Mean Field Asymptotics in Statistical Learning.

Mar 29

Lecture 17. Random matrices and Stieltjes transforms

Tao. Topics in RMT
AGZ. RMT
[Anderson, Guionnet, Zeitouni]

① Motivation.

Many quantities of interest are of form

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E} f(\lambda_i(A)) = \mathbb{E}\int_{\mathbb{R}} f(\lambda)\, \hat{\mu}_A(d\lambda), \qquad \hat{\mu}_A(d\lambda) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)}.$$

where A is some random matrix.

Example: In double descent of linear model (Lecture 16).

$X_{ij} \sim N(0,1)$.
$A = X^T X/d$

Bias: $$B(\lambda) = \lambda^2 \cdot \mathbb{E}\int_{\mathbb{R}} \frac{1}{(s+\lambda)^2}\, \hat{\mu}_{X^TX/d}(ds).$$

Var: $$V(\lambda) = \sigma^2 \cdot \mathbb{E}\int_{\mathbb{R}} \frac{s}{(s+\lambda)^2}\, \hat{\mu}_{X^TX/d}(ds).$$

Hope to characterize $\hat{\mu}_{X^TX/d}(ds)$ for large n.

② Stieltjes transforms and Stieltjes functions.

[Marchenko, Pastur, 1967]

Def: The Stieltjes transform $\mathcal{S}$.

$$\mathcal{S} : \mathcal{P}(\mathbb{R}) \longrightarrow \mathcal{F}(\mathbb{C})$$
$$\mu \longmapsto m = \mathcal{S}(\mu).$$

Prob. measure $\mu \in \mathcal{P}(\mathbb{R})$.

$m : \mathbb{C} \setminus \mathrm{supp}(\mu) \to \mathbb{C}$ is the Stieljtes function associated with μ.

$$m(z) = \int \frac{1}{x - z}\, \mu(dx).$$

Rmk:

$z = -\lambda$
$A = X^T X$

$$\mu = \hat{\mu}_A = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)}.$$

$$m(z) = \int \frac{1}{x-z}\, \hat{\mu}_A(dx) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i(A) - z} = \frac{1}{n}\mathrm{tr}\left((A - zI)^{-1}\right).$$

$$\frac{1}{n}\mathrm{tr}\left((X^TX + \lambda I)^{-1}\right).$$

Properties of Stieltjes functions (fix μ, talk about properties of m).

⊗ Let $z = \lambda + i\eta$, $\lambda, \eta \in \mathbb{R}$.

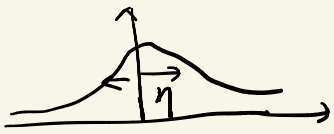
$$\mathrm{Im}\, m(z) = \int \mathrm{Im}\left(\frac{1}{x-z}\right)\mu(dx) = \int \mathrm{Im}\, \frac{(x-\bar{z})}{(x-z)(x-\bar{z})}\, \mu(dx)$$
$$= -\int \frac{\mathrm{Im}\,(x-z)}{(x-\lambda)^2 + \eta^2}\, \mu(dx) = \int \frac{\eta}{(x-\lambda)^2 + \eta^2}\, \mu(dx)$$

$$\mathrm{Re}\, m(z) = \int_{\mathbb{R}} \frac{x-\lambda}{(x-\lambda)^2 + \eta^2}\, \mu(dx).$$

⊗ When $\eta > 0$, $\mathrm{Im}\, m(z) > 0$ $\qquad \mathrm{Im}\, m(\bar{z}) = \mathrm{Im}\, \overline{m(z)}$

So m maps $\mathbb{C}_+ \to \mathbb{C}_+ = \{z \in \mathbb{C} : \mathrm{Im}\, z > 0\}$.

⊗ Interpretation of $\operatorname{Im} m(z)$.



Cauchy$(0,\eta)$ density $\quad p(x) = \frac{1}{\pi}\frac{\eta}{x^2+\eta^2}$

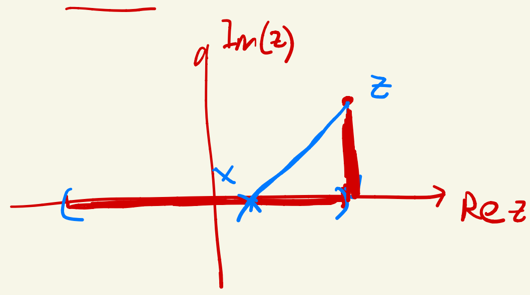
$$\frac{1}{\pi}\operatorname{Im} m(z) = \int p(x-z)\,\mu(dx). = \text{density of } \mu * \text{Cauchy}(0,\eta).$$

$X \sim \mu$, $\quad W \sim \text{Cauchy}(0,\eta)$, $\quad \frac{1}{\pi}\operatorname{Im} m(z) = \text{density of } X+W$

$\operatorname{Im} m(z)$ is a smoothified version of density of μ.

