Spectral Analysis of Large Dimensional Random Matrices

Lecture Notes for LDRM Seminar

(2018 Fall – 2019 Spring, SUSTech)

(Scribed and Edited by QIU Jiaxin & YANG Xuzhi)

Lastest Updated: April 22, 2019, 16:39

Contents

Lecture 1

Wigner Matrices and Semicircular Law

1.1 Wigner's Semicircular Law (iid Case)

1.1.1 Complex Random Variable

Definition 1.1.1. *A* **complex random variable** Z *on the probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *is a function* $Z : \Omega \to \mathbb{C}$ *such that both its part* $\Re(Z)$ *and its imaginary part* $\Im(Z)$ *are real random variables on* $(\Omega, \mathcal{F}, \mathsf{P})$.

Definition 1.1.2. *The* **expectation** *of a complex random variable is defined as*

$$\mathsf{E}Z = \mathsf{E}[\Re Z] + i\mathsf{E}[\Im Z].$$

Definition 1.1.3. *The* **variance** *of a complex random variable* Z *i defined as*

$$\mathsf{Var}Z = \mathsf{E}[|Z - \mathsf{E}Z|^2] = \mathsf{E}|Z|^2 - |\mathsf{E}Z|^2.$$

1.1.2 Empirical Spectral Distribution (ESD)

Definition 1.1.4. *Let* $\mathbf{A}$ *be a* $p \times p$ *Hermitian matrix with eigenvalues* $\lambda_j, j = 1, 2, \ldots, p$. *The* **empirical spectral distribution (ESD)** *is defined as*

$$F^{\mathbf{A}}(x) = \frac{1}{p} \sum_{i=1}^{p} I(\lambda_i \le x),$$

where I *is the indicator function.*

Let $\{\mathbf{A}_n\}$ *be a sequence of* $p_n \times p_n$ *matrices. The* **limit spectral distribution (LSD)** F *is the weak limit of* $F^{\mathbf{A}_n}$.

1.1.3 Weak Convergence

Definition 1.1.5. *A sequence of d.f.s* $\{F_n, n \ge 1\}$ *is said to* **converge weakly** *to a d.f.* F, *written as* $F_n \xrightarrow{\mathsf{w}} F$, *if* $F_n(x) \longrightarrow F(x)$ *for all* $x \in C(F)$.

1.1.4 Metrics on Cumulative Distribution Functions

Let F and G be two cumulative distribution functions.

Definition 1.1.6. *The* **Kolmodorov or supremum metric** *is*

$$\|F - G\| = \sup_x |F(x) - G(x)|.$$

Throughout this note, $\|f\| = \sup_x |f(x)|$.

Definition 1.1.7. *The* **Lévy metric** *is*

$$L(F, G) = \inf\{\varepsilon > 0 \mid F(x - \varepsilon) - \varepsilon \le G(x) \le F(x + \varepsilon) + \varepsilon, \forall x \in \mathbb{R}\}$$

▶ **Theorem 1.1.8.** $\{F_n, n \ge 1\}$ *is a sequence of d.f.s. If* $L(F_n, F) \to 0$ *as* $n \to \infty$, *then* $F_n \xrightarrow{w} F$.

Proof. $\forall x_0 \in C(F), \forall \varepsilon > 0, \exists \delta > 0$, such that $\forall x \in (x_0 - \delta, x_0 + \delta)$, we have $|F(x) - F(x_0)| < \varepsilon/2$. Since $L(F_n, F) \to 0$, then for $\varepsilon_1 := \min(\delta, \varepsilon/2)$, $\exists n_0$, if $n \ge n_0$, we have $L(F_n, F) < \varepsilon_1$, that is,

$$\inf\{a \mid F(x - a) - a \le F_n(x) \le F(x + a) + a, \forall x \in \mathbb{R}\} < \varepsilon_1.$$

So there are $a < \varepsilon_1$, such that $\forall x \in \mathbb{R}$, we have

$$F(x - a) - a \le F_n(x) \le F(x + a) + a.$$

For x_0, we have

$$\begin{aligned} F_n(x_0) &\ge F(x_0 - a) - a > F(x_0) - \frac{\varepsilon}{2} - a > F(x_0) - \varepsilon, \\ F_n(x_0) &\le F(x_0 + a) + a < F(x_0) + \frac{\varepsilon}{2} + a > F(x_0) + \varepsilon, \end{aligned}$$

which implies that $|F_n(x_0) - F(x_0)| < \varepsilon$. □

Remark 1.1.9. *In fact,* $F_n \xrightarrow{w} F$ *can also imply that* $L(F_n, F) \to 0$ *as* $n \to \infty$. *See Page 20 in* [4]. *Therefore,*

$$L(F_n, F) \to 0 \iff F_n \xrightarrow{w} F.$$

The following lemmas are two useful tools in the proof of the semicircular law.

▶ **Lemma 1.1.10.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $n \times n$ *Hermitian matrices, then*

$$L^3(F^{\mathbf{A}}, F^{\mathbf{B}}) \le \frac{1}{n}\mathrm{tr}[(\mathbf{A} - \mathbf{B})^2].$$

▶ **Lemma 1.1.11.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $n \times n$ *Hermitian matrices, then*

$$\left\|F^{\mathbf{A}} - F^{\mathbf{B}}\right\| \le \frac{1}{n}\mathrm{rank}(\mathbf{A} - \mathbf{B}).$$

1.1.5 Wigner Matrix

Definition 1.1.12. *Let $\{Z_{ij}\}_{1\le i<j}$ be a family of i.i.d., zero mean random variable on $\mathbb{C}$, independent from a family $\{Y_i\}_{i\ge 1}$ of i.i.d., zero mean random variables on $\mathbb{R}$. Consider the $n \times n$ matrix with entries*

$$X_{ij} = \overline{X}_{ji} = \begin{cases} Y_i, & \text{if } i = j, \\ Z_{ij}, & \text{if } i < j. \end{cases}$$

We call such a matrix a **Wigner matrix**.

1.1.6 Wigner's Semicircular Law

►►► **Theorem 1.1.13** (Semicircular Law). *Suppose that $\mathbf{X}_n = \{X_{ij}\}_{i,j=1}^n$ is an $n \times n$ Hermitian matrix with $X_{ij} = \overline{X}_{ji}$. If $\{X_{ii}\}$ are i.i.d., $\{X_{ij}, i \neq j\}$ are i.i.d. with variance $\sigma^2 = 1$, $\{X_{ii}\}$ and $\{X_{ij}, i \neq j\}$ are independent, then, with probability 1, the ESD of $\mathbf{W}_n = n^{-1/2}\mathbf{X}_n$ tends to the semicircular law, i.e.,*

$$F^{\mathbf{W}_n}(x) \to F(x), \quad \text{a.s.},$$

where

$$F'(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & \text{if } |x| \le 2, \\ 0, & \text{otherwise}. \end{cases}$$

1.2 Moment Convergence Theorem

Suppose $\{F_n\}$ denotes a sequence of distribution functions with finite moments of all orders. Let the k-th moment of the distribution F_n be denoted by

$$\beta_{n,k} = \beta_k(F_n) := \int x^k \, \mathrm{d}F_n(x).$$

The MCT incestigates under what conditions the convergence of moments of all fixed orders implies the weak convergence of $\{F_n\}$.

$$\boxed{\beta_{n,k} \longrightarrow \beta_k \qquad \xrightarrow{\text{what conditions}} \qquad F_n \xrightarrow{w} F \qquad n \to \infty.}$$

1.2.1 Moment Convergence Theorem

►►► **Theorem 1.2.1** (MCT). *A sequence of distribution functions $\{F_n\}$ converges weakly to a limit if the following conditions are satisfied:*

1. *Each F_n has finite moments of all orders.*
2. *For each fixed integer $k \ge 0$, $\beta_{n,k}$ converges to a finite limit β_k as $n \to \infty$.*
3. *If two right-continuous nondecreasing functions F and G have the same moment sequence $\{\beta_k\}$, then $F = G + const$.*

When we apply MCT, one needs to verify condition (3) of the theorem. The following lemmas give conditions that imply (3).

►► **Lemma 1.2.2** (M.Riesz). *Let $\{\beta_k\}$ be the sequence of moments of the distribution function F. If*

$$\liminf_{k\to\infty} \frac{1}{k}\beta_{2k}^{1/2k} < \infty,$$

then F is uniquely determined by the moment sequence $\{\beta_k, k = 0, 1, \ldots\}$.

Lemma 1.2.3 (Carleman). *Let $\{\beta_k = \beta_k(F)\}$ be the sequence of moments of the distribution function F. If the Carleman condition*

$$\sum_{k=0}^{\infty} \beta_{2k}^{-1/2k} = \infty$$

is satisfied, then F id uniquely determined by the moment sequence $\{\beta_k, k = 0, 1 \ldots\}$.

Remark 1.2.4. *Lemma 1.2.2 is a corollary of the lemma 1.2.3 due to Carleman. However, the proof of lemma 1.2.2 is much easier and it is powerful enough in spectral analysis of large dimensional random matrices.*

1.2.2 The Moment of Semicicular Law

In order to apply the moment method to prove the Theorem 1.1.13, we calculate the moment of the semicircular law and show that they satisfy the Carleman condition.

Let β_k be the k-th moment of the semicircular law. We have the following lemma.

Lemma 1.2.5. *For $k = 0, 1, 2, \ldots$, the moments of the semicircular law are given by*

$$\beta_{2k} = \frac{1}{k+1}\binom{2k}{k}, \qquad \beta_{2k+1} = 0.$$

Proof. Since the semicircular distribution is symmetric about 0, thus we have $\beta_{2k+1} = 0$. Also, we have

$$\begin{aligned}
\beta_{2k} &= \frac{1}{2\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}\,\mathrm{d}x \\
&= \frac{1}{\pi}\int_{0}^{2} x^{2k}\sqrt{4-x^2}\,\mathrm{d}x \\
&= \frac{2^{2k+1}}{\pi}\int_{0}^{1} y^{k-1/2}(1-y)^{1/2}\,\mathrm{d}y \quad (\text{by setting } x = 2\sqrt{y}) \\
&= \frac{2^{2k+1}}{\pi}\frac{\Gamma(k+1/2)\Gamma(3/2)}{\Gamma(k+2)} = \frac{1}{k+1}\binom{2k}{k}.
\end{aligned}$$

Here, we use the fact that $\Gamma(k+1/2) = \frac{(2k)!}{4^k k!}\sqrt{\pi}$. □

Moments of the semicircular distribution satisfy M.Riesz condition.

Using $\left(\frac{k}{e}\right)^k \le k! \le k^k$, we have

$$
\begin{aligned}
0 \le \frac{1}{k}\beta_{2k}^{1/2k} &= \frac{1}{k}\left[\frac{1}{k+1}\frac{(2k)!}{(k!)^2}\right]^{1/2k} \\
&\le \frac{1}{k}\left[\frac{1}{k}\frac{(2k)^{2k}}{(k/e)^{2k}}\right]^{1/2k} \\
&= \frac{2e}{k}\left(\frac{1}{k}\right)^{1/2k} \longrightarrow 0 \quad (k \to \infty) \\
\Longrightarrow \qquad & \liminf_{k\to\infty} \frac{1}{k}\beta_{2k}^{1/2k} = 0 < \infty
\end{aligned}
$$

1.3 Proof of Semicircular Law (iid Case)

Before applying MCT to the proof of the Theorem 1.1.13, we first remove the diagonal entries of $\mathbf{X}_n$, truncate the off-disgonal entries of the matrix, and renormalize them, without changing the LSD.

$$\boxed{F_n \overset{\text{a.s.}}{\to} F,\ G_n \overset{\text{a.s.}}{\to} G,\ \|F_n - G_n\| \overset{\text{a.s.}}{\to} 0 \implies F = G\ \text{a.s.}}$$

$$\boxed{F_n \overset{\text{a.s.}}{\to} F,\ G_n \overset{\text{a.s.}}{\to} G,\ L(F_n, G_n) \overset{\text{a.s.}}{\to} 0 \implies F = G\ \text{a.s.}}$$

Before we proceed, we point out two commom methods for proving almost sure convergence.

Proposition 1.3.1. *Let $\{X_n\}$ be a sequence of random variables, not necessarily independent. Then*

1. *If $\sum_{n=1}^{\infty} \mathsf{E}[|X|^s] < \infty$ for some $s > 0$, then $X_n \overset{\text{a.s.}}{\to} 0$.*
2. *If $\sum_{n=1}^{\infty} \mathsf{P}(|X_n| > \varepsilon) < \infty$ for any $\varepsilon > 0$, then $X_n \overset{\text{a.s.}}{\to} 0$.*

Step 1. Removing the Diagonal Elements

Let $\widehat{\mathbf{W}}_n$ be the matrix obtained from $\mathbf{W}_n$ by replacing the diagonal elements with zero, i.e.,

$$
(\widetilde{\mathbf{W}}_n)_{ij} = \begin{cases} (\mathbf{W})_{ij}, & i \neq j, \\ 0, & i = j. \end{cases}
$$

We shall show that the two matrices are asymptotically equivalent; i.e.

$$
F^{\widetilde{\mathbf{W}}_n} = F^{\mathbf{W}_n} \quad \text{a.s.}.
$$

Let $N_n = \#\{|x_{ii}| \ge \sqrt[4]{n}\}$. Replace the diagonal elements of $\mathbf{W}_n$ by $\frac{1}{\sqrt{n}} x_{ii} I(|x_{ii}| < \sqrt[4]{n})$, and denote the resulting matrix by $\widehat{\mathbf{W}}_n$, i.e.,

$$(\widehat{\mathbf{W}}_n)_{ij} = \begin{cases} (\mathbf{W})_{ij}, & i \neq j, \\ \frac{1}{\sqrt{n}} x_{ii} I(|x_{ii}| < \sqrt[4]{n}), & i = j. \end{cases}$$

Then, by Lemma 1.1.10, we have

$$\begin{aligned} L^3\left(F^{\widehat{\mathbf{W}}_n}, F^{\widetilde{\mathbf{W}}_n}\right) &\leq \frac{1}{n} \operatorname{tr}\left[\left(\widetilde{\mathbf{W}}_n - \widehat{\mathbf{W}}_n\right)^2\right] \\ &\leq \frac{1}{n^2} \sum_{i=1}^{n} |x_{ii}|^2 I\left(|x_{ii}| < \sqrt[4]{n}\right) \leq \frac{1}{n^2} n \cdot (\sqrt[4]{n})^2 = \frac{1}{\sqrt{n}}. \end{aligned}$$

On the other hand, by Lemma 1.1.11, we obtain

$$\left\| F^{\mathbf{W}_n} - F^{\widehat{\mathbf{W}}_n} \right\| \leq \frac{N_n}{n}.$$

Write $p_n = \mathsf{P}\left(|x_{11}| \geq \sqrt[4]{n}\right) \to 0$.

Letting $Y_i = I(|x_{ii}| \geq \sqrt[4]{n})$, then $\sum_{i=1}^n Y_i \sim \text{Binomial}(n, p_n)$. By Bernstein's inequality [1], we have, $\forall \varepsilon > 0$,

$$\begin{aligned} \mathsf{P}\left(N_n \geq \varepsilon n\right) &= \mathsf{P}\left(\sum_{i=1}^{n} \left(I\left(|x_{ii}| \geq \sqrt[4]{n}\right) - p_n\right) \geq (\varepsilon - p_n)\, n\right) \\ &\leq 2\exp\left(-\left(\varepsilon - p_n\right)^2 n^2 / 2\left[np_n(1-p_n) + (\varepsilon - p_n)\, n\right]\right) \\ &\leq 2\exp\left(-\left(\varepsilon - p_n\right)^2 n^2 / 2\left[np_n + (\varepsilon - p_n)\, n\right]\right) \\ &= 2\exp\left(-\left(\varepsilon - p_n\right)^2 n / (2\varepsilon)\right) \leq 2e^{-bn} \qquad \text{(summable)} \end{aligned}$$

for some positive constant $b > 0$, which implies that

$$\frac{N_n}{n} \to 0, \quad \text{a.s.}.$$

In the following steps, we shall assume that the diagonal elements of $\mathbf{W}_n$ are all zero.

[1] Bernstein's inequality states that if $X_1, \ldots, X_n$ are independent random variables with mean zero and uniformly bounded by c, then, $\forall \varepsilon > 0$,

$$\mathsf{P}(|S_n| \geq \varepsilon) \leq 2\exp(-\varepsilon^2 / [2(B_n^2 + c\varepsilon)]),$$

where $S_n = X_1 + \cdots + X_n$ and $B_n^2 = \mathsf{E} S_n^2$.

Step 2. Truncation

For a fixed positive constant C, truncate the variables at C and write $x_{ij(C)} = x_{ij} I(|x_{ij}| \le C)$. Denote a truncated Wigner matrix $\mathbf{W}_{n(C)}$ as following:

$$(\mathbf{W}_{n(C)})_{ij} = \begin{cases} 0, & i = j, \\ \frac{1}{\sqrt{n}} x_{ij(C)}, & i \neq j. \end{cases}$$

Lemma 1.3.2. *Suppose that the assumptions of Theorem 1.1.13 are true. Truncate the off-diagonal elements of $\mathbf{X}_n$ at C, and denote the matrix by $\mathbf{X}_{n(C)}$. Write $\mathbf{W}_{n(C)} = n^{-1/2}\mathbf{X}_{n(C)}$. Then, for any fixed contant C,*

$$\limsup_n L^3 \left(F^{\mathbf{W}_n}, F^{\mathbf{W}_{n(C)}} \right) \le \mathsf{E} \left(|x_{12}|^2 I \left(|x_{12}| > C \right) \right), \quad \text{a.s.} \tag{1.1}$$

Proof. By Lemma 1.1.10 and the law of large numbers, we have

$$\begin{aligned} L^3 \left(F^{\mathbf{W}_n}, F^{\mathbf{W}_{n(C)}} \right) &\le \frac{2}{n^2} \left(\sum_{1 \le i < j \le n} |x_{ij}|^2 I \left(|x_{ij}| > C \right) \right) \\ &\to \mathsf{E} \left(|x_{12}|^2 I \left(|x_{12}| > C \right) \right). \end{aligned}$$

□

Remark 1.3.3. *The RHS of (1.1) can be made arbitrarily small by making C large. Therefore, in the proof of Theorem 1.1.13, we can assume that the entries of $\mathbf{X}_n$ are uniformly bounded.*

Step 3. Centralization

Remove the real part of $\mathsf{E}\left(x_{ij(C)}\right)$

Applying Lemma 1.1.11, we have

$$\left\| F^{\mathbf{W}_{n(C)}} - F^{\mathbf{W}_{n(C)} - a\mathbf{1}\mathbf{1}'} \right\| \le \frac{1}{n}, \tag{1.2}$$

where $a = \frac{1}{\sqrt{n}} \Re(\mathsf{E}(x_{12(C)}))$. Furthermore, by Lemma 1.1.10, we have

$$L^3 \left(F^{\mathbf{W}_{n(C)} - \Re(\mathsf{E}(\mathbf{W}_{n(C)}))}, F^{\mathbf{W}_{n(C)} - a\mathbf{1}\mathbf{1}'} \right) \le \frac{\left| \Re \left(\mathsf{E} \left(x_{12(C)} \right) \right) \right|^2}{n} \to 0 \tag{1.3}$$

This shows that we can assume that the real parts of the mean values of the off-diagonal elements are 0.

Remove the imaginary part of $\mathsf{E}\left(x_{ij(C)}\right)$

Lemma 1.3.4. *Let $\mathbf{A}_n$ be an $n \times n$ skew-symmetric matrix whose elements above the diagonal are 1 and those below the diagonal are -1. Then, the eigenvalues of $\mathbf{A}_n$ are*

$$\lambda_k = i \cot\left(\frac{(2k-1)\pi}{2n}\right), k = 1, 2, \ldots, n.$$

Proof. Omitted. $\square$

Let $b = \Im(\mathsf{E}(x_{12(C)}))$. Then, $\Im\left(\mathsf{E}\left(\mathbf{W}_{n(C)}\right)\right) = \frac{1}{\sqrt{n}} b\mathbf{A}_n$. By Lemma 1.3.4, the eigenvalues of the matrix $i\Im\left(\mathsf{E}\left(\mathbf{W}_{n(C)}\right)\right) = ib\mathbf{A}_n/\sqrt{n}$ is

$$\frac{ib\lambda_k}{\sqrt{n}} = -\frac{b}{\sqrt{n}} \cot\left(\frac{(2k-1)\pi}{2n}\right), \qquad k = 1, 2, \ldots, n.$$

If the spectral decomposition of $\mathbf{A}_n$ is $\mathbf{U}_n\mathbf{D}_n\mathbf{D}_n^{\mathsf{H}}$, then we write

$$i\Im\left(\mathsf{E}\left(\mathbf{W}_{n(C)}\right)\right) = \mathbf{B}_1 + \mathbf{B}_2,$$

where

$$\mathbf{B}_j = -\frac{1}{\sqrt{n}} b\mathbf{U}_n\mathbf{D}_{nj}\mathbf{U}_n^{\mathsf{H}}, \qquad j = 1, 2,$$

where $\mathbf{U}_n$ is a unitary matrix, $\mathbf{D}_n = \mathrm{diag}[\lambda_1, \ldots, \lambda_n]$, and

$$\mathbf{D}_{n1} = \mathbf{D}_n - \mathbf{D}_{n2} = \mathrm{diag}\left[0, \cdots, 0, \lambda_{[n^{3/4}]}, \lambda_{[n^{3/4}]+1}, \cdots, \lambda_{n-[n^{3/4}]}, 0, \cdots, 0\right].$$

For any $n \times n$ Hermitian matrix $\mathbf{C}$, by Lemma 1.1.10, we have

$$\begin{aligned} L^3\left(F^{\mathbf{C}}, F^{\mathbf{C}-\mathbf{B}_1}\right) &\le \frac{b^2}{n^2} \sum_{n^{3/4} \le k \le n - n^{3/4}} \cot^2(\pi(2k-1)/2n) \\ &< \frac{2}{n \sin^2\left(n^{-1/4}\pi\right)} \to 0 \end{aligned} \qquad ?$$

and, by Lemma 1.1.11,

$$\left\| F^{\mathbf{C}} - F^{\mathbf{C}-\mathbf{B}_2} \right\| \le \frac{2n^{3/4}}{n} \to 0. \tag{1.4}$$

Summing up equation (1.2)-(1.4), we established the following centralization lemma.

