > 1. \end{cases}$$

almost surely convergence.

$$\lim_{n\to\infty} \| \hat\theta_{spec}\hat\theta_{spec}^T - \theta\theta^T \|_F^2/n = \begin{cases} 2 & \text{for } \lambda \le 1. \\ \frac{2}{\lambda^2} & \text{for } \lambda > 1 \end{cases}$$

To show this
① Show concentration.
② Calculate limit.

(2) Bayes estimator.

$$\lim_{n\to\infty} \langle \theta, \hat\Theta_{Bayes}\theta \rangle/n^2 = \begin{cases} 0, & \text{for } \lambda \le 1 \\ q_*(\lambda)^2 & \lambda > 1 \end{cases}$$

$$\lim_{n\to\infty} \| \hat\Theta_{Bayes} - \theta\theta^T \|_F^2/n^2 = \begin{cases} 1, & \text{for } \lambda \le 1 \\ 1-q_*(\lambda)^2, & \text{for } \lambda > 1. \end{cases}$$

where $q_*(\lambda)$ is the unique non-negative solution of

$$q = \mathbb{E}_{G\sim N(0,1)}[\tanh(\lambda^2 q + \lambda\sqrt{q}\,G)^2].$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

② Derivation of asymptotic risk.

Quantity of interest : $\lim_{n\to\infty} \mathbb{E}[\| \hat\Theta_{Bayes}(Y) - \theta\theta^T\|_F^2]/n^2$

$$= 1 - 2\overset{q_*(\lambda)^2}{a_*} + \overset{q_*(\lambda)^2}{b_*}.$$

$$a_* = \lim_{n\to\infty} \mathbb{E}[\langle \theta, \hat\Theta_{Bayes}\theta \rangle]/n^2.$$

$$b_* = \lim_{n\to\infty} \mathbb{E}[\| \hat\Theta_{Bayes}(Y)\|_F^2]/n^2$$

General recipe for for calculating $m_* = \lim_{n\to\infty} \mathbb{E}[\;?\;]$.

Free energy appoach.

(1). Find a configuration space Ω, ν_0.
a observable $M : \Omega \to \mathbb{R}$.
a perturbed Hamiltonian $H_\lambda : \Omega \to \mathbb{R}$.

$$H_\lambda(\sigma) = H_0(\sigma) + \lambda M(\sigma)$$

$$P_{\beta,\lambda}(\sigma) \propto \exp\{-\beta H_\lambda(\sigma)\}.$$

s.t. $\mathbb{E}[\;?\;] = \langle M \rangle_{\beta,\lambda}/n$ for some β and λ.

(2) $f(\beta,\lambda) = \lim_{n\to\infty} -\frac{1}{n\beta}\mathbb{E}[\log\int \exp\{-\beta H_\lambda(\sigma)\}\nu_0(d\sigma)]$.

Calculate this analytic.

(3) $m_* = \lim_{n\to\infty} \mathbb{E}[\langle M \rangle_{\beta,\lambda}]/n = \partial_\lambda f(\beta,\lambda)$.

a). (1) $$a_* = \lim_{n\to\infty} \mathbb{E}[\langle \theta, \hat{\Theta}_{Bayes}\, \theta \rangle]/n^2$$

$$= \lim_{n\to\infty} \mathbb{E}[\underbrace{\langle \theta, \sum_{\sigma\in\{\pm1\}^n} \sigma\sigma^T p(\sigma|Y)\, \theta \rangle}_{\downarrow}]/n^2.$$

$$= \sum_{\sigma\in\{\pm1\}^n} (\langle \sigma, \theta \rangle^2/n^2)\, p(\sigma|Y).$$

$$p(\sigma|Y) \propto \exp\{\lambda \langle \sigma, Y\sigma \rangle/2\}$$

$$= \exp\{-\lambda[-\langle \sigma, W\sigma \rangle/2 - \tfrac{\lambda}{2n}\langle \theta, \sigma \rangle^2]\}.$$

Define $\Omega = \{\pm 1\}^n$, $\nu_0 = \text{Unif}$.

$$M(\sigma) = -\langle \sigma, \theta \rangle^2/(2n)$$

$$H_\lambda(\sigma) = -\langle \sigma, W\sigma \rangle/2 + \lambda M(\sigma).$$

$$P_{\beta,\lambda}(\sigma) \propto \exp\{-\beta H_\lambda(\sigma)\}.$$

$$p(\sigma|Y) = P_{\beta,\lambda}(\sigma)\Big|_{\beta=\lambda}.$$

$$a_* = \lim_{n\to\infty} -2\cdot \mathbb{E}[\langle M \rangle_{\lambda,\lambda}]/n$$

