|WEEK 3|

Limiting spectral distributions.

$\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n$ n random samples of dimension p.

Sample covariance matrix

$$S_n = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbb{X}_i - \overline{\mathbb{X}})(\mathbb{X}_i - \overline{\mathbb{X}})^*$$

$\overline{\mathbb{X}}$ is sample mean given by $\overline{\mathbb{X}} := \frac{1}{n} \sum_{i=1}^{n} \mathbb{X}_i$

Many traditional multivariate statistics are functions of the eigenvalues (λ_i) of S_n.

In the most basic form, $T_n = \frac{1}{p} \sum_{k=1}^{p} \varphi(\lambda_k)$, $\varphi : \mathbb{C} \to \mathbb{R}$

Example: The generalised variance can be written

$$T_n = \frac{1}{p} \log |S_n| = \frac{1}{p} \sum_{k=1}^{p} \log(\lambda_k)$$

T_n is a "linear spectral statistic of the sample covariance matrix S_n with test function $\varphi(x) = \log(x)$"

First order Random matrix limits are concerned with <u>when</u> and <u>how</u> shall $T_n \to c$ as $p,n \to \infty$.

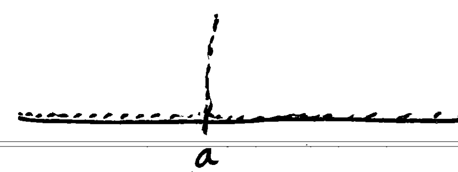
It concerns the "joint limit" of the p eigenvalues. $(\lambda_k)_{k=1}^{p}$

Empirical distributions and their limits

Let $\mathbb{M}_p(\mathbb{C})$ be $p \times p$ matrices with $\mathbb{C}$-valued entries.

and let $(\lambda_k)_{k=1}^{p}$ be the eigenvalues of $A \in \mathbb{M}_p(\mathbb{C})$.

Let $\delta_a(x) = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{otherwise} \end{cases}$



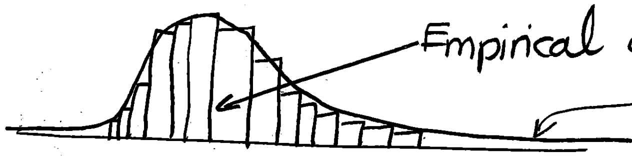
The <u>empirical spectral distribution</u> (ESD) of A is given by

$$F^A_{(x)} := \frac{1}{p} \sum_{k=1}^{p} \delta_{\lambda_k}(x)$$

Generally, F^A takes $\mathbb{C}$ values. If $A \in \mathbb{H}_p$ then $F^A_{(x)} \in \mathbb{R}$.

<u>Example:</u> $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ eigenvalues are $-i$, $+i$.

$$F^A = \frac{1}{2}(\delta_i + \delta_{-i})$$



Take a sequence of matrices $(A_n)_{n\geq 1} \in \mathbb{M}_p(\mathbb{C})$, if the sequence of corresponding ESD F^{A_n} <u>vaguely converges</u> to a (possibly defective) measure F, we call F the <u>limiting spectral distribution</u> (LSD) of $(A_n)_{n\geq 1}$.

<u>Vague convergence</u> means that for <u>any</u> continuous function that is compactly supported, called φ,

$$F^{A_n}(\varphi) \to F(\varphi) \quad \text{as } n \to \infty.$$

Here, we use the notation

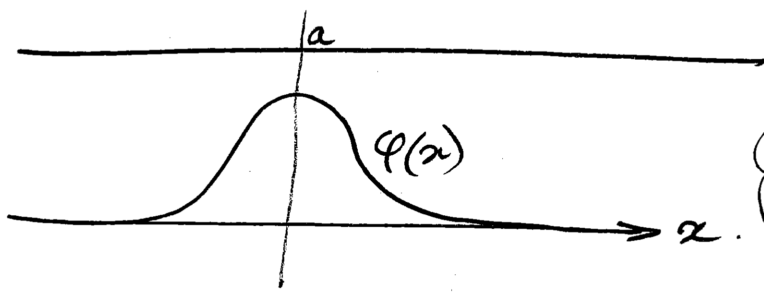
$$F(\varphi) := \int_{\mathbb{R}^p} \varphi(x) F(dx)$$

COMPACT SUPPORTED.

