Mean Field Asymptotics in Statistical Learning.

Mar 29

Lecture 17. Random matrices and Stieltjes transforms

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).

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

$A = X^TX/d$

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

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

② Stieltjes transforms and Stieltjes functions.

Def: Prob. measure $\mu \in \mathcal{P}(\mathbb{R})$, the Stieltjes transform $\mathcal{S}$

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

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

singular in support

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

Rmk: ST is not transforming "z" to "$m(z)$"

ST is transforming "$\mu \in \mathcal{P}(\mathbb{R})$" to "$m \in \mathcal{F}(\mathbb{C})$".

Stieltjes trans. $\longleftrightarrow$ Fourier trans.

Stieltjes func. $\longleftrightarrow$ Characteristic functions

Will consider $\mu = \hat{\mu}_A$. $\Rightarrow$ $m(z) = \frac{1}{n}\mathrm{tr}[(A - z\mathbf{I})^{-1}]$.

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

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

$$\mathrm{Im}\, m(z) = \int \mathrm{Im}\frac{1}{x-z}\,\mu(dx) = \int_{\mathbb{R}} \frac{\eta}{(x-\lambda)^2+\eta^2}\,\mu(dx).$$

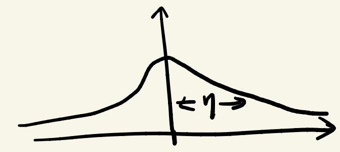
$$\mathrm{Re}\, m(z) = \int \mathrm{Re}\frac{1}{x-z}\,\mu(dx) = \int_{\mathbb{R}} \frac{x-\lambda}{(x-\lambda)^2+\eta^2}\,\mu(dx).$$

By Cauchy-Riemann, one almost determines the other.

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

So m maps $\mathbb{C}_+ \longrightarrow \mathbb{C}_+$

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



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

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

Imaginery part is the smoothified version of μ.

Less and less smooth as $\eta \to 0$

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

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

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

$$|m'(z)| \leq \int_{\mathbb{R}} \frac{1}{|x-z|^2}\mu(dx) \leq \frac{1}{\operatorname{dist}(z, \operatorname{supp}(\mu))^2} \leq \frac{1}{\operatorname{Im} z^2} = \frac{1}{\eta^2}.$$

m on $\{z \in \mathbb{C}_+ : \operatorname{Im} z \geq \eta_0)$ is $\frac{1}{\eta_0^2}$ - Lipschitz. (no matter what is μ).

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

Meaning: a) Its taylor series locally converge to itself

b) If $(z_n)_{n\geq 1} \subseteq D$ open set, $z_n \to z_0 \in D$.

$m(z_n) = \bar{m}(z_n) \implies m(z) = \bar{m}(z)$ on D.

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

$$\Rightarrow \quad m(z) = \int_{\mathbb{R}} \frac{1}{x-z}\mu(dx) = -z^{-1}\int_{\mathbb{R}} (1 - \tfrac{x}{z})^{-1}\mu(dx)$$

$$= -z^{-1}\int_{\mathbb{R}} \sum_{k=0}^{\infty} (\tfrac{x}{z})^k \mu(dx) = -\sum_{k=0}^{\infty} z^{-(k+1)} \int x^k \mu(dx).$$

In particular, $m(z) = -\frac{1}{z} + O(\frac{1}{|z|^2})$ for large $|z|$.

Prop (Stieltjes inversion): For any $a<b$ where the CDF of μ is continous, $\mu([a,b]) = \lim_{\eta \to 0} \int_a^b \frac{1}{\pi} \operatorname{Im} m(\lambda + i\eta)\, d\lambda$

If for $x \in \mathbb{R}$, $\lim_{z \in \mathbb{C}^+ \to x} \operatorname{Im} m(z)$ exists, the μ has a density at x,

which is $f(x) = \lim_{z \in \mathbb{C}^+ \to x} \frac{1}{\pi} \operatorname{Im} m(z)$.

