Mean Field Asymptotics in Statistical Learning.

Feb 10th

Lecture 7: Concentration inequalities. Field theoretic calculations.

① Concentration of ensemble average of an observable.

$$H_\lambda(\sigma) = H_0(\lambda) + \lambda M(\sigma). \qquad H_0(\sigma) = \langle \sigma, W\sigma \rangle. \qquad M(\sigma) = \langle \sigma, \theta \rangle^2/n$$

Show the concentration of $\langle M \rangle_{\beta_*, \lambda_*}/n$

$$\langle M \rangle_{\beta,\lambda} \equiv \int_\Omega M(\sigma)\, P_{\beta,\lambda}(d\sigma), \qquad P_{\beta,\lambda}(d\sigma) \propto \exp\{-\beta H_\lambda(\sigma)\}.$$

Step a). $F_n(\beta, \lambda)/n \equiv -\frac{1}{n\beta} \log \int_\Omega \exp\{-\beta H_\lambda(\sigma)\} \nu_0(d\sigma).$

Show concentration of $U_n(\lambda) \equiv F_n(\beta_*, \lambda)/n$

(Gaussian concentration inequality).

$$\lim_{n\to\infty} \mathbb{P}\left(|U_n(\lambda) - \mathbb{E}[U_n(\lambda)]| \geq \varepsilon \right) = 0 \qquad ①$$

Step b). Show that

limiting free energy.

$$\lim_{n\to\infty} \mathbb{E}\, U_n(\lambda) = \underline{u(\lambda)} \qquad \forall \lambda$$

and that $u(\lambda)$ is differentiable at λ_*.

i.e. defining

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

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

we have $\lim_{\delta\to 0^+} \Delta^+(\lambda_*, \delta) = \lim_{\delta\to 0^+} \Delta^-(\lambda_*, \delta) = u'(\lambda_*).$ ②

Step c). $\Delta_n(\lambda) \equiv \langle M \rangle_{\beta_*,\lambda}/n = U_n'(\lambda).$

$$\Delta_n^+(\lambda, \delta) \equiv (U_n(\lambda+\delta) - U_n(\lambda))/\delta, \quad \Delta_n^-(\lambda, \delta) \equiv (U_n(\lambda) - U_n(\lambda-\delta))/\delta.$$

Note $U_n''(\lambda) = -\beta\, \mathrm{Var}_{\beta_*,\lambda}(M)/c \leq 0.$ concave.

$$\Delta_n^+(\lambda, \delta) \leq \Delta_n(\lambda) \leq \Delta_n^-(\lambda, \delta).$$

Final Step: $\Delta_n^+(\lambda_*, \delta) \xrightarrow[n\to\infty]{①} \Delta^+(\lambda_*, \delta) \xrightarrow[\delta\to 0]{②} u'(\lambda_*)$

$\Delta_n^-(\lambda_*, \delta) \xrightarrow{n\to\infty} \Delta^-(\lambda_*, \delta) \xrightarrow{\delta\to 0} u'(\lambda_*)$.

$$\Rightarrow \quad \Delta_n(\lambda_*) \xrightarrow{n\to\infty} u'(\lambda_*).$$
$$= \langle M \rangle_{\beta_*, \lambda_*}$$

Remark:

If we directly apply Gaussian concentration inequality on $\langle M \rangle_{\beta_*, \lambda_*}$, we cannot get a tight concentration bound.

② Trace of resolvent of Wishart matrix. (Stieltjes transform)

Let $X \in \mathbb{R}^{n\times d}$ with $X_{ij} \sim_{i.i.d.} N(0, \frac{1}{d})$.

Denote $S(\lambda) \equiv \mathrm{tr}\left[(X^T X + \lambda I_d)^{-1}\right]/d$. $\quad \lambda > 0$.

Prop 2: For any $\delta > 0$, we have $\sqrt{L^2 c \log(2/\delta)}$ $\quad \frac{n}{d} \to r$

$$\mathbb{P}\left(|S(\lambda) - \mathbb{E}[S(\lambda)]| \leq \sqrt{\frac{c\log(2/\delta)}{\lambda^3 d^2}} \right) \geq 1 - \delta.$$

fixed λ, this is $\Theta(\frac{1}{d})$.

Proof: Want to show S is $\frac{1}{d}$-Lipschitz in Gaussian r.v.

