Mean Field Asymptotics in Statistical Learning.

Feb 17th

Lecture 8: Field theoretic calculations.

Large deviations.

Replica methods.

Feedback form

① Large deviation of overlap matrix and field theoretic calculations.

$\sigma_1, \cdots, \sigma_k \in \mathbb{R}^n$ $\quad k$ fixed, $\quad n \to \infty$

Let $(\sigma_i)_{i\in[k]} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(S^{n-1}(\sqrt{n}))$.

$\sigma = [\sigma_1, \sigma_2, \cdots, \sigma_k] \in \mathbb{R}^{n\times k}$.

$$\bar{Q}(\sigma) = \sigma^T\sigma/n = \begin{bmatrix} \|\sigma_1\|_2^2/n, & \langle\sigma_1,\sigma_2\rangle/n, & \cdots \\ & & \\ & & \|\sigma_k\|_2^2/n \end{bmatrix} \in \mathbb{R}^{k\times k}.$$

$\bar{Q}(\sigma)$ symmetric $\qquad \bar{Q}(\sigma)_{ii} = 1$

Let $Q \in \mathbb{R}^{k\times k}$ be a symmetric matrix with $Q_{ii}=1$.

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\bar{Q}(\sigma)_{ij} \in [-\varepsilon+Q_{ij}, Q_{ij}+\varepsilon], \forall i,j\right)$$

$$\mathbb{P}(\bar{Q}(\sigma) \approx Q).$$

$$\doteq \frac{\int_{\mathbb{R}^{n\times k}} \delta(\hat{Q}(\sigma)-Q)\prod_{i=1}^k d\sigma_i}{\int_{\mathbb{R}^{n\times k}} \prod_{i=1}^{n}\delta(\|\sigma_i\|_2^2/n - 1)\prod_{i=1}^k d\sigma_i} \doteq \frac{S_n(Q)}{T_n}$$

$(Q_{ii}=1)$.

$$S_n(Q) = \int_{\mathbb{R}^{n\times k}} \prod_{1\le i\le j\le k} \delta(\langle\sigma_i,\sigma_j\rangle - nQ_{ij}) \prod_{i=1}^k d\sigma_i$$

$$\delta(x-a) = \int_{\mathbb{R}} e^{ip(x-a)}dp$$

$\frac{1}{n}\log S_n(Q)$

$$= \int_{\mathbb{R}^{n\times k}} \prod_{i=1}^k d\sigma_i \frac{1}{(2\pi)^{k(k+1)/2}} \int_{\mathbb{R}^{k(k+1)/2}} \prod_{1\le i\le j\le k} \exp\{-i\lambda_{ij}\langle\sigma_i,\sigma_j\rangle + i\lambda_{ij}nQ_{ij}\} \prod_{1\le i<j\le k} d\lambda_{ij}$$
$$\prod_{1\le i\le k}\exp\{-i\lambda_{ii}\|\sigma_i\|_2^2/2 + i\lambda_{ii}nQ_{ii}/2\}\prod_{1\le i\le k}d\lambda_{ii}$$

↓ saddle point approx.

$$\doteq \underset{\Lambda\in\mathbb{R}^{k\times k}}{\overset{\text{ext}}{\inf}} \int_{\mathbb{R}^{n\times k}} \left(\prod_{i=1}^k d\sigma_i\right)\exp\left\{-\sum_{ij=1}^k \lambda_{ij}\langle\sigma_i,\sigma_j\rangle/2 + n\sum_{ij=1}^k\lambda_{ij}Q_{ij}/2\right\}$$

$\mathrm{ext}\{\Lambda\in\mathbb{C}^{k\times k}\}$

$$= \inf_{\Lambda} \int_{\mathbb{R}^{n\times k}} \prod_{i=1}^{k}\prod_{\alpha=1}^{n} d\sigma_i^\alpha \exp\left\{-\sum_{i,j=1}^{k}\lambda_{ij}\sum_{\alpha=1}^{n}\sigma_i^\alpha\sigma_j^\alpha/2\right\}\times\exp\left\{n\sum_{i,j=1}^{k}\lambda_{ij}Q_{ij}/2\right\}.$$

