The Stieltjes transform

Def: For a probability measure μ on $\mathbb{R}$ (or a random variable $X \sim \mu$), its Stieltjes transform $m : \mathbb{C} \setminus \operatorname{supp}(\mu) \to \mathbb{C}$ is the function $m(z) = \mathbb{E}\left[\frac{1}{X-z}\right] = \int \frac{1}{x-z}\, d\mu(x)$.

Note: Some authors use $\mathbb{E}\left[\frac{1}{z-X}\right]$. Also sometimes called the Cauchy transform.

(Here, $\operatorname{supp}(\mu) = \{x \in \mathbb{R} : \text{Every open nbd } U \text{ of } x \text{ has } \mu(U) > 0\}$.
Easy to check $\operatorname{supp}(\mu)$ is a closed subset of $\mathbb{R}$.)

Compare: For a probability measure μ on $\mathbb{R}$, its characteristic function is the function $\varphi : \mathbb{R} \to \mathbb{C}$ given by

$$\varphi(t) = \mathbb{E}[e^{itX}] = \int e^{itx}\, d\mu(x).$$

- If two measures μ_1, μ_2 are such that $\varphi_1(t) = \varphi_2(t)$ for all $t \in \mathbb{R}$, then $\mu_1 = \mu_2$.
- If a sequence of prob. measures $\mu_1, \mu_2, \ldots$ are such that $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}$, and φ is the characteristic function of some probability measure μ on $\mathbb{R}$, then $\mu_n \to \mu$ weakly.

We'll show the same two properties for the Stieltjes transform $m(z)$.

For a random matrix with ESD μ_n,

$$\int e^{itx}\, d\mu_n(x) = \frac{1}{n}\sum_{j=1}^{n} e^{it\lambda_j(X)} \quad \text{is typically hard to study,}$$

but

$$\int \frac{1}{x-z}\, d\mu_n(x) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\lambda_j(X) - z} = \frac{1}{n}\operatorname{Tr}(X - zI)^{-1} \quad \text{is tractable.}$$

Basic properties:

- Let $z = \lambda + i\eta$, where $\lambda = \operatorname{Re} z$ and $\eta = \operatorname{Im} z$. Then

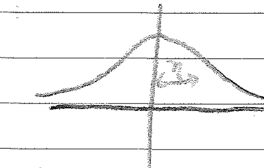
$$\operatorname{Im} m(z) = \int \operatorname{Im} \frac{1}{x - \lambda - i\eta}\, d\mu(x) = \int \operatorname{Im} \frac{x - \lambda + i\eta}{(x-\lambda)^2 + \eta^2}\, d\mu(x) = \int \frac{\eta}{(x-\lambda)^2 + \eta^2}\, d\mu(x).$$

If $z \in \mathbb{C}^+$ so $\eta > 0$, then $m(z) \in \mathbb{C}^+$.

$$\operatorname{Re} m(z) = \int \operatorname{Re} \frac{1}{x - \lambda - i\eta}\, d\mu(x) = \int \frac{x - \lambda}{(x-\lambda)^2 + \eta^2}\, d\mu(x).$$

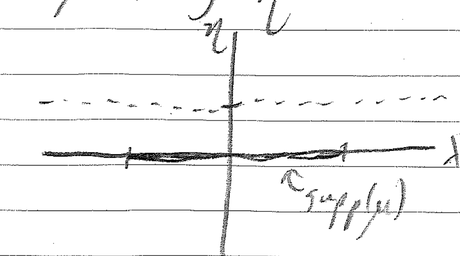
- Recall the Cauchy$(0,\eta)$ distribution has density

$$p(x) = \frac{1}{\pi} \cdot \frac{\eta}{x^2 + \eta^2}.$$



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

I.e., fixing $\eta = \operatorname{Im} z > 0$, the map $\lambda \mapsto \frac{1}{\pi}\operatorname{Im} m(\lambda + i\eta)$ is this density.



- Letting $\bar{z} = \lambda - i\eta$,

$$m(\bar{z}) = \int \frac{1}{x - \bar{z}}\, d\mu(x) = \int \overline{\left(\frac{1}{x-z}\right)}\, d\mu(x) = \overline{m(z)}.$$

Sometimes $m(z)$ is just defined as a function $m : \mathbb{C}^+ \to \mathbb{C}^+$.