⊗ For $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$.

$$|m(z)| \le \int_{\mathbb{R}} \frac{1}{|x-z|}\,\mu(dx) \le \frac{1}{\operatorname{dist}(z, \operatorname{supp}(\mu))} \le \frac{1}{\operatorname{Im}(z)} = \frac{1}{\eta}.$$



$$m'(z) = \int_{\mathbb{R}} \frac{1}{(x-z)^2}\,\mu(dx)$$

$$|m'(z)| \le \int_{\mathbb{R}} \frac{1}{|x-z|^2}\,\mu(dx) \le \frac{1}{\eta^2}.$$

$$|m^{(k)}(z)| \le (k-1)!\,\frac{1}{\eta^k}$$

⊗ m is an analytic function on $\mathbb{C} \setminus \operatorname{supp}(\mu)$.

Mean: a) Its taylor expansion converges to itself locally.

b) If $(z_n)_{n\ge 1} \subseteq D \subseteq \mathbb{C}_+$ (D open set)

$z_n \to z_* \in D$.

$m(z_n) = \tilde{m}(z_n) \quad \Rightarrow \quad m(z) = \tilde{m}(z)$ on D.

⊗ $\operatorname{Supp}(\mu) \subseteq [-M, M]$. then $\forall z \in \mathbb{C}$ with $|z| > M$,

$$m(z) = \int_{\mathbb{R}} \frac{1}{x-z}\,\mu(dx) = -z^{-1}\int_{\mathbb{R}} \left(1 - \frac{x}{z}\right)^{-1}\mu(dx)$$

$$= -z^{-1}\int_{\mathbb{R}} \sum_{k=0}^{\infty}\left(\frac{x}{z}\right)^k \mu(dx) = -\sum_{k=0}^{\infty} z^{-(k+1)}\int x^k\,\mu(dx)$$

$$\tilde{m}(y) = m(1/y) = -\sum_{k=0}^{\infty} y^{(k+1)}\int x^k\,\mu(dx)$$

$$m(z) = -\frac{1}{z} + O\left(\frac{1}{|z|^2}\right)$$

Prop (Stieltjes inversion) For $a < b$ where the CDF of μ is continous.

$$\mu([a,b]) = \lim_{\eta\to 0}\int_a^b \frac{1}{\pi}\operatorname{Im} m(\lambda + i\eta)\,d\lambda.$$

Intuition: $\quad \frac{1}{\pi}\operatorname{Im} m(\lambda + i\eta) = \mu * \text{Cauchy}(0,\eta) \to \mu \quad$ as $\eta \to 0$.

Implication: If we know two distributions have the same Stieltjes function, then they are equal.

"same" $\longleftrightarrow$ "approximately same"

"equal" $\longleftrightarrow$ "approx. equal".

Thm: Let $(\mu_n)_{n\geq 1} \subseteq \mathcal{P}(\mathbb{R})$ with Stieltjes functions $(m_n)_{n\geq 1} \subseteq \mathcal{F}(\mathbb{C}_+)$

If there exists $m: \mathbb{C}_+ \to \mathbb{C}_+$ s.t.

- For any $z \in \mathbb{C}_+$, $m_n(z) \to m(z)$.
- m is Stieltjes function of some prob. measure μ.

then $\mu_n \to \mu$ weakly.

$\Updownarrow$

$\forall$ fixed $f \in C(\mathbb{R})$. $\int f(x)\mu_n(dx) \to \int f(x)\mu(dx)$ as $n\to\infty$

Rmk: Goal: $\lim_{n\to\infty} F_n = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n f(\lambda_i(A))$ A random matrix.