$$= \inf_{\Lambda}\left(\int_{\mathbb{R}^k}\prod_{i=1}^{k}d\sigma_i\exp\left\{-\sum_{i,j=1}^{k}\lambda_{ij}\,\sigma_i\sigma_j/2\right\}\right)^n\times\exp\left\{n\sum_{i,j=1}^{k}\lambda_{ij}Q_{ij}/2\right\}.$$

$$\left(\begin{aligned}&\int_{\mathbb{R}^k}\prod_{i=1}^{k}d\sigma_i\exp\left\{-\sum_{i,j=1}^{k}\lambda_{ij}\sigma_i\sigma_j/2\right\}=\int_{\mathbb{R}^k}\exp\{-\langle\bar\sigma,\Lambda\bar\sigma\rangle/2\}d\bar\sigma\\&=\det(\Lambda)^{-\frac{1}{2}}\left(\sqrt{2\pi}\right)^k\end{aligned}\right)$$

$$= \inf_{\Lambda}\left(\det(\Lambda)^{-\frac{1}{2}}\cdot\left(\sqrt{2\pi}\right)^k\right)^n\times\exp\{n\langle\Lambda,Q\rangle/2\}.$$

$$= \inf_{\Lambda}\exp\left\{n\left[\langle\Lambda,Q\rangle/2-\frac{1}{2}\log\det(\Lambda)+\frac{k}{2}\log(2\pi)\right]\right\}.$$

$$\frac{1}{n}\log S_n(Q) = \inf_{\Lambda}\left[\langle\Lambda,Q\rangle/2-\frac{1}{2}\log\det(\Lambda)+\frac{k}{2}\log 2\pi\right]$$

$$= \frac{1}{2}\log\det(Q)+\frac{k}{2}\log 2\pi.$$

$$\frac{1}{n}\log T_n = \frac{k}{2}\log 2\pi.$$

$$Q \approx \mathbb{E}[\bar{Q}(\sigma)] = \mathbb{1}$$

$$\frac{1}{n}\log\mathbb{P}(\bar{Q}(\sigma)\approx Q) \doteq \frac{1}{2}\log\det(Q).$$

$$\sup_{Q}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\bar{Q}(\sigma)\approx Q) = 0.$$

② Field theoretic calculation for general large deviations.

$x_i \in \mathbb{R}^k$, $x_i \sim$ i.i.d. $\mathbb{P}_x$, $\quad M: \mathbb{R}^k \to \mathbb{R}^p$.

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} M(x_i) \approx m\right) \qquad \lim_{\varepsilon\to 0}\lim_{n\to 0}\frac{1}{n}\log \mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{n} M(x_i) - m\right\|^2 \le \varepsilon\right)$$

$$\doteq \mathbb{E}_x\left[\delta\left(\sum_{i=1}^{n} M(x_i) - nm\right)\right]$$

$$\doteq \mathbb{E}_x\left[\frac{1}{(2\pi)^p}\int_{\mathbb{R}^p} \exp\left\{ i\left< \sum_{i=1}^{n} M(x_i) - nm, \Lambda\right>\right\} d\Lambda\right]$$

$$= \frac{1}{(2\pi)^p}\int_{\mathbb{R}^p} \exp\{-in\langle m,\Lambda\rangle\}\ \mathbb{E}_x\left[\exp\left\{i\left<\sum_{i=1}^{n} M(x_i), \Lambda\right>\right\}\right] d\Lambda$$

$$= \frac{1}{(2\pi)^p}\int_{\mathbb{R}^p} \exp\{-in\langle m,\Lambda\rangle\}\left(\mathbb{E}\left[e^{i\langle\Lambda, M(x)\rangle}\right]\right)^n d\Lambda$$

$$\doteq \operatorname*{ext}_{\Lambda}\ \exp\{-n\langle m, m\rangle\} \times \mathbb{E}\left[e^{\langle\Lambda, M(x)\rangle}\right]^n$$

$$\Rightarrow \lim_{n\to\infty}\frac{1}{n}\log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} M(x_i) \approx m\right)$$

$$\doteq \operatorname*{exp}_{\lambda}\left\{-\langle m,\Lambda\rangle + \log \mathbb{E}\left[e^{\langle\Lambda, M(x)\rangle}\right]\right\}.$$

③ Cramer's thm.