- $|m(z)| \le \int \frac{1}{|x-z|} d\mu(x) \le \frac{1}{dist(z, supp(\mu))}$.

  In particular, if $z = \lambda \pm i\eta \in \mathbb{C}^{\pm}$, $|m(z)| \le \frac{1}{\eta}$.

- Can differentiate under the integral:
  $\frac{d}{dz} m(z) = \int \frac{1}{(x-z)^2} d\mu(z)$.
  Then also $\left|\frac{d}{dz} m(z)\right| \le \int \frac{1}{|x-z|^2} d\mu(z) \le \frac{1}{dist(z, supp(\mu))^2}$.

  In particular, $m(z)$ restricted to $\{z \in \mathbb{C}^+ : Im\, z \ge \eta_0\}$ is $\left(\frac{1}{\eta_0^2}\right)$-Lipschitz.

- $m(z)$ is complex-analytic on $\mathbb{C} \setminus supp(\mu)$.

(Recall: This means for any $z_0 \in \mathbb{C} \setminus supp(\mu)$, the Taylor series
$$m(z) = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{d^k}{dz^k} m(z_0)\right) \cdot (z-z_0)^k$$
is convergent in a neighborhood of z_0.
If $D \subseteq \mathbb{C}$ is a connected open set and f, g are two complex analytic functions such that $f(z_i) = g(z_i)$ for all z_i in some sequence of distinct values $z_1, z_2, z_3, \ldots$ which have a limit $z \in D$, then in fact $f \equiv g$ on all of D.)

- Suppose $supp(\mu) \subseteq [-M, M]$. Then for $z \in \mathbb{C}$ with $|z| > M$, applying $(x-z)^{-1} = -z^{-1}\left(1 - \frac{x}{z}\right)^{-1} = -z^{-1} \cdot \sum_{k \ge 0} \left(\frac{x}{z}\right)^k$ for $|x| \le M$,
$$m(z) = -\sum_{k=0}^{\infty} z^{-(k+1)} \int x^k d\mu(x)$$
as a convergent series. In particular, $m(z) = -\frac{1}{z} + O\left(\frac{1}{|z|^2}\right)$ for large $|z|$.

One may recover μ and its density from $m(z)$:

Prop (Stieltjes inversion): For any $a<b$ where the CDF of μ is continuous,
$$\mu([a,b]) = \lim_{\eta \to 0} \int_a^b \frac{1}{\pi} Im\, m(\lambda + i\eta) d\lambda$$
$$= \lim_{\eta \to 0} \int_a^b \frac{1}{2\pi i}\left(m(\lambda + i\eta) - m(\lambda - i\eta)\right) d\lambda.$$

If, at some $x \in \mathbb{R}$, $\lim_{z \in \mathbb{C}^+ \to x} Im\, m(z)$ exists, then μ has a density at x, which is $f(x) = \lim_{z \in \mathbb{C}^+ \to x} \frac{1}{\pi} Im\, m(z)$.

Proof: Let $X \sim \mu$ and $Z_\eta \sim Cauchy(0, \eta)$ be independent. Then
$\int_a^b \frac{1}{\pi} Im\, m(\lambda + i\eta) d\lambda = \mathbb{P}[X + Z_\eta \in [a,b]]$.
Since $Z_\eta \xrightarrow{P} 0$ as $\eta \to 0$, Slutsky's Lemma implies $X + Z_\eta \to X$ in distribution.
Then $\mathbb{P}[X + Z_\eta \in [a,b]] \to \mathbb{P}[X \in [a,b]] = \mu([a,b])$.