Lemma 1.3.5. *Under the conditions assumed in Lemma 1.3.2, we have*

$$L\left(F^{\mathbf{W}_{n(C)}}, F^{\mathbf{W}_{n(C)} - E\left(\mathbf{W}_{n(C)}\right)}\right) = o(1).$$

Step 4. Rescaling

Write $\sigma^2(C) = \mathrm{Var}(x_{12}(C))$, and define

$$\widetilde{\mathbf{W}}_n = \sigma^{-1}(C)\left(\mathbf{W}_{n(C)} - \mathsf{E}\left(\mathbf{W}_{n(C)}\right)\right),$$

note that the off-diagonal entries of $\sqrt{n}\widetilde{\mathbf{W}}_n$ are

$$\tilde{x}_{ij} = \sigma^{-1}(C)\left(x_{ij(C)} - \mathsf{E}(x_{ij(C)})\right).$$

Applying Lemma 1.1.10, we have

$$\begin{aligned} L^3\left(F^{\widetilde{\mathbf{W}}}, F^{\mathbf{W}_{n(C)} - \mathsf{E}(\mathbf{W}_{n(C)})}\right) &\le \frac{1}{n}\mathrm{tr}\left\{\left[\widetilde{\mathbf{W}} - \mathbf{W}_{n(C)} + \mathsf{E}(\mathbf{W}_{n(C)})\right]^2\right\} \\ &\le \frac{2(\sigma(C)-1)^2}{n^2\sigma^2(C)} \sum_{1\le i<j\le n} \left|x_{kj(C)} - \mathsf{E}\left(x_{kj(C)}\right)\right|^2 \\ &\to (\sigma(C)-1)^2 \quad \text{a.s.} \end{aligned}$$

Note that $(\sigma(C)-1)^2$ can be made arbitrarily small if C is large.

To prove the semicircular law, we may assume that

1. The entries of $\mathbf{X}_n$ are bounded by C.
2. $x_{ii} = 0$.
3. $\mathsf{E}(x_{ij}) = 0,\ \mathrm{Var}(x_{ij}) = 1,\ i \ne j$.

Some Lemmas in Combinatorics

Lemma 1.3.6. *Each isomorphic class contains $n(n-1)\cdots(n-t+1)$ $\Gamma(k,t)$ graphs.*

Tree is a connected graph without cycles. A single edge is a edge not coincident with any other edges.

Three Categories of canonical $\Gamma(k,t)$-graphs

1. $\Gamma_1(k)$:
   - each edge is coincident with exactlly one other edge of opposite direction.
   - the graph of noncoincident edges forms a tree
2. $\Gamma_2(k,t)$: at least one single edge
3. $\Gamma_3(k,t)$: all other canoncial $\Gamma(k,t)$-graphs. If we classify the k edges into coincidence classes, then there two kinds of $\Gamma_3(k,t)$-graphs:
   - every coincident class with at least 3 edges.

- a cycle of noncoincident edges.

Lemma 1.3.7. *In a* $\Gamma_3(k,t)$*-graph,* $t \le (k+1)/2$.

Lemma 1.3.8. *The number of* $\Gamma_1(2m)$*-graphs is* $\frac{1}{m+1}\binom{2m}{m}$.

Step 5. Proof of the Semicircular Law

For simplicity, we will use $\mathbf{W}_n$ and x_{ij} to denote the Winger matrix and basic variables after truncation, centralization, and rescaling.

The k-th moment of the ESD of $\mathbf{W}_n$:

$$\begin{aligned}\beta_k(\mathbf{W}_n) = \beta_k\left(F^{\mathbf{W}_n}\right) &= \int x^k\, \mathrm{d}F^{\mathbf{W}_n}(x)\\ &= \lim_{m\to\infty}\sum_{i=1}^{m-1}\lambda_i^k\left(F^{\mathbf{W}_n}(x_{j+1}) - F^{\mathbf{W}_n}(x_j)\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\lambda_i^k = \frac{1}{n}\mathrm{tr}(\mathbf{W}_n^k) = \frac{1}{n^{1+k/2}}\mathrm{tr}(\mathbf{X}_n^k)\\ &= \frac{1}{n^{1+k/2}}\sum_{\mathbf{i}} X(\mathbf{i}),\end{aligned}$$

where $\lambda_i's$ are the eigenvalues of the matrix $\mathbf{W}_n$, $X(\mathbf{i}) = x_{i_1i_2}x_{i_2i_3}\cdots x_{i_ki_1}$, $\mathbf{i} = (i_1,\dots,i_k)$, and the summation $\sum_{\mathbf{i}}$ runs over all possibilities that $\mathbf{i} \in \{1,\dots,n\}^k$.

Remark 1.3.9.

$$\begin{aligned}(\mathbf{X}_n^2)_{i_1i_1} &= \sum_{i_2=1}^{n} x_{i_1i_2}x_{i_2i_1}\\ (\mathbf{X}_n^3)_{i_1i_1} &= \sum_{i_3=1}^{n}(\mathbf{X}_n^2)_{i_1i_3}\cdot x_{i_3i_1} = \sum_{i_2=1}^{n}\sum_{i_3=1}^{n} x_{i_1i_2}x_{i_2i_3}x_{i_3i_1}\\ &\vdots\\ (\mathbf{X}_n^k)_{i_1i_1} &= \sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n} x_{i_1i_2}x_{i_2i_3}\cdots x_{i_ki_1}\end{aligned}$$

$$\Rightarrow \mathrm{tr}(\mathbf{X}_n^k) = \sum_{i_1=1}^{n}(\mathbf{X}_n^k)_{i_1i_1} = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n} x_{i_1i_2}x_{i_2i_3}\cdots x_{i_ki_1} = \sum_{\mathbf{i}} X(\mathbf{i})$$

By applying the moment convergence theorem, we complete the proof of the semicircular law for the iid case by showing the following:

(1) $\mathrm{E}[\beta_k(\mathbf{W}_n)] \to \beta_k$ as $n \to \infty$.

(2) For each fixed k, $\sum_n \mathrm{Var}[\beta_k(\mathbf{W}_n)] < \infty$.

The Proof of (1):

We have

$$\mathsf{E}[\beta_k(\mathbf{W}_n)] = \frac{1}{n^{1+k/2}} \sum_{\mathbf{i}} \mathsf{E}(X(\mathbf{i})).$$

For each vector $\mathbf{i}$, construct a graph $G(\mathbf{i})$. To specify the graph, we rewrite $X(\mathbf{i}) = X(G(\mathbf{i}))$. The summation is taken over all sequences $\mathbf{i} = (i_1, i_2, \ldots, i_k) \in \{1, 2 \ldots, n\}^k$.

Note that isomorphic graphs corresponds to equal terms in $\sum_{\mathbf{i}} \mathsf{E}(X(\mathbf{i}))$. Thus, we first group the terms according to isomorphic classes and then split $\mathsf{E}[\beta_k(\mathbf{W}_n)]$ into three sums according to categories. Then

$$\mathsf{E}[\beta_k(\mathbf{W}_n)] = S_1 + S_2 + S_3,$$

where

$$S_j = \frac{1}{n^{1+k/2}} \sum_{\Gamma(k,t) \in C_j} \sum_{G(\mathbf{i}) \in \Gamma(k,t)} \mathsf{E}[X(G(\mathbf{i}))], \quad j = 1, 2, 3.$$

$\sum_{\Gamma(k,t) \in C_j}$: sum over all canonical $\Gamma(k,t)$ graphs in category j.

$\sum_{G(\mathbf{i}) \in \Gamma(k,t)}$: sum over all isomorphic graphs for a given canonical graph.

By the definition of the categories and by the assumptions on the entries of the random matrices, i.e. $\mathsf{E}(X_{ij}) = 0$, we have

$$S_2 = 0.$$

Since the random variables are bounded by C, the number of isomorphic graphs is less than n^t by Lemma 1.3.6, and $t \le (k+1)/2$ by Lemma 1.3.7, we conclude that

$$S_3 \le n^{-1-k/2} O(n^t) = o(1).$$

If $k = 2m - 1$, then $S_1 = 0$ since there are no terms in S_1. We consider the case where $k = 2m$. Since each edge coincides wioth edge of opposote direction, each term in S_1 is $(\mathsf{E}|x_{12}|^2)^m = 1$. So, by Lemma 1.3.8,

$$\begin{aligned} S_1 =& n^{-1-m} \sum_{\Gamma(2m,t) \in C_1} n(n-1)\cdots(n-m) \\ =& \beta_{2m} \left(1 - \frac{1}{m}\right) \cdots \left(1 - \frac{m}{n}\right) \to \beta_{2m}. \end{aligned}$$

Assertion (1) is then proved.

The proof of (2):

We have

$$\begin{aligned}\mathrm{Var}\left(\beta_k\left(\mathbf{W}_n\right)\right) &= \mathsf{E}\left[\left|\beta_k\left(\mathbf{W}_n\right)\right|^2\right] - \left|\mathsf{E}\left[\beta_k\left(\mathbf{W}_n\right)\right]\right|^2 \\ &= \frac{1}{n^{2+k}}\sum_{\mathbf{i},\mathbf{j}}\{\mathsf{E}[X(\mathbf{i})X(\mathbf{j})] - \mathsf{E}[X(\mathbf{i})]\mathsf{E}[X(\mathbf{j})]\},\end{aligned} \tag{1.5}$$

where $\mathbf{i} = (i_1, \ldots, ß_k), \mathbf{j} = (j_1, \ldots, j_k)$, and $\sum$ is taken over all possibilities for $\mathbf{i}, \mathbf{j} \in \{1, 2, \ldots, n\}^k$.

Using $\mathbf{i}$ and $\mathbf{j}$, we can construct two graphs $G(\mathbf{i})$ and $G(\mathbf{j})$, as in the proof of (1). There are two cases that some terms in (1.5) are zero:

- No coincident edges between $G(\mathbf{i})$ and $G(\mathbf{j})$

$$\Longrightarrow \quad X(\mathbf{i}) \perp X(\mathbf{j}).$$

- $G = G(\mathbf{i}) \cup G(\mathbf{j})$ has a single edge

$$\Longrightarrow \quad \mathsf{E}[X(\mathbf{i})\mathsf{E}(\mathbf{j})] = \mathsf{E}[X(\mathbf{i})]\mathsf{E}[X(\mathbf{j})] = 0.$$

Now, let us consider the nonzero terms in (1.5).

- G conatins no single edges and the graph of noncoincident edges has a cycle. Then the noncoincident vertices of G are not more than k.
- G contains no single edges and the graph of noncoincident edges has no cycles. Then there is at least one edge with coincidence multiplicity greater than or equal to 4, thus the number of noncoincident vertices is not larger than k.

Also, each term is not larger than $2C^{2k}n^{-2-k}$. Consequently, we can conclude that

$$\mathrm{Var}\left(\beta_k\left(\mathbf{W}_n\right)\right) \leq K_k C^{2k} n^{-2},$$

where K_k is a constant taht depends on k only. This completes the proof of assertion (2).

The proof of Theorem 1.1.13 is then complete.

1.4 Generalizations to the Non-iid Case

Theorem 1.4.1. *Suppose that $\mathbf{W}_n = \frac{1}{\sqrt{n}}\mathbf{X}_n$ is a Winger matrix and the extries above or on the disgonal of $\mathbf{X}_n$ are independent but may be dependent on n and may not necessarily be identically distributed. Assume that all the entries of $\mathbf{X}_n$ are of mean 0 and variance 1 and satisfy the condition that, for any constant $\eta > 0$,*

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{ij}\mathsf{E}\left|x_{ij}^{(n)}\right|^2 I\left(\left|x_{ij}^{(n)}\right| \geq \eta\sqrt{n}\right) = 0. \tag{1.6}$$

Then, the ESD of $\mathbf{W}_n$ converges to the semicircular law almost surely.

Again, we need to truncate, remove diagonal entries, and renormalize before we use the MCT. Because the entries are not iid, we cannot truncate the entries at constant position. Instead, we shall truncate them at $\eta_n \sqrt{n}$ for some sequence $\eta_n \downarrow 0$.

Step 1. Truncation

We use the rank inequality (Lemma 1.1.11) to truncate the variables.

Note that condition (1.6) is equivalent to: for any $\eta > 0$,

$$\lim_{n\to\infty} \frac{1}{\eta^2 n^2} \sum_{ij} \mathsf{E} \left| x_{ij}^{(n)} \right|^2 I\left(\left| x_{ij}^{(n)} \right| \geq \eta\sqrt{n} \right) = 0. \tag{1.7}$$

Thus, one can select a sequence $\eta_n \downarrow 0$ such that (1.7) remain true when η is replace by η_n.

Define

$$\mathbf{W}_{n(\eta_n\sqrt{n})} = \frac{1}{\sqrt{n}} \left(x_{ij}^{(n)} I\left(\left| x_{ij}^{(n)} \right| \leq \eta_n \sqrt{n} \right) \right)$$

Using rank inequality, we obtain

$$\begin{aligned} \left\| F^{\mathbf{W}_n} - F^{\mathbf{W}_{n(\eta_n\sqrt{n})}} \right\| &\leq \frac{1}{n} \mathsf{rank} \left(\mathbf{W}_n - \mathbf{W}_{n(\eta_n\sqrt{n})} \right) \\ &\leq \frac{2}{n} \sum_{1\leq i\leq j\leq n} I\left(\left| x_{ij}^{(n)} \right| \geq \eta_n\sqrt{n} \right). \end{aligned} \tag{1.8}$$

By condition (1.7), we have

$$\begin{aligned} \mathsf{E}\left(\frac{1}{n} \sum_{1\leq i\leq j\leq n} I\left(\left| x_{ij}^{(n)} \right| \geq \eta_n\sqrt{n} \right) \right) &= \frac{1}{n} \sum_{1\leq i\leq j\leq n} \mathsf{E}\, I\left(\left| x_{ij}^{(n)} \right| \geq \eta_n\sqrt{n} \right) \\ &\leq \frac{1}{n} \sum_{ij} \mathsf{E}\, I\left(\left| x_{ij}^{(n)} \right| \geq \eta_n\sqrt{n} \right) \\ &\leq \frac{1}{n} \sum_{ij} \mathsf{E} \frac{|x_{ij}^{(n)}|^2}{\eta_n^2 n} I\left(\left| x_{ij}^{(n)} \right| \geq \eta_n\sqrt{n} \right) \\ &= \frac{1}{\eta_n^2 n^2} \sum_{ij} \mathsf{E} |x_{ij}^{(n)}|^2 I\left(\left| x_{ij}^{(n)} \right| \geq \eta_n\sqrt{n} \right) \\ &= o(1), \end{aligned}$$

and

$$\begin{aligned}
\mathsf{Var}\left(\frac{1}{n}\sum_{1\le i\le j\le n} I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right)\right) &= \frac{1}{n^2}\sum_{1\le i\le j\le n} \mathsf{Var}\left[\, I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right)\right] \\
&\le \frac{1}{n^2}\sum_{1\le i\le j\le n} \mathsf{E}\left[\, I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right)\right]^2 \\
&\le \frac{1}{n^2}\sum_{ij} \mathsf{E}\frac{|x_{ij}^{(n)}|^2}{\eta_n^2 n} I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right) \\
&\le \frac{1}{\eta_n^2 n^3}\sum_{jk} \mathsf{E}\left|x_{ij}^{(n)}\right|^2 I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right) \\
&= o(1/n).
\end{aligned}$$

Then, applying Bernstein's inequality, for all small $\varepsilon > 0$ and large n, we have

$$\mathsf{P}\left(\frac{1}{n}\sum_{1\le i\le j\le n} I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right) \ge \varepsilon\right) \le 2e^{-\varepsilon n}, \tag{1.9}$$

which is summable. Thus, by (1.8) and (1.9), to prove $F^{\mathbf{W}_n}$ converges to the semicircular law a.s., it suffices to show that $F^{\mathbf{W}_{n(\eta_n\sqrt{n})}}$ converges to the semicircular law a.s..

<u>My result:</u> Write $p_{ij}^{(n)} = \mathsf{P}(|x_{ij}^{(n)}| \ge \eta_n\sqrt{n})$,

$$S_n = \frac{1}{n}\sum_{1\le i\le j\le n}\left[I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right) - p_{ij}^{(n)}\right],$$

then

$$\mathsf{E}(S_n) = 0, \quad B_n^2 = \mathsf{E}S_n^2 = \mathsf{Var}(S_n) = o(1/n).$$

By Bernstein's inequality,

$$\begin{aligned}
&\mathsf{P}\left(\frac{1}{n}\sum_{1\le i\le j\le n} I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right) \ge \varepsilon\right)\\
&= \mathsf{P}\left(|S_n| \ge \varepsilon - \frac{1}{n}\sum_{1\le i\le j\le n} p_{ij}^n\right)\\
&\le 2\exp\left\{\frac{-(\varepsilon - \frac{1}{n}\sum_{1\le i\le j\le n} p_{ij}^n)^2}{2(B_n^2 + \varepsilon - \frac{1}{n}\sum_{1\le i\le j\le n} p_{ij}^n)}\right\}\\
&\le 2\exp\left\{\frac{-(\varepsilon - \frac{n+1}{2})^2}{2(1+\varepsilon)}\right\}\\
&= 2\exp\left\{-\frac{(n+1-2\varepsilon)^2}{8(1+\varepsilon)}\right\}.
\end{aligned}$$

Step 2. Removing diagonal elements

Let $\widehat{\mathbf{W}}_n$ be the matrix $\mathbf{W}_{n(\eta_n\sqrt{n})}$ with diagonal elements replaced by 0. Then,

$$L^3\left(F^{\mathbf{W}_{n(\eta_n\sqrt{n})}}, F^{\widehat{\mathbf{W}}_n}\right) \le \frac{1}{n^2}\sum_{k=1}^n \left|x_{kk}^{(n)}\right|^2 I\left(\left|x_{kk}^{(n)}\right| \le \eta_n\sqrt{n}\right) \le \eta_n^2 \to 0.$$

Step 3. Centralization

$$\begin{aligned}
&L^3\left(F^{\widehat{\mathbf{W}}_n}, F^{\widehat{\mathbf{W}}_n - \mathsf{E}\widehat{\mathbf{W}}_n}\right)\\
&\le \frac{1}{n^2}\sum_{ij}\left|\mathsf{E}\left(x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right)\right)\right|^2\\
&\le \frac{1}{n^2}\sum_{ij}\left|\mathsf{E}\left(x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right)\right)\right|^2\\
&\le \frac{1}{n^3\eta_n^2}\sum_{ij}\mathsf{E}\left|x_{jk}^{(n)}\right|^2 I\left(\left|x_{jk}^{(n)}\right| \ge \eta_n\sqrt{n}\right) \to 0.
\end{aligned}$$

Step 4. Rescaling

Write $\widetilde{\mathbf{W}}_n = \frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n$, where

$$\widetilde{\mathbf{X}}_n = \left(\frac{x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right) - \mathsf{E}\left(x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right)\right)}{\sigma_{ij}}(1-\delta_{ij})\right),$$

$$\sigma_{ij}^2 = \mathsf{E}\left|x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right) - \mathsf{E}\left(x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right)\right)\right|^2$$

and δ_{ij} is Kronecker's delta.[2]

Note that

$$\begin{aligned}
\sigma_{ij}^2 &= \mathsf{E}\left|x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right) - \mathsf{E}\left(x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right)\right)\right|^2 \\
&= \mathsf{Var}\left[x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right) - \mathsf{E}\left(x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right)\right)\right] \\
&= \mathsf{Var}\left[x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right)\right] \\
&\le \mathsf{Var}(x_{ij}^{n}) = 1. \qquad \text{(by the assumption of Theorem 1.4.1)}
\end{aligned}$$

By Lemma 1.1.10, it follows that

$$\begin{aligned}
& L^3\left(F^{\widetilde{\mathbf{W}}_n}, F^{\widehat{\mathbf{W}}_n - \mathsf{E}\widehat{\mathbf{W}}_n}\right) \\
\le & \frac{1}{n^2}\sum_{i\ne j}\left(1-\sigma_{ij}^{-1}\right)^2 \left|x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right) - \mathsf{E}\left(x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right)\right)\right|^2,
\end{aligned}$$

Note that

$$\begin{aligned}
& \mathsf{E}\left(\frac{1}{n^2}\sum_{i\ne j}\left(1-\sigma_{ij}^{-1}\right)^2 \left|x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right) - \mathsf{E}\left(x_{ij}^{(n)} I\left(\left|x_{ij}^{(n)}\right| \le \eta_n\sqrt{n}\right)\right)\right|^2\right) \\
= & \frac{1}{n^2}\sum_{ij}(1-\sigma_{ij})^2 \le \frac{1}{n^2\eta_n^2}\sum_{ij}(1-\sigma_{ij})^2 \qquad (\because \eta_n \downarrow 0) \\
\le & \frac{1}{n^2\eta_n^2}\sum_{ij}\left(1-\sigma_{ij}^2\right) \\
\le & \frac{1}{n^2\eta_n^2}\sum_{ij}\left[\mathsf{E}\left|x_{ij}^{(n)}\right|^2 I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right) + \mathsf{E}^2\left|x_{ij}^{(n)}\right| I\left(\left|x_{ij}^{(n)}\right| \ge \eta_n\sqrt{n}\right)\right] \\
? \quad & \to 0.
\end{aligned}$$

[2] $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Also, we have [3]

$$\mathsf{E}\left|\frac{1}{n^2}\sum_{i\neq j}\left(1-\sigma_{ij}^{-1}\right)^2\left|x_{ij}^{(n)}I\left(\left|x_{ij}^{(n)}\right|\leq\eta_n\sqrt{n}\right)-\mathsf{E}\left(x_{ij}^{(n)}I\left(\left|x_{ij}^{(n)}\right|\leq\eta_n\sqrt{n}\right)\right)\right|^2\right|^4$$

$$\leq\frac{C}{n^8}\left[\sum_{i\neq j}\mathsf{E}\left|x_{ij}^{(n)}\right|^8 I\left(\left|x_{ij}^{(n)}\right|\leq\eta_n\sqrt{n}\right)+\left(\sum_{i\neq j}\mathsf{E}\left|x_{ij}^{(n)}\right|^4 I\left(\left|x_{ij}^{(n)}\right|\leq\eta_n\sqrt{n}\right)\right)^2\right]$$

$$\leq Cn^{-2}\left[n^{-1}\eta_n^6+\eta_n^4\right],$$

which is summable. From the two estimates above, we conclude that

$$L\left(F^{\widetilde{\mathbf{W}}_n},F^{\widehat{\mathbf{W}}_n-\mathsf{E}\widehat{\mathbf{W}}_n}\right)\to 0,\quad \text{a.s.}. \qquad ?$$

Step 5. Proof by MCT

Up to here, we have proved that we may truncate, centralize, and rerscalse the entries of the Winger matrix at $\eta_n\sqrt{n}$ and remove the diagonal elements without changing the LSD.

Noe, we assume that the variables are truncated at $\eta_n\sqrt{n}$ and then centralized and rescaled.

Again for simplicity, the truncated and centralized variables are still denoted by x_{ij} with properties as following:

1. The variables $\{x_{ij}, 1\leq i\leq j\leq n\}$ are independent and $x_{ii}=0$.
2. $\mathsf{E}(x_{ij})=0$ and $\mathsf{Var}(x_{ij})=1$.
3. $|x_{ij}|\leq\eta_n\sqrt{n}$.

In order to prove the Theorem 1.4.1, we need to show that

(1) $\mathsf{E}[\beta_k(\mathbf{W}_n)]\to\beta_k$ as $n\to\infty$.

(2) For each fixed k, $\sum_n \mathsf{E}|\beta_k(\mathbf{W}_n)-\mathsf{E}(\beta_k(\mathbf{W}_n))|^4<\infty$.

