Chapter 2
Wigner Matrices and Semicircular Law

A Wigner matrix is a symmetric (or Hermitian in the complex case) random matrix. Wigner matrices play an important role in nuclear physics and mathematical physics. The reader is referred to Mehta [212] for applications of Wigner matrices to these areas. Here we mention that they also have a strong statistical meaning. Consider the limit of a normalized Wishart matrix. Suppose that $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are iid samples drawn from a p-dimensional multivariate normal population $N(\boldsymbol{\mu}, \mathbf{I}_p)$. Then, the sample covariance matrix is defined as

$$\mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})',$$

where $\overline{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$. When n tends to infinity, $\mathbf{S}_n \to \mathbf{I}_p$ and $\sqrt{n}\,(\mathbf{S}_n - \mathbf{I}_p) \to \sqrt{p}\mathbf{W}_p$. It can be seen that the entries above the main diagonal of $\sqrt{p}\mathbf{W}_p$ are iid $N(0,1)$ and the entries on the diagonal are iid $N(0,2)$. This matrix is called the (standard) Gaussian matrix or Wigner matrix.

A generalized definition of Wigner matrix only requires the matrix to be a Hermitian random matrix whose entries on or above the diagonal are independent. The study of spectral analysis of the large dimensional Wigner matrix dates back to Wigner's [295] famous **semicircular law**. He proved that the expected ESD of an $n \times n$ standard Gaussian matrix, normalized by $1/\sqrt{n}$, tends to the semicircular law F whose density is given by

$$F'(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2}, & \text{if } |x| \le 2, \\ 0, & \text{otherwise.} \end{cases} \tag{2.0.1}$$

This work has been extended in various aspects. Grenander [136] proved that $\|F^{\mathbf{W}_n} - F\| \to 0$ in probability. Further, this result was improved as in the sense of "almost sure" by Arnold [8, 7]. Later on, this result was further generalized, and it will be introduced in the following sections.

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_2,

2.1 Semicircular Law by the Moment Method

In order to apply the moment method (see Appendix B, Section B.1) to prove the convergence of the ESD of Wigner matrices to the semicircular distribution, we calculate the moments of the semicircular distribution and show that they satisfy the Carleman condition. In the remainder of this section, we will show the convergence of the ESD of the Wigner matrix by the moment method.

2.1.1 Moments of the Semicircular Law

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

Lemma 2.1. *For* $k = 0, 1, 2, \cdots,$ *we have*

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

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

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

2.1.2 Some Lemmas in Combinatorics

In order to calculate the limits of moments of the ESD of a Wigner matrix, we need some information from combinatorics. This is because the mean and variance of each empirical moment will be expressed as a sum of expectations of products of matrix entries, and we need to be able to systematically count the number of significant terms. To this end, we introduce some concepts from graph theory and establish some lemmas.

A graph is a triple (E, V, F), where E is the set of edges, V is the set of vertices, and F is a function, $F : E \mapsto V \times V$. If $F(e) = (v_1, v_2)$, the vertices v_1, v_2 are called the ends of the edge e, v_1 is the initial of e, and v_2 is the terminal of e. If $v_1 = v_2$, edge e is a loop. If two edges have the same set of ends, they are said to be coincident.

Let $\mathbf{i} = (i_1, \cdots, i_k)$ be a vector valued on $\{1, \cdots, n\}^k$. With the vector $\mathbf{i}$, we define a Γ-graph as follows. *Draw a horizontal line and plot the numbers* $i_1, \cdots, i_k$ *on it. Consider the distinct numbers as vertices, and draw* k *edges* e_j *from* i_j *to* i_{j+1}, $j = 1, \cdots, k$, *where* $i_{k+1} = i_1$ *by convention. Denote the number of distinct* i_j*'s by* t. Such a graph is called a $\Gamma(k, t)$-graph. An example of $\Gamma(6, 4)$ is shown in Fig. 2.1.

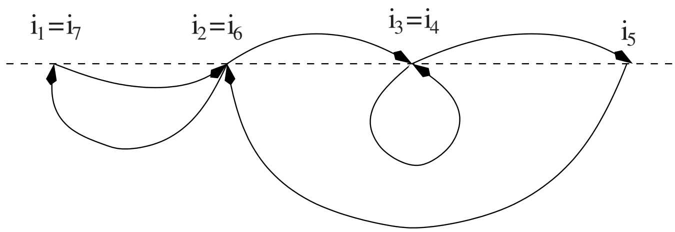

Fig. 2.1 A Γ-graph

By definition, a $\Gamma(k, t)$-graph starts from vertex i_1, and the k edges consecutively connect one after another and finally return to vertex i_1. That is, a $\Gamma(k, t)$-graph forms a cycle.