For the second claim, let $F(x) = \mu((-\infty, x])$ be the CDF. Denote $Im\, m_0(x) = \lim_{z \in \mathbb{C}^+ \to x} \frac{1}{\pi} Im\, m(z)$. It suffices to show
$$\lim_{x_i \nearrow x,\, y_i \searrow x} \frac{F(y_i) - F(x_i)}{y_i - x_i} = \frac{1}{\pi} Im\, m_0(x) \qquad (*)$$
along any sequences $x_i \nearrow x$ and $y_i \searrow x$ where x_i, y_i are continuity points of F.
For any $\varepsilon > 0$:
- Pick i large enough and $\delta > 0$ small enough s.t. for all $\eta < \delta$,
$$\left|\frac{1}{y_i - x_i} \int_{x_i}^{y_i} \frac{1}{\pi} Im\, m(\lambda + i\eta) d\lambda - \frac{1}{\pi} Im\, m_0(x)\right| < \frac{\varepsilon}{2}.$$
- Then pick $\eta < \delta$ small enough s.t.
$$\left|\int_{x_i}^{y_i} \frac{1}{\pi} Im\, m(\lambda + i\eta) d\lambda - \mu([x_i, y_i])\right| < \frac{\varepsilon}{2} \cdot (y_i - x_i),$$

Then $\left|\frac{F(y_i)-F(x_i)}{y_i-x_i} - \frac{1}{\pi}\operatorname{Im} m_0(x)\right|$

$$\leq \left|\frac{1}{y_i-x_i}\int_{x_i}^{y_i}\frac{1}{\pi}\operatorname{Im} m_0(\lambda+i\eta)\,d\lambda - \frac{1}{\pi}\operatorname{Im} m_0(x)\right|$$

$$+\left|\frac{1}{y_i-x_i}\left(\mu((x_i,y_i]) - \int_{x_i}^{y_i}\frac{1}{\pi}\operatorname{Im} m_0(\lambda+i\eta)\,d\lambda\right)\right| \leq \varepsilon.$$

As ε is arbitrary, this shows $(*)$.

Corollary: If μ_1, μ_2 have Stieltjes transforms m_1, m_2 s.t. $m_1(z)=m_2(z)$ for all $z\in\mathbb{C}^+$ (or for a sequence of $z\in\mathbb{C}^+$ having a limit point), then $\mu_1=\mu_2$.

Thm: Let $\mu_1,\mu_2,\dots$ be a sequence of probability measures on $\mathbb{R}$ with Stieltjes transforms $m_1, m_2, \dots$. If there exists $m:\mathbb{C}^+\to\mathbb{C}^+$ such that

- For each $z\in\mathbb{C}^+$, $m_n(z)\to m(z)$, and
- m is the Stieltjes transform of a probability measure μ,

then $\mu_n\to\mu$ weakly.

Example: If $\mu_n=\frac{1}{2}\delta_0+\frac{1}{2}\delta_n$ (discrete mixture), then $m_n(z)=\frac{1}{2}\left(-\frac{1}{z}\right)+\frac{1}{2}\left(\frac{1}{n-z}\right)\to-\frac{1}{2z}$ for fixed $z\in\mathbb{C}^+$. But $m(z)=-\frac{1}{2z}$ is not the Stieltjes transform of a probability measure μ on $\mathbb{R}$, and indeed μ_n has no weak limit.

To prove the theorem:

Def: A measure μ on $\mathbb{R}$ is a sub-probability measure if $\mu(\mathbb{R})\leq 1$.

Def: Sub-probability measures $\mu_1,\mu_2,\dots$ converge vaguely to a limit μ if, for any continuous $f:\mathbb{R}\to\mathbb{R}$ with $\lim_{x\to\pm\infty} f(x)=0$,

$$\int f(x)\,d\mu_n(x)\to\int f(x)\,d\mu(x).$$

Equivalently, letting $F_n(x)=\mu_n((-\infty,x])$ and $F(x)=\mu((-\infty,x])$ be the "CDFs", $\mu_n(x)\to\mu(x)$ vaguely iff $F_n(x)\to F(x)$ at each continuity point x of F. (If μ is a probability measure, this is same as $\mu_n\to\mu$ weakly.)

Example: If $\mu_n=\frac{1}{2}\delta_0+\frac{1}{2}\delta_n$, then $\mu_n\to\frac{1}{2}\delta_0$ vaguely.

Lemma: For any sequence of sub-probability measures μ_n, there exists a subsequence $\{\mu_{n_k}\}_{k=1}^{\infty}$ and a limit μ s.t. $\mu_{n_k}\to\mu$ vaguely.