The Proof of (1):

Let $\mathbf{i}=(i_1,\ldots,i_k)\in\{1,\ldots,n\}^k$. As in the iid case, we write

$$\mathsf{E}\left[\beta_k\left(\mathbf{W}_n\right)\right]=n^{-1-k/2}\sum_{\mathbf{i}}\mathsf{E}X(G(\mathbf{i})),$$

where $X(\mathbf{i})=x_{i_1i_2}x_{i_2i_3}\cdots x_{i_ki_1}$ and $G(\mathbf{i})$ is the graph defined by $\mathbf{i}$.

[3] Here we use the elementary inequality

$$\mathsf{E}\left|\sum X_i\right|^{2k}\leq C_k\left(\sum\mathsf{E}\left|X_i\right|^{2k}+\left(\sum\mathsf{E}\left|X_i\right|^2\right)^k\right)$$

for some constant C_k if the X_i's are independent with zero mean.

As same as the iid case, we split $\mathsf{E}\left[\beta_k\left(\mathbf{W}_n\right)\right]$ into 3 sums according to the categories of graphs:

$$\mathsf{E}\left[\beta_k\left(\mathbf{W}_n\right)\right] = S_1 + S_2 + S_3.$$

We know that the terms in S_2 are all 0, so $S_2 = 0$.

We now show that $S_3 \to 0$. Split S_3 as $S_{31} + S_{32}$, where S_{31} consists of the terms corresponding to a $\Gamma_3(k,t)$-graph that contains a coincident class with at least 3 edges and S_{32} is the sum of the remaining terms in S_3.

To estimate S_{31}, assume that the $\Gamma_3(k,t)$-graph contains ℓ noncoincident edges with multiplicity $\nu_1, \ldots, \nu_\ell$ among which at least one is greater than 2. Note that the multiplicities are subjects to $\nu_1 + \cdots + \nu_\ell = k$. Also, each term in S_{31} is bounded by

$$n^{-1-k/2} \prod_{i=1}^{\ell} E\left|x_{a_i,b_i}\right|^{\nu_i} \leq n^{-1-k/2}\left(\eta_n\sqrt{n}\right)^{\sum_{i=1}^{\ell}(\nu_i-2)} = n^{-1-\ell}\eta_n^{k-2\ell}.$$

Since the graph is connected and the number of its noncoincident edges is ℓ, the number of noncoincident vertieces is not more than $\ell+1$, which implies that the number of terms in S_{31} is not more than $n^{\ell+1}$. Therefore,

$$|S_{31}| \leq C_k \eta_n^{k-2\ell} \to 0$$

since $k - 2\ell \geq 1$.

To estimate S_{32}, we note that the $\Gamma_3(k,t)$-graph contains exactlly $k/2$ noncoincident edges, each with multiplicity 2. Then each term in S_{32} is bounded by $n^{-1-k/2}$. Since the graph is not in category 1, the graph of noncoincident edges must contain a cycle, and hence the number of noncoincident vertices is not more than $k/2$ and therefore

$$|S_{32}| \leq Cn^{-1} \to 0.$$

Then, the evaluation of S_1 is exactlly the same as in the iid case and hence is omitted. Hence, we complete the proof of $\mathsf{E}[\beta_k(\mathbf{W}_n)] \to \beta_k$ as $n \to \infty$.

The Proof of (2):

Unlike in the proof of (1.5), the almost sure convergence cannot follow by estimating the variance of $\beta_k(\mathbf{W}_n)$. We need to estimate its fourth moment as

$$
\begin{aligned}
&\mathsf{E}[\beta_k(\mathbf{W}_n)-\mathsf{E}(\beta_k(\mathbf{W}_n))]^4\\
&=n^{-4-2k}\cdot\mathsf{E}\left[\sum_{\mathbf{i}}[X(\mathbf{i})-\mathsf{E}X(\mathbf{i})]\right]^4\\
&=n^{-4-2k}\cdot\mathsf{E}\Bigg\{\sum_{\mathbf{i}_1}[X(\mathbf{i}_1)-\mathsf{E}X(\mathbf{i}_1)]+\sum_{\mathbf{i}_2}[X(\mathbf{i}_2)-\mathsf{E}X(\mathbf{i}_2)]\\
&\qquad\qquad+\sum_{\mathbf{i}_3}[X(\mathbf{i}_3)-\mathsf{E}X(\mathbf{i}_3)]+\sum_{\mathbf{i}_4}[X(\mathbf{i}_4)-\mathsf{E}X(\mathbf{i}_4)]\Bigg\}^4\\
&=n^{-4-2k}\sum_{\mathbf{i}_j,\,j=1,2,3,4}\left\{\mathsf{E}\prod_{j=1}^{4}[X(\mathbf{i}_j)-\mathsf{E}X(\mathbf{i}_j)]\right\},
\end{aligned}
\tag{1.10}
$$

where $\mathbf{i}_j$ is a vector of k integers not larger than n, $j=1,2,3,4$. As in the last section, for each $\mathbf{i}_j$, we construct a graph $G_j=G(\mathbf{i}_j)$.

There are two cases that some terms in (1.10) are zero:

- For some j, $G(\mathbf{i}_j)$ does not have any edges coincident with edges of the other three graphs.
- $G=\cup_{j=1}^{4}G_j$ has a single edge.

Now, let us estimate the nonzero terms in (1.10). Assume that G has ℓ noncoincident edges with multiplicities $\nu_1,\dots,\nu_\ell$, sunject to the constraint $\nu_1+\cdots+\nu_\ell=4k$. Then, the term corresponding to G is bounded by

$$16\cdot n^{-4-2k}\prod_{j=1}^{\ell}(\eta_n\sqrt{n})^{\nu_j-2}=16\cdot\eta_n^{4k-2\ell}n^{-4-\ell}.$$

Suppose the number of noncoincident vertices in G is t. It is obvious that $t\le\ell+1$ and $\ell\le 2k$. For a fixes k, we have

$$
\begin{aligned}
&\mathsf{E}[\beta_k(\mathbf{W}_n)-\mathsf{E}(\beta_k(\mathbf{W}_n))]^4\\
&=\sum_{\ell\le 2k}\sum_{t\le\ell+1}4k\cdot\mathrm{C}_t^2\cdot n^t\cdot\left(16\eta_n^{4k-2\ell}n^{-4-\ell}\right)\\
&=32\sum_{\ell\le 2k}\sum_{t\le\ell+1}k\cdot t(t-1)\cdot n^{t-4-\ell}\cdot\eta_n^{4k-2\ell}\\
&\le 32\sum_{\ell\le 2k}k\eta_n^{4k-2\ell}\ell(\ell+1)^2n^{-3}\\
&\le C_k\eta_n^{4k}n^{-3},
\end{aligned}
$$

which is summable, and thus (2) is proved. Consequently, the proof of Theorem 1.4.1 is complete.

1.5 Semicircular Law by the Stieltjes Transform

1.5.1 Cauchy's Residue Theorem

Theorem 1.5.1 (Residue Theorem). *Let f be holomorphic inside and on a simple closed, positively oriented path γ except at points $a_1, \ldots, a_n$ inside γ. Then*

$$\int_\gamma f(z)\, \mathrm{d}z = 2\pi i \sum_{k=1}^{n} \mathsf{Res}(f(z); a_k).$$

Theorem 1.5.2 (Resideus at Simple Poles). *Suppose taht $f(z)$ has a simple pole at a. Then*

$$\mathsf{Res}(f(z); a) = \lim_{z\to a}(z-a)f(z).$$

1.5.2 Stieltjes Transform

Stieltjes transform (or Cauchy transformation) is another important transformation in mathematics. Compared with Fourier transform, it offers a easier way to obtain the density function of a signed measure via its stieltjes transform.

Definition 1.5.3. *If $G(x)$ is a function of bounded variation on the real line, then its stieltjes transform is defined by*

$$s_G(z) = \int \frac{1}{x-z}\, \mathrm{d}G(x),$$

where $z \in D \equiv \{z \in \mathbb{C} : \Im z > 0\}$

Remark 1.5.4. *Note the integration here is Lebesgue-Stieltjes integration, which generalizes Riemann-Stieltjes integration. Here we give some explaination about Lebesgue-Stieltjes. Firstly, we need to generate Lebesgue-Stieltjes measure, which may be associated to any function of bounded variation on the real line, such as some $G(x)$. And we define $G((a,b]) = G(b) - G(a)$ for any $a, b \in \mathbb{R}$, we can verify that this definition follow Caratheodory-Hahn Theorem, which means we could obtain a measure μ_G is an extension of G on $(a,b]$, $a, b \in \mathbb{R}$. Secondly, by using the classical process we could construct L-S integral.*

Fristly, we give a Fundamental Theorem of Calculus for Lebesgue-Stieltjes integrals.

Theorem 1.5.5. *If g is a Lebesgue measurable function on $\mathbb{R}$, f is a nonnegative Lebesgue integrable function on $\mathbb{R}$, and $F(x) = L\int_{-\infty}^{x} f d\mu$, then:*

1. *F is bounded, monotone increasing, absolutely, continous, and differeiable almost every where, and $F' = f$ a.e.*
2. *We have Lebesgue-Stieltjes measure μ_f so that, for any Lebesgue measurable set E, $\mu_f(E) = L\int_E f\, \mathrm{d}\mu$, and μ_f is absolutley continous w.r.t. Lebesgue measure.*
3. *$L - S\int_{\mathbb{R}} g\, \mathrm{d}\mu_f = L\int_{\mathbb{R}} gf\, \mathrm{d}\mu = L\int_{\mathbb{R}} gF'\, \mathrm{d}\mu$*

Theorem 1.5.6. *For any continuity points $a < b$ of G, we have*

$$\mu_G((a,b]) = G((a,b]) = \lim_{\epsilon \to 0+} \frac{1}{\pi} \int_a^b \Im s_G(x+i\epsilon)\,\mathrm{d}x$$

Proof. Note that

$$\begin{aligned}
&\frac{1}{\pi}\int_a^b \Im s_G(x+i\epsilon)\,\mathrm{d}x \\
=& \frac{1}{\pi}\int_a^b \int \frac{\epsilon dG(y)}{(x-y)^2+\epsilon^2}\,\mathrm{d}x \\
=& \frac{1}{\pi}\int \int_a^b \frac{\epsilon dG(y)}{(x-y)^2+\epsilon^2}\,\mathrm{d}x \\
=& \int \frac{1}{\pi}[\arctan(\epsilon^{-1}(b-y)) - \arctan(\epsilon^{-1}(a-y))]\,\mathrm{d}G(y)
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{\epsilon \to 0+} [\arctan(\epsilon^{-1}(b-y)) - \arctan(\epsilon^{-1}(a-y))] \\
=& \begin{cases} 0, & \text{if } y < a, \\ \frac{2}{\pi}, & \text{if } y = a, \\ \pi, & \text{if } a < y < b, \\ \frac{2}{\pi}, & \text{if } y = b, \\ 0, & \text{if } y > b. \end{cases}
\end{aligned}$$

By using Lebesgue's Dominated Convergence Theorem, we find that the RHS tends to $G[a,b]$. □

From this theorem and the definition of Stieltjes transform we note that there is a one-to-one correspondence between the finite signed measures and their Stieltjes transforms.

The importance of Stieltjes transforms also relies on the next theorem, which shows that to establish the convergence of ESD of a sequence of matrices, one needs only to show that convergence of their Stieltjes transforms and the LSD can be found by the limit Stieltjes transform.

Theorem 1.5.7. *Assume that $\{G_n\}$ is a sequence of functions of bounded variation and $G_n(-\infty) = 0$ for all n.Then*

$$\lim_{n\to\infty} s_{G_n}(z) = s(z), \forall z \in D$$

if and only if there is a function of bounded variation G with $G(-\infty) = 0$ and Stieltjes transforms $s(z)$ and such that $G_n \to G$ vaguely.

Proof. $\Leftarrow$: By observing that $\frac{1}{x-z}$ is continous and bounded and according to the definition of weakly convergence we complete this part immediately.
$\Rightarrow$: By Helly's Selection Theorem, for any subsequenc $\mu_{G_{n_k}}$ of μ_{G_n}, there exist a further sub-

sequence $\mu_{G_{n_k}{}'}$ and a signed measure μ_{G^k} s.t.

$$\mu_{G_{n_k}{}'} \xrightarrow{W} \mu_{G^k}.$$

Therefore, we have

$$s_{G_{n_k}{}'}(z) \to s_{G^k}(z),$$

and since

$$s_{G_n}(z) \to s(z),$$

we know that

$$s_{G^k}(z) = s(z).$$

Therefore, we have proved that for any subsequence $\mu_{G_{n_k}}$ of μ_{G_n} there exist a further subsequence $\mu_{G_{n_k}{}'}$, such that

$$\mu_{G_{n_k}{}'} \xrightarrow{W} \mu_{G^k}$$

and the Stieltjes transform of μ_{G^k} is $s(z)$.
The preceding theorem tells us all these μ_{G^k} are the same, say some μ_G. Here, we complete the proof. $\square$

Theorem 1.5.8. *Let G be a function of bounded variation and $x_0 \in \mathbb{R}$. Suppose that $\lim_{z\in D\to x_0} \Im s_G(x_0)$ exists. Call it $\Im s_G(x_0)$. Then G is differentiable at x_0, and its derivative is $\frac{1}{\pi}\Im s_G(x_0)$.*

1.5.3 Stieltjes Transform of the Semicircular Law

Let $z = u + iv$ with $v > 0$, let $s(z)$ be the Stieltjes transform of the semicircular law. We consider

$$\begin{aligned}
s(z) &= \frac{1}{2\pi\sigma^2}\int_{-2\sigma}^{2\sigma} \frac{1}{x-z}\sqrt{4\sigma^2 - x^2}\,\mathrm{d}x \\
&= \frac{2}{\pi}\int_0^{\pi} \frac{1}{2\sigma\cos y - z}\sin^2 y\,\mathrm{d}y \quad (\text{setting } x = 2\sigma\cos y) \\
&= \frac{1}{\pi}\int_0^{2\pi} \frac{1}{2\sigma\frac{e^{iy}+e^{-iy}}{2} - z}\left(\frac{e^{iy}-e^{-iy}}{2i}\right)^2 \mathrm{d}y \\
&= -\frac{1}{4i\pi}\oint_{|\zeta|=1} \frac{1}{\sigma(\zeta+\zeta^{-1}) - z}\left(\zeta - \zeta^{-1}\right)^2 \zeta^{-1}\,\mathrm{d}\zeta \quad (\text{setting } \zeta = e^{iy}) \\
&= -\frac{1}{4i\pi}\oint_{|\zeta|=1} \frac{(\zeta^2-1)^2}{\zeta^2(\sigma\zeta^2 + \sigma - z\zeta)}\,\mathrm{d}\zeta.
\end{aligned}$$

We will use Residue Theorem to evaluate this integral. We need three steps:

(1) Find all poles of the integrand;

(2) Determine which ones falls inside the integral area;

(3) Evaluate residues.

Step 1

By letting $\zeta^2(\sigma\zeta^2+\sigma-z\zeta)=0$, we got three roots: $\zeta_0=0$, $\zeta_1=(z+\sqrt{z^2-4\sigma^2})/(2\sigma)$ and $\zeta_2=(z-\sqrt{z^2-4\sigma^2})(2\sigma)$. Note that the square root of a compelx number is not unique, it depends on its argument, however here, and throughout this lecture, the square root of a complex number is specified as the one with the positive imaginary part.

Step 2

Lemma 1.5.9. *If $z=u+iv\in\mathbb{C}$, we have:*

$$\sqrt{z}=\mathrm{sign}(\Im z)\frac{|z|+z}{\sqrt{2(|z|+\Re z)}}. \tag{1.11}$$

Proof. Let θ denotes a given argument of z, then we have $z=|z|e^{i\theta}$. When$\theta\in(0,\pi)$,

$$\begin{aligned}\sqrt{z}&=\sqrt{|z|e^{\frac{i\theta}{2}}}\\&=\sqrt{|z|}\left(\cos\frac{\theta}{2}+i\sin\frac{\theta}{2}\right)\\&=\sqrt{|z|}\left(\sqrt{\frac{1+\cos\theta}{2}}+i\sqrt{\frac{1-\cos\theta}{2}}\right)\\&=\sqrt{|z|}\left(\sqrt{\frac{1+\Re z/|z|}{2}}+i\sqrt{\frac{1-\Re z/|z|}{2}}\right)\\&=\sqrt{\frac{|z|+\Re z}{2}}+i\sqrt{\frac{|z|-\Re z}{2}}\\&=\frac{|z|+z}{\sqrt{2(|z|+\Re z)}}.\end{aligned}$$

Similarily, when $\theta\in(\pi,2\pi]$, we gain

$$\sqrt{z}=\frac{-|z|-z}{\sqrt{2(|z|+\Re z)}}.$$

This lemma is proved. □

Remark 1.5.10. *By the lemma above, we have*

$$\Re\left(\sqrt{z}\right)=\frac{1}{\sqrt{2}}\mathrm{sign}(\Im z)\sqrt{|z|+\Re z}=\frac{\Im z}{\sqrt{2(|z|-\Re z)}}$$

and

$$\Im\left(\sqrt{z}\right)=\frac{1}{\sqrt{2}}\mathrm{sign}(\Im z)\sqrt{|z|-\Re z}=\frac{|\Im z|}{\sqrt{2(|z|+\Re z)}}.$$

Throughout the lecture note, the square root of any complex number has positive imaginary part.

Now, we're ready to determine which poles falls inside the integral area. Applying 1.11 to

ζ_1 and ζ_2, we find that the real part of $\sqrt{z^2-4\sigma^2}$ has the same sign as the real part of z.(Since the real part of $\sqrt{z^2-4\sigma^2}$ has the same sign as the imaginary part of $z^2-4\sigma^2$.) This implies that $|\zeta_1|>|\zeta_2|$. But we have $\zeta_1\zeta_2=1$, we conclude that $\zeta_2<1$ and thus the two poles 0 and ζ_2 of the integrand are in the disk $|\zeta|<1$.

Step 3

By simple calculation, we find the residues at there two poles are

$$\frac{z}{\sigma^2} \quad \text{and} \quad -\sigma^{-1}\sqrt{z^2-4\sigma^2}.$$

Hence, we have the following lemma.

Lemma 1.5.11. *The Stieltjes transform for the semicircular law with scale parameter $\sigma^2=1$ is*

$$s(z)=-\frac{1}{2}(z-\sqrt{z^2-4}). \tag{1.12}$$

1.5.4 Proof of Theorem 4.1

At first, we truncate the underlying variables at $\eta_n\sqrt{n}$ and remove the diagonal elements and then centralize and rescale the off-diagonal elements as done in Step $1-4$ in the last section. That is, we assume that:

1. The variables $\{x_{ij}, 1\le i\le j\le n\}$ are independent and $x_{ii}=0$.
2. $\mathsf{E}(x_{ij})=0$ and $\mathsf{Var}(x_{ij})=1$.
3. $|x_{ij}|\le\eta_n\sqrt{n}$.

By definition, the Stieltjes transform of $F^{\mathbf{W}_n}$ is given by

$$s_n(z)=\frac{1}{n}\mathrm{tr}(\mathbf{W}_n-z\mathbf{I}_n)^{-1}.$$

We shall then proceed in our proof by taking the following three steps:

(1) For any fixed $z\in\mathbb{C}^+$,$s_n-\mathsf{E}s_n(z)\to 0$, a.s.

(2) For any fixed $z\in\mathbb{C}^+$, $\mathsf{E}s_n(z)\to s(z)$, the Stieltjes transform of the semicircular law.

(3) Outside a null set, $s_n(z)\to s(z)$ for every $z\in\mathbb{C}^+$.

Then, apply Theorem 1.5.7, it follows that, except for this null set, $F^{\mathbf{W}_n}\to F(x)$ weakly.

Step 1. Almost sure convergence of the random part

In this part we want to prove $s_n - \mathsf{E}s_n(z) \to 0$, a.s. For the first step, we need the extended Burkholder inequality.

Lemma 1.5.12. *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $p > 1$,*

$$\mathsf{E}|\sum X_k|^p \le K_p \mathsf{E}\left(\sum |X_k|^2\right)^{p/2}.$$

Proof. The lemma can be proved by C_r inequality, we shall omit the proof. □

Similarly, we introduce here another inequality without proving it.

Lemma 1.5.13. *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$, and let E_k denote conditional expectation w.r.t. $\mathcal{F}_k$.Then, for $p \ge 2$,*

$$\mathsf{E}|\sum X_k|^p \le K_p \left(\mathsf{E}(\sum \mathsf{E}_{k-1}|X_k|^2)^{p/2} + \mathsf{E}\sum |X_k|^p\right).$$

And we need two lemmas from linear algebra:

Lemma 1.5.14. *If matrix $\mathbf{A}$ and $\mathbf{A}_k$, the k-th major submatrix of $\mathbf{A}$ of order (n-1), are both nonsingular and symmetric, then*

$$\operatorname{tr}\left(\mathbf{A}^{-1}\right) - \operatorname{tr}\left(\mathbf{A}_k{}^{-1}\right) = \frac{1 + \boldsymbol{\alpha}_k' \mathbf{A}_k{}^{-2} \boldsymbol{\alpha}_k}{a_{kk} - \boldsymbol{\alpha}_k' \mathbf{A}_k{}^{-1} \boldsymbol{\alpha}_k}$$

If $\mathbf{A}$ is Hermitian, then $\boldsymbol{\alpha}_k'$ is replaced by $\boldsymbol{\alpha}_k^{\mathsf{H}}$.