$G = \sqrt{d}\, X \in \mathbb{R}^{n\times d} \qquad (G_{ij}) \sim_{i.i.d.} N(0,1)$.

$$\bar{S}(G) = \mathrm{tr}\left[(\lambda I_d + G^T G/d)^{-1}\right]/d$$
$$= \mathrm{tr}\left[(\lambda d \cdot I_d + G^T G)^{-1}\right].$$

$\left[\text{Goal: } |\bar{S}(G_1) - \bar{S}(G_2)| \leq \frac{c}{d} \cdot \|G_1 - G_2\|_F. \right]$ $\quad E_{ij} \in \mathbb{R}^{n\times d}$: $\begin{bmatrix} 0 & 0 \\ 0 & 1\ 0 \end{bmatrix}$ (j)

$$\partial_{G_{ij}} \bar{S}(G) = 2\,\mathrm{tr}\left((\lambda d I_d + G^T G)^{-1} G^T E_{ij} (\lambda d I_d + G^T G)^{-1} \right)$$

$\frac{d}{dt}[A(t)^{-1}] = A(t)^{-1} \frac{dA(t)}{dt} A(t)^{-1}$

$\mathrm{tr}(AB) = \mathrm{tr}(BA)$ $\qquad = 2 \cdot \mathrm{tr}\left((\lambda d I_d + G^T G)^{-2} G^T E_{ij} \right)$

$$= 2 \langle G(\lambda d I_d + G^T G)^{-2}, E_{ij} \rangle$$

$$\nabla_G \bar{S}(G) = 2\, G(\lambda d I_d + G^T G)^{-2}$$

$n > d.$

$$L^2 \equiv \sup_G \| \nabla_G \bar{S}(G) \|_F^2$$

$G = U \Sigma V^T \in \mathbb{R}^{n \times d}.$

$U \in \mathbb{R}^{n \times d}$

$\Sigma \in \mathbb{R}^{d \times d}$

$V \in \mathbb{R}^{d \times d}$

$$= \sup_G 4 \| G(\lambda d I_d + G^T G)^{-2} \|_F^2$$

$$= \sup_G 4 \| \Sigma (\lambda d I_d + \Sigma^2)^{-2} \|_F^2$$

$$U \Sigma V^T (\lambda d I_d + V \Sigma^2 V^T)^{-2}$$

$$= U \Sigma (\lambda d I_d + \Sigma^2)^{-2} V^T$$

$$= \sup_G 4 \cdot \sum_{i=1}^{n} \frac{\sigma_i(G)^2}{(\lambda d + \sigma_i(G)^2)^4}$$

$$\sigma_i(G)^2 = c \cdot \lambda d$$

$$\leq 4 \sum_{i=1}^{n} \frac{c \lambda d}{(\lambda d + c \lambda d)^4} = \frac{n}{\lambda^3 d^3}.$$

$\square$

③ Field theoretic calculation (Heuristic physics calculation)

$$Z_n = \int_\Omega \exp\{ - \beta H_\lambda(\sigma) \} d\sigma$$

$$\lim_{n \to \infty} \mathbb{E}\left[\frac{1}{n} \log Z_n\right]$$

a) Delta function $\delta(x-a)$.

i) $\mu(\{a\}) = 1$ $\qquad \mu((-\infty, +\infty) / \{a\}) = 0.$

ii) $\delta(x-a) = \lim_{\sigma \to 0} \underbrace{\varphi_\sigma(x-a)}$ ← Gausian density with mean a and variance σ^2.

$$\int f(x) \delta(x-a) dx = \lim_{\sigma \to 0} \int f(x) \varphi_\sigma(x-a) dx$$

— $f(a) = \int_{\mathbb{R}} f(x) \delta(x-a) dx.$

— $a \in \mathbb{R}$ $\delta(x-a) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ip(x-a)} dp$ (Delta identity formula).