Proof: Billingsley Probability and Measure, Chapter 28.

Proof of Thm: Suppose $m_n(z)\to m(z)$ for each $z\in\mathbb{C}^+$, but $\mu_n\not\to\mu$ weakly. There exists a subsequence n_k where $d(\mu_n,\mu)>\varepsilon$ along this subsequence, for some $\varepsilon>0$. Take a further subsequence n_k where $\mu_{n_k}\to\nu$ vaguely for some sub-probability limit ν. Since $x\mapsto\frac{1}{x-z}$ is continuous and satisfies $\lim_{x\to\pm\infty}\frac{1}{x-z}=0$, by definition of vague convergence

$$m_n(z)=\int\frac{1}{x-z}\,d\mu_n(x)\to\int\frac{1}{x-z}\,d\nu(x).$$

Thus $m(z)$ is the Stieltjes transform of ν. This implies $\nu=\mu$. But then $\mu_{n_k}\to\mu$ weakly, contradicting $d(\mu_{n_k},\mu)>\varepsilon$ for all k.

Prop: If μ_n are prob. measures s.t. $m_n(z)\to m(z)$, then $m(z)$ is the Stieltjes transform of a probability measure μ iff

$$\lim_{\eta\to\infty} i\eta\cdot m(i\eta)=-1.$$

Proof: Previous argument shows $m(z)$ is the Stieltjes transform of a sub-probability measure μ. For $z=i\eta$, by Dominated Convergence,

$$\lim_{\eta\to\infty} z\cdot m(z) = \lim_{\eta\to\infty}\int \frac{z}{x-z}\, d\mu(x) = \int (-1)\, d\mu(x) = -\mu(\mathbb{R}).$$

So μ is a probability measure iff this limit is -1.

Semicircle law via Stieltjes transform

Heuristic: Fix $z\in\mathbb{C}^+$. Let X be a Wigner matrix with ESD μ_n.

Let $m_n(z) = \int \frac{1}{x-z}\, d\mu_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{\lambda_i(X)-z} = \frac{1}{n}\mathrm{Tr}(X-zI)^{-1}$.

Recall the Schur complement formula for block inversion:

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

Applying with A being the (i,i) coordinate,

$$[(X-zI)^{-1}]_{ii} = \frac{1}{X_{ii} - z - x_i^T (X^{(i)}-zI)^{-1} x_i}$$

where: x_i is column i of X with entry i removed $(\in\mathbb{R}^{n-1})$.

$X^{(i)}$ is the minor of X with row/column i removed $(\in\mathbb{R}^{(n-1)\times(n-1)})$.

Note: x_i is independent of $X^{(i)}$. For any fixed A,

$$\mathbb{E}[x_i^T A x_i] = \mathbb{E}\Big[\sum_{j,k} x_{ij} x_{ik} A_{jk}\Big] = \frac{1}{n}\sum_j A_{jj} = \frac{1}{n}\mathrm{Tr}\,A.$$

We will show $x_i^T (X^{(i)}-zI)^{-1} x_i$ concentrates around $\frac{1}{n}\mathrm{Tr}(X^{(i)}-zI)^{-1}$.

Make the approximations:

- $X_{ii} \approx 0$
- $x_i^T (X^{(i)}-zI)^{-1} x_i \approx \frac{1}{n}\mathrm{Tr}(X^{(i)}-zI)^{-1}$
- $\frac{1}{n}\mathrm{Tr}(X^{(i)}-zI)^{-1} \approx \frac{1}{n}\mathrm{Tr}(X-zI)^{-1} = m_n(z)$.

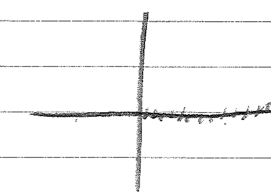
Then $m_n(z) = \frac{1}{n}\sum_{i=1}^n \frac{1}{X_{ii}-z-x_i^T(X^{(i)}-zI)^{-1}x_i} \approx \frac{1}{-z-m_n(z)}$.