Lemma 1.5.15. *Let $z = u + iv$, $v > 0$, and let $\mathbf{A}$ be an $n \times n$ Hermitian matrix. Then*

$$|\operatorname{tr}(\mathbf{A} - z\mathbf{I}_n)^{-1} - \operatorname{tr}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1}| \le v^{-1}.$$

Proof. According to Lemma 1.5.14, we have

$$\operatorname{tr}(\mathbf{A} - z\mathbf{I}_n)^{-1} - \operatorname{tr}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1} = \frac{1 + \boldsymbol{\alpha}_k^{\mathsf{H}}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k}{a_{kk} - z - \boldsymbol{\alpha}_k^{\mathsf{H}}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}$$

Since $\mathbf{A}_k$ is Hermitian, there exist an $(n-1) \times (n-1)$ unitary matrix $\mathbf{E}$ such that

$$\mathbf{A}_k = \mathbf{E}^{\mathsf{H}} \operatorname{diag}[\lambda_1, \lambda_2, \ldots, \lambda_{n-1}]\, \mathbf{E}$$

and let $\boldsymbol{\alpha}_k^{\mathsf{H}} \left(\mathbf{E}^{\mathsf{H}}\right)^2 = (y_1, y_2, \ldots, y_{n-1})$. Then we have

$$\begin{aligned}
|1+\boldsymbol{\alpha}_k^{\mathsf{H}} (\mathbf{A}_k - z\mathbf{I}_{n-1})^{-2} \boldsymbol{\alpha}_k| &= |1+\boldsymbol{\alpha}_k^{\mathsf{H}} \left(\mathbf{E}^{\mathsf{H}}\boldsymbol{\Lambda}\mathbf{E} - z\mathbf{I}_{n-1}\right)^{-2} \boldsymbol{\alpha}_k| \\
&= |1+\boldsymbol{\alpha}_k^{\mathsf{H}} \left(\mathbf{E}^{\mathsf{H}}\right)^2 (\boldsymbol{\Lambda} - z\mathbf{I}_{n-1})^{-2} \mathbf{E}^2\boldsymbol{\alpha}_k| \\
&\leq 1 + \left| \sum_{\ell=1}^{n-1} |y_l|^2 \frac{1}{(\lambda_l - z)^2} \right| \\
&= 1 + \sum_{\ell=1}^{n-1} |y_l|^2 \left((\lambda_l - u)^2 + v^2\right)^{-1} \\
&= 1 + \sum_{\ell=1}^{n-1} |y_l| \left((\lambda_l - u)^2 + v^2\right)^{-1} |y_l| \\
&= 1 + \boldsymbol{\alpha}_k^{\mathsf{H}} ((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1} \boldsymbol{\alpha}_k,
\end{aligned}$$

on the other hand, we have

$$\begin{aligned}
&\Im(a_{kk} - z - \boldsymbol{\alpha}_k^{\mathsf{H}} (\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1} \boldsymbol{\alpha}_k) \\
&= -v(1 + \boldsymbol{\alpha}_k^{\mathsf{H}}((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k).
\end{aligned}$$

Thus,

$$\begin{aligned}
&|\mathsf{tr}\,(\mathbf{A} - z\mathbf{I}_n)^{-1} - \mathsf{tr}\,(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1}| \\
&\leq \frac{1 + \boldsymbol{\alpha}_k^{\mathsf{H}}((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})\boldsymbol{\alpha}_k}{v(1 + \boldsymbol{\alpha}_k^{\mathsf{H}}((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k)} \\
&= 1/v.
\end{aligned}$$

□

Remark 1.5.16. *In the proof of the Lemma 1.5.15, we can obtain two useful formulas:*

$$\Im(-z - \boldsymbol{\alpha}_k^{\mathsf{H}} (\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1} \boldsymbol{\alpha}_k) = -v(1 + \boldsymbol{\alpha}_k^{\mathsf{H}}((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k)$$

and

$$\boldsymbol{\alpha}_k^{\mathsf{H}} (\mathbf{A}_k - z\mathbf{I}_{n-1})^{-2} \boldsymbol{\alpha}_k \leq \boldsymbol{\alpha}_k^{\mathsf{H}}((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k.$$

Now, we are ready to prove Theorem 1.4.1. Denote by $\mathsf{E}_k(\cdot)$ conditional expectation w.r.t. the σ-field generated by the random variables $\{x_{ij}, i, j > k\}$, with the convention that $\mathsf{E}_n s_n(z) = \mathsf{E} s_n(z)$ and $\mathsf{E}_0 s_n(z) = s_n(z)$. Then, we have

$$s_n(z) - \mathsf{E}\,(s_n(z)) = \sum_{k=1}^{n} [\mathsf{E}_{k-1}\,(s_n(z)) - \mathsf{E}_k\,(s_n(z))] := \sum_{k=1}^{n} \gamma_k.$$

And we consider

$$\begin{aligned}\gamma_k &= \frac{1}{n}\left(\mathsf{E}_{k-1}\mathrm{tr}\left(\mathbf{W}_n - z\mathbf{I}\right)^{-1} - \mathsf{E}_k\mathrm{tr}\left(\mathbf{W}_n - z\mathbf{I}\right)^{-1}\right) \\ &= \frac{1}{n}\left(\left[\mathsf{E}_{k-1}\mathrm{tr}\left(\mathbf{W}_n - z\mathbf{I}\right)^{-1} - \mathsf{E}_{k-1}\mathrm{tr}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\right]\right. \\ &\qquad\qquad \left. - \left[\mathsf{E}_k\mathrm{tr}\left(\mathbf{W}_n - z\mathbf{I}\right)^{-1} - \mathsf{E}_k\mathrm{tr}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\right]\right)\end{aligned}$$

where $\mathbf{W}_k$ is the matrix obtained from $\mathbf{W}_n$ with the k-th row and column removed and $\boldsymbol{\alpha}_k$ is the k-th column of $\mathbf{W}_n$ with the k-th element removed.

By Lemma 1.5.15, we know that

$$|\mathsf{E}_{k-1}\mathrm{tr}\left(\mathbf{W}_n - z\mathbf{I}\right)^{-1} - \mathsf{E}_{k-1}\mathrm{tr}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}| \leq 2v^{-1},$$

hence,

$$|\gamma_k| \leq 2/nv.$$

Note that $\{\gamma_k\}$ is a martingale difference sequence, thus, by Lemma 1.5.12, we have

$$\begin{aligned}\mathsf{E}|s_n(z) - \mathsf{E}(s_n(z))|^4 &\leq K_4\mathsf{E}\left(\sum_{k=1}^{n}|\gamma_k|^2\right)^2 \\ &\leq K_4\mathsf{E}\left(\sum_{k=1}^{n}\frac{2}{n^2v^2}\right)^2 \\ &\leq \frac{4K_4}{n^2v^4}.\end{aligned}$$

By the Borel-Cantelli lemma, we complete the proof.

Step 2. Convergence of the expected Stieltjes transform

In this part, we want to prove $\mathsf{E}s_n(z) \to s(z)$. We will proceed this part by some estimations. Firstly, we have a lemma about the trace of an inverse matrix.

Lemma 1.5.17. *If both* $\mathbf{A}$ *and* $\mathbf{A}_k$, $k = 1, 2, \ldots, n$, *are nonsingular, and if we write* $\mathbf{A}^{-1} = \left(a^{kl}\right)$, *then*

$$a^{kk} = \frac{1}{a_{kk} - \boldsymbol{\alpha}_k{}'\mathbf{A}_k^{-1}\boldsymbol{\beta}_k},$$

and hence

$$\mathrm{tr}\left(\mathbf{A}^{-1}\right) = \sum_{k=1}^{n}\frac{1}{a_{kk} - \boldsymbol{\alpha}_k'\mathbf{A}_k^{-1}\boldsymbol{\beta}_k},$$

where a_{kk} *is the k-th diagonal entry of* $\mathbf{A}$, $\mathbf{A}_k$ *is defined above,* $\boldsymbol{\alpha}_k'$ *is the vector obtained from the k-th row of* $\mathbf{A}$ *by deleting the k-th entry, and* $\boldsymbol{\beta}_k$ *is the vector from the k-th column by deleting the k-th entry.*

From this lemma, if $\mathbf{A}$ is an $n \times n$ Hermitian nonsinguar matrix, it follows immediately

that
$$\operatorname{tr}\left(\mathbf{A}^{-1}\right)=\sum_{k=1}^{n}\frac{1}{a_{kk}-\boldsymbol{\alpha}_k^{\mathsf{H}}\mathbf{A}_k^{-1}\boldsymbol{\alpha}_k}.$$

By Lemma 1.5.17, we have
$$\begin{aligned}s_n(z)&=\frac{1}{n}\operatorname{tr}\left(\mathbf{W}_n-z\mathbf{I}_n\right)^{-1}\\&=\frac{1}{n}\sum_{k=1}^{n}\frac{1}{-z-\boldsymbol{\alpha}_k{}^{\mathsf{H}}\left(\mathbf{W}_k-z\mathbf{I}_{n-1}\right)^{-1}\boldsymbol{\alpha}_k}.\end{aligned}$$

Let $\varepsilon_k=\mathsf{E}s_n(z)-\boldsymbol{\alpha}_k{}^{\mathsf{H}}\left(\mathbf{W}_k-z\mathbf{I}_{n-1}\right)^{-1}\boldsymbol{\alpha}_k$. Then we have

$$\begin{aligned}\mathsf{E}s_n(z)&=\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\frac{1}{-z-\mathsf{E}s_n(z)+\varepsilon_k}\\&=\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\frac{z+\mathsf{E}s_n(z)}{(-z-\mathsf{E}s_n(z)+\varepsilon_k)(z+\mathsf{E}s_n(z))}\\&=\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\frac{z+\varepsilon_k+\boldsymbol{\alpha}_k{}^{\mathsf{H}}\left(\mathbf{W}_k-z\mathbf{I}_{n-1}\right)^{-1}\boldsymbol{\alpha}_k}{(-z-\mathsf{E}s_n(z)+\varepsilon_k)(z+\mathsf{E}s_n(z))}\\&=-\frac{1}{z+\mathsf{E}s_n(z)}+\delta_n,\end{aligned}\tag{1.13}$$

Where
$$\delta_n=\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left(\frac{\varepsilon_k}{(z+\mathsf{E}s_n(z))\left(-z-\mathsf{E}s_n(z)+\varepsilon_k\right)}\right).$$

Soving equation (1.13), we obtain two solutions:
$$\frac{1}{2}\left(-z+\delta_n\pm\sqrt{(z+\delta_n)^2-4}\right).$$

Thus, there might be three cases:
$$\mathsf{E}s_n(z)=\frac{1}{2}\left(-z+\delta_n+\sqrt{(z+\delta_n)^2-4}\right),\qquad \mathsf{E}s_n(z)=\frac{1}{2}\left(-z+\delta_n-\sqrt{(z+\delta_n)^2-4}\right)$$

or $\mathsf{E}s_n(z)$ takes both values on different sets. We show that only the first case will occur.

For the second case, note that

$$|\,\mathsf{E}s_n(z)|\le\frac{1}{n}\mathsf{E}\sum_{k=1}^{n}\frac{1}{|-z-\boldsymbol{\alpha}_k{}^{\mathsf{H}}\left(\mathbf{W}_k-z\mathbf{I}_{n-1}\right)^{-1}\boldsymbol{\alpha}_k|}.$$

And we consider,

$$\begin{aligned} |-z-\boldsymbol{\alpha}_k^{\mathsf{H}}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\boldsymbol{\alpha}_k| &\geq \left|\Im\left(-z-\boldsymbol{\alpha}_k^{\mathsf{H}}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\boldsymbol{\alpha}_k\right)\right| \\ &= \left|v\left(1+\boldsymbol{\alpha}_k^{\mathsf{H}}\left[\left(\mathbf{W}_k - u\mathbf{I}_{n-1}\right)^2 + v^2\mathbf{I}_{n-1}\right]^{-1}\boldsymbol{\alpha}_k\right)\right|, \end{aligned}$$

which indicates that if we fix $\Re z$ and let $\Im z = v \to \infty$, we have $\mathsf{E}s_n(z) \to 0$ and $\delta_n \to 0$. Consequently,

$$\begin{aligned} \Im\left(\frac{1}{2}\left(-z+\delta_n-\sqrt{(z+\delta_n)^2-4}\right)\right) &= -\frac{v}{2}+\frac{1}{2}\Im(\delta_n)-\frac{1}{2}\Im\left(\sqrt{(z+\delta_n)^2-4}\right) \\ &\leq -\frac{v}{2}+\frac{1}{2}|\left(\delta_n\right)| \to -\infty. \end{aligned}$$

which cannot be $\mathsf{E}s_n(z)$ since this is a contradiction with the property that $\Im\mathsf{E}s_n(z) \geq 0$. Thus, we proved that the second case is impossible, now, se claim that the third case is also impossible.

It's easy to see that $\mathsf{E}s_n(z)$ and $\frac{1}{2}\left(-z+\delta_n \pm \sqrt{(z+\delta_n)^2-4}\right)$ are continous functions on the upper half plane $\mathbb{C}^+$. Then, we know that if $\mathsf{E}s_n(z)$ takes both values on different sets, there must exist some point $z_0 \in \mathbb{C}^+$ such that the two branches intersect at z_0. That is, at this point, we would have

$$\frac{1}{2}\left(-z_0+\delta_n+\sqrt{(z_0+\delta_n)^2-4}\right) = \frac{1}{2}\left(-z_0+\delta_n-\sqrt{(z_0+\delta_n)^2-4}\right),$$

hence $\mathsf{E}s_n(z_0)$ has to be one of tue following:

$$\frac{1}{2}\left(-z_0+\delta_n\right) = \frac{1}{2}\left(-2z_0 \pm 2\right).$$

However, both of the two values above have negative imaginary parts. This also contradicts with $\Im\mathsf{E}s_n(z) \geq 0$. Thus, we proved that

$$\mathsf{E}s_n(z) = \frac{1}{2}\left(-z+\delta_n+\sqrt{(z+\delta_n)^2-4}\right). \tag{1.14}$$

From 1.14, to prove $\mathsf{E}s_n(z) \to s(z)$, it suffices to show that for any fixed $z \in \mathbb{C}^+$,

$$\delta_n(z) \to 0.$$

Since ε_k is related to n, it will be denoted by $\varepsilon_{k,n}$ in the following part. Note that

$$|z+\mathsf{E}s_n(z)| \geq \Im\left(z+\mathsf{E}s_n(z)\right) = v + \mathsf{E}\left(\Im\left(s_n(z)\right)\right) \geq v$$

and

$$\begin{aligned}
\left|-z - \mathsf{E}s_n(z) + \varepsilon_k\right| &= \left|-z - \alpha_k^{\mathsf{H}} \left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1} \boldsymbol{\alpha}_k\right| \\
&\geq \Im\left(z + \alpha_k^{\mathsf{H}} \left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1} \boldsymbol{\alpha}_k\right) \\
&\geq v.
\end{aligned}$$

Now, we consider

$$\begin{aligned}
\left|\delta_n\left(z\right)\right| &= \left|\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left(\frac{\varepsilon_{k,n}}{\left(z + \mathsf{E}s_n(z)\right)\left(-z - \mathsf{E}s_n(z) + \varepsilon_{k,n}\right)}\right)\right| \\
&\leq \frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\frac{\left|\varepsilon_{k,n}\right|}{\left|z + \mathsf{E}s_n(z)\right|\left|-z - \mathsf{E}s_n(z) + \varepsilon_{k,n}\right|} \\
&\leq \frac{1}{n}\sum_{k=1}^{n} \frac{\mathsf{E}\left|\varepsilon_{k,n}\right|}{v^2} \\
&\leq \frac{\max\limits_{1\leq k\leq n} \mathsf{E}\left|\varepsilon_{k,n}\right|}{v^2}.
\end{aligned}$$

Hence, to prove $\delta_n\left(z\right) \to 0$, it is sufficient to show that

$$\max_{1\leq k\leq n} \mathsf{E}\left|\varepsilon_{k,n}\right| \to 0 \quad (n \to +\infty). \tag{1.15}$$

Moreover, using Lemma 1.5.15, we have

$$\left|\frac{1}{n}\mathsf{E}\left(\mathrm{tr}\left(\mathbf{W}_n - z\mathbf{I}\right)^{-1} - \mathrm{tr}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\right)\right| \leq \frac{1}{nv},$$

which indicates that $\mathsf{E}s_n(z) \approx \frac{1}{n}\mathsf{E}\,\mathrm{tr}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}$ when n is large. Therefore, $\varepsilon_{k,n}$ could be approximated by

$$\frac{1}{n}\mathrm{tr}\left(\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\right) - \boldsymbol{\alpha}_k^{\mathsf{H}}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\boldsymbol{\alpha}_k.$$

By elementary calculations, we have

$$\begin{aligned}
&\mathsf{E}\left|\boldsymbol{\alpha}_k^{\mathsf{H}}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\boldsymbol{\alpha}_k - \frac{1}{n}\,\mathrm{tr}\left(\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\right)\right|^2 \\
=&\frac{1}{n^2}\sum_{ij\neq k} \mathsf{E}\left|b_{ij}\right|^2 + \frac{1}{n^2}\sum_{i\neq k} \mathsf{E}\left|b_{ii}\right|^2\left(\mathsf{E}\left|x_{ik}\right|^4 - 1\right) \\
\leq&\frac{1}{n^2}\mathsf{E}\mathrm{tr}\left(\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)\left(\mathbf{W}_k - \bar{z}\mathbf{I}_{n-1}\right)\right)^{-1} + \frac{\eta_n^2}{n}\sum_{i\neq k} \mathsf{E}\left|b_{ii}\right|^2 \\
\leq&\frac{1}{nv^2} + \eta_n^2 \to 0.
\end{aligned}$$

Thus, for any fixed k, we have

$$\begin{aligned}\lim_{n\to+\infty} \mathsf{E}\left|\varepsilon_{k,n}\right|^2 &= \lim_{n\to+\infty} \mathsf{E}\left|\mathsf{E}s_n(z) - \boldsymbol{\alpha}_k^{\mathsf{H}}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\boldsymbol{\alpha}_k\right|^2 \\ &= \lim_{n\to+\infty} \mathsf{E}\left|\boldsymbol{\alpha}_k^{\mathsf{H}}\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\boldsymbol{\alpha}_k - \frac{1}{n}\operatorname{tr}\left(\left(\mathbf{W}_k - z\mathbf{I}_{n-1}\right)^{-1}\right)\right|^2 \\ &= 0\end{aligned}$$

And by using Holder's inequality, we conclude that

$$\mathsf{E}\left|\varepsilon_{k,n}\right| \le \left(\mathsf{E}\left|\varepsilon_{k,n}\right|^2\right)^{1/2} \to 0$$

holds for all ks, which leads to (1.15).

Step 3. Completion of the proof of Theorem 1.4.1

Lemma 1.5.18. *Let $f_1, f_2, \dots,$ be analytic in D, a connected open set of $\mathbb{C}$, satisfying $|f_n(z)| \le M$ for every n and z in D, and f_n converges as $n \to \infty$ for each z in a subset of D having a limit point in D. Then there exists a function f analytic in D for which $f_n(z) \to f(z)$ and $f_n'(z) \to f'(z)$ for all $z \in D$.*

By Step 1 and Step 2, we have proved that for any fixed $z \in \mathbb{C}^+$, there exists a set N_z such that $P(N_z) = 0$ and

$$s_n(z,\omega) \to s(z) \quad \text{for all } \omega \in N_z^c.$$

However, by Theorem 1.5.7 we need to find an null set N that is uniform w.r.t. all $z \in \mathbb{C}^+$. This process will need the lemma above.

Now, let $\mathbb{C}_0^+ = \{z_\ell\}$ be a set that consists of all z of rational real and imaginary parts, and let $N = \cup N_{z_\ell}$. Then

$$s_n(z,\omega) \to s(z) \quad \text{for all } \omega \in N^c \text{ and } z \in \mathbb{C}_0^+.$$

Let

$$\mathbb{C}_m^+ = \{z \in \mathbb{C}^+, \Im z > 1/m, |z| \le m\}.$$

By the definition of Stieltjes transformation we have $|s_n(z)| \le m$, when $z \in \mathbb{C}_m^+$. Morever, we have

$$s_n(z) \to s(z) \quad \text{for all } \omega \in N^c \text{ and } z \in \mathbb{C}_m^+ \cap \mathbb{C}_0^+$$

and $\mathbb{C}_m^+ \cap \mathbb{C}_0^+$ has a limit point in $\mathbb{C}_m$. Therefore, by applying Lemma 1.5.18, we have

$$s_n(z) \to s(z) \quad \text{for all } \omega \in N^c \text{ and } z \in \mathbb{C}_m^+.$$

Let $m \to \infty$, we conclude that

$$s_n(z) \to s(z) \quad \text{for all } \omega \in N^c \text{ and } z \in \mathbb{C}^+.$$

Applying Theorem 1.5.7, we complete the proof.

Lecture 2

Sample Covariance Matrices and Marăenko-Pastur Law

2.1 Marăenko-Pastur Law

2.1.1 Sample Covariance Matrix

Suppose that $\{x_{ij},\ i,j = 1,2,\ldots\}$ is a double array of i.i.d complex random variables with mean zero and variance σ^2. Write $\mathbf{x}_k = (x_{1k},\ldots,x_{pk})^\top$ and $\mathbf{X} = (\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n)$. The sample covariance matrix is usually defined by

$$\mathbf{S}_0 = \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{x}_k-\overline{\mathbf{x}})(\mathbf{x}_k-\overline{\mathbf{x}})^{\mathsf{H}} = \frac{1}{n-1}(\mathbf{X}-\overline{\mathbf{X}})(\mathbf{X}-\overline{\mathbf{X}})^{\mathsf{H}}, \tag{2.1}$$

where $\overline{\mathbf{X}} = (\overline{\mathbf{x}},\overline{\mathbf{x}},\ldots,\overline{\mathbf{x}})$ and $\overline{\mathbf{x}} = \sum_{k=1}^{n}\mathbf{x}_k/n$. However, in spectral analysis of LDRM, the sample covariance matrix is simply defined as

$$\mathbf{S} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^{\mathsf{H}} = \frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}. \tag{2.2}$$

Indeed, both $\mathbf{S}_0$ and $\mathbf{S}_1$ have a same LSD (when it exists). Denote that $\mathbf{S}_1 = \frac{1}{n}(\mathbf{X}-\overline{\mathbf{X}})(\mathbf{X}-\overline{\mathbf{X}})^{\mathsf{H}}$, then it is easy to see that $\mathbf{S}_0$ and $\mathbf{S}_1$ have the same LSD since $(n-1)/n \to 1$. In more detail, suppose that $F(x)$ is the weak limit of

$$F^{S_0} = \frac{1}{p}\sum_{k=1}^{p} I\left(\frac{1}{n-1}\lambda_k \le x\right),$$

then

$$F^{S_1} = \frac{1}{p}\sum_{k=1}^{p} I\left(\frac{1}{n}\lambda_k \le x\right) \xrightarrow{\mathsf{w}} F(x), \tag{2.3}$$

where λ_k's are the eigenvalues of $(n-1)\mathbf{S}_n$.

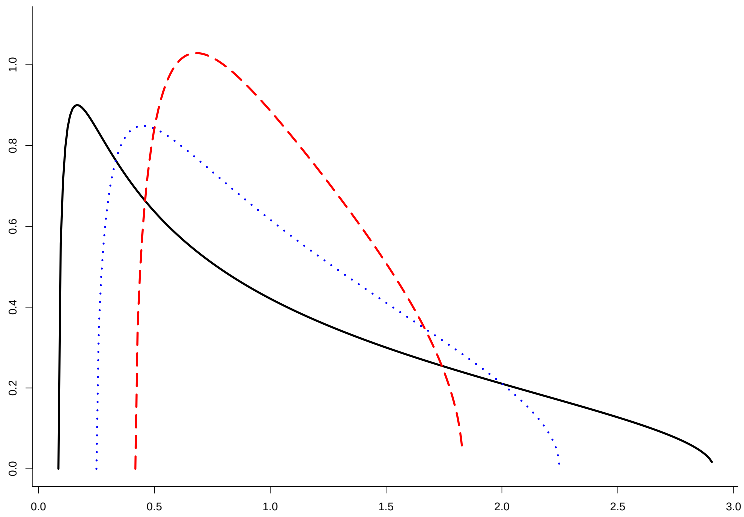


Figure 2.1: Density plots of the M-P distributions with indexes $\sigma^2 = 1$ and $y = 1/8$ (dashed line), $1/4$ (dotted line) and $1/2$ (solid line).