If the distribution F is <u>non-defective</u> (i.e. $\int F(dx) = 1$.)
then vague convergence becomes <u>weak convergence</u>,
that is,

$$F^{A_n}(\varphi) \longrightarrow F(\varphi) \qquad \text{as } n \to \infty$$

for all φ continuous and <u>bounded</u>.



φ is bounded below the value a.

In our situation, we shall be dealing with sample covariance matrices (S_n). This means that:

- support of F^{S_n} is $\mathbb{R}_+$ since S_n are Hermitian and non-negative definite.
- $F^{S_n}(x) = \frac{1}{p}\sum_{k=1}^{p} 1(\lambda_k < x)$ ESD.
- Eigenvalues are random variables and ESDs (F^{S_n}) are <u>random</u> probability distributions on $\mathbb{R}_+$.

The <u>fundamental question</u> is: Does the limit of (F^{S_n}) exist?

How can we show this?

The eigenvalues of a matrix are continuous functions of the entries of the matrix.

There is no closed-form solution for eigenvalues when dimension of a square matrix is greater than 4.

There are three main techniques used in RMT:

- Method of moments.
- Orthogonal polynomial decomposition.
- Stieltjes transform. (ST)

We shall focus on the ST approach.

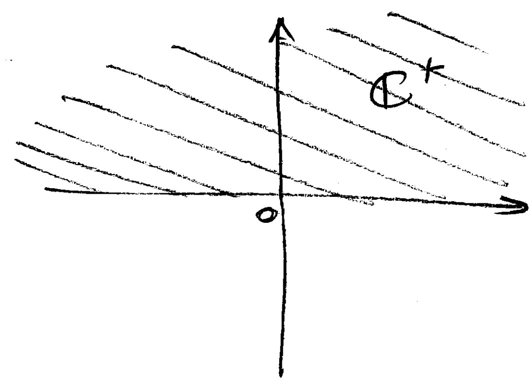
Stieltjes transform (ST)

The ST plays nearly as useful role in RMT as the Moment-generating function (MGF) or characteristic function (CF) in classic probability theory.

It is defined for a measure μ as.

$$S_\mu(z) = \int \frac{1}{x-z}\,\mu(dx), \qquad z \in \mathbb{C}^+$$

where $\mathbb{C}^+ = \{x+iy : y>0\}$.



The following lemma allows us to reconstruct the distribution function from its Stieltjes transform.

<u>Lemma</u> (Inversion): Let μ be a probability measure on $\mathbb{R}$. If $a<b$ are points of continuity of the associated dist, then $$\mu((a,b)) = \lim_{v\to 0^+} \frac{1}{\pi} \int_a^b \operatorname{Im}(S_\mu(x+iv))\,dx.$$

The following lemma gives a necessary and sufficient condition for a sequence of ST to be the ST of a probability measure.

<u>Lemma</u>(Geronimo and Hill, 2003): Suppose that (μ_n) is a sequence of probability measures on $\mathbb{R}$ with Stieltjes transforms (S_{μ_n}). If $\lim_{n\to\infty} S_{\mu_n}(z) = S_\mu(z)$ for all $z \in \mathbb{C}^+$, then there exists a probability measure μ with ST given by S_μ if and only if

$$\lim_{y\to\infty} iy S_\mu(iy) = -1.$$

In which case, $\mu_n \to \mu$ in distribution.

There are some more technical results that I will now state without proof.

First, we say that a function f is <u>holomorphic</u> if it is complex differentiable at every point of its domain, ie.

$$f'(z_0) = \lim_{z\to z_0} \frac{f(z)-f(z_0)}{z-z_0} \quad \text{exists.}$$

Holomorphic functions are very nice:

- Infinitely differentiable.
- Equals to its Taylor series.