Thm (Cramer) Let $X_i \sim$ i.i.d. μ_X. Let $f: \mathcal{X} \to \mathbb{R}$.

If $A \subseteq \mathbb{R}$ is a closed interval of real axis, then

$$\lim_{n\to\infty}\frac{1}{n}\log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i) \in A\right) = -\inf_{a\in A} I(a) \qquad \text{rate function}$$

where $I(a) = \sup_{\lambda}\left\{\lambda a - \log \mathbb{E}_{x\sim\mu_x}\left[e^{\lambda f(x)}\right]\right\}$.

We also write $\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i) \in A\right) \doteq \exp\left\{-n \inf_{a\in A} I(a)\right\}$.

Intuition: $\quad A = \{t : t \ge a\}$.

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i) \in A\right)$$

$$= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i) \ge a\right) = \mathbb{P}\left(e^{\lambda\sum_{i=1}^{n} f(x_i)} \ge e^{\lambda n a}\right)$$

$$\le \inf_{\lambda}\frac{\mathbb{E}\left[e^{\lambda\sum_{i=1}^{n} f(x_i)}\right]}{e^{n\lambda a}} = \inf_{\lambda}\frac{\mathbb{E}_x\left[e^{\lambda f(x)}\right]^n}{e^{n\lambda a}}$$

④ Replica method.

The Spiked GOE matrix.

$u \sim \mathrm{Unif}(S^{n-1})$.

Let $u \in S^{n-1} \equiv \{ x \in \mathbb{R}^n : \|x\|_2 = 1 \}$.

$UWU^T \overset{d}{=} W$

$Y = \lambda u u^T + W \in \mathbb{R}^{n\times n}$.

$\lambda \geq 0$. $\tilde{Y} = \lambda U u u^T U^T + W \overset{d}{=} Y$.

$W \sim \mathrm{GOE}(n)$. $W_{ij} \sim_{i.i.d.} N(0, \frac{1}{n})$ $\perp$ $W_{ij} = W_{ji}$

$W_{ii} \sim_{i.i.d} N(0, \frac{2}{n})$

Our interest:

$$\varphi(\lambda) = \lim_{n\to\infty} \mathbb{E}\left[\sup_{\sigma \in S^{n-1}} \langle \sigma, Y\sigma \rangle\right]$$

$$\hat{\theta} = \operatorname{argmax}_{\sigma \in S^{n-1}} \langle \sigma, Y\sigma \rangle$$

MLE to estimate u.

$$m(\lambda) = \lim_{n\to\infty} \mathbb{E}[\langle \hat{\theta}, u \rangle^2].$$

HW1:

$$\boxed{\varphi(\lambda) = \begin{cases} 2, & \lambda \leq 1. \\ \lambda + \frac{1}{\lambda}, & \lambda > 1. \end{cases} \qquad m(\lambda) = \begin{cases} 0, & \lambda \leq 1. \\ 1 - \frac{1}{\lambda^2}, & \lambda > 1. \end{cases}}$$

← BBP phase transition

$\lambda = 1$.

[Baik, Ben Arous, Peche. 2005]

[CS 85] ←

↑ Sommer

[SK 1975]

The free energy approach. (Lecture 5).

$\Omega = S^{n-1}$. ν_0 is uniform on Ω.

$$H_\lambda(\sigma) = -n \langle \sigma, W\sigma \rangle - n\lambda \langle \sigma, u \rangle^2.$$

$$Z_n(\beta, \lambda) = \int_{S^{n-1}} \exp\{-\beta H_\lambda(\sigma)\} \nu_0(d\sigma).$$

$$\Phi_n(\beta, \lambda) = \log Z_n(\beta, \lambda).$$

$$\varphi(\beta, \lambda) = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[\log Z_n(\beta, \lambda)] \quad \leftarrow ?$$

$$\begin{cases} \varphi(\lambda) = \lim_{\beta\to\infty} \frac{1}{\beta} \varphi(\beta, \lambda) \to \mathbb{E}[\max_\sigma H_\lambda(\sigma)/n] \\ m(\lambda) = \varphi'(\lambda) \end{cases}$$