We expect $m_n(z)\to m_0(z)$ where $m_0(z)$ is the root of

$$m_0(z) = \frac{1}{-z-m_0(z)}$$

in $\mathbb{C}^+$. Solving this quadratic equation,

$$m_0(z) = \frac{-z+\sqrt{z^2-4}}{2} = \frac{-\lambda - i\eta + \sqrt{\lambda^2-\eta^2-4+2i\lambda\eta}}{2} \qquad z=\lambda+i\eta$$



$$\lim_{\eta\downarrow 0} m_0(z) = \begin{cases} \frac{-\lambda+\sqrt{\lambda^2-4}}{2} & \lambda>2 \\ \frac{-\lambda-\sqrt{\lambda^2-4}}{2} & \lambda<-2 \\ \frac{-\lambda+i\sqrt{4-\lambda^2}}{2} & \lambda\in[-2,2] \end{cases}$$

So $\lim_{\eta\downarrow 0}\frac{1}{\pi}\mathrm{Im}\, m_0(z) = \frac{1}{2\pi}\sqrt{4-\lambda^2}\,\mathbb{1}\{\lambda\in[-2,2]\}$, the semicircle density.

Lemma 1: Let $A\in\mathbb{C}^{n\times n}$ be deterministic. Let $x=\frac{1}{\sqrt{n}}z\in\mathbb{R}^n$ where z_i are independent, $\mathbb{E}[z_i]=0$, $\mathrm{Var}[z_i]=1$, and z_i satisfy LSI with constant $c>0$. Then there exist constants $C, d>0$ s.t.

$$\mathbb{P}\Big[|x^TAx - \tfrac{1}{n}\mathrm{Tr}\,A| \geq \tfrac{t}{\sqrt{n}}\Big] \leq 2e^{-c't^2/\|A\|^2} + e^{-dn} \quad \text{for all } t>C.$$

Note: There is a sharper result called the Hanson-Wright inequality. If $z_1,\dots,z_n$ are sub-Gaussian

$$\mathbb{P}\Big[|x^TAx - \tfrac{1}{n}\mathrm{Tr}\,A| \geq \tfrac{t}{\sqrt{n}}\Big] \leq 2\exp\Big(-c\min\Big(\frac{t^2 n}{\|A\|_F^2}, \frac{t\sqrt{n}}{\|A\|}\Big)\Big).$$

See Rudelson, Vershynin '13 "Hanson-Wright ineq. and sub-Gaussian concentration".

Lemma 2: For any deterministic, symmetric $X\in\mathbb{R}^{n\times n}$ and minor $X'\in\mathbb{R}^{(n-1)\times(n-1)}$,

$$\Big|\frac{1}{n}\mathrm{Tr}(X'-zI)^{-1} - \frac{1}{n}\mathrm{Tr}(X-zI)^{-1}\Big| \leq \frac{1}{n}\Big(\frac{2\|X\|}{(\mathrm{Im}\,z)^2} + \frac{1}{\mathrm{Im}\,z}\Big).$$

Prop (Lipschitz extension): Let $G \subseteq \mathbb{R}^n$ and let $f: G \to \mathbb{R}$ be L-Lipschitz: for all $x, y \in G$, $|f(x)-f(y)| \le L\|x-y\|$. Then there exists an L-Lipschitz function $F: \mathbb{R}^n \to \mathbb{R}$ s.t. $F(x) = f(x)$ for all $x \in G$.

Note: If G is convex and f differentiable, suffices to check $\|\nabla f(x)\| \le L$ for all $x \in G$. In general need to check $|f(x)-f(y)| \le L\|x-y\|$ when G not convex.

Proof: Let $F(x) = \inf_{y \in G} f(y) + L\|x-y\|$. For any $x, y \in G$,

$f(x) \le f(y) + L\|x-y\|$, so $F(x) = f(x)$ when $x \in G$.

For any $x, x' \in \mathbb{R}^n$, $\varepsilon > 0$, there is $y' \in G$ where

$F(x') \ge f(y') + L\|x'-y'\| - \varepsilon$. For this y', $F(x) \le f(y') + L\|x-y'\|$.

Then $F(x) - F(x') \le L\|x-y'\| - L\|x'-y'\| + \varepsilon \le L\|x-x'\| + \varepsilon$.