Furthermore, it follows from Theorem A.44 in [1] that

$$\left\| F^{\mathbf{S}_1} - F^{\mathbf{S}} \right\| \leq \frac{1}{p}\mathsf{rank}(\overline{\mathbf{X}}) = \frac{1}{p} \quad \to \quad 0. \tag{2.4}$$

By (2.3) and (2.4), we conclude that (2.1) and (2.2) have the same LSD.

2.1.2 Maršenko-Pastur Law

Theorem 2.1.1 (M-P Law)**.** *Suppose that $p/n \to y \in (0, \infty)$. Under the assumptions stated at the begining of this section, the ESD of $\mathbf{S}$ tends to a limiting distribution with density*

$$p_y(x) = \begin{cases} \frac{1}{2\pi x y \sigma^2}\sqrt{(b-x)(x-a)}, & \textit{if } a \leq x \leq b, \\ 0, & \textit{otherwise}, \end{cases}$$

and a point mass $1 - 1/y$ at the origin if $y > 1$, where $a = a(y) = \sigma^2(1-\sqrt{y})^2$ and $b = b(y) = \sigma^2(1+\sqrt{y})^2$.

Theorem 2.1.2. *Suppose that, for each n, the entries of $\mathbf{X}$ are independent complex variables with a common mean μ and variance σ^2. Assume that $p/n \to y \in (0,\infty)$ and that, for any $\eta > 0$,*

$$\frac{1}{\eta^2 np}\sum_{jk} \mathsf{E}\left(\left| x_{jk}^{(n)} \right|^2 I\left(\left| x_{jk}^{(n)} \right| \geq \eta\sqrt{n} \right) \right) \to 0. \tag{2.5}$$

Then, with probability one, $F^{\mathbf{S}}$ tends to the Maršenko-Pastur Law with ratio index y and scale index σ^2.

Remark 2.1.3 (Assumptions)**.** *As in Section 1.4, by condition (2.5), we further assume that*

1. *There is a sequence $\eta_n \downarrow 0$ such that condition (2.5) holds true when η is replaced by η_n.*

$$|x_{ij}| < \eta_n \sqrt{n}. \tag{2.6}$$

2. *Without loss of generality, we assume $\mu = 0$, $\sigma^2 = 1$ and*

$$\mathsf{E}(x_{ij}) = 0, \qquad \mathsf{Var}(x_{ij}) = 1. \tag{2.7}$$

By (2.7), we have

$$\mathsf{E}(x_{ki}\overline{x}_{kj}) = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}$$

2.1.3 M-P Law and Large-Dimensional Statistics

This section comes from the Section 2.3.1 of [2].

The M-P Law was found as early as in the late sixties (convergence in expectation). However its importance for large-dimensional statistics has been recognised only recently at the beginning of this century. To understand its deep ifluence on multivariate analysis, we plot in Figure 2.2 sample eigenvalues from i.i.d. Gaussian variables $\{x_{ij}\}$. In other words, we use $n = 320$ i.i.d. random vectors $\{\mathbf{x}_i\}$, each with $p = 40$ i.i.d. standard Gaussian coordinates. The histogram of $p = 40$ sample eigenvalues of $\mathbf{S}_n$ displays a wide dispersion from the unit value 1. According to the classical large-sample asymptotic (assuming $n = 320$ is large enough), the sample covariance matrix $\mathbf{S}_n$ should be close to the population covariance matrix $\boldsymbol{\Sigma} = \mathbf{I}_p$. As eigenvalues are continuous functions of matrix entries, the sample eigenvalues of $\mathbf{S}_n$ should converge to 1 (unique eigenvalue of $\mathbf{I}_p$). The plot clearly assesses that this convergence is far from the reality. On the same graph is also plotted the Marăenko-Pastur density function with $y = 40/320 = 1/8$. The closeness between this density and the sample histogram is striking.

Since the sample eigenvalues deviate significantly from the population eigenvalues, the sample covariance matrix $\mathbf{S}_n$ is no more a reliable estimator of its population counter-part $\boldsymbol{\Sigma}$. This observation is indeed the fundamental reason for that classical multivariate methods break down when the data dimension is a bit large compared to the sample size. As an example, consider Hotelling's T^2 statistic which relies on $\mathbf{S}_n^{-1}$. In large-dimensional context (as $p = 40$ and $n = 320$ above), $\mathbf{S}_n^{-1}$ deviates significantly from $\boldsymbol{\Sigma}^{-1}$. In particular, the wider spread of the sample eigenvalues implies that $\mathbf{S}_n$ may have many small eigenvalues, especially when p/n is close to 1. For example, for $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_p$ and $y = 1/8$, the smallest eigenvalue of $\mathbf{S}_n$ is close to $a = (1 - \sqrt{y})^2\sigma^2 = 0.42\sigma^2$ so that the largest eigenvalue of $\mathbf{S}_n^{-1}$ is close to $a^{-1} = 2.38\sigma^{-2}$. When the data to sample size increases to $y = 0.9$, the largest eigenvalue of $\mathbf{S}_n^{-1}$ becomes close to $380\sigma^{-2}$! Clearly, $\mathbf{S}_n^{-1}$ is completely unreliable as an estimator of $\boldsymbol{\Sigma}^{-1}$.

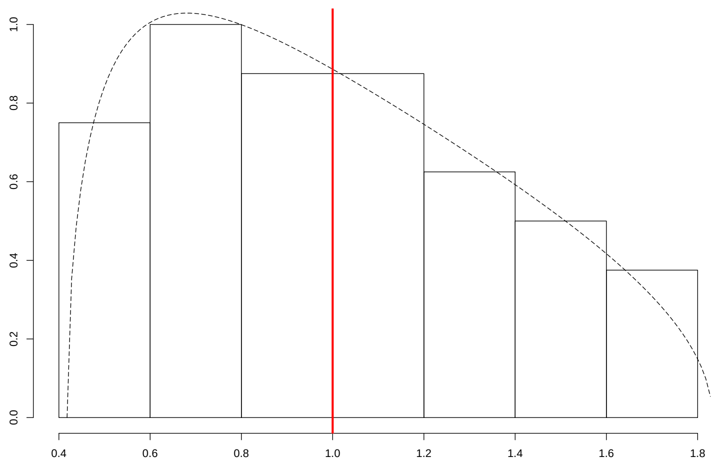


Figure 2.2: Eigenvalues of a sample covariance matrix with standard Gaussian entries, $p = 40$ and $n = 320$. The dashed curve plots the M-P density with $y = 1/8$ and the vertical bar shows the unique population unit eigenvalue.

2.2 M-P Law by the Stieltjes Transform

2.2.1 Stieltjes Transform of the M-P Law

Let $z = u + iv$ with $v > 0$ and $s(z)$ be the Stieltjes transform of the M-P law.

Lemma 2.2.1.

$$s(z) = \frac{\sigma^2(1-y) - z + \sqrt{(z - \sigma^2 - y\sigma^2)^2 - 4y\sigma^4}}{2yz\sigma^2}. \tag{2.8}$$

Proof. When $y < 1$, we have

$$s(z) = \int_a^b \frac{1}{x-z} \frac{1}{2\pi x y \sigma^2} \sqrt{(b-x)(x-a)}\, \mathrm{d}x,$$

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

Letting $x = \sigma^2(1 + y + 2\sqrt{y}\cos w)$ and $\zeta = e^{iw}$, then

$$\begin{aligned}
s(z) &= \int_0^\pi \frac{2}{\pi} \frac{1}{(1+y+2\sqrt{y}\cos w)\left(\sigma^2(1+y+2\sqrt{y}\cos w) - z\right)} \sin^2 w \,\mathrm{d}w \\
&= \frac{1}{\pi}\int_0^{2\pi} \frac{\left(\left(e^{iw} - e^{-iw}\right)/2i\right)^2}{\left(1+y+\sqrt{y}\left(e^{iw}+e^{-iw}\right)\right)\left(\sigma^2\left(1+y+\sqrt{y}\left(e^{iw}+e^{-iw}\right)\right) - z\right)} \mathrm{d}w \\
&= -\frac{1}{4i\pi} \oint_{|\zeta|=1} \frac{\left(\zeta - \zeta^{-1}\right)^2}{\zeta\left(1+y+\sqrt{y}\left(\zeta+\zeta^{-1}\right)\right)\left(\sigma^2\left(1+y+\sqrt{y}\left(\zeta+\zeta^{-1}\right)\right) - z\right)} \mathrm{d}\zeta \\
&= -\frac{1}{4i\pi} \oint_{|\zeta|=1} \frac{\left(\zeta^2 - 1\right)^2}{\zeta\left((1+y)\zeta+\sqrt{y}\left(\zeta^2+1\right)\right)\left(\sigma^2(1+y)\zeta+\sqrt{y}\sigma^2\left(\zeta^2+1\right) - z\zeta\right)} \mathrm{d}\zeta.
\end{aligned}$$

Denote the integrand function as $f(\zeta)$, which has five simple poles at

$$\zeta_0 = 0,$$

$$\zeta_1 = \frac{-(1+y)+(1-y)}{2\sqrt{y}} = -\sqrt{y},$$

$$\zeta_2 = \frac{-(1+y)-(1-y)}{2\sqrt{y}} = -1/\sqrt{y},$$

$$\zeta_3 = \frac{-\sigma^2(1+y)+z+\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}}{2\sigma^2\sqrt{y}},$$

$$\zeta_4 = \frac{-\sigma^2(1+y)+z-\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}}{2\sigma^2\sqrt{y}}.$$

Rewrite $f(\zeta)$ as

$$f(\zeta) = \frac{1}{\sigma^2 y}\cdot\frac{(\zeta^2-1)^2}{\zeta(\zeta-\zeta_1)(\zeta-\zeta_2)(\zeta-\zeta_3)(\zeta-\zeta_4)}.$$

By Theorem 1.5.2, we find that the residues at these five poles are

$$\mathsf{Res}(f;\zeta_0) = \frac{1}{\sigma^2 y}\cdot\frac{(\zeta_0^2-1)^2}{(\zeta_0-\zeta_1)(\zeta_0-\zeta_2)(\zeta_0-\zeta_3)(\zeta_0-\zeta_4)} = \frac{1}{y\sigma^2},$$

$$\mathsf{Res}(f;\zeta_1) = \frac{1}{\sigma^2 y}\cdot\frac{(\zeta_1^2-1)^2}{\zeta_1(\zeta_1-\zeta_2)(\zeta_1-\zeta_3)(\zeta_1-\zeta_4)} = -\frac{1-y}{yz},$$

$$\mathsf{Res}(f;\zeta_2) = \frac{1}{\sigma^2 y}\cdot\frac{(\zeta_2^2-1)^2}{\zeta_2(\zeta_2-\zeta_1)(\zeta_2-\zeta_3)(\zeta_2-\zeta_4)} = \frac{1-y}{yz},$$

$$\mathsf{Res}(f;\zeta_3) = \frac{1}{\sigma^2 y}\cdot\frac{(\zeta_3^2-1)^2}{\zeta_3(\zeta_3-\zeta_1)(\zeta_3-\zeta_2)(\zeta_3-\zeta_4)} = \frac{1}{\sigma^2 yz}\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2},$$

$$\mathsf{Res}(f;\zeta_4) = \frac{1}{\sigma^2 y}\cdot\frac{(\zeta_4^2-1)^2}{\zeta_4(\zeta_4-\zeta_1)(\zeta_4-\zeta_2)(\zeta_4-\zeta_3)} = -\frac{1}{\sigma^2 yz}\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}.$$

We are now in a position to determine which poles fall inside the curve $|\gamma| = 1$. We claim that both the real part and imaginary part of

$$\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2} \qquad \text{and} \qquad -\sigma^2(1+y)+z$$

have the same signs, i.e.,

$$\mathsf{sign}\left\{\Re\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}\right\} = \mathsf{sign}\left\{\Re[-\sigma^2(1+y)+z]\right\}$$

and

$$\text{sign}\left\{\Im\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}\right\}=\text{sign}\left\{\Im[-\sigma^2(1+y)+z]\right\}.$$

By Remark 1.5.10, the imaginary part of $\sqrt{\sigma^2(1-y)^2-2\sigma^2(1+y)z+z^2}$ is positive, so it has same sign as $\Im(-\sigma^2(1+y)+z)=v$. Note that

$$\begin{aligned}&\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}\\ =&\sqrt{[\sigma^4(1-y)^2-2\sigma^2(1+y)u+u^2-v^2]+i\cdot 2v[-\sigma^2(1+y)+u]},\end{aligned}$$

by Remark 1.5.10, the real part of $\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}$ has the same sign as $2v[-\sigma^2(1+y)+u]$, where $v>0$.

Noting that $\zeta_3\zeta_4=1$, so we have $|\zeta_3|>1$ and $|\zeta_4|<1$. (See Remark 2.2.2) Also, $|\zeta_1|<1$ and $|\zeta_2|>1$. By Cauchy's Residue Theorem, we obtain

$$\begin{aligned}s(z)&=-\frac{1}{2}\left(\frac{1}{y\sigma^2}-\frac{1}{\sigma^2 yz}\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}-\frac{1-y}{yz}\right)\\&=\frac{\sigma^2(1-y)-z+\sqrt{(z-\sigma^2-y\sigma^2)^2-4y\sigma^4}}{2yz\sigma^2}.\end{aligned}$$

This proves equation (2.8) when $y<1$.

When $y>1$, since the M-P law has a point mass $1-1/y$ at zero, $s(z)$ equals the integral above plus $-(y-1)/yz$. In this case, $|\zeta_1|>1$ and $|\zeta_2|<1$, and thus the residue at ζ_2 should be counted into the integral. Finally, one find that equation (2.8) still holds.

When $y=1$, the equation is still true by continuity in y. □

Remark 2.2.2. *Suppose that $z_1=a_1+ib_1, z_2=a_2+ib_2$, where $a_1a_2>0$ and $b_1b_2>0$. Then*

$$|z_1+z_2|=\sqrt{(a_1+a_2)^2+(b_1+b_2)^2}>\sqrt{(a_1-a_2)^2+(b_1-b_2)^2}=|z_1-z_2|.$$

2.2.2 Proof of Theorem 2.1.2

Let the Stieltjes transform of the ESD of $\mathbf{S}_n$ be denoted by $s_n(z)$. Define

$$s_n(z)=\frac{1}{p}\text{tr}(\mathbf{S}_n-z\mathbf{I}_p)^{-1}.$$

As in Section 1.5, we shall complete the proof by the following steps:

1. For any fixed $z\in\mathbb{C}^+$, $s_n(z)\to \mathrm{E}s_n(z)$, a. s..
2. For any fixed $z\in\mathbb{C}^+$, $\mathrm{E}s_n(z)\to s(z)$, the Stieltjes transform of the M-P Law.
3. Except for a null set, $s_n\to s(z)$ for every $z\in\mathbb{C}^+$.

Similar to Section 1.5, the last step is implied by the first two steps and thus its proof is omitted. We now proceed with the first two steps.

Step 1. Almost sure convergence of the random part

In step 1, we shall use the martingale decomposition method to prove that

$$s_n(z) - \mathsf{E}s_n(z) \to 0, \quad \text{a.s.} \tag{2.9}$$

The following lemma is useful.

Theorem 2.2.3 (Sherman-Morrison)**.**

$$\left(\mathbf{A} + \boldsymbol{\alpha}\boldsymbol{\beta}^{\mathsf{H}}\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathsf{H}}\mathbf{A}^{-1}}{1 + \boldsymbol{\beta}^{\mathsf{H}}\mathbf{A}^{-1}\boldsymbol{\alpha}}.$$

Let $\mathsf{E}_k(\cdot)$ denote the conditional expectation given $\{\mathbf{x}_i,\ k+1 \le i \le n\}$. Note that $s_n(z) = \mathsf{E}_0 s_n(z)$ and $\mathsf{E}s_n(z) = \mathsf{E}_n s_n(z)$. Then, we have

$$\begin{aligned} s_n(z) - \mathsf{E}s_n(z) &= \frac{1}{p}\sum_{k=1}^{n}\left[\mathsf{E}_{k-1}\mathsf{tr}\left(\mathbf{S}_n - z\mathbf{I}_p\right)^{-1} - \mathsf{E}_k\mathsf{tr}\left(\mathbf{S}_n - z\mathbf{I}_p\right)^{-1}\right] \\ &\triangleq \frac{1}{p}\sum_{k=1}^{n}\gamma_k. \end{aligned}$$

Let $\mathbf{S}_{nk} = \mathbf{S}_n - \frac{1}{n}\mathbf{x}_k\mathbf{x}_k^{\mathsf{H}}$, then

$$\begin{aligned} \gamma_k &= \mathsf{E}_{k-1}\mathsf{tr}\left(\mathbf{S}_n - z\mathbf{I}_p\right)^{-1} - \mathsf{E}_k\mathsf{tr}\left(\mathbf{S}_n - z\mathbf{I}_p\right)^{-1} \\ &= \mathsf{E}_{k-1}\mathsf{tr}\left(\mathbf{S}_n - z\mathbf{I}_p\right)^{-1} - \mathsf{E}_{k-1}\mathsf{tr}(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1} + \mathsf{E}_k\mathsf{tr}(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1} - \mathsf{E}_k\mathsf{tr}\left(\mathbf{S}_n - z\mathbf{I}_p\right)^{-1} \\ &= (\mathsf{E}_{k-1} - \mathsf{E}_k)\ \mathsf{tr}\ \left[(\mathbf{S}_n - z\mathbf{I}_p)^{-1} - (\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1}\right] \qquad \text{(Lemma 1.5.14 does NOT work)} \\ &= -(\mathsf{E}_{k-1} - \mathsf{E}_k)\ \mathsf{tr}\ \frac{(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1}\frac{1}{n}\mathbf{x}_k\mathbf{x}_k^{\mathsf{H}}(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1}}{1 + \frac{1}{n}\mathbf{x}_k^{\mathsf{H}}(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1}\mathbf{x}_k} \qquad \text{(Sherman-Morrison)} \\ &= -(\mathsf{E}_{k-1} - \mathsf{E}_k)\ \frac{\mathbf{x}_k^{\mathsf{H}}(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-2}\mathbf{x}_k}{n + \mathbf{x}_k^{\mathsf{H}}(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1}\mathbf{x}_k} \qquad (\because\ \mathsf{tr}(\mathbf{AB}) = \mathsf{tr}(\mathbf{BA})). \end{aligned}$$

An argument similar to the one used in Lemma 1.5.15 shows that

$$\frac{\mathbf{x}_k^{\mathsf{H}}(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-2}\mathbf{x}_k}{n + \mathbf{x}_k^{\mathsf{H}}(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1}\mathbf{x}_k} \le \frac{\mathbf{x}_k^{\mathsf{H}}\left((\mathbf{S}_{nk} - u\mathbf{I}_p)^2 + v^2\mathbf{I}_p\right)^{-1}\mathbf{x}_k}{\Im\left(n + \mathbf{x}_k^{\mathsf{H}}\left(\mathbf{S}_{nk} - z\mathbf{I}_p\right)^{-1}\mathbf{x}_k\right)} = \frac{1}{v}.$$

Therefore, we have

$$|\gamma_k| \le 2/v.$$

Noting that $\{\gamma_k\}$ forms a sequence of bounded martingale differences, by Lemma 1.5.12 with

$p = 4$, we obtain

$$\mathsf{E}\left|s_n(z) - \mathsf{E}s_n(z)\right|^4 \le \frac{K_4}{p^4}\mathsf{E}\left(\sum_{k=1}^{n}|\gamma_k|^2\right)^2 \le \frac{4K_4 n^2}{v^4 p^4} = O\left(n^{-2}\right),$$

which is summable. By Proposition 1.3.1, the inequality above implies (2.9). The proof is complete.

Step 2. Mean convergence

We will show that

$$\mathsf{E}s_n(z) \to s(z), \tag{2.10}$$

where $s_n(z)$ is defined in (2.8) with $\sigma^2 = 1$.

For simplicity of presentation, we need some notations. Let $\mathbf{A}$ be a $n \times n$ matrix, we denote:

$(\mathbf{A})_{ij}$	The (i,j)-th entry of the matrix $\mathbf{A}$,
$[\mathbf{A}]_{ij}$	The ij-submatrix, i.e. $\mathbf{A}$ with i-th row and j-th column deleted,
$(\mathbf{A})_{i\cdot}$	The i-th row of matrix $\mathbf{A}$,
$(\mathbf{A})_{\cdot j}$	The j-th column of matrix $\mathbf{A}$,
$[\mathbf{A}]_{i\cdot}$	Matrix $\mathbf{A}$ with i-th row deleted,
$[\mathbf{A}]_{\cdot j}$	Matrix $\mathbf{A}$ with j-th column deleted.

By Lemma 1.5.17, we can rewrite $s_n(z)$ as the following form:

Lemma 2.2.4.

$$s_n(z) = \frac{1}{p}\sum_{k=1}^{p}\frac{1}{\frac{1}{n}\boldsymbol{\alpha}_k^\top\overline{\boldsymbol{\alpha}}_k - z - \frac{1}{n^2}\boldsymbol{\alpha}_k^\top\mathbf{X}_k^{\mathsf{H}}\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}} - z\mathbf{I}_{p-1}\right)^{-1}\mathbf{X}_k\overline{\boldsymbol{\alpha}}_k}. \tag{2.11}$$

Proof. Let $\boldsymbol{\alpha}_k^\top$ be the k-th row of $\mathbf{X}$, then

$$\mathbf{X}^\top = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n) \quad \text{and} \quad \mathbf{X}^{\mathsf{H}} = (\overline{\boldsymbol{\alpha}}_1, \ldots, \overline{\boldsymbol{\alpha}}_n).$$

The k-th row of $\mathbf{S}_n - z\mathbf{I}_p$ deleting the k-th entry is

$$\frac{1}{n}\boldsymbol{\alpha}_k^\top\mathbf{X}_k^{\mathsf{H}} = \frac{1}{n}\left(\boldsymbol{\alpha}_k^\top\overline{\boldsymbol{\alpha}}_1, \ldots, \boldsymbol{\alpha}_k^\top\overline{\boldsymbol{\alpha}}_{k-1}, \boldsymbol{\alpha}_k^\top\overline{\boldsymbol{\alpha}}_{k+1}, \ldots, \boldsymbol{\alpha}_k^\top\overline{\boldsymbol{\alpha}}_n\right),$$

where $\mathbf{X}_k$ is the kk-submatrix of $\mathbf{X}$. Since $\mathbf{S}_n - z\mathbf{I}_p$ is symmetric, then the k-th column of $\mathbf{S}_n - z\mathbf{I}_p$ deleting the k-th entry is $\frac{1}{n}\mathbf{X}_k\overline{\boldsymbol{\alpha}}_k$.