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}[\log Z_n(\beta, \lambda)] = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left[\log \int_{S^{n-1}} \exp\{-\beta H_\lambda(\sigma)\} \nu_0(d\sigma)\right].$$

$$\mathbb{E}\left[\int_{S^{n-1}} \exp\{-\beta H_\lambda(\sigma)\} \nu_0(d\sigma)\right]$$

$$= \int_{S^{n-1}} \underline{\mathbb{E}[\exp\{-\beta H_\lambda(\sigma)\}]} \nu_0(d\sigma)$$

$$\frac{1}{n} \mathbb{E}[\log Z_n(\beta, \lambda)] \overset{?}{\asymp} \frac{1}{n} \log \mathbb{E}[Z_n(\beta, \lambda)]$$

Lemma (Replica trick). $\mathbb{E}[\log Z] \equiv \lim_{k\to 0} \frac{1}{k} \log \mathbb{E}[Z^k]$.

Heuristic derivation:

$$\mathbb{E}[\log Z] = \mathbb{E}[(\log Z^k)/k] = \lim_{k\to 0} \mathbb{E}[\log(1+\underbrace{Z^k-1}_{})/k]$$

$Z^k-1 \to 0$ as $k\to 0$.

$\log(1+x) \approx x, \quad x\to 0$

$$= \lim_{k\to 0} \mathbb{E}[(Z^k-1)k]$$

$$= \lim_{k\to 0} (\mathbb{E}[Z^k]-1)/k = \lim_{k\to 0} \frac{1}{k} \log(1+\mathbb{E}[Z^k]-1)$$

$$= \lim_{k\to 0} \frac{1}{k} \log \mathbb{E}[Z^k].$$

Sherrington - Kirkpatrick [1975].

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}[\log Z_n] = \lim_{n\to\infty} \lim_{k\to 0} \frac{1}{nk} \mathbb{E}[Z_n^k] = \lim_{k\to 0} \lim_{n\to\infty} \frac{1}{nk} \mathbb{E}[Z_n^k].$$

$\mathbb{E}[Z_n^k]$ when k is integer.

$$\mathbb{E}\left[\left(\int_{S^{n-1}} \exp\{-\beta H_\lambda(\sigma)\} \nu_0(d\sigma)\right)^k\right].$$

$$= \mathbb{E}\left[\int_{(S^{n-1})^{\otimes k}} \exp\left\{-\beta \sum_{a=1}^{k} H_\lambda(\sigma^a)\right\} \prod_{a=1}^{k} \nu_0(d\sigma^a)\right].$$

$$= \int_{(S^{n-1})^{\otimes k}} \mathbb{E}\left[\exp\left\{-\beta \sum_{a=1}^{k} H_\lambda(\sigma^a)\right\}\right] \prod_{a=1}^{k} \nu_0(d\sigma^a).$$

---

a). $S(k,\beta,\lambda) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[Z_n^k] = \sup_Q U(Q)$.

b). $\varphi(\beta,\lambda) = \lim_{k\to 0} \frac{1}{k} S(k,\beta,\lambda)$.

c). $\varphi(\lambda) = \lim_{\beta\to\infty} \frac{1}{\beta} \varphi(\beta,\lambda)$.

a). Lemma: For $k \in \mathbb{N}_+$. $\quad Q \in \mathbb{R}^{(k+1)\times(k+1)}$

$$S(k,\beta,\lambda) = \sup_{Q} U(Q) \quad \text{s.t. } Q \succeq 0, \; Q_{ii} = 1. \qquad Q = (q_{ij})_{0\le i,j\le k}.$$

$$U(Q) = \underbrace{\beta\lambda \sum_{i=1}^{k} q_{0i}^2 + \beta^2 \sum_{ij=1}^{k} q_{ij}^2}_{} + \underbrace{\frac{1}{2}\log\det(Q)}_{}.$$

σ^a : replicas.