Take $\varepsilon \to 0$: $F(x) - F(x') \le L\|x-x'\|$. Similarly $F(x') - F(x) \le L\|x-x'\|$.

Proof of Lemma 1: Assume WLOG $\|A\| = 1$.

① Bound $\mathbb{P}[\|x\| > 1+t]$, $t > 0$.

The map $x \mapsto \|x\|$ is 1-Lipschitz in x, so $\frac{1}{\sqrt{n}}$-Lipschitz in z.

Then $\mathbb{P}[\,|\|x\| - \mathbb{E}\|x\|| > t] \le e^{-c't^2 n}$. $\mathbb{E}[\|x\|] \le \sqrt{\mathbb{E}\|x\|^2} = 1$.

So $\mathbb{P}[\|x\| > 1+t] \le \mathbb{P}[\|x\| - \mathbb{E}\|x\| > t] \le e^{-c't^2 n}$

② Note $\|\nabla_x (x^T A x)\| = \|(A + A^T)x\| \le 2\|A\| \cdot \|x\| \le 4$ for $\|x\| \le 2$.

Thus $x \mapsto x^T A x$ is 2-Lipschitz (in x) over $B = \{x : \|x\| \le 2\}$.

Let $F(x)$ be the 2-Lipschitz extension to B^c.

Then $\mathbb{P}[|F(x) - \mathbb{E}F(x)| > t] \le 2e^{-c't^2 n}$.

So $\mathbb{P}[|x^T A x - \mathbb{E}F(x)| > t] \le 2e^{-c't^2 n} + \mathbb{P}[\|x\| > 2]$

$\le 2e^{-c't^2 n} + e^{-cn}$

We have $|\mathbb{E}F(x) - \mathbb{E}\, x^T A x| \le \mathbb{E}[\,|F(x) - x^T A x|\,]$

$= \mathbb{E}[\,|F(x) - x^T A x| \cdot \mathbb{1}\{\|x\| > 2\}]$

$\le \mathbb{E}[(L+1)\|x\|^2 \mathbb{1}\{\|x\| > 2\}]$

since $|F(x)| = |F(x) - F(0)| \le L\|x\|$, and $|x^T A x| \le \|x\|^2$.

By same argument as last lecture, $|\mathbb{E}F(x) - \mathbb{E}\, x^T A x| \le Ce^{-cn}$.

Set $t = \frac{s}{\sqrt{n}}$ where $s > C'$, C' large enough s.t. $Ce^{-cn} < \frac{C'}{2\sqrt{n}}$.

Then $\mathbb{P}[|x^T A x - \mathbb{E}\, x^T A x| > \frac{s}{\sqrt{n}}]$

$\le \mathbb{P}[|x^T A x - \mathbb{E}F(x)| > \frac{s}{2\sqrt{n}}] \le 2e^{-c's^2} + e^{-cn}$

Prop (eigenvalue interlacing): If $X \in \mathbb{R}^{n \times n}$ is symmetric, with eigenvalues $\lambda_1(X) \ge \dots \ge \lambda_n(X)$, and X' is a minor with one row and column removed, with eigenvalues $\lambda_1(X') \ge \dots \ge \lambda_{n-1}(X')$, then

$$\lambda_1(X) \ge \lambda_1(X') \ge \lambda_2(X) \ge \dots \ge \lambda_{n-1}(X) \ge \lambda_{n-1}(X') \ge \lambda_n(X).$$

o o o ... o Eigs of X

x x x ... x x Eigs of X.

Proof: Recall the Courant-Fischer min-max principle:

$$\lambda_k(X) = \max\left(\min_{v \in U: \|v\|=1} v^T X v : U \subseteq \mathbb{R}^n \text{ a subspace of dim. } k\right)$$

$$= \min\left(\max_{v \in U: \|v\|=1} v^T X v : U \subseteq \mathbb{R}^n \text{ a subspace of dim. } n-k+1\right).$$

(Proof: Let $X = \sum \lambda_i v_i v_i^T$. Take $U = \text{span}(v_1, \dots, v_k)$. Then

$$\min_{v \in U : \|v\|=1} v^T X v = v_k^T X v_k = \lambda_k. \text{ So } \lambda_k \le \max_U \min_v v^T X v.$$

On the other hand, for any U of dim. k, it has non-empty intersection w/ $\text{span}(v_k, \dots, v_n)$. Let v be a unit vector in this intersection. Then $v^T X v \le v_k^T X v_k = \lambda_k$.