⊗ Figure out the limit

$$m(z) = \lim_{n\to\infty} \mathbb{E}[m_n(z)] = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\int \frac{1}{x-z}\hat{\mu}_A(dx)\right]$$

$$= \lim_{n\to\infty} \mathbb{E}[\operatorname{tr}((A-zI)^{-1})]/n.$$

Show $m_n(z)$ concentrate

⊗ Express F_n as a function of m_n i.e. $\mathcal{G}: \mathcal{F}(\mathbb{C}_+) \to \mathbb{R}$

$$F_n = \mathcal{G}(m_n) = \lim_{\eta\to 0} \int f(\lambda) \frac{1}{\pi} \operatorname{Im} m_n(\lambda + i\eta)\, d\lambda$$

$$B_n(\lambda) = \lambda^2 \cdot m_n'(-\lambda).$$

⊗ Show continuity of $\mathcal{G}$.

So that $\lim_{n\to\infty} F_n = \mathcal{G}(m)$.

③ Limiting Stieltjes transform for the GOE matrix.

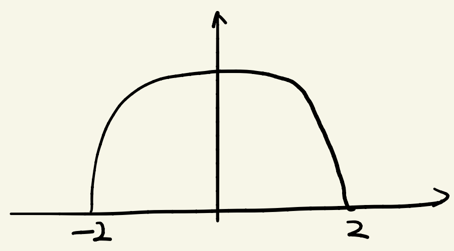
$W \sim GOE(n)$, $\quad W_{ij} \sim N(0, \frac{1}{n}) \quad 1 \le i < j \le n$

$\quad\quad\quad\quad\quad\quad W_{ii} \sim N(0, \frac{2}{n}) \quad 1 \le i \le n.$

$\quad\quad\quad\quad\quad\quad W_{ji} = W_{ij}$

Semicircle law:

$$\mu(dx) = \frac{1}{2\pi}\sqrt{4-x^2}\, 1\{x \in [-2, 2]\}\, dx$$



Thm: Let $W \sim GOE(n)$. Then almost surely as $n \to \infty$,

$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(W)}$ converges weakly to the semicircle law μ.

Rmk: Generalizable to Wigner matrix:

symmetric matrix having independent entries (lower-left triangle part) with matching first and second moment with $W \sim GOE(n)$.

(and other Higher moment conditions).

Idea: Show the convergence of $m_n(z)$ to $m(z)$ stieltjes function of Semi-circle law,

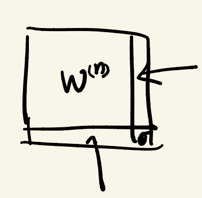
$$m(z) = \frac{\sqrt{z^2-4} - z}{2} \quad (\text{satisfy } m^2 + zm + 1 = 0).$$

Heuristic:

⊗ $m_n(z) = \frac{1}{n}\operatorname{tr}((W - zI)^{-1}).$

$W \in \mathbb{R}^{n\times n} \quad\quad W^{(n)} \in \mathbb{R}^{(n-1)\times(n-1)}.$

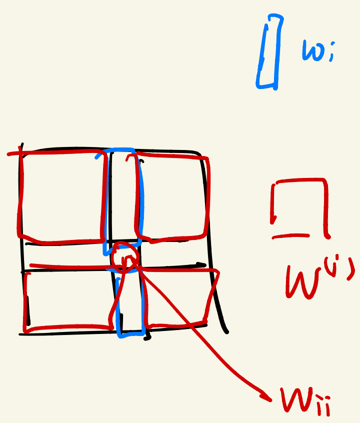
$W^{(n)}_{ij} = W_{ij}$



$W^{(n)} \sim \frac{\sqrt{n}}{\sqrt{n-1}} GOE(n-1)$

$\frac{1}{n}\operatorname{tr}((W^{(n)} - zI_{n-1})^{-1}) \approx \frac{1}{n}\operatorname{tr}((W - zI_n)^{-1}).$

⊗ $(W - zI)^{-1}.$



$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & * \\ * & * \end{bmatrix}$$

$w_i \triangleq (W_{si})_{s\ne i} \in \mathbb{R}^{n-1}, \quad W^{(i)} = (W_{st})_{s,t\ne i} \in \mathbb{R}^{(n-1)\times(n-1)}$

$A = W_{ii} - z, \quad B^T = C = w_i, \quad D = W^{(i)} - zI_{n-1}$

$$(W - zI)^{-1}_{ii} = \frac{1}{W_{ii} - z - w_i^T (W^{(i)} - zI_{n-1})^{-1} w_i}.$$