Derivation:

$$\mathbb{E}[Z_n(\beta,\lambda)^k] = \int_{(S^{n-1})^{\otimes k}} \mathbb{E}\Big[\exp\{-\beta \sum_{a=1}^{k} H_\lambda(\sigma^a)\}\Big] \prod_{a=1}^{k} \nu_0(d\sigma^a).$$

$$= \int_{(S^{n-1})^{\otimes k}} \mathbb{E}\Big[\exp\{\beta n (\sum_{a=1}^{k} \lambda <\sigma^a, u>^2 + <\sigma^a, W\sigma^a>)\}\Big] \prod_{a=1}^{k} \nu_0(d\sigma^a).$$

$$= \int_{(S^{n-1})^{\otimes k}} \exp\{\beta n (\sum_{a=1}^{k} \lambda <\sigma^a, u>^2)\} \times$$

$$\underbrace{\mathbb{E}[\exp\{\beta n \sum_{a=1}^{k} <\sigma^a, W\sigma^a>\}]}_{E} \prod_{c=1}^{k} \nu_0(d\sigma^a),$$

$W \sim GOE(n)$. $\quad W = (G+G^T)/\sqrt{2n}$, $\quad G \in \mathbb{R}^{n\times n}$.

$G_{ij} \sim_{i.i.d.} N(0,1)$. $\quad 1 \le i,j \le n$.

$$E = \mathbb{E}[\exp\{\beta n \sum_{a=1}^{k} <\sigma^a, (G+G^T)\sigma^a>/\sqrt{2n}\}]$$

$$= \mathbb{E}[\exp\{\beta n \,\mathrm{tr}(\sum_{a=1}^{k} \sigma^a(\sigma^a)^T (G+G^T))/\sqrt{2n}\}]$$

$$= \mathbb{E}[\exp\{\beta \cdot \sqrt{2n} < \sum_{a=1}^{k} \sigma^a(\sigma^a)^T, G>\}]$$

$\mathbb{E}_{G\sim N(0,1)} e^{\lambda G} = e^{\frac{1}{2}\lambda^2}$

$$= \exp\{\beta^2 n \,\| \sum_{a=1}^{k} \sigma^a(\sigma^a)^T \|_F^2\}.$$

$$= \exp\{\beta^2 n < \sum_{a=1}^{k} \sigma^a(\sigma^a)^T, \sum_{a=1}^{k} \sigma^a(\sigma^a)^T >\}$$

$$= \exp\{\beta^2 n \sum_{a,b=1}^{k} <\sigma^a(\sigma^a)^T, \sigma^b(\sigma^b)^T>\}$$

$$= \exp\{\beta^2 n \sum_{ab=1}^{k} <\sigma^a, \sigma^b>^2\}.$$

$$\Rightarrow \quad \mathbb{E}[Z_n(\beta,\lambda)^k]$$

$$= \int_{(S^{n-1})^{\otimes k}} \exp\{\beta n \sum_{a=1}^{k} \lambda <u, \sigma^a>^2 + \beta^2 n \sum_{ab=1}^{k} <\sigma^a, \sigma^b>^2\} \prod_{a=1}^{k} \nu_0(d\sigma^a).$$

$$\Big(1 = \int \prod_{1\le a\le k} \delta(<\sigma^a, u> - q_{0a}) \prod_{1\le a<b\le k} \delta(<\sigma^a, \sigma^b> - q_{ab}) \prod dq_{0a} \prod dq_{ab}\Big)$$

$k(k+1)/2$

$$= \int_{(S^{n-1})^{\otimes k}} \exp\{\beta n \sum_{a=1}^{k} \lambda q_{0a}^2 + \beta^2 n \sum_{ab=1}^{k} q_{ab}^2\}$$

$$\Big(\int \prod_{1\le a\le k} \delta(<\sigma^a, u> - q_{0a}) \prod_{1\le a<b\le k} \delta(<\sigma^a, \sigma^b> - q_{ab}) \prod dq_{0a} \prod dq_{ab}\Big) \prod \nu_0(d\sigma^a),$$

$$= \sup_{Q} \exp\{\beta n \sum_{a=1}^{k} \lambda q_{0a}^2 + \beta^2 n \sum_{ab=1}^{k} q_{ab}^2\}.$$

$$= [q_{ij})_{0\le i,j\le k} \quad \times \int_{(S^{n-1})^{\otimes k}} \prod \delta \quad \prod \delta \quad \prod \nu_0(d\sigma^a).$$

$$= \sup_{Q} \exp\left[n\left(\beta \sum_{a=1}^{k} \lambda q_{0a}^2 + \beta^2 \sum_{a,b=1}^{k} q_{ab}^2 + \frac{1}{2}\log\det(Q)\right)\right]$$