(2) $$f(\beta,\lambda) = \lim_{n\to\infty} -\frac{1}{n\beta}\mathbb{E}\left[\log\int \exp(-\beta H_\lambda(\sigma))\right]$$

$$f(\beta,\lambda) = \max_{b,q} f_{mf}(b, q; \beta, \lambda)$$

$$f_{mf}(b, q; \beta, \lambda) \equiv -\frac{1}{4}\beta(1-q)^2 + \frac{1}{2}\lambda b^2 - \frac{1}{\beta}\mathbb{E}[\log 2\cosh(\beta(\lambda b + \sqrt{q}\,G))]$$

$G \sim N(0,1)$.

$$\begin{cases} b_* = \mathbb{E}[\tanh(\beta(\lambda b_* + \sqrt{q_*}\,G))] \\ q_* = \mathbb{E}[\tanh(\beta(\lambda_* + \sqrt{q_*}\,G))^2] \end{cases}$$

$$m_*(\beta,\lambda) = \partial_\lambda f(\beta,\lambda) = \frac{1}{2}b^2 - \frac{1}{\cancel{\beta}}\underline{\mathbb{E}[\tanh(\beta(\lambda b + \sqrt{q}\,G))]}\cdot \cancel{\beta}\underline{b}\Big|_{b_*, q_*}$$

$$= \frac{1}{2}b_*^2 - b_*^2 = -\frac{1}{2}b_*^2$$

$$a_* = -2\, m_*(\beta,\lambda)\Big|_{\beta=\lambda} = b_*^2(\lambda,\lambda) = q_*(\lambda)^2$$

$$q_* = \mathbb{E}[\tanh(\beta(\lambda_* + \sqrt{q_*})G)^2].$$

$$f_*(\lambda) = \max_q f(q,\lambda)$$

implicit differentiation

$$\partial_\lambda f_*(\lambda) = \partial_\lambda\left[\max_q f(q,\lambda)\right] = \partial_\lambda f(q,\lambda)\Big|_{q=q_*}.$$

$$q_* = \arg\max_q f(q,\lambda)$$

(b) $$b_* = \lim_{n\to\infty} \mathbb{E}[\ \| \hat{\Theta}_{Bayes}(Y) \|_F^2] / n^2.$$

$$\| \hat{\Theta}_{Bayes} \|_F^2 / n^2 \qquad \Theta = \{\pm 1\}^n$$

$$= \langle \sum_{\sigma_1 \in \Theta} \sigma_1 \sigma_1^T p(\sigma_1 | Y), \sum_{\sigma_2 \in \Theta} \sigma_2 \sigma_2^T p(\sigma_2 | Y) \rangle / n^2$$

$$= \sum_{\substack{\sigma_1 \in \Theta \\ \sigma_2 \in \Theta}} \langle \sigma_1 \sigma_1^T, \sigma_2 \sigma_2^T \rangle\ p(\sigma_1 | Y)\, p(\sigma_2 | Y) / n^2.$$

$$= \sum_{\substack{\sigma_1 \in \Theta \\ \sigma_2 \in \Theta}} \langle \sigma_1, \sigma_2 \rangle^2\ p(\sigma_1 | Y)\, p(\sigma_2 | Y) / n^2.$$

$$= \sum_{(\sigma_1, \sigma_2) \in \Theta \times \Theta} \langle \sigma_1, \sigma_2 \rangle^2\ \mu(\sigma_1, \sigma_2 | Y) / n^2$$

$$\sigma = (\sigma_1, \sigma_2) \in \Theta \times \Theta.$$

$$\mu(\sigma_1, \sigma_2 | Y) = p(\sigma_1 | Y)\, p(\sigma_2 | Y)$$
$$\propto \exp\{ \lambda \langle \sigma_1, Y \sigma_1 \rangle / 2 + \lambda \langle \sigma_2, Y \sigma_2 \rangle / 2 \}.$$

Define $\Omega = \Theta \times \Theta$. $\quad \nu_0 \sim \mathrm{Unif}(\Theta) \times \mathrm{Unif}(\Theta).$