$\otimes$ $\quad m_n(z) = \frac{1}{n}\operatorname{tr}\left((W - zI)^{-1}\right) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{W_{ii} - z - w_i^\top (W^{(i)} - zI_{n-1})^{-1} w_i}$

$$\mathbb{E}\left[w_i^\top (W^{(i)} - zI_{n-1})^{-1} w_i \mid W^{(i)}\right] = \frac{1}{n}\operatorname{tr}\left((W^{(i)} - zI_{n-1})^{-1}\right) \approx m_n(z). \quad ②$$

$$w_i^\top (W^{(i)} - zI_{n-1})^{-1} w_i \quad ①$$

$z = O(1)$. $\qquad W_{ii} \sim N(0, \frac{2}{n})$. $\qquad |W_{ii}| = O(\frac{1}{\sqrt{n}})$.

$$m_n(z) \approx \frac{1}{n}\sum_{i=1}^{n} \frac{1}{-z - m_n(z)} + o_n(1).$$

$$m_n(z)\,(z + m_n(z)) + 1 \approx 0.$$

$$m_n(z)^2 + z \cdot m_n(z) + 1 \approx 0. \quad ③$$

①: Hanson–Wright inequality. $\qquad A = (W^{(i)} - zI)^{-1}$, $\quad g = \sqrt{n}\cdot w_i$.

$$\mathbb{P}_{g\sim N(0, I_n)}\left(|\langle g, Ag\rangle/n - \operatorname{tr}(A)/n| \geq \frac{t}{n}\right)$$

$$\leq 2\cdot \exp\left(-c\cdot \min\left(\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|_{op}}\right)\right).$$

$$\|(W^{(i)} - zI)^{-1}\|_{op} \leq \frac{1}{\eta} \qquad \eta = \operatorname{Im}(z).$$

$$\|(W^{(i)} - zI)\|_F \leq \frac{\sqrt{n}}{\eta}.$$

$$\sup_{i\in[n]} |\langle g^{(i)}, A^{(i)} g^{(i)}\rangle/n - \operatorname{tr}(A^{(i)})/n| \approx O\left(\sqrt{\frac{\log n}{n}}\right)$$

② Eigenvalue interpolacing.

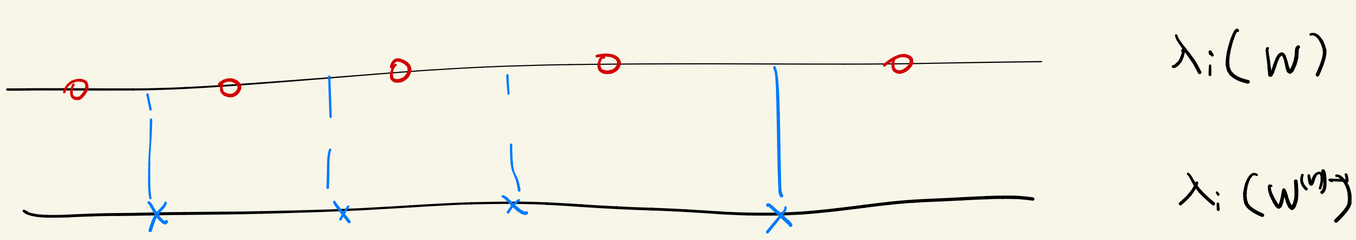
$W \in \mathbb{R}^{n\times n}$. $\qquad W^{(n)} \in \mathbb{R}^{(n-1)\times(n-1)}$ $\qquad$ symmetrix

$$W^{(n)}_{ij} = W_{ij} \qquad 1 \leq i, j \leq n-1$$

$$\lambda_1(W) \geq \lambda_2(W) \geq \cdots \geq \lambda_n(W)$$

$$\lambda_1(W^{(n)}) \geq \cdots \geq \lambda_{n-1}(W^{(n)}).$$

$$\Rightarrow \lambda_1(W) \geq \lambda_1(W^{(n)}) \geq \lambda_2(W) \geq \cdots \geq \lambda_{n-1}(W) \geq \lambda_{n-1}(W^{(n)}) \geq \lambda_n(W)$$



③ $$m_n(z) = \frac{1}{-z - m_n(z) - r_n(z)} \qquad r_n(z) \longrightarrow 0 \quad a.s.$$

$$m_n(z) = \frac{-(z + r_n(z)) + \sqrt{(z + r_n(z))^2 - 4}}{2} \in \mathbb{C}_+$$

$$\longrightarrow m(z) = \frac{\sqrt{z^2 - 4} - z}{2}$$