So $\lambda_k \ge \min_{v \in U : \|v\|=1} v^T X v$ for any U of dim. k. Then $\lambda_k \ge \max_U \min_v v^T X v$. The second statement is similar.)

Let X' remove the last row & column of X.

Denote $\mathbb{R}^{n-1} = \{ v \in \mathbb{R}^n : v_n = 0 \}$. For each $k \in \{1, \dots, n-1\}$,

$$\lambda_k(X) = \max \left(\min_{v \in U : \|v\|=1} v^T X v : U \subseteq \mathbb{R}^n, \dim(U) = k \right)$$

$$\ge \max \left(\quad " \quad " \quad : U \subseteq \mathbb{R}^{n-1}, \dim(U) = k \right)$$

$$= \max \left(\min_{v \in U : \|v\|=1} v^T X' v : U \subseteq \mathbb{R}^{n-1}, \dim(U) = k \right) = \lambda_k(X').$$

For each $k \in \{2, \dots, n\}$:

$$\lambda_k(X) = \min \left(\max_{v \in U : \|v\|=1} v^T X v : U \subseteq \mathbb{R}^n, \dim(U) = n-k+1 \right)$$

$$\le \min \left(\quad " \quad " \quad : U \subseteq \mathbb{R}^{n-1}, \dim(U) = n-k+1 \right)$$

$$= \min \left(\max_{v \in U : \|v\|=1} v^T X' v : U \subseteq \mathbb{R}^{n-1}, \dim(U) = n-k+1 \right) = \lambda_{k-1}(X').$$

Proof of Lemma 2: Let $\lambda_1, \dots, \lambda_n$ be eigs of X, $\lambda_1', \dots, \lambda_{n-1}'$ those of X'.

$$\left| \frac{1}{n} \operatorname{Tr}(X - zI)^{-1} - \frac{1}{n} \operatorname{Tr}(X' - zI)^{-1} \right|$$

$$= \left| \frac{1}{n} \left(\frac{1}{\lambda_1 - z} - \frac{1}{\lambda_1' - z} + \frac{1}{\lambda_2 - z} - \frac{1}{\lambda_2' - z} + \dots + \frac{1}{\lambda_{n-1} - z} - \frac{1}{\lambda_{n-1}' - z} + \frac{1}{\lambda_n - z} \right) \right|$$

$$\le \frac{1}{n} \sum_{i=1}^{n-1} \left| \frac{1}{\lambda_i - z} - \frac{1}{\lambda_i' - z} \right| + \frac{1}{n} \cdot \frac{1}{|\lambda_n - z|}$$

$$= \frac{1}{n} \sum_{i=1}^{n-1} \frac{|\lambda_i' - \lambda_i|}{|\lambda_i - z| \cdot |\lambda_i' - z|} + \frac{1}{n} \cdot \frac{1}{|\lambda_n - z|}$$

Apply $|\lambda_i - z| \ge |\operatorname{Im}(\lambda_i - z)| = |\operatorname{Im} z|$, and by interlacing $\sum_{i=1}^{n-1} |\lambda_i - \lambda_i'| \le \lambda_1 - \lambda_n \le 2\|X\|$. Get LHS $\le \frac{1}{n} \left(\frac{2\|X\|}{(\operatorname{Im} z)^2} + \frac{1}{|\operatorname{Im} z|} \right)$.

Proof #2 of semicircle law (when X_{ij} satisfy LSI):

Recall $m_n(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{X_{ii} - z - x_i^T (X^{(i)} - zI)^{-1} x_i}$.