It is easy to see that $[\mathbf{S}_n]_{k\cdot} = \frac{1}{n}\mathbf{X}_k\mathbf{X}^{\mathsf{H}}$ and $[\mathbf{S}_n]_{\cdot k} = \frac{1}{n}\mathbf{X}\mathbf{X}_k^{\mathsf{H}}$, then we have $[\mathbf{S}_n]_{kk} = \frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}$ and hence

$$[\mathbf{S}_n - z\mathbf{I}_p]_{kk} = \frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}} - z\mathbf{I}_{p-1}.$$

By Lemma 1.5.17, we obtain (2.11). □

Set

$$\varepsilon_k = \frac{1}{n}\boldsymbol{\alpha}_k^\top \bar{\boldsymbol{\alpha}}_k - 1 - \frac{1}{n^2}\boldsymbol{\alpha}_k^\top \mathbf{X}_k^{\mathsf{H}} \left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}} - z\mathbf{I}_{p-1}\right)^{-1} \mathbf{X}_k\bar{\boldsymbol{\alpha}}_k + y_n + y_n z \mathsf{E} s_n(z),$$

where $y_n = p/n$. Then, by (2.11), we have

$$\mathsf{E}s_n(z) = \frac{1}{1 - z - y_n - y_n z \mathsf{E}s_n(z)} + \delta_n, \tag{2.12}$$

where

$$\delta_n = -\frac{1}{p}\sum_{k=1}^{p} \mathsf{E}\left(\frac{\varepsilon_k}{(1 - z - y_n - y_n z\, \mathsf{E}s_n(z))\,(1 - z - y_n - y_n z\, \mathsf{E}s_n(z) + \varepsilon_k)}\right). \tag{2.13}$$

Solving $\mathsf{E}s_n(z)$ from equation (2.12), we get two solutions:

$$s_1(z) = \frac{1}{2y_n z}\left(1 - z - y_n + y_n z\delta_n + \sqrt{(1 - z - y_n - y_n z\delta_n)^2 - 4y_n z}\right),$$
$$s_2(z) = \frac{1}{2y_n z}\left(1 - z - y_n + y_n z\delta_n - \sqrt{(1 - z - y_n - y_n z\delta_n)^2 - 4y_n z}\right).$$

Comparing this with (2.8), it suffices to show that

$$\mathsf{E}s_n(z) = s_1(z) \tag{2.14}$$

and

$$\delta_n \to 0. \tag{2.15}$$

The Proof of (2.14):

Lemma 2.2.5. *Making $v \to \infty$, we have*

$$\mathsf{E}s_n(z) \to 0 \qquad \textit{and} \qquad \delta_n \to 0.$$

Proof.

$$\begin{aligned} |\mathsf{E}s_n(z)| &\le \mathsf{E}|s_n(z)| \\ &\le \frac{1}{p}\mathsf{E}\sum_{k=1}^{p} \frac{1}{\left|\frac{1}{n}\boldsymbol{\alpha}_k^\top\bar{\boldsymbol{\alpha}}_k - z - \frac{1}{n^2}\boldsymbol{\alpha}_k^\top\mathbf{X}_k^{\mathsf{H}}\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}} - z\mathbf{I}_{p-1}\right)^{-1}\mathbf{X}_k\bar{\boldsymbol{\alpha}}_k\right|} \\ &\triangleq \frac{1}{p}\mathsf{E}\sum_{k=1}^{p}\frac{1}{|D_n(v)|}, \end{aligned}$$

where

$$
\begin{aligned}
|D_n(v)| &\geq |\Im(D_n(v))| \\
&= \left| v \left\{ 1 + \frac{1}{n^2} \boldsymbol{\alpha}_k^\top \mathbf{X}_k^{\mathsf{H}} \underbrace{\left[\left(\frac{1}{n} \mathbf{X}_k \mathbf{X}_k^{\mathsf{H}} - u \mathbf{I}_{p-1} \right)^2 + v^2 \mathbf{I}_{p-1} \right]^{-1}}_{\text{positive definite}} \mathbf{X}_k \bar{\boldsymbol{\alpha}}_k \right\} \right| \qquad \text{(Remark 1.5.16)} \\
&\geq |v| \quad \to \quad \infty,
\end{aligned}
$$

which implies that $\mathsf{E}s_n(z) \to 0$ when $v \to \infty$.

The equation (2.12) gives us that

$$
\begin{aligned}
|\delta_n| &= \left| \mathsf{E}s_n(z) - \frac{1}{1 - z - y_n - y_n z \mathsf{E}s_n(z)} \right| \\
&\leq |\mathsf{E}s_n(z)| + \frac{1}{|1 - z - y_n - y_n z \mathsf{E}s_n(z)|}.
\end{aligned}
$$

Let $\mathsf{E}s_n(z) = A + iB$, then

$$
|1 - z - y_n - y_n z \mathsf{E}s_n(z)| \geq |\Im(1 - z - y_n - y_n z \mathsf{E}s_n(z))| = |v + u y_n B| > v - 1,
$$

which implies that $\delta_n \to 0$ as $v \to \infty$. □

By the Lemma 2.2.5 implies, we have

$$
\begin{aligned}
s_1(z) &= \frac{1}{2y_n z} - \frac{1}{2y_n} - \frac{1}{2z} + \frac{\delta_n}{2} - \frac{1}{2} \sqrt{\left(\frac{1}{y_n z} - \frac{1}{y_n} - \frac{1}{z} - \delta_n \right)^2 - \frac{4}{y_n z}} \\
&\to -\frac{1}{2y_n} + \frac{1}{2y_n} = 0, \\
s_2(z) &= \frac{1}{2y_n z} - \frac{1}{2y_n} - \frac{1}{2z} + \frac{\delta_n}{2} + \frac{1}{2} \sqrt{\left(\frac{1}{y_n z} - \frac{1}{y_n} - \frac{1}{z} - \delta_n \right)^2 - \frac{4}{y_n z}} \\
&\to -\frac{1}{2y_n} - \frac{1}{2y_n} = -\frac{1}{y_n} \neq 0.
\end{aligned}
$$

Therefore, $\mathsf{E}s_n(z) = s_1(z)$ for all z with large imaginary part.

If (2.14) is not true for all $z \in \mathbb{C}^+$, then by the continuity of $s_1(z)$ and $s_2(z)$, there exists $z_0 \in \mathbb{C}^+$ such that $s_1(z_0) = s_2(z_0)$, which implies that

$$
(1 - z_0 - y_n + y_n z \delta_n)^2 - 4 y_n z_0 = 0. \tag{2.16}
$$

Thus,

$$
\mathsf{E}s_n(z_0) = s_1(z_0) = \frac{1 - z_0 - y_n + y_n z_0 \delta_n}{2 y_n z_0}. \tag{2.17}
$$

Substituting the solution δ_n of equation (2.12) into the identity above, we obtain

$$\mathsf{E}s_n(z_0) = \frac{1 - z_0 - y_n}{y_n z_0} + \frac{1}{y_n + z_0 - 1 + y_n z_0 \mathsf{E}s_n(z_0)}. \tag{2.18}$$

Noting that for any Stieltjes transform $s_n(z)$ of probability F defined on $\mathbb{R}^+$ and positive y, we have

$$\begin{aligned}
&\Im(y + z - 1 + yzs(z)) \\
&\qquad = \Im\left(z - 1 + \int_0^\infty \frac{yx\,\mathrm{d}F(x)}{x - z}\right) \qquad \left(\because \int \frac{x}{x-z}\,\mathrm{d}F(x) = 1 + zs(z)\right) \\
&\qquad = \Im\left(z - 1 + \int_0^\infty \frac{yx(x - u + iv)}{(x-u)^2 + v^2}\,\mathrm{d}F(x)\right) \\
&\qquad = v\left(1 + \int_0^\infty \frac{yx\,\mathrm{d}F(x)}{(x-u)^2 + v^2}\right) > v > 0.
\end{aligned} \tag{2.19}$$

In view of this, it follows that

$$\Im\left(\frac{1}{y_n + z_0 - 1 + y_n z_0 \mathsf{E}s_n(z_0)}\right) < 0.$$

If $y_n \le 1$, it can be easily seen that

$$\Im\left(\frac{1 - z_0 - y_n}{y_n z_0}\right) = \Im\left(\frac{1 - y_n}{y_n} \cdot \frac{1}{z_0}\right) < 0.$$

Then we conclude that $\Im \mathsf{E}s_n(z_0) < 0$, which is impossible since the imaginary part of the Stieltjes transform should be positive. Ths contradiction leads to the truth of (2.14) for the case $y_n \le 1$.

Remark 2.2.6. *Suppose that $z = u + iv \in \mathbb{C}$. The imaginary parts of z and $1/z$ have different signs, the real parts of z and $1/z$ have the same sign.*

$$\Re(1/z) = \frac{u}{u^2 + v^2}, \quad \Im(1/z) = \frac{-v}{u^2 + v^2}.$$

Therefore,

$$\mathsf{sign}[\Re(1/z)] = \mathsf{sign}(\Re z), \quad \mathsf{sign}[\Im(1/z)] = -\mathsf{sign}(\Im z).$$

In view of (2.16), (2.17) and (2.19), we have

$$y_n + z_0 - 1 + y_n z_0 \mathsf{E}s_n(z_0) \xlongequal{(2.17)} \frac{1}{2}(z_0 + y_n - 1 + y_n z_0 \delta_n) \underset{(2.19)}{\overset{(2.16)}{=\!=\!=}} \sqrt{y_n z_0}. \tag{2.20}$$

Now, let $\underline{s}_n(z)$ be the Stieltjes transform of the matrix $\frac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X}$. We have the relation between $s_n(z)$ and $\underline{s}_n(z)$ given by

Lemma 2.2.7.
$$s_n(z) = \frac{\underline{s}_n(z)}{y_n} - \frac{1-1/y_n}{z}, \qquad \forall y_n > 0.$$

In order to prove this lemma, we need a result from linear algebra:

Lemma 2.2.8. *Suppose that* $\mathbf{A} \in \mathbb{C}^{n\times m}, \mathbf{B} \in \mathbb{C}^{m\times n}, n \geq m$, *then*

$$\sigma(\mathbf{AB}) = \sigma(\mathbf{BA}) \cup \{\underbrace{0,\ldots,0}_{n-m}\},$$

where $\sigma(\mathbf{M}) = \sigma\{\lambda_1,\ldots,\lambda_n\}$ *is the set of all the eigenvalues of* $\mathbf{M} \in \mathbb{C}^{n\times n}$.

Proof of Lemma 2.2.8. The result follows immediately from a well-known identity:

$$|\lambda\mathbf{I} - \mathbf{AB}| = \lambda^{n-m}|\lambda\mathbf{I} - \mathbf{BA}|,$$

the proof of which may be found in standard Linear Algebra textbooks, see, e.g. [5]. □

Proof of Lemma 2.2.7. Without loss of generality, we may assume that $p \geq n$, then

$$\sigma\left(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}\right) = \sigma\left(\frac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X}\right) \cup \{\underbrace{0,\ldots,0}_{p-n}\}.$$

So we have
$$\begin{aligned} y_n s_n(z) &= \frac{1}{n}\mathrm{tr}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_p\right)^{-1} \\ &= \frac{1}{n}\mathrm{tr}\left(\frac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n\right)^{-1} + \frac{1}{n}(p-n)\left(-\frac{1}{z}\right) \\ &= \underline{s}_n(z) + \frac{1-y_n}{z}. \end{aligned}$$

This complete the proof. □

Using Lemma 2.2.7, we get

$$y_n - 1 + y_n z_0 \mathsf{E} s_n(z_0) = z_0 \mathsf{E}\underline{s}_n(z_0).$$

Substituting this into (2.20), we obtain

$$1 + \mathsf{E}\underline{s}_n(z_0) = \sqrt{y}/\sqrt{z_0},$$

which leads to controdiction that

$$\Im(1 + \mathsf{E}\underline{s}_n(z_0)) > 0 \quad \text{but} \quad \Im(\sqrt{y}/\sqrt{z_0}) < 0.$$

This completes the proof of (2.14).

The Proof of (2.15):

Rewrite δ_n as

$$\begin{aligned}\delta_n &= -\frac{1}{p}\sum_{k=1}^{p}\left[\frac{\mathsf{E}\varepsilon_k}{(1-z-y_n-y_n z\mathsf{E}s_n(z))^2}\right]\\ &\quad+\frac{1}{p}\sum_{k=1}^{p}\mathsf{E}\left[\frac{\varepsilon_k^2}{(1-z-y_n-y_n z\mathsf{E}s_n(z))^2\,(1-z-y_n z\mathsf{E}s_n(z)+\varepsilon_k)}\right]\\ &\triangleq J_1+J_2.\end{aligned}$$

At first, by assumptions given in (2.7), we note that

$$\mathsf{E}(\boldsymbol{\alpha}_k^\top\overline{\boldsymbol{\alpha}}_k)=\sum_{k=1}^{n}\mathsf{E}(x_{ki}\overline{x}_{ki})=\sum_{k=1}^{n}\mathsf{Var}(x_{ki})=n$$

and

$$\begin{aligned}\mathsf{E}(\boldsymbol{\alpha}_k^\top\mathbf{M}\overline{\boldsymbol{\alpha}}_k) &= \mathsf{E}\Big(\sum_{i,j}M_{ij}x_{ki}\overline{x}_{kj}\Big)\\ &=\sum_{i,j}\mathsf{E}(M_{ij})\cdot\mathsf{E}(x_{ki}\overline{x}_{kj}) \qquad (\because \text{ independent})\\ &\overset{(2.7)}{=}\sum_{i=1}^{n}\mathsf{E}(M_{ii})=\mathsf{E}[\mathsf{tr}(\mathbf{M})],\end{aligned}\tag{2.21}$$

where $\mathbf{M}=\mathbf{X}_k^{\mathsf{H}}(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}-z\mathbf{I}_{p-1})^{-1}\mathbf{X}_k$. Therefore, we have

$$\begin{aligned}|\mathsf{E}\varepsilon_k| &= \left|-\frac{1}{n^2}\mathsf{E}\,\mathsf{tr}\left[\mathbf{X}_k^{\mathsf{H}}\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}-z\mathbf{I}_{p-1}\right)^{-1}\mathbf{X}_k\right]+y_n+y_n z\mathsf{E}s_n(z)\right|\\ &=\left|-\frac{1}{n}\mathsf{E}\,\mathsf{tr}\left[\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}-z\mathbf{I}_{p-1}\right)^{-1}\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}\right]+y_n+y_n z\mathsf{E}s_n(z)\right| \quad (\mathsf{tr}(\mathbf{AB})=\mathsf{tr}(\mathbf{BA}))\\ &=\left|-\frac{1}{n}\mathsf{E}\,\mathsf{tr}\left[\mathbf{I}_{p-1}+z\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}-z\mathbf{I}_{p-1}\right)^{-1}\right]+\frac{p}{n}+\frac{z}{n}\mathsf{E}\,ps_n(z)\right|\\ &\le\frac{1}{n}+\frac{|z|}{n}\mathsf{E}\left|\mathsf{tr}\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}-z\mathbf{I}_{p-1}\right)^{-1}-ps_n(z)\right|\\ &\le\frac{1}{n}+\frac{|z|}{nv}\to 0, \qquad (\text{Lemma } 1.5.15)\end{aligned}$$

Furthermore, using (2.19), we conclude that

$$|J_1|\le\frac{|\mathsf{E}\varepsilon_k|}{pv^2}\quad\to\quad 0.$$

Now we prove $J_2 \to 0$. Since

$$\begin{aligned}
&\Im\left(1 - z - y_n - y_n z \mathsf{E} s_n(z) + \varepsilon_k\right) \\
= &\Im\left(\frac{1}{n}\boldsymbol{\alpha}_k^\top \overline{\boldsymbol{\alpha}}_k - z - \frac{1}{n^2}\boldsymbol{\alpha}_k^\top \mathbf{X}_k^{\mathsf{H}}\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}} - z\mathbf{I}_{p-1}\right)^{-1}\mathbf{X}_k\overline{\boldsymbol{\alpha}}_k\right) \\
= &-v\left(1 + \frac{1}{n^2}\boldsymbol{\alpha}_k^\top \mathbf{X}_k^{\mathsf{H}}\left[\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}} - u\mathbf{I}_{p-1}\right)^2 + v^2\mathbf{I}_{p-1}\right]^{-1}\mathbf{X}_k\overline{\boldsymbol{\alpha}}_k\right) < -v,
\end{aligned}$$

the last '$<$' follows from the fact that $\left(\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}/n - u\mathbf{I}_{p-1}\right)^2 + v^2\mathbf{I}_{p-1}$ is positive definite. Combining this with (2.19), we obtain

$$\begin{aligned}
|J_2| &\le \frac{1}{pv^3}\sum_{k=1}^{p}\mathsf{E}|\varepsilon_k|^2 \\
&= \frac{1}{pv^3}\sum_{k=1}^{p}\left\{\mathsf{E}\,\widetilde{\mathsf{E}}_k|\varepsilon_k - \widetilde{\mathsf{E}}_k\varepsilon_k|^2 + \mathsf{E}|\widetilde{\mathsf{E}}_k\varepsilon_k - \mathsf{E}\varepsilon_k|^2 + |\mathsf{E}\varepsilon_k|^2\right\},
\end{aligned}$$

where $\widetilde{\mathsf{E}}_k(\cdot)$ denotes the conditional expectation given $\{\boldsymbol{\alpha}_j, j = 1, \ldots, k-1, k+1, \ldots, p\}$, and the second '$=$' follows from the fact that

$$\mathsf{E}|\varepsilon_k|^2 = \mathsf{E}|\varepsilon_k - \mathsf{E}\varepsilon_k|^2 + |\mathsf{E}\varepsilon_k|^2, \tag{2.22}$$

in more detail, we have

$$\begin{aligned}
\mathsf{E}|\varepsilon_k^2| &= \mathsf{E}(\widetilde{\mathsf{E}}_k|\varepsilon_k|^2) \\
&= \mathsf{E}\left(\widetilde{\mathsf{E}}_k|\varepsilon_k - \widetilde{\mathsf{E}}_k\varepsilon_k|^2 + |\widetilde{\mathsf{E}}_k\varepsilon_k|^2\right) \\
&= \mathsf{E}\,\widetilde{\mathsf{E}}_k|\varepsilon_k - \widetilde{\mathsf{E}}_k\varepsilon_k|^2 + \mathsf{E}|\widetilde{\mathsf{E}}_k\varepsilon_k - \mathsf{E}\varepsilon_k|^2 + |\mathsf{E}\varepsilon_k|^2,
\end{aligned}$$

here we have used (2.22) twice and the fact that $\mathsf{E}(\widetilde{\mathsf{E}}_k\varepsilon_k) = \mathsf{E}\varepsilon_k$.

In the estimation of J_1, we have proved that

$$|\mathsf{E}\varepsilon_k| \le \frac{1}{n} + \frac{|z|}{nv} \to 0.$$

Write $\mathbf{A} = \mathbf{I}_n - \frac{1}{n}\mathbf{X}_k^{\mathsf{H}}(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}} - z\mathbf{I}_{p-1})^{-1}\mathbf{X}_k$. Note that $\mathbf{A}$ is indenpendent of $\boldsymbol{\alpha}_k$. Then, we have

$$\frac{1}{n^2}\boldsymbol{\alpha}_k^\top \mathbf{X}_k^{\mathsf{H}}\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}} - z\mathbf{I}_{p-1}\right)^{-1}\mathbf{X}_k\overline{\boldsymbol{\alpha}}_k = \frac{1}{n}\boldsymbol{\alpha}_k^\top(\mathbf{I}_n - \mathbf{A})\overline{\boldsymbol{\alpha}}_k = \frac{1}{n}\boldsymbol{\alpha}_k^\top\overline{\boldsymbol{\alpha}}_k - \frac{1}{n}\boldsymbol{\alpha}_k^\top\mathbf{A}\overline{\boldsymbol{\alpha}}_k,$$

and hence

$$\varepsilon_k = -1 + \frac{1}{n}\boldsymbol{\alpha}_k^\top\mathbf{A}\overline{\boldsymbol{\alpha}}_k + y_n + y_n z\mathsf{E}s_n(z).$$

Then, we have

$$
\begin{aligned}
\varepsilon_k - \widetilde{\mathsf{E}}_k \varepsilon_k &= \frac{1}{n} \boldsymbol{\alpha}_k^\top \mathbf{A} \overline{\boldsymbol{\alpha}}_k - \frac{1}{n} \widetilde{\mathsf{E}}_k \, \boldsymbol{\alpha}_k^\top \mathbf{A} \overline{\boldsymbol{\alpha}}_k \\
&= \frac{1}{n} \boldsymbol{\alpha}_k^\top \mathbf{A} \overline{\boldsymbol{\alpha}}_k - \frac{1}{n} \sum_{i,j} a_{ij} \widetilde{\mathsf{E}}_k (x_{ki} \overline{x}_{kj}) \\
&= \frac{1}{n} \sum_{i,j} a_{ij} x_{ki} \overline{x}_{kj} - \frac{1}{n} \mathsf{tr}(\mathbf{A}) \\
&= \frac{1}{n} \left[\sum_{i=1}^{n} a_{ii} (|x_{ki}|^2 - 1) + \sum_{i \neq j} a_{ij} x_{ki} \overline{x}_{kj} \right].
\end{aligned} \tag{2.23}
$$

Note that

$$
a_{ij} x_{ki} \overline{x}_{kj} \times \overline{a}_{ij} \overline{x}_{ki} x_{kj} = |a_{ij}|^2 |x_{ki}|^2 |x_{kj}|^2 \qquad \text{and} \qquad a_{ij} x_{ki} \overline{x}_{kj} \times \overline{a}_{ji} \overline{x}_{kj} x_{ki} = a_{ij}^2 x_{ki}^2 \overline{x}_{kj}^2.
$$

Then

$$
\begin{aligned}
\widetilde{\mathsf{E}}_k |\varepsilon_k - \widetilde{\mathsf{E}}_k \varepsilon_k|^2 &= \frac{1}{n^2} \left(\sum_{i=1}^{n} |a_{ii}|^2 \mathsf{E}(|x_{ki}|^2 - 1)^2 + \sum_{i \neq j} |a_{ij}|^2 \mathsf{E}|x_{ki}|^2 \mathsf{E}|x_{kj}|^2 + \sum_{i \neq j} a_{ij}^2 \mathsf{E} x_{ki}^2 \mathsf{E} \overline{x}_{kj}^2 \right) \\
&= \frac{1}{n^2} \left(\sum_{i=1}^{n} |a_{ii}|^2 (\mathsf{E}|x_{ki}|^4 - 1) + \sum_{i \neq j} |a_{ij}|^2 \mathsf{E}|x_{ki}|^2 \mathsf{E}|x_{kj}|^2 + \sum_{i \neq j} a_{ij}^2 \mathsf{E} x_{ki}^2 \mathsf{E} \overline{x}_{kj}^2 \right) \\
&= \frac{1}{n^2} \left[\sum_{i=1}^{n} |a_{ii}|^2 (\mathsf{E}|x_{ki}|^4 - 1) + \sum_{i \neq j} |a_{ij}|^2 + \Re \left(\sum_{i \neq j} a_{ij}^2 \mathsf{E} x_{ki}^2 \mathsf{E} \overline{x}_{kj}^2 \right) \right] \\
&\leq \frac{1}{n^2} \left(\sum_{i=1}^{n} |a_{ii}|^2 (\eta_n^2 n) + 2 \sum_{i \neq j} |a_{ij}|^2 \right) && ? \\
&\leq \frac{\eta_n^2}{v^2} + \frac{2}{n v^2}. && ?
\end{aligned}
$$

Here, have used the fact that $|a_{ii}| < v^{-1}$. ?