$\sigma \in \Omega \qquad \sigma = (\sigma_1, \sigma_2) \qquad \sigma_1, \sigma_2 \in \Theta.$

$$M(\sigma) = \langle \sigma_1, \sigma_2 \rangle^2 / n.$$

$$H_{\lambda,h}(\sigma) = -\frac{1}{2} \langle \sigma_1, W \sigma_1 \rangle - \frac{\lambda}{2n} \langle \sigma_1, \theta \rangle^2$$
$$- \frac{1}{2} \langle \sigma_2, W \sigma_2 \rangle - \frac{\lambda}{2n} \langle \sigma_2, \theta \rangle^2$$
$$+ h \langle \sigma_1, \sigma_2 \rangle^2 / n.$$

$$P_{\beta,\lambda,h}(\sigma) \propto \exp\{ -\beta H_{\lambda,h}(\sigma) \}$$

$$b_* = \lim_{n\to\infty} \mathbb{E}[\langle M \rangle_{\beta,\lambda,h}] \Big|_{\substack{\beta=\lambda \\ h=0}}.$$

(2). $$f(\beta, \lambda, h) = \text{formula} = \max_{?\,?} f_{\text{mf}}(?, ?\ ; \beta, \lambda, h).$$

$$b_* = \partial_h f(\beta, \lambda, h) \Big|_{\substack{\beta=\lambda \\ h=0}}$$

$$b_* = q_*(\lambda)^2.$$

③ $$m_* = \lim_{n\to\infty} \mathbb{E}\left[\langle \hat{\theta}_{spec}(Y), \theta \rangle^2 / n^2 \right].$$

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

$$M(\sigma) = \langle \sigma, \theta \rangle^2 / n.$$

$$H_\lambda(\sigma) = - \langle \sigma, Y\sigma \rangle / 2$$

$$P_{\beta,\lambda,h}(\sigma) \propto \exp\{ -\beta [H_\lambda(\sigma) + h M(\sigma)] \}.$$

$$\begin{aligned}
m_* &= \lim_{n\to\infty} \mathbb{E}[M(\hat{\theta}_{spec})] / n \\
&= \lim_{n\to\infty} \mathbb{E}[M(\arg\min_\sigma H_\lambda(\sigma))] / n \\
&= \lim_{n\to\infty} \lim_{\beta\to\infty} \mathbb{E}[\langle M \rangle_{\beta,\lambda,h}] / n \Big|_{h=0} \\
&\overset{?}{=} \lim_{\beta\to\infty} \lim_{n\to\infty} \mathbb{E}[\langle M \rangle_{\beta,\lambda,h}] / n \Big|_{h=0} \\
&= \lim_{\beta\to\infty} \lim_{n\to\infty} \mathbb{E}[\partial_h F(\beta,\lambda,h)] / n \Big|_{h=0} \\
&= \lim_{\beta\to\infty} \partial_h f(\beta,\lambda,h) \Big|_{h=0}.
\end{aligned}$$

Exercise: Suppose $\beta_0 \sim N(0, \sigma_0^2 I_d)$.

$y = X\beta_0 + \varepsilon$, $\quad X_{ij} \sim_{i.i.d.} N(0, \frac{1}{d})$. $\quad \varepsilon_i \sim_{i.i.d.} N(0, \sigma^2)$.

Denote $\hat{\beta} = \arg\min_\beta \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 \quad = (X^T X + \lambda I)^{-1} X^T y$

① Figure out a configuration space Ω, ν_0

an observable $M: \Omega \to \mathbb{R}$

and a perturbed Hamiltonian $H_\lambda: \Omega \to \mathbb{R}$.

s.t. defining $F(\beta,\lambda) = -\frac{1}{\beta} \log \int \exp\{-\beta H_\lambda(\sigma)\} \nu_0(d\sigma)$.

we have $\langle \hat{\beta}, \beta_0 \rangle / n = \lim_{\beta\to 0} \partial_\lambda \lim_{n\to\infty} F(\beta,\lambda)/n$

② Do similar things for $\|\hat{\beta}\|_2^2 / n$.

③ Hopefully, your $H_\lambda(\sigma)$ is a quadratic function of $\sigma \in \mathbb{R}^d$ and $\nu_0(\sigma)$ is Lebesgue measure.

In this case, $\int \exp\{-\beta H_\lambda(\sigma)\} \nu_0(d\sigma)$ is a Gaussian integration, and can be written out explicitly.

Please simplify $\mathbb{E}_{X,\beta}[F(\beta,\lambda)]$ as much as possible