Two $\Gamma(k, t)$-graphs are said to be isomorphic if one can be converted to the other by a permutation of $(1, \cdots, n)$. By this definition, all Γ-graphs are classified into isomorphism classes.

We shall call the $\Gamma(k, t)$-graph canonical if it has the following properties:

1. Its vertex set is $V = \{1, \cdots, t\}$.
2. Its edge set is $E = \{e_1, \cdots, e_k\}$.
3. There is a function g from $\{1, 2, \cdots, k\}$ onto $\{1, 2, \cdots, t\}$ satisfying $g(1) = 1$ and $g(i) \le \max\{g(1), \cdots, g(i-1)\} + 1$ for $1 < i \le k$.
4. $F(e_i) = (g(i), g(i+1))$, for $i = 1, \cdots, k$, with convention $g(k+1) = g(1) = 1$.

It is easy to see that each isomorphism class contains one and only one canonical Γ-graph that is associated with a function g, and a general graph in this class can be defined by $F(e_j) = (i_{g(j)}, i_{g(j+1)})$. Therefore, we have the following lemma.

Lemma 2.2. *Each isomorphism class contains* $n(n-1)\cdots(n-t+1)$ $\Gamma(k, t)$ *graphs.*

The canonical $\Gamma(k,t)$-graphs can be classified into three categories.

Category 1 (denoted by $\Gamma_1(k)$): A canonical graph $\Gamma(k,t)$ is said to belong to category 1 if each edge is coincident with exactly one other edge of opposite direction and the graph of noncoincident edges forms a tree (i.e., a connected graph without cycles). It is obvious that there is no $\Gamma_1(k)$ if k is odd.

Category 2 ($\Gamma_2(k,t)$) consists of all those canonical $\Gamma(k,t)$-graphs that have at least one single edge; i.e., an edge not coincident with any other edges.

Category 3 ($\Gamma_3(k,t)$) consists of all other canonical $\Gamma(k,t)$-graphs. If we classify the k edges into coincidence classes, a $\Gamma_3(k,t)$-graph contains either a coincidence class of at least three edges or a cycle of noncoincident edges. In both cases, $t \le (k+1)/2$. Then, in fact we have proved the following lemma.

Lemma 2.3. *In a $\Gamma_3(k,t)$-graph, $t \le (k+1)/2$.*

Now, we begin to count the number of $\Gamma_1(k)$-graphs for $k = 2m$. We have the following lemma.

Lemma 2.4. *The number of $\Gamma_1(2m)$-graphs is $\frac{1}{m+1}\binom{2m}{m}$.*

Proof. Suppose G is a graph of $\Gamma_1(2m)$. We define a function $H : E \to \{-1,1\}$; $H(e) = +1$ if e is single up to itself (called an innovation) and $= -1$ otherwise (called a Type 3 (T_3) edge, the edge that coincides with an innovation that is single up to it). Corresponding to the graph G, we call the sequence $(H(e_1),\cdots,H(e_k)) = (a_1 = 1, a_2,\cdots,a_{2m-1}, a_{2m} = -1)$ the characteristic sequence of the graph G. By definition, all partial sums of the characteristic sequence are nonnegative; i.e., for all $1 \le \ell \le 2m$,

$$a_1 + a_2 + \cdots + a_\ell \ge 0. \tag{2.1.1}$$

We show that there is a one-to-one correspondence between $\Gamma_1(2m)$-graphs and the characteristic sequences. That is, we need to show that any sequence of ± 1 satisfying (2.1.1) corresponds to a $\Gamma_1(2m)$-graph. Suppose $(a_1,\cdots,a_{2m})$ is a given sequence satisfying (2.1.1). We construct a $\Gamma_1(2m)$-graph with the given sequence as its characteristic sequence.

By (2.1.1), $a_1 = 1$ and $F(e_1) = (1,2)$; i.e., $g(1) = 1$, $g(2) = 2$. Suppose $g(1), g(2),\cdots,g(s)$ $(2 \le s < 2m)$ have been defined with the following properties:

(i) For each $i \le s$, we have $g(i) \le \max\{g(1),\cdots,g(i-1)\} + 1$.
(ii) If we define $(g(i), g(i+1))$, $i = 1,\cdots,s-1$, as edges, then from $g(1) = 1$ to $g(s)$ there is a path of single innovations if $g(s) \ne 1$. All other edges not on the path must coincide with another edge of opposite direction. If $g(s) = 1$, then each edge coincides with another edge of opposite direction.
(iii) $H(g(i), g(i+1)) = a_i$ for all $i < s$.