Using the martingale decomposition method in the proof of (2.9), we can show that

$$
\begin{aligned}
|\widetilde{\mathsf{E}}_k \varepsilon_k - \mathsf{E} \varepsilon_k|^2 &= \left| \frac{1}{n} \widetilde{\mathsf{E}}_k \boldsymbol{\alpha}_k^\top \mathbf{A} \overline{\boldsymbol{\alpha}}_k - \frac{1}{n} \mathsf{E} \boldsymbol{\alpha}_k^\top \mathbf{A} \overline{\boldsymbol{\alpha}}_k \right|^2 \\
&= \frac{1}{n^2} |\mathsf{tr}(\mathbf{A}) - \mathsf{E}\mathsf{tr}(\mathbf{A})|^2 \qquad [(2.21) \ \& \ (2.23)] \\
&= \frac{|z|^2}{n^2} \left| \mathsf{tr} \left(\frac{1}{n} \mathbf{X}_k \mathbf{X}_k^{\mathsf{H}} - z \mathbf{I}_{p-1} \right)^{-1} - \mathsf{E}\mathsf{tr} \left(\frac{1}{n} \mathbf{X}_k \mathbf{X}_k^{\mathsf{H}} - z \mathbf{I}_{p-1} \right)^{-1} \right|^2 \\
&\leq \frac{|z|^2}{n^2 v^2} \to 0. \qquad \text{(Martingale decomposition method)}
\end{aligned}
$$

Combining the three estimations above, we have completed the proof of the mean convergence of the Stieltjes transform of the ESD of $\mathbf{S}_n$.

Consequently, Theorem 2.1.2 is proved by the method of Stieltjes trandforms.

2.3 M-P Law by the Moment Method

2.3.1 Moments of the M-P Law

To use moment method, the explicity form of $k-th$ moment $\beta_k = \beta_k\left(y, \sigma^2\right) = \int_a^b x^k p_y(x)dx$ need to be deduced firstly. Since $\beta_k\left(y, \sigma^2\right) = \sigma^{2k}\beta_k(y, 1)$, we need only compute β_k for the standard M-P Law.

Lemma 2.3.1. *We have*

$$\beta_k = \sum_{r=0}^{k-1} \frac{1}{r+1} \binom{k}{r} \binom{k-1}{r} y^r.$$

Proof. By definition,

$$\begin{aligned}
\beta_k &= \frac{1}{2\pi y} \int_a^b x^{k-1} \sqrt{(b-x)(x-a)}dx \\
&= \frac{1}{2\pi y} \int_{-2\sqrt{y}}^{2\sqrt{y}} (1+y+z)^{k-1} \sqrt{4y-z^2}dz \quad (\text{ with } \quad x = 1+y+z) \\
&= \frac{1}{2\pi y} \sum_{\ell=0}^{k-1} \binom{k-1}{\ell} (1+y)^{k-1-\ell} \int_{-2\sqrt{y}}^{2\sqrt{y}} z^\ell \sqrt{4y-z^2}dz \\
&= \frac{1}{2\pi y} \sum_{\ell=0}^{[(k-1)/2]} \binom{k-1}{2\ell} (1+y)^{k-1-2\ell} (4y)^{\ell+1} \int_{-1}^{1} u^{2\ell} \sqrt{1-u^2}du \\
&= \frac{1}{2\pi y} \sum_{\ell=0}^{[(k-1)/2]} \binom{k-1}{2\ell} (1+y)^{k-1-2\ell} (4y)^{\ell+1} \int_{0}^{1} w^{\ell-1/2} \sqrt{1-w}dw \\
&= \sum_{\ell=0}^{[(k-1)/2]} \frac{(k-1)!}{\ell!(\ell+1)!(k-1-2\ell)!} y^\ell (1+y)^{k-1-2\ell} \\
&= \sum_{\ell=0}^{[(k-1)/2]} \sum_{s=0}^{k-1-2\ell} \frac{(k-1)!}{\ell!(\ell+1)!s!(k-1-2\ell-s)!} y^{\ell+s} \\
&= \sum_{\ell=0}^{[(k-1)/2]} \sum_{r=\ell}^{k-1-\ell} \frac{(k-1)!}{\ell!(\ell+1)!(r-\ell)!(k-1-r-\ell)!} y^r \\
&= \frac{1}{k} \sum_{r=0}^{k-1} \binom{k}{r} y^r \sum_{\ell=0}^{\min(r, k-1-r)} \binom{r}{\ell} \binom{k-r}{k-r-\ell-1} \qquad (2.24) \\
&= \frac{1}{k} \sum_{r=0}^{k-1} \binom{k}{r} \binom{k}{r+1} y^r = \sum_{r=0}^{k-1} \frac{1}{r+1} \binom{k}{r} \binom{k-1}{r} y^r.
\end{aligned}$$

The tricky of exchanging the order of summation in 2.24 is analogic to the problem of exchanging the order of integral in the Riemann Integral, it can be shown in 2.3.

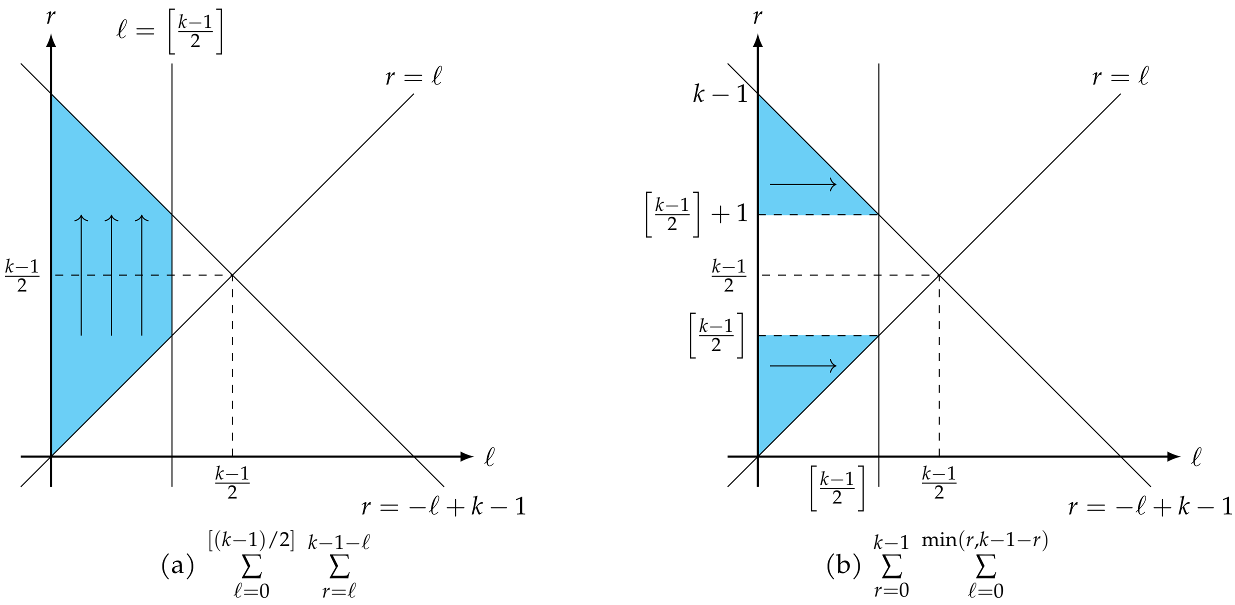


Figure 2.3: Change the order of summation

□

By definition, we have

$$\beta_{2k} = \frac{1}{2\pi y}\int_b^a x^{2k-1}\sqrt{(b-x)(x-a)}dx \;\; \leq \frac{1}{2\pi y}\int_a^b x^{2k-1}\frac{b-a}{2}dx \leq \frac{b^{2k}}{2k\pi\sqrt{y}},$$

implies that $\beta_{2k} \leq b^{2k}$ for large k.Thus,it's easy to see the Carleman condition is satistied and we can use 1.2.1 to derive the MP-Law.

2.3.2 Some lemmas on Graph Theory and Combinatorics

Definition 2.3.2. *Suppose that $i_1, i_2, \ldots, i_k$ are k positive integers (not necessarily distinct) not greater than p and $j_1, j_2, \ldots, j_k$ are k positive integers (not necessarily distinct) not larger than n. We draw two paralle lines: I line and J line, and plot $i_1, i_2, \ldots, i_k$ on I line, plot $j_1, j_2, \ldots, j_k$ on J line. Then, we draw k down edges from vertices i_u to $j_u, u = 1, 2, \ldots, k$, and k up edges from vertices j_u to $i_{u+1}, u = 1, 2, \ldots, k$(with the convention $i_{k+1} = i_1$). By this process, we get a Δ-graph and this graph is denoted by $G(\mathbf{i}, \mathbf{j})$. An example of a Δ-graph is shown in Figure 2.4.*

Remark 2.3.3. *Two graphs are said to be isomorphic is one becomes the other by a suitable permutation on $(1, 2, \ldots, p)$ and a suitable permutation on $(1, 2, \ldots, n)$. The following two graphs G and G' in Figure 2.5 are isomorphic through pemutation (1) and (12) with respect to J line and I line.*

Definition 2.3.4. *We say a Δ-graph is **canonical**, if it satisties*

1. $i_1 = j_1 = 1$;
2. $i_u \leq max\{i_1, \ldots, i_{u-1}\} + 1$ and $j_u \leq max\{j_1, \ldots, j_{u-1}\} + 1$.

Figure 2.4: A Δ-graph.

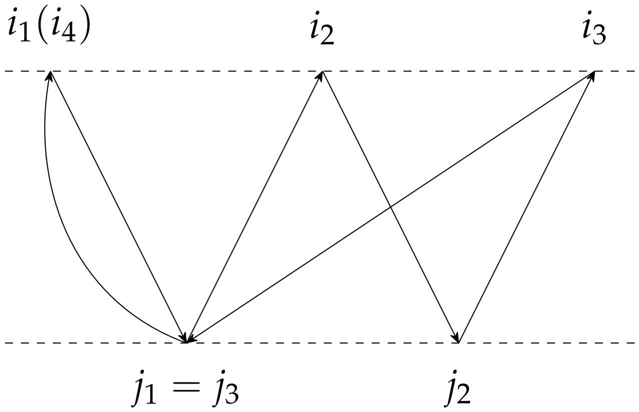
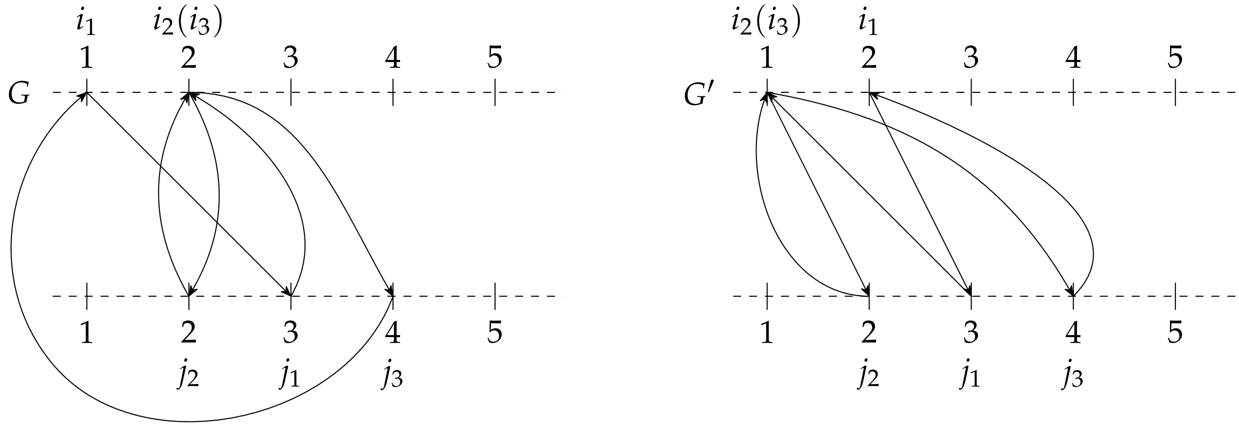


Figure 2.5: Two isomorphic graphs.



Remark 2.3.5. *Note that any permutations on I line or J line will make a jump in the graph, which betrays the second condtion of canonical graph. Therefore, for each isomorphism class, there is only one canonical graph.*

A canonical Δ-graph $G(\mathbf{i}, \mathbf{j})$ is denoted by $\Delta(k, r, s)$ if G has $r+1$ noncoincident I-vertices and s noncoincident J-vertices. It is obviously that there is only one graph in $\Delta(k, k-1, k)$. Moreover, we have the following:

1. Its vertex set $V = V_J + V_I$, where the I-vertices $V_I = 1, \ldots, r+1$ and the J-vertices $V_J = 1, \ldots, s$.
2. There are two functions, $f : \{1, \ldots, k\} \mapsto \{1, \ldots, r+1\}$ and $g : \{1, \ldots, k\} \mapsto \{1, \ldots, s\}$, satisfying
$$\begin{aligned} &f(1) = 1 = g(1) = f(k+1), \\ &f(i) \leq max\{f(1), \ldots, f(i-1)\} + 1, \\ &g(j) \leq max\{g(1), \ldots, g(i-1)\} + 1. \end{aligned}$$

Remark 2.3.6. *We can see f and g as two maps from the number of vertix to its coordinate. And we have edge set $E = \{e_{1d}, e_{1u}, \cdots, e_{kd}, e_{ku}\}$, where $e_{1d}, \cdots, e_{kd}$ are called **down edges** and $e_{1u}, \cdots, e_{ku}$ are called **up edges**.*

Definition 2.3.7. *If $f(j+1) = \max\{f(1), \cdots, f(j)\} + 1$, the dege $e_{j,u}$ is called an **up innovation**, and in the case where $g(j) = \max\{g(1), \cdots, g(j-1)\} + 1$ the edge $e_{j,d}$ is called a **down innovation**.*

Remark 2.3.8. *Intuitively, an up innovation leads to a new I-vertex and a down innovation leads to a new J-vertex. We make the convention that the first down edge is a down innovation and the last up edge is not an innovation.*

Similar to Chapter2, we may need to compute a sophisticated summation of expectation in latter section. To determine the number of terms in the summation, we will divide one summation into three summations, and each summation is corresponding to a class of $\Delta(k, r, s)$-graph. Thus, we classify $\Delta(k, r, s)$-graph into three categories:

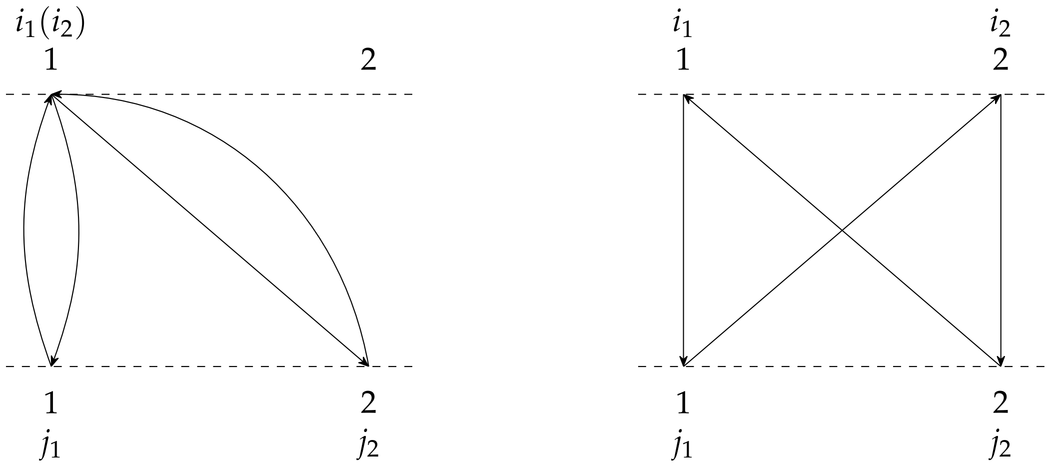
Category 1 (denoted by $\Delta_1(k, r)$): Δ-graphs in which each down edge must coincide with one and only one up edge. And if we glue the conincident deges the resulting graph is a tree of k edges. An example is given in 2.6(a).

Category 2 (denoted by $\Delta_2(k, r, s)$): Δ-graph that contain at least one single edge. An example is given in 2.6(b).

Catagory 3 (denoted by $\Delta_3(k, r, s)$): Δ-graphs that do not belong to $\Delta_1(k, r)$ and $\Delta_2(k, r, s)$.

Remark 2.3.9. *For a given Δ_1-graph, if we glue the conincident edges, the resulting graph is a tree and contains $r+s+1$ vertices and $r+s$ edges [3]. Thus, $k = r+s$ and s is suppressed for simplicity.*

The number of graphs in each isomorphism class for a given canonical $\Delta(k, r, s)$ is given by the following lemma.



(a) A $\Delta_1(k,r)$-graph.

(b) A $\Delta_2(k,r,s)$-graph.

Figure 2.6: $\Delta_1(k,r)$-graph and $\Delta_2(k,r,s)$-graph.

Lemma 2.3.10. *For a given k, r, and s, the number of graphs in the isomorphism class for each canonical $\Delta(k,r,s)$-graph is*

$$p(p-1)\cdots(p-r)n(n-1)\cdots(n-s+1) = p^{r+1}n^s\left[1+O\left(n^{-1}\right)\right].$$

Proof. For given k,r,s, let $G_1 \in \Delta(k,r,s)$. Thus, G_1 has $r+1$ I-vertices and s J-vertices. Since the isomorphism class for G_1 could be generated by permuting the I-vertices and J-vertices of G_1 over I line and J line, respectively. Thus, we only need to choose ordered $r+1$ positions from p coordinates(I line) with no repetitions allowed and choose ordered s positions from n coordinates(J line) with no repetitions allowed. Therefore, the number of graphs in the isomorphism class for each canonical $\Delta(k,r,s)$-graph is

$$p(p-1)\cdots(p-r)n(n-1)\cdots(n-s+1).$$

□

Remark 2.3.11. *Firstly, we can not use $\binom{p}{r+1}$ or $\binom{n}{k}$, since $i_1=1, i_2=2$ and $i_1=2, i_2=1$ are two diferent kinds of cases. Secondly, G_1 does not generated all $\Delta-$graph with $r+1$ I-vertices and s J-vertices, since different canonical graphs have different patterns.*

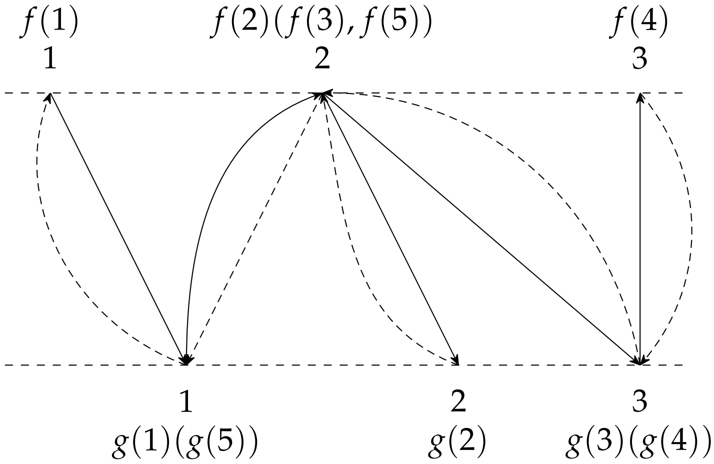
For a Δ_3-graph, we have the following lemma.

Lemma 2.3.12. *For a given $\Delta_3(k,r,s)$-graph we have $k \geq r+s$.*

Proof. Let G be a graph of $\Delta_3(k,r,s)$. Since G is not in category 2, it does not contain single edges and hence the number of noncoincident edges is not larger than k. Note that noncoincident edges of G forms a connected graph $\tilde{G}$ in undirected sense. Thus, the number of edges of $\tilde{G}$ is larger than or equal to the numbers of vertices of $\tilde{G}$ [3], that is, $k \geq E\{\tilde{G}\} \geq r+1+s-1 = r+s$. □

The next lemma is more difficult to convey.

Figure 2.7: Definition of (u, d) sequence.



Lemma 2.3.13. *For k and r, the number of $\Delta_1(k, r)$-graph is*

$$\frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}.$$

Proof. Define two characteristic sequences $\{u_1, \cdots, u_k\}$ and $\{d_1, \cdots, d_k\}$ of a graph $G \in \Delta_1(k, r)$ by

$$u_\ell = \begin{cases} 1, & \text{if } f(\ell+1) = \max\{f(1), \cdots, f(\ell)\} + 1 \\ 0, & \text{otherwise,} \end{cases}$$

and

$$d_\ell = \begin{cases} -1, & \text{if } f(\ell) \notin \{1, f(\ell+1), \cdots, f(k)\} \\ 0, & \text{otherwise.} \end{cases}$$

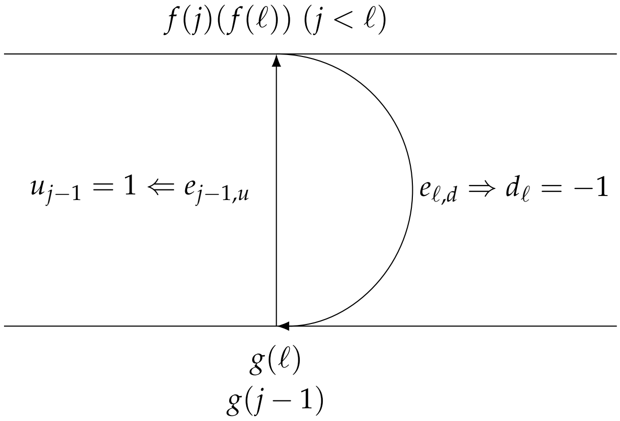
An example with $r = 2$ and $s = 3$ is given in Figure2.7. And we give some interpretions:

1. $u_l = 1$ if and only if $e_{l,u}$ is an up innovation, thus $u_k = 0$;

2. $d_l = -1$ if and only if $e_{l,d}$ conincides with an up innovation, thus $d_1 = 0$;
   If for some l, we have $d_l = -1$, then there won't be another $I-$vertices after i_l come back to $f(l)$. Since for every down edge in $Delta_1(k, r)-$graph there must exists one and only one up edge conincides with it, we know that the $\Delta_1(k, r)-$graph should contain a structure like Fig.2.8.

3. $d_l = 0$ indicates that $(f(l), g(l)) = e_{l,d}$ is a down innovation.
   Since there are r noncoincident vertices, except 1, on I line. Thus there are r up innovations, which means $\sum_l u_l = r$. We will show the number of $d_l = -1$ is equal to the number of $u_l = 1$. If $d_l = -1$, then there exists a $j < l$, s.t. $u_j = 1$, hence, we have $\#\{l | d_l = -1\} \le \#\{l | u_l = 1\}$. On the other hand, if $u_l = 1$, then there exists exactly one down edge $e_{m,d}(m > l)$ coincides with $e_{l,u}$.(Since G is belong to Δ_1-graph.) Then, $d_m = 1$, which implies that $\#\{l | d_l = -1\} \ge \#\{l | u_l = 1\}$. Therefore, we proved that $\sum_l u_l = -\sum_l d_l = r$. If $d_l = -1$, the graph must contain a structure like Fig.2.8, thus it's impossible for $e_{l,u}$ to be a down innovation. But there are $s = k - r$ noncoincident

Figure 2.8: A special structure in $\Delta_1(k,r)$-graph when $d_l = -1$.