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

Similar to characteristic function $\hat{f}(\lambda) = \int_{\mathbb{R}} f(x) e^{-i\lambda x} dx$

" Fourier transform ". $f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} f(\lambda) e^{i\lambda x} d\lambda$.

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

Actually, we can replace "same" by "approximately same".
"equal" by "approximate 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 each $z \in \mathbb{C}^+$, $m_n(z) \to m(z)$
- m is the Stieltjes transform of a prob. measure μ.

then $\mu_n \to \mu$ weakly.

Rmk: If our goal is to compute the limit

$\lim_{n\to\infty} F_n = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[f(\lambda_i(A))]$, where A is R.M.

We can proceed with the following steps.

⊗ 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 the concentration of m_n. (so that a.s convergence).

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

$$F_n = \mathcal{F}(m_n) = \underbrace{\lim_{\eta\to 0} \int f(\lambda) \frac{1}{\pi} \operatorname{Im} m_n(\lambda + i\eta)\, d\lambda}_{\text{Sometimes has a simpler form.}}$$

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

⊗ Show continuity property of $\mathcal{F}$,

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

③ Limiting Stieltjes transform for GOE matrix.

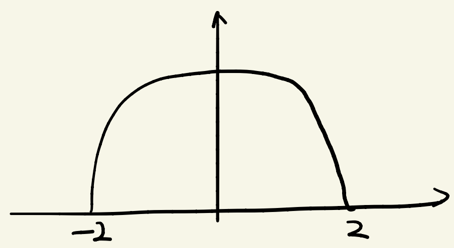
$W \sim GOE(n)$, $\quad W_{ij} \sim N(0, \frac{1}{n}) \quad 1 \le i < j \le n$

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

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

Semicircle law :

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



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.

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

Heuristics :

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

⊗ Block matrix inversion formula.

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

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

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

$$(W - zI)^{-1}_{ii} = W_{ii} - z - w_i^T (W^{(i)} - zI)^{-1} w_i$$

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

Note : W_{ii}, w_i, $W^{(i)}$ are mutually independent,

$z = O(1)$, $\quad W_{ii} = O(\frac{1}{\sqrt{n}}) = o(1)$.

$$\mathbb{E}[w_i^T (W^{(i)} - zI)^{-1} w_i \mid W^{(i)}] = \operatorname{tr}[(W^{(i)} - zI)^{-1}]/n$$

Concentration : $w_i^T (W^{(i)} - zI)^{-1} w_i / n$

$\approx \operatorname{tr}[(W^{(i)} - zI)^{-1}]/n \approx m_n(z).$

⊗ Self consistent equation

$$m_n(z) \approx \frac{1}{-z - m_n(z)} \Rightarrow m_n(z)^2 + z \cdot m_n(z) + 1 \approx 0.$$

$$m_n(z) \longrightarrow m(z), \qquad m(z)^2 + z \cdot m(z) + 1 = 0.$$

Proof of thm:

⊗ Lemma 1: We have a.s. convergence.

$$\lim_{n\to\infty} \sup_{i\in[n]} \left| \langle w_i, (W^{(i)} - zI)^{-1} w_i \rangle - \mathrm{tr}((W^{(i)} - zI)^{-1})/n \right| = 0$$

Proof:

> Lemma (Hanson - Wright inequality).
>
> Let $A \in \mathbb{C}^{n\times n}$ be deterministic,
>
> $$\Rightarrow \mathbb{P}_{g\sim N(0,I_n)}\left(|\langle g, Ag\rangle - \mathrm{tr}(A)| \geq t \right)$$
>
> $$\leq 2\exp\left[-c \cdot \min\left(\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|_{op}} \right) \right].$$

$$A = (W^{(i)} - zI)^{-1}, \qquad g = \sqrt{n}\, w_i$$

$$\|A\|_{op} \leq \max_{j\in[n-1]} \left| (\lambda_j(W^{(i)}) - z)^{-1} \right| \leq \frac{1}{\mathrm{Im}\, z} = \frac{1}{\eta}.$$