<u>Proposition:</u> The Stieltjes transform has the following properties:

- S_μ is holomorphic on $\mathbb{C} \setminus \Gamma_\mu$ where $\Gamma_\mu := \mathrm{Supp}(\mu)$.
- $z \in \mathbb{C}^+ \iff S_\mu(z) \in \mathbb{C}^+$
- if $\Gamma_\mu \subset \mathbb{R}_+$ and $z \in \mathbb{C}^+$, then $z S_\mu(z) \in \mathbb{C}^+$
- $|S_\mu(z)| \leq \dfrac{\mu(1)}{\mathrm{dist}(z, \Gamma_\mu) \vee |\mathrm{Im}(z)|}$

<u>Proposition:</u> The mass $\mu(1)$ can be recovered through the formula

$$\mu(1) = \lim_{y \to \infty} -iy S_\mu(iy)$$

Moreover, for all continuous and compactly supported $\varphi : \mathbb{R} \to \mathbb{R}$

$$\mu(\varphi) = \int \varphi(x) \mu(dx) = \lim_{y \downarrow 0} \frac{1}{\pi} \int \varphi(x) \, \mathrm{Im}[S_\mu(x+iy)] \, dx$$

<u>Proposition:</u> Assume that the following conditions hold for a complex-valued $g(z)$:

- g is holomorphic on $\mathbb{C}^+$.
- $g(z) \in \mathbb{C}^+$ for all $z \in \mathbb{C}^+$.
- $\lim\limits_{v \to \infty} \sup |ivg(iv)| < \infty$.

The g is a ST of a bounded measure on $\mathbb{R}$.

<u>Theorem:</u> A sequence of measures (μ_n) converges vaguely to some positive measure $\mu \iff (S_{\mu_n})$ converges to S_μ on $\mathbb{C}^+$.

The idea is that we show $S_{\mu_n} \to S_\mu$ (vague conv.) and then show that μ is a probability measure by checking that $\mu(1) = 1$.

We have A positive semidefinite and symmetric. Then ESD of A is $F^A = \frac{1}{p}\sum_{j=1}^{p} \delta_{\lambda_j}$ $\quad$ A $p \times p$ matrix

$$S_A(z) = \int \frac{1}{x-z} F^A(dx)$$

$$= \frac{1}{p} \sum_{k=1}^{p} \int \frac{1}{x-z} \delta_{\lambda_k}(dx).$$

$$= \frac{1}{p} \sum_{k=1}^{p} \frac{1}{\lambda_k - z}$$

$$= \frac{1}{p} tr[(A - zI)^{-1}]$$

$$tr(A+B) = tr(A) + tr(B)$$

$$tr(A^k) = \sum_{i=1}^{p} \lambda_i^k$$

Trace of an inverse matrix: For $n \times n$ matrix Q, define Q_k to be the submatrix obtained by deleting k'th row and column.

Theorem (Bai & Silverstein, Thm A.4): If B and B_k, $k=1, \dots, n$, are nonsingular and writing $B^{-1} = [b^{kl}]$, then

$$tr(B^{-1}) = \sum_{k=1}^{n} \frac{1}{b_{kk} - \beta_k' B_k^{-1} \beta_k}.$$

b_{kk} : k'th diagonal entry of B.

β_k' : Vector obtained from k'th row of B by deleting k'th entry

β_k : ——— column ———

Applying this theorem

$$S_A(z) = \frac{1}{p}\sum_{k=1}^{p} \frac{1}{a_{kk} - z - \alpha_k^*(A_k - zI)^{-1}\alpha_k}. \qquad (*)$$

We would like to show that denominator is equal to

$$g(z, S_A(z)) + o(1)$$

Then we can solve for $S_A(z) = \dfrac{1}{g(z, S_A(z))}$.

to obtain the ST of the ESD.

Marčenko–Pastur distributions

The Marčenko–Pastur distribution F_{y,σ^2} with index y and scale parameter σ has density

$$P_{y,\sigma^2}(x) = \begin{cases} \dfrac{1}{2\pi x y \sigma^2}\sqrt{(b-x)(x-a)} & \text{if } a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

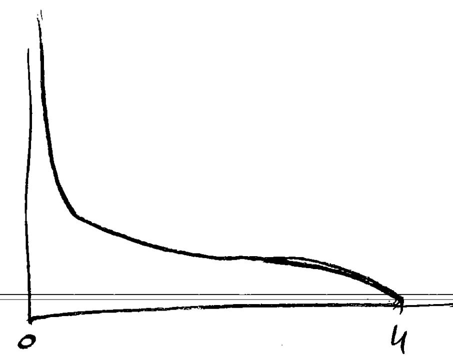
$a = \sigma^2(1-\sqrt{y})^2$
$b = \sigma^2(1+\sqrt{y})^2$