On an event of probability $\le C e^{-c\sqrt{n}}$:

- Applying Lemma 1 cond'tional on $X^{(i)}$,

$$\left| x_i^T (X^{(i)} - zI)^{-1} x_i - \frac{1}{n} \operatorname{Tr}(X^{(i)} - zI)^{-1} \right| \le \frac{\|(X^{(i)} - zI)^{-1}\|}{n^{1/4}}.$$

Note $\|(X^{(i)} - zI)^{-1}\| = \max_j \frac{1}{|\lambda_j^{(i)} - z|} \le \frac{1}{\operatorname{Im} z}$.

- Applying Lemma 2 and $\|X\| \le C$,

$$\left| \frac{1}{n} \operatorname{Tr}(X^{(i)} - zI)^{-1} - m_n(z) \right| \le \frac{C}{n} \left(\frac{1}{(\operatorname{Im} z)^2} + \frac{1}{\operatorname{Im} z} \right).$$

- $|X_{ii}| = \frac{1}{\sqrt{n}} |Z_{ii}| \le \frac{C}{n^{1/4}}$.

So on this event, $m_n(z) = \frac{1}{-z - m_n(z) + r_n}$

where $|r_n| \le C n^{-1/4}$. (C depends on $\operatorname{Im} z$.) Solving this equation,

$m_n(z) \in \frac{-(z + r_n) \pm \sqrt{(z + r_n)^2 - 4}}{2}$. Then $|m_n(z) - \tilde{m}(z)| \le C n^{-1/4}$,

for $\tilde{m}(z) \in \left\{ \frac{-z + \sqrt{z^2 - 4}}{2}, \frac{-z - \sqrt{z^2 - 4}}{2} \right\}$. Let $m(z)$ be the root w/ positive imaginary part. Since $m_n(z)$ also has positive imaginary part, for large enough n we must have

$|m_n(z)-m(z)| \le Cn^{-1/4}$, on this event.

By Borel-Cantelli, $m_n(z) \to m(z)$ a.s. This holds for each fixed $z \in \mathbb{C}^+$, so $\mu_n \to \mu$ weakly where μ has Stieltjes transform $m(z)$, i.e. μ is the semi-circle law.

<u>Problem 2.</u> For each $z \in \mathbb{C}^+$, let $m(z)$ be the Stieltjes transform of the semicircle law, i.e. the root of

$$m^2 + zm + 1 = 0 \qquad (*)$$

with positive imaginary part.

(a) Let $z = \lambda + i\eta$, where $\eta > 0$, $\lambda \in [-3,3]$. Set $\kappa = \min(|\lambda-2|, |\lambda+2|)$. Show there are universal constants $C, c > 0$ such that for any such z,

- $c < |m(z)| < C$
- If $\lambda \in [-2,2]$, then $c\sqrt{\eta+\kappa} < \operatorname{Im} m(z) < C\sqrt{\eta+\kappa}$.
- If $\lambda \in [-3,-2] \cup [2,3]$ then $\frac{c\eta}{\sqrt{\eta+\kappa}} < \operatorname{Im} m(z) < \frac{C\eta}{\sqrt{\eta+\kappa}}$.

(b) Fix any $\varepsilon \in (0, \frac{1}{4})$ and set

$$D_n = \{z = \lambda + i\eta \in \mathbb{C}^+ : \lambda \in [-3,3],\ \eta \in [n^{-1/4+\varepsilon}, 1]\}.$$

Using Lemmas 1 and 2 of today's class, show a.s. as $n \to \infty$,

$$\sup_{z \in D_n} |m_n(z) - m(z)| \to 0.$$

(Two things to be careful about:

(1) How close is $m_n(z)$ to some root of $(*)$, if $z = z_n$ is s.t. $|z_n - z| \to 0$?

(2) How close are the imaginary parts of the two roots of $(*)$ to 0?)

(c) If you restrict to the domain

$$D_n^{in} = \{z = \lambda + i\eta \in \mathbb{C}^+ : \lambda \in [-2+\varepsilon, 2-\varepsilon],\ \eta \in [n^{-\alpha+\varepsilon}, 1]\},$$

for what α can you prove a similar statement? What about the domain

$$D_n^{out} = \{z = \lambda + i\eta \in \mathbb{C}^+ : \lambda \in [-3,-2-\varepsilon] \cup [2+\varepsilon, 3],\ \eta \in [n^{-\alpha+\varepsilon}, 1]\}?$$