$a \in \mathbb{R}^d$ $\delta(\vec{x} - \vec{a}) = \frac{1}{(2\pi)^d} \int_{-\infty}^{+\infty} e^{i \langle p, x-a \rangle} dp$

Intuition: $\quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ipx} \left(\int_{-\infty}^{+\infty} e^{-i\alpha x} f(\alpha) d\alpha \right) dp$

$$= \int_{-\infty}^{+\infty} \underline{\left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ip(x-\alpha)} dp \right)} f(\alpha) d\alpha$$

$$f(x) = \int_{-\infty}^{+\infty} f(\alpha) \underline{\delta(x-\alpha)} d\alpha.$$

— $1 = \int_{\mathbb{R}} \delta(x-a)\, dx$

b) Gaussian identity formula. $\mathbb{E}[e^{<p, x-a>}]$ $p \sim N(0, I_d)$

① $\det(A)^{-\frac{1}{2}} \exp\{ \|x-a\|_A^2/2 \} = \int \frac{1}{\sqrt{2\pi}^d} \exp\{ <p, x-a> - \frac{1}{2}\|p\|_{A^{-1}}^2 \} dp.$

② $\det(A)^{-\frac{1}{2}} \exp\{ -\|x-a\|_A^2/2 \} = \int \frac{1}{\sqrt{2\pi}^d} \exp\{ i<p, x-a> - \frac{1}{2}\|p\|_{A^{-1}}^2 \} dp.$

$\mathbb{E}[e^{i<p, x-a>}]$

① Gaussian moment generating function

② Gaussian characteristic formula.

$$\int \frac{1}{(2\pi)^{\frac{d}{2}} \det(A)^{\frac{1}{2}}} \exp\{ -(x-a-Ap)^T A^{-1} (x-a-Ap)/2 \} dp = 1$$

c) Laplace method.

— If $f_n(\lambda) \to f(\lambda)$ as $n \to \infty$. then

$$\int_{\mathbb{R}^k} \exp\{ n f_n(\lambda) \} d\lambda \doteq \sup_{\lambda \in \mathbb{R}^k} \exp\{ n f(\lambda) \}.$$

$a_n \doteq b_n \iff \lim_{n\to\infty} \frac{1}{n} \log a_n = \lim_{n\to\infty} \frac{1}{n} \log b_n.$

— $f_n(\lambda) = \frac{1}{n} \log \int_{\mathbb{R}^n} \exp\{ \sum_{i=1}^n h(\sigma_i; \lambda) \} \prod_{i=1}^n d\sigma_i$

— $\int_{\mathbb{R}} d\lambda \int_{\mathbb{R}^n} \exp\{ \sum_{i=1}^n h(\sigma_i, \lambda) \} \prod_{i=1}^n d\sigma_i$

$\doteq \sup_{\lambda} \int_{\mathbb{R}^n} \exp\{ \sum_{i=1}^n h(\sigma_i, \lambda) \} \prod_{i=1}^n d\sigma_i \overset{\times}{\doteq} \sup_{\sigma} \exp\{ \sum_{i=1}^n h(\sigma_i, \lambda) \}$

d). Saddle point approximation. (Method of steepest descent).

— If $f_n(i\lambda) \to f(i\lambda)$ as $n \to \infty$. then

$$\int_{\mathbb{R}} \exp\{ n f_n(i\lambda) \} d\lambda \doteq \operatorname{ext}_{\lambda \in \mathbb{C}} \exp\{ n \cdot f(\lambda) \}$$

$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}} \exp\{ n f_n(i\lambda) \} d\lambda = \operatorname{ext}_{\lambda \in \mathbb{C}} f(\lambda).$

"$\operatorname{ext}_{\lambda \in \mathbb{C}}$" $\operatorname{ext}_{\lambda \in \mathbb{C}} f(\lambda) = \{ f(\lambda_*) : f'(\lambda_*) = 0 \}.$

— $f_n(i\lambda) = \frac{1}{n} \log \int_{\mathbb{R}^n} \exp\{ \sum_{i=1}^n h(\sigma_i; i\lambda) \} d\sigma$

$$- \int_{\mathbb{R}} d\lambda \int_{\mathbb{R}^n} \exp\left\{ \sum_{i=1}^n h(\sigma_i ; i\lambda) \right\} d\sigma$$

$$\doteq \operatorname*{ext}_{\lambda \in \mathbb{R}} \int_{\mathbb{R}^n} \exp\left\{ \sum_{i=1}^n h(\sigma_i ; \lambda) \right\} d\sigma .$$

④ Large deviation of overlap matrix.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\Rightarrow \frac{1}{n} \log \mathbb{P}\left(\hat{Q}(\bar{\sigma}) \approx Q \right) \doteq \frac{1}{2} \log\det(Q) - \frac{1}{2} \log\det(I)$$

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