J-vertices need to generate, thus all of these noncoincident J-vertices are generated by the rest of $k-r$ down edges. Here,we complete the proof of 3.

From the argument above, one sees that $d_l = -1$ must follows a $u_j = 1$ for some $j < l$. Therefore, the two sequences shouls satisfy the restriction

$$u_1 + \cdots + u_{\ell-1} + d_2 + \cdots + d_\ell \geq 0, \quad \ell = 2, \cdots, k. \tag{2.25}$$

Next we will prove that if two sequences of number satisties 2.25, then a $\Delta_1(k,r)$-graph could be determined uniquely.

At first, we notice that $u_l = 1$ implies that $e_{l,u}$ is an up innovation and thus

$$f(l+1) = 1 + \#\{j \leq l, u_j = 1\}.$$

Similarly, $d_l = 0$ implies that $e_{l,d}$ is a down innovation and thus

$$g(l) = \#\{j \leq l, d_j = 0\}.$$

However, it is not easy to define the values of f and g at other points. We will directly create the $\Delta_1(k,r)$-graph from these two characteristic sequences by ploting every pair of down-up edges.

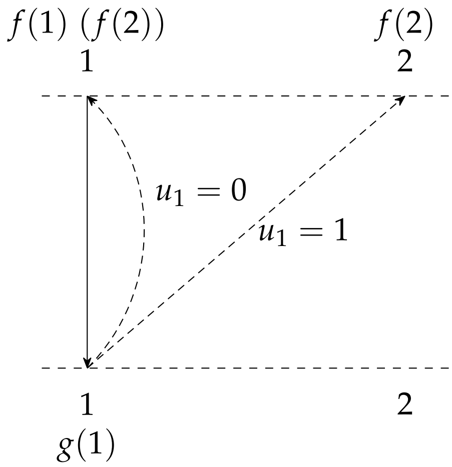
Firstly, it easy to see that the first pair of down-up edges are uniquely determined by u_1 and d_1. We only need to consider the cases of $u_1 = 0$ and $u_1 = 1$. See Fig. 2.3.2.

Suppose that the first l pairs of the down and up edges are uniquely determined by the sequence $\{u_1, u_2, \ldots, u_l\}$ and $\{d_1, \ldots, d_l\}$. Also ,suppose that the subgraph G_l of the first l pairs of down-up edges satisfies the following properties

1. G_l is connected, and the undirectional noncoincident edges of G_l form a tree.

2. If the end vertex $f(l+1)$ of e_l, u is the I-vertex 1, then each down edge of G_l conincides with an up edge of G_l. Thus, G_l does not have single innovations.
   If the end vertex $f(l+1)$ of e_l, u is not the $I-$vertex 1, then from the $I-$vertex 1 to the

Figure 2.9: First pair of down-up edges are uniquely determined by (u_1, d_1).



I-vertex $f(l+1)$ there is only one path (chain without cycles) of down-up-down-up single innovations and all other down edges coincide with an up edge.

We only need to show that the $(l+1)-$st pair of down-up edges will also satisfies these two properties. We consider the following four cases, then let $l = k$ we can see $\{u_l, d_l\}$ must determine a $\Delta_1(k, r)-$graph one the ground that $f(k+1) = 1$.

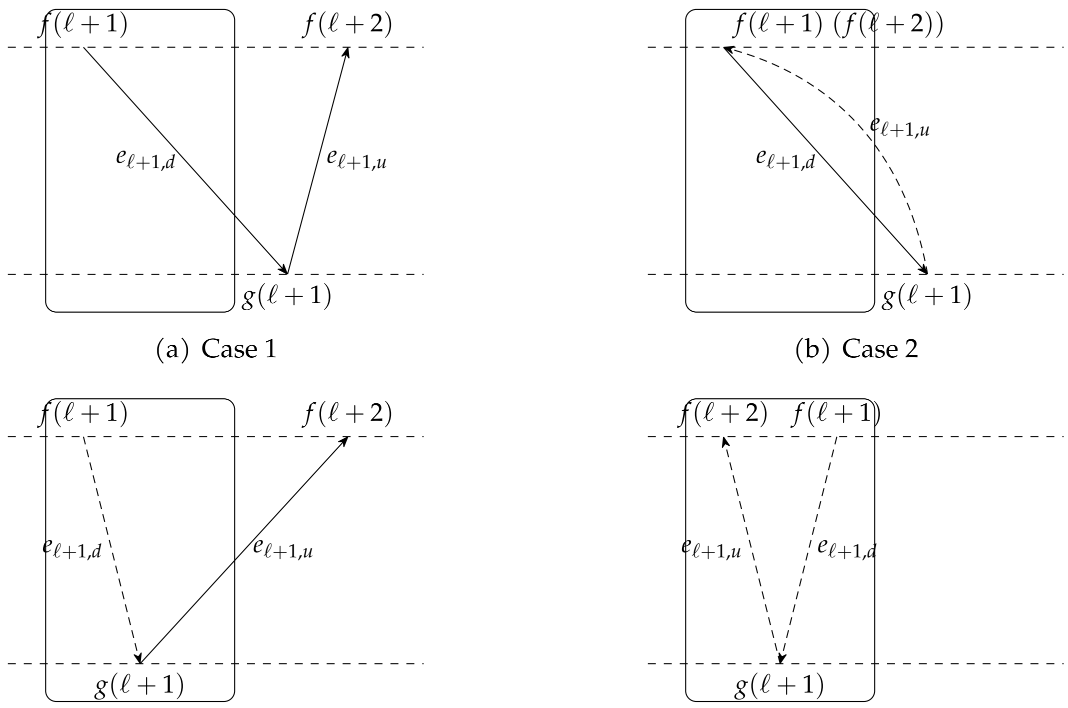
Case 1. $d_{l+1} = 0$ and $u_{l+1} = 1$. Then both edges of the $(l+1)-$st pair are innovations. Thus, we only need to add two innovations to G_l. And the down-up-down-up single innovations path will be these two innovations(if $f(l+1) = 1$) or the original path of single inovations and these two new innovations(if $f(l+1) \neq 1$). See Fig 2.10(a).

Case 2. $d_{l+1} = 0$ and $u_{l+1} = 0$. Then, $e_l + 1, d$ di a down innovation and $e_l + 1, u$ is not an up innovation. Let $e_{l+1,u}$ conincidew with $e_{l+1,d}$. See Case 2. in Fig 2.10(b). Thus, if $f(l+1) = f(l+2) = 1$ the first point in the second property will be met. If $f(l+1) = f(l+2) \neq 1$, the down-up-down-up single innovations path is exactly the same as the original path in G_l.

Case 3. $d_{l+1} = -1$ and $u_{l+1} = 1$. In this case, $e_{l+1,d}$ will coincide with an up innovation and $e_{l+1,u}$ will be an up innovation.by 2.25 we have

$$u_1 + \cdots + u_\ell + d_2 + \cdots + d_\ell \geq 1$$

which implies that the total number of $I-$vertices of G_l(i.e. $u_1 + \cdots + u_\ell$) other than 1 is greater than the number of $I-$vertices of G_l from which the graph ultimatly leaves(i.e. $d_2 + \cdots + d_\ell$). Thus G_l must contain single up-innovations. Therefore, $f(l+1) \neq 1$ by the first point in propery 2. As there must be a single up innovation leading to the end vertex $f(l+1)$, we can draw the down edge $e_{l+1,d}$ conincident with this single up innovation. Then, draw $e_{l+1,u}$ as the next innovation from the vertex $g(l+1)$. See Case 3. in Fig 2.10(c). And it is easy to see that the new down-up-down-up single innovations path is the original one with the last up innovation replaced by $e_{l+1,u}$.



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Figure 2.10: Examples of the four cases. In the four graphs, the rectangle denotes the subgraph G_ℓ, solid arrows are new innovations, and broken arrows are new T_3 edges.

Case 4. $d_{l+1} = -1$ and $u_{l+1} = 0$.

By induction and let $l = k$, it is By induction, it is shown that two sequences subject to restriction 2.25 uniquely determine a $\Delta_1(k,r)-$graph. Therefore,counting the number of $\Delta_1(k,r)-$graph is equivalent to counting the number of pairs of characteristic sequences.

Now, we count the number of characteristic sequences for given k and r. Ignoring the restriction 2.25, we have $\left(\begin{array}{c} k-1 \\ r \end{array}\right)\left(\begin{array}{c} k-1 \\ r \end{array}\right)$ ways to arrange r ones in the $k-1$ positions $u_1, \cdots, u_{k-1}$ and to arrange r minus ones in the $k-1$ positions $d_2, \cdots, d_k$. If there is an interger $2 \le l \le k$such that

$$u_1 + \cdots + u_{\ell-1} + d_1 + \cdots + d_\ell = -1,$$

we define a one-to-one transform,

$$\tilde{u}_j = \begin{cases} u_j, & \text{if } j < \ell \\ -d_{j+1}, & \text{if } \ell \le j < k, \end{cases}$$

and

$$\tilde{d}_j = \begin{cases} d_j, & \text{if } 1 < j \le \ell \\ -u_{j-1}, & \text{if } \ell < j \le k. \end{cases}$$

Then we have $r-1$ $\tilde{u}$'s equal to one and $r+1$ $\tilde{d}$'s equal to minus one. There are $\begin{pmatrix} k-1 \\ r-1 \end{pmatrix}\begin{pmatrix} k-1 \\ r+1 \end{pmatrix}$ ways to arrange $r-1$ ones in the $k-1$ positions $\tilde{u}_1, \cdots, \tilde{u}_{k-1}$, and to arrange $r+1$ minus ones in the $k-1$ positions $\tilde{d}_2, \cdots, \tilde{d}_k$.

Therefore, the numer of paris of characteristic sequences with indices k and r satistying 2.25 is

$$\begin{pmatrix} k-1 \\ r \end{pmatrix}^2 - \begin{pmatrix} k-1 \\ r-1 \end{pmatrix}\begin{pmatrix} k-1 \\ r+1 \end{pmatrix} = \frac{1}{r+1}\begin{pmatrix} k \\ r \end{pmatrix}\begin{pmatrix} k-1 \\ r. \end{pmatrix}$$

Then we have $r-1$ Here we complete the proof. □

2.3.3 M-P Law for the iid Case

We shall give a proof of the following theorem

Theorem 2.3.14. *Suppose that $\{x_{ij}\}$ are iid complex random varianbles with variance σ^2.Also assume that $p/n \to y \in (0, \infty)$. Then, with probablity one, F^S tends to the M-P Law.*

By using the same technique in chapter 2, we can assume that the variables x_{jk} are uniformly bounded with mean zero and variance 1.The process will be omitted here, more details could look [1]. Firstly, we have

$$\begin{aligned}
\beta_k(\mathbf{S}_n) &= \int x^k F^{\mathbf{S}_n}(dx) \\
&= \frac{1}{p}\mathrm{tr}(\mathbf{S}_n^k) \\
&= \frac{1}{pn^k}\mathrm{tr}((\mathbf{X}\mathbf{X}^H)^k).
\end{aligned}$$

To derive $\mathrm{tr}((\mathbf{X}\mathbf{X}^H)^k)$, we consider

$$\begin{aligned}
(\mathbf{X}\mathbf{X}^H)_{i_1 i_1} &= \sum_{i_2=1}^{n} x_{i_1 i_2}\overline{x}_{i_1 i_2} \\
[(\mathbf{X}\mathbf{X}^H)^2]_{i_1 i_1} &= (\mathbf{X}\mathbf{X}^H)_{i_1 \cdot}(\mathbf{X}\mathbf{X}^H)_{\cdot i_1}; \\
&= (\sum_{i_2=1}^{n} x_{i_1 i_2}\overline{x}_{1 i_2}, \sum_{i_2=1}^{n} x_{i_1 i_2}\overline{x}_{2 i_2}, \ldots, \sum_{i_2=1}^{n} x_{i_1 i_2}\overline{x}_{p i_2})(\sum_{i_2=1}^{n} \overline{x}_{i_1 i_2}x_{1 i_2}, \sum_{i_2=1}^{n} \overline{x}_{i_1 i_2}x_{2 i_2}, \ldots, \sum_{i_2=1}^{n} \overline{x}_{i_1 i_2}x_{p i_2})^T \\
&= \sum_{i_2}\sum_{j_1, j_2} x_{i_1 j_1}\overline{x}_{i_2 j_1}x_{i_2 j_2}\overline{x}_{i_1 j_2}; \\
[(\mathbf{X}\mathbf{X}^H)^3]_{i_1 i_1} &= \sum_{i_3}(\mathbf{X}\mathbf{X}^H)^2_{i_1 i_3}(\mathbf{X}\mathbf{X}^H)_{i_1 i_3} \\
&= \sum_{i_2, i_3}\sum_{j_1, j_2, j_3} x_{i_1 j_1}\overline{x}_{i_2 j_1}x_{i_2 j_2}\overline{x}_{i_3 j_2}x_{i_3 j_3}\overline{x}_{i_1 j_3}.
\end{aligned}$$

Thus, by elmentary calculus, we have

$$\begin{aligned}\beta_k(\mathbf{S}_n) &= p^{-1}n^{-k}\sum_{\{i_1,\cdots,i_k\}}\sum_{\{j_1,\cdots,j_k\}} x_{i_1j_1}\overline{x}_{i_2j_1}x_{i_2j_2}\cdots x_{i_kj_k}\overline{x}_{i_1j_k} \\ &= p^{-1}n^{-k}\sum_{\mathrm{i,j}}\mathbf{X}_{G(\mathrm{i,j})},\end{aligned} \tag{2.26}$$

where the summation runs over all $G(\mathbf{i},\mathbf{j})$-graphs, the indices in $\mathbf{i} = (i_1,\cdots,i_k)$ run over $1,2,\ldots,p$, and the indices in $\mathbf{j} = (j_1,\cdots,j_k)$ run over $1,2,\ldots,n$. To complete the proof of the almost sure convergence of the ESD of $\mathbf{S}_n$ we need only show the following two assertions:

$$\begin{aligned}\mathsf{E}\left(\beta_k\left(\mathbf{S}_n\right)\right) &= p^{-1}n^{-k}\sum_{\mathbf{i},\mathbf{j}}\mathsf{E}\left(x_{G(\mathbf{i},\mathbf{j})}\right) \\ &= \sum_{r=0}^{k-1}\frac{y_n^r}{r+1}\binom{k}{r}\binom{k-1}{r}+O\left(n^{-1}\right),\end{aligned} \tag{2.27}$$

and

$$\begin{aligned}&\mathrm{Var}\left(\beta_k\left(\mathbf{S}_n\right)\right) \\ &= p^{-2}n^{-2k}\sum_{\mathbf{i}_1,\mathbf{j}_1,\mathbf{i}_2,\mathbf{j}_2}\left[\mathsf{E}\left(x_{G_1(\mathbf{i}_1,\mathbf{j}_1)}x_{G_2(\mathbf{i}_2,\mathbf{j}_2)}-\mathsf{E}\left(x_{G_1(\mathbf{i}_1,\mathbf{j}_1)}\right)\mathsf{E}\left(x_{G_2(\mathbf{i}_2,\mathbf{j}_2)}\right)\right)\right] \\ &= O\left(n^{-2}\right),\end{aligned} \tag{2.28}$$

<u>**The Proof of 2.27**</u>: We claim that on the left hand of 2.27, two terms are equal if their corresponding graphs are isomorphic. It is easy to see that $x_{i_kj_k}$ represents $e_{k,d}$ and $\overline{x}_{i_l,j_{l-1}}$ represents $e_{l-1,u}$ in Δ-graph. And the isomorpism will not change the structure of a graph, thus isomorphic transform will not change the exponent of the following formula:

$$\mathsf{E}|x_{i_1'j_1'}|^{k_1}\mathsf{E}|x_{i_2'j_2'}|^{k_2}\ldots\mathsf{E}|x_{i_m'j_m'}|^{k_m},$$

Therefore, by 2.3.10, we may rewrite

$$\mathsf{E}\left(\beta_k\left(\mathbf{S}_n\right)\right) = p^{-1}n^{-k}\sum_{\Delta(k,r,s)}p(p-1)\cdots(p-r)n(n-1)\cdots(n-s+1)\mathrm{E}\left(X_{\Delta(k,r,s)}\right). \tag{2.29}$$

Now, split the sum in 2.29 into three parts according to $\Delta_1(k,r)$, $\Delta_2(k,r)$ and $\Delta_3(k,r)$. Since the graph in $\Delta_2(k,r,s)$ contains at least one single edge, thus

$$S_2 = p^{-1}n^{-k}\sum_{\Delta_2(k,r,s)}p(p-1)\cdots(p-r)n(n-1)\cdots(n-s+1)\mathrm{E}\left(X_{\Delta_2(k,r,s)}\right) = 0.$$

By 2.3.12, we have $r+s\le k$ for a graph of $\Delta_3(k,r,s)$. And since x_{jk} are uniformly bounded by C, we have

$$\begin{aligned} S_3 &= p^{-1}n^{-k}\sum_{\Delta_3(k,r,s)} p(p-1)\cdots(p-r)n(n-1)\cdots(n-s+1)\mathrm{E}\left(X_{\Delta(k,r,s)}\right) \\ &= \sum_{\Delta_3(k,r,s)} (\frac{p-1}{n})\ldots(\frac{p-r}{n})(1-\frac{1}{n})\ldots(1-\frac{s-1}{n})\frac{C^{2k}}{n^l} \quad (Here\ l\ \geq 0.) \\ &= O\left(n^{-1}\right) \end{aligned}$$

Now let us evaluate S_1. For a graph in $\Delta_1(k,r)$, each pair of coincident edges consists of exactly one down edge and an up edge. Therefore, we have

$$\mathsf{E}\mathbf{X}_{\Delta_1(k,r)} = \mathsf{E}\mathbf{X}^2_{i'_1j'_1}\mathsf{E}\mathbf{X}^2_{i'_2j'_2}\ldots\mathsf{E}\mathbf{X}^2_{i'_kj'_k} = 1.$$

And by 2.3.13,

$$\begin{aligned} S_1 &= p^{-1}n^{-k}\sum_{\Delta_1(k,r)} p(p-1)\cdots(p-r)n(n-1)\cdots(n-s+1)\mathrm{E}\left(X_{\Delta_1(k,r)}\right) \\ &= \sum_{r=0}^{k-1}\frac{y_n^r}{r+1}\left(\begin{array}{c}k\\r\end{array}\right)\left(\begin{array}{c}k-1\\r\end{array}\right)+O\left(n^{-1}\right) \\ &\to \beta_k. \end{aligned}$$

<u>The Proof of 2.28</u>: Recall

$$\begin{aligned} &\operatorname{Var}\left(\beta_k\left(\mathbf{S}_n\right)\right) \\ =&\mathsf{E}|\beta_k(\mathbf{S}_n)|^2-|\mathsf{E}\beta_k(\mathbf{S}_n)|^2 \\ =&p^{-2}n^{-2k}\sum_{\mathbf{i},\mathbf{j}}\left[\mathrm{E}\left(\mathbf{X}_{G_1(\mathbf{i}_1,\mathbf{j}_1)}\mathbf{X}_{G_2(\mathbf{i}_2,\mathbf{j}_2)}\right)-\mathrm{E}\left(\mathbf{X}_{G_1(\mathbf{i}_1,\mathbf{j}_1)}\right)\mathrm{E}\left(\mathbf{X}_{G_2(\mathbf{i}_2,\mathbf{j}_2)}\right)\right]. \end{aligned} \tag{2.30}$$

Here G_i $(i=1,2)$denote two Δ-graph. Note that if G_1 has no edges coincident with edges of G_2 or $G=G_1\cup G_2$ has an single edge, then

$$\mathsf{E}\left(\mathbf{X}_{G_1(\mathbf{i}_1,\mathbf{j}_1)}\mathbf{X}_{G_2(\mathbf{i}_2,\mathbf{j}_2)}\right)-\mathsf{E}\left(\mathbf{X}_{G_1(\mathbf{i}_1,\mathbf{j}_1)}\right)\mathsf{E}\left(\mathbf{X}_{G_2(\mathbf{i}_2,\mathbf{j}_2)}\right)=0$$

by independence between $\mathbf{X}_{G_1}$ and $\mathbf{X}_{G_2}$. On the other hand, if G has no single edge, then, we can see the number of noncoincident edges of G is not more than $2k$. Then, we must have the following expression:

$$\mathsf{E}\left(\mathbf{X}_{G_1(\mathbf{i}_1,\mathbf{j}_1)}\mathbf{X}_{G_2(\mathbf{i}_2,\mathbf{j}_2)}\right)=\mathsf{E}|x_{i'_1j'_1}|^{k_1}\mathsf{E}|x_{i'_2j'_2}|^{k_2}\ldots\mathsf{E}|x_{i'_lj'_l}|^{k_l}$$

here $k_1 + k_2 + \dots k_l = 4k$, $l \le 2k$. And $\mathsf{E}\left(\mathbf{X}_{G_1(\mathrm{i}_1,\mathrm{j}_1)}\right)\mathsf{E}\left(\mathbf{X}_{G_2(\mathrm{i}_2,\mathrm{j}_2)}\right)$ will become

$$\mathsf{E}\left(\mathbf{X}_{G_1(\mathrm{i}_1,\mathrm{j}_1)}\right)\mathsf{E}\left(\mathbf{X}_{G_2(\mathrm{i}_2,\mathrm{j}_2)}\right) = \left(\mathsf{E}|x_{i_1'j_1'}|\right)^{k_1}\left(\mathsf{E}|x_{i_2'j_2'}|\right)^{k_2}\dots\left(\mathsf{E}|x_{i_l'j_l'}|\right)^{k_l},$$

here we still have $k_1 + k_2 + \dots k_l = 4k$, $l \le 2k$. Thus, we have each term in 2.30 is smaller than $2C^{4k}p^{-2}n^{-2k}$. Consequently, we have

$$\begin{aligned}|\operatorname{Var}(\beta_k(\mathbf{S}_n))| &\le \sum_{\mathbf{i},\mathbf{j}} 2C^{4k}n^{-2k}p^{-2} \\ &= \binom{p}{k}k!\binom{n}{k}k!\dot{2}C^{4k}n^{-2k}p^{-2} = O(n^{-2})\end{aligned}$$

Here, we proved 2.3.14.

Bibliography

[1] Zhidong Bai and Jack W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer, second edition, 2010.

[2] Shurong Zheng Jianfeng Yao and Zhidong Bai. *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press, 2015.

[3] Brian Lian. *A first course in discrete mathematics*. Springer Science & Business Media, 2000.

[4] Yan Shijian and Liu Xiufang. *Measure and Probaility* (*in Chinese*). second edition, 2005.

[5] Zhang Xianke. *Advanced Linear Algebra* (*in Chinese*). Higher Education Press, first edition, 2012.