If $\sigma^2 = 1$: standard MP dist

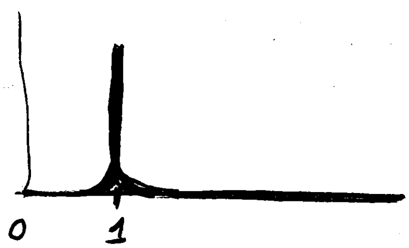
Special case $y = 1$, and $\sigma^2 = 1$.

$$P_1(x) = \begin{cases} \frac{1}{2\pi x}\sqrt{x(4-x)} & 0 < x \leq 4. \\ 0 & \text{otherwise.} \end{cases}$$

$\Rightarrow$ density is unbounded in region.



As $y \to 0$, $F_y \to \delta_1$



MP distribution for independent vectors without cross-correlation

$$S_n = \frac{1}{n-1}\sum_{i=1}^{n} x_i x_i^* - \cancel{\frac{n}{n-1}\bar{x}\bar{x}^*} \quad \text{ignore}$$

$$\simeq \frac{1}{n-1}\sum x_i x_i^*$$

We shall sometimes write n sample vectors as $p \times n$ random matrix

$$\mathbb{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_n)$$

$$\Longrightarrow \quad S_n = \frac{1}{n} \mathbb{X}\mathbb{X}^*$$

Marčenko & Pastur found the LSD of the large sample covariance matrix $\mathbf{S}_n$.

<u>Theorem</u>: (MP) Suppose that the entries $\{x_{ij}\}$ of $\mathbb{X}$ are iid complex random variables with mean zero and variance σ^2, and $p/n \to y \in (0, \infty)$. Then, almost surely, $F^{S_n} \to F_{y,\sigma^2}$ ← MP dist.

This theorem was shown in a special case in 1960s but its influence in statistics was only recognized recently.

How does the MP dist. appear in the limit?

$\sigma^2 = 1$

$$P_y(x) = \frac{1}{2\pi x y}\sqrt{(b-x)(x-a)} \qquad a \leq x \leq b.$$

$$a = (1-\sqrt{y})^2 \qquad b = (1+\sqrt{y})^2.$$

The Stieltjes transform is.

$$S(z) = \int_a^b \frac{1}{x-z} P_y(x)\, dx.$$

$$= \frac{(1-y) - z + \sqrt{(z-1-y)^2 - 4y}}{2yz}.$$

rearanging notice that $s = s(z)$ satisfies the quadratic equation

$$yzs^2 + (z+y-1)s + 1 = 0.$$

The ST of the ESD of S_n is $s_n(z) = \frac{1}{p} tr\left[(S_n - zI_p)^{-1}\right]$

If we can show $S_n(z) \to S(z)$ as $n \to \infty$ for every $z \in \mathbb{C}^+$, then $F^{S_n} \to F_y$.

By (*) on page 11,

$$S_n(z) = \frac{1}{p} \sum_{k=1}^{p} \frac{1}{\frac{1}{n}\alpha_k' \overline{\alpha_k} - z - \frac{1}{n^2}\alpha_k' \mathbb{X}_k^* \left(\frac{1}{n}\mathbb{X}_k \mathbb{X}_k^* - zI_{p-1}\right)^{-1} \mathbb{X}_k \overline{\alpha_k}}$$

$\mathbb{X}_k = \mathbb{X}$ with k'th row removed

$\alpha_k' =$ k'th row of $\mathbb{X}$, size $n \times 1$.

Assume $\mathbb{E}$["denominator terms with rows removed"] $\to$ $\mathbb{E}$["terms with rows intact"].

ie. random error caused by approx is small for large p and n.

$$\mathbb{E}\left[\frac{1}{n}\alpha_k' \overline{\alpha_k}\right] = \frac{1}{n}\sum_{j=1}^{n} |x_{kj}|^2 = 1.$$

<u>Lemma</u>: Let u be a $n\times 1$ random vector with ~~entries~~ u_i that are all independent with mean 0 and unit variance. Let Q be a (non-random) $n\times n$ complex matrix. Then

$$\mathbb{E}\left[u^* Q u\right] = \operatorname{tr} Q.$$

A, B matrices
$$(AB)_{ik} = \sum_{j=1}^{m} a_{ij} b_{jk}$$

<u>Proof</u>: As $u^* Q u = \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{u_i}\, Q_{ij}\, u_j$.