$$\|A\|_F \leq \sqrt{n}\, \|A\|_{op} \leq \frac{\sqrt{n}}{\eta}.$$

$$\Rightarrow \mathbb{P}\left(|\langle w_i, (W^{(i)} - zI)^{-1} w_i \rangle - \mathrm{tr}((W^{(i)} - zI)^{-1})/n| \geq \frac{s}{\sqrt{n}} \right)$$

$$\leq 2\exp\left\{ -C \cdot n \cdot \min(\eta^2 s^2, \eta s) \right\}.$$

$$\Rightarrow \mathbb{P}\left(\forall i \in [n],\ |\langle w_i, (W^{(i)} - zI)^{-1} w_i \rangle - \mathrm{tr}((W^{(i)} - zI)^{-1})/n| \geq \frac{s}{\sqrt{n}} \right)$$

$$\leq 2n \cdot \exp\left\{ -c \cdot n \cdot \min(\eta^2 s^2, \eta s) \right\}. \qquad \square$$

⊗ Lemma 2: For any deteministic symmetric $W \in \mathbb{R}^{n\times n}$ and minor matrix $W^{(i)} \in \mathbb{R}^{(n-1)\times(n-1)}$, we have

$$\left| \frac{1}{n}\mathrm{tr}(W - zI) - \frac{1}{n}\mathrm{tr}(W^{(i)} - zI) \right| \leq \frac{1}{n}\left(\frac{2\|W\|_{op}}{(\mathrm{Im}\, z)^2} + \frac{1}{|\mathrm{Im}\, z|} \right).$$

When $W \sim GOE(n)$, this $\to 0$ a.s.

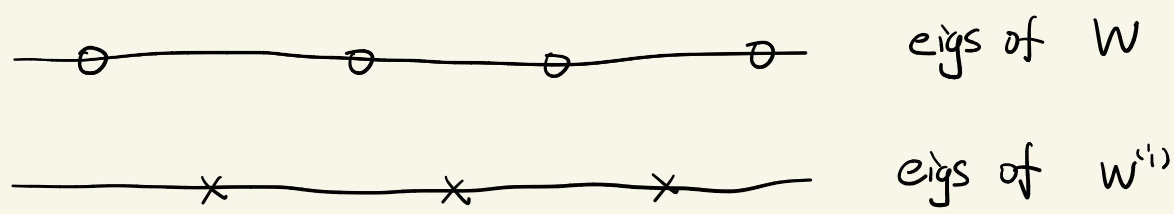
To prove this, we use eigenvalue interpolacing property.

Lemma: If $W \in \mathbb{R}^{n\times n}$ symmetric and $W^{(i)} \in \mathbb{R}^{(n-1)\times(n-1)}$ is the minor matrix

Let $\lambda_1(W) \geq \lambda_2(W) \geq \cdots \geq \lambda_n(W)$ be eigs of W

$\lambda_1(W^{(i)}) \geq \lambda_2(W^{(i)}) \geq \cdots \geq \lambda_{n-1}(W^{(i)})$ be the eigs of $W^{(i)}$

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



Proof use Courant-Fischer min-mex formula (variational formula for eigs).

(*) This gives
$$m_n(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{W_{ii} - z - \langle w_i, (W^{(i)} - zI)^{-1} w_i \rangle}$$
$$= \frac{1}{-z - m_n(z) - r_n(z)}$$
where $r_n(z) \to 0$ a.s. (for fixed $z \in \mathbb{C}^+$)

$$m_n(z) = \frac{-(z + r_n(z)) \pm \sqrt{(z+r_n)^2 - 4}}{2}.$$

Wher $r_n(z)$ small, choose "+" for $m_n(z) \in \mathbb{C}_+$.

$$\Rightarrow \lim_{n\to\infty} m_n(z) = \frac{-z + \sqrt{z^2 - 4}}{2} \quad \text{a.s.} \qquad \square$$