$$\mathbb{E}\left[u^* Q u\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left[Q_{ij}\,\overline{u_i}\, u_j\right].$$

$$= \sum_{i=1}^{n} Q_{ii}\,\mathbb{E}\left[\overline{u_i}\, u_i\right]$$

$$= \operatorname{tr} Q. \qquad \text{as } \mathbb{E}\left[\overline{u_i}\, u_i\right] = 1.$$

<u>Corollary</u>: $\mathbb{E}\left[u^* u\right] = n$.

Proof: Take $Q = I_n$, then $\operatorname{tr} Q = \operatorname{tr} I_n = n$.

$$\mathbb{E}\left[\frac{1}{n^2}\alpha_k' \mathbb{X}_k^*\left(\frac{1}{n}\mathbb{X}_k\mathbb{X}_k^* - zI_{p-1}\right)^{-1}\mathbb{X}_k\overline{\alpha_k}\right]$$

$$= \frac{1}{n^2}\mathbb{E}\left[\operatorname{tr}\left\{\mathbb{X}_k^*\left(\frac{1}{n}\mathbb{X}_k\mathbb{X}_k^* - zI_{p-1}\right)^{-1}\mathbb{X}_k\overline{\alpha_k}\alpha_k'\right\}\right].$$

$$= \frac{1}{n^2}\operatorname{tr}\left\{\mathbb{E}\left[\mathbb{X}_k^*\left(\frac{1}{n}\mathbb{X}_k\mathbb{X}_k^* - zI_{p-1}\right)^{-1}\mathbb{X}_k\right]\mathbb{E}\left[\overline{\alpha_k}\alpha_k'\right]\right\}.$$

$$= \frac{1}{n^2}\operatorname{tr}\left\{\mathbb{E}\left[\mathbb{X}_k^*\left(\frac{1}{n}\mathbb{X}_k\mathbb{X}_k^* - zI_{p-1}\right)^{-1}\mathbb{X}_k\right]\right\}.$$

$$= \frac{1}{n^2}\mathbb{E}\left[\operatorname{tr}\left\{\mathbb{X}_k^*\left(\frac{1}{n}\mathbb{X}_k\mathbb{X}_k^* - zI_{p-1}\right)^{-1}\mathbb{X}_k\right\}\right].$$

$$= \frac{1}{n^2}\mathbb{E}\left[\operatorname{tr}\left\{\left(\frac{1}{n}\mathbb{X}_k\mathbb{X}_k^* - zI_{p-1}\right)^{-1}\mathbb{X}_k\mathbb{X}_k^*\right\}\right].$$

We note that $\frac{1}{n}\mathbb{X}_k\mathbb{X}_k^* \approx S_n$ (only 1 vector removed).

So
$$\frac{1}{n^2}\mathbb{E}\left[\operatorname{tr}\left\{\left(\frac{1}{n}\mathbb{X}_k\mathbb{X}_k^* - zI_{p-1}\right)^{-1}\mathbb{X}_k\mathbb{X}_k^*\right\}\right].$$

$$\approx \frac{1}{n^2}\mathbb{E}\left[\operatorname{tr}\left\{\left(\frac{1}{n}\mathbb{X}\mathbb{X}^* - zI_p\right)^{-1}\mathbb{X}\mathbb{X}^*\right\}\right]$$

$$= \frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left\{I_p + z\left(\frac{1}{n}\mathbb{X}\mathbb{X}^* - zI_p\right)^{-1}\right\}\right]$$

$$= \frac{p}{n} + z\frac{p}{n}\mathbb{E}[s_n(z)].$$

so denominator is roughly

$$1 - z - \left\{ \frac{p}{n} + z \frac{p}{n} \mathbb{E}\left[S_n(z) \right] \right\}.$$

as $p \to \infty$, $n \to \infty$ and $p/n \to y > 0$.

$$\mathbb{E}[S_n(z)] \to S(z).$$

lots of hand waving here!

so denominator

$$\to 1 - z - (y + zyS(z))$$

and $$S(z) = \frac{1}{1 - z - (y + yzS(z))}.$$

This is ST of MP dist F_y!