Now, we define $g(s+1)$ in the following way:

Case 1. If $a_s = 1$, define $g(s+1) = \max\{g(1), \cdots, g(s)\} + 1$. Obviously, the edge $(g(s), g(s+1))$ is a single innovation that, combining the original path of single innovations, forms the new path of single innovations from $g(1) = 1$ to $g(s+1)$ if $g(s) \neq 1$. If $g(s) = 1$, then $g(s+1) \neq 1$ and the edge $(g(s), g(s+1))$ forms the new path of single innovations. Also, all other edges coincide with an edge of opposite directions. That is, conditions (i)–(iii) are satisfied.

Case 2. If $a_s = -1$, then $g(s) \neq 1$ for otherwise condition (2.1.1) will be violated. Hence, there is an $i < s$ such that $(g(i), g(s))$ is a single innovation (the last edge of a path of single innovations). Then, define $g(s+1) = g(i)$. If $g(i) = 1$, then the new graph has no single edges. If $g(i) \neq 1$, the original path of single innovations has at least two single innovations. Then, the new path of single innovations is obtained by cutting the last edge from the original path of single innovations. Also, conditions (i)–(iii) are satisfied.

By induction, the functions $g(1), \cdots, g(2m)$ are well defined, and hence a $\Gamma_1(2m)$ with characteristic sequence $(a_1, \cdots, a_{2m})$ is defined.

Therefore, to count the number of $\Gamma_1(2m)$-graphs is equivalent to counting the number of characteristic sequences of isomorphism classes.

Arbitrarily arrange m ones and m minus ones. The total number of possibilities is obviously $\binom{2m}{m}$. We shall use the symmetrization principle to count the number of noncharacteristic sequences. Write the sequence of ± 1s as $(a_1, \cdots, a_{2m})$ and $S_0 = 0$ and $S_i = S_{i-1} + a_i$, for $i = 1, 2, \cdots, 2m$. Plot the graph of $(i, S(i))$ on the plane. The graph should start from $(0, 0)$ and return to $(2m, 0)$. If for all i, $S_i \geq 0$ (that is, the figure is totally above or on the horizontal axis), then $(a_1, \cdots, a_{2m})$ is a characteristic sequence. Otherwise, if $(a_1, \cdots, a_{2m})$ is not a characteristic sequence, then there must be an $i \geq 1$ such that $S_i = -1$. Then we turn over the rear part after i along the line $S = -1$ and we get a new graph $(0, 0)$ to $(2m, -2)$, as shown in Fig. 2.2.

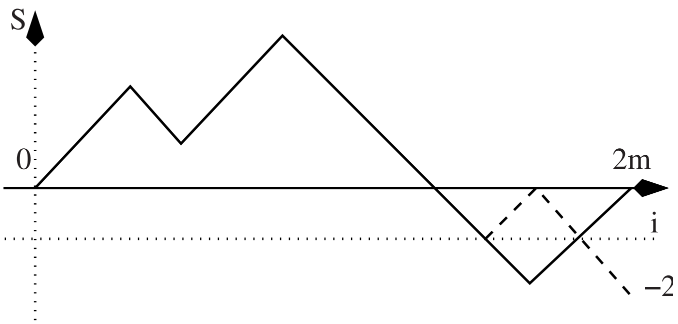

Fig. 2.2 Symmetrization principle

This is equivalent to defining $b_j = a_j$ for $j \leq i$ and $b_j = -a_j$ for $j > i$. Then, the sequence $(b_1, \cdots, b_{2m})$ contains $m - 1$ ones and $m + 1$ minus ones.

Conversely, for any sequence of $m-1$ ones and $m+1$ minus ones, there must be a smallest integer $i < 2m$ such that $b_1 + \cdots + b_i = -1$. Then the sequence $(b_1, \cdots, b_k, -b_{k+1}, \cdots, -b_{2m})$ contains m ones and m minus ones which is a noncharacteristic sequence. The number of b-sequences is $\binom{2m}{m-1}$. Thus, the number of characteristic sequences is

$$\binom{2m}{m} - \binom{2m}{m-1} = \frac{1}{m+1}\binom{2m}{m}.$$

The proof of the lemma is complete.

2.1.3 Semicircular Law for the iid Case

In this subsection, we will show the semicircular law for the iid case; that is, we shall prove the following theorem. For brevity of notation, we shall use $\mathbf{X}_n$ for an $n \times n$ Wigner matrix and save the notation $\mathbf{W}_n$ for the normalized Wigner matrix, i.e., $\frac{1}{\sqrt{n}}\mathbf{X}_n$.

Theorem 2.5. *Suppose that $\mathbf{X}_n$ is an $n \times n$ Hermitian matrix whose diagonal entries are iid real random variables and those above the diagonal are iid complex random variables with variance $\sigma^2 = 1$. Then, with probability 1, the ESD of $\mathbf{W}_n = \frac{1}{\sqrt{n}}\mathbf{X}_n$ tends to the semicircular law.*

Before applying the MCT to the proof of Theorem 2.5, we first remove the diagonal entries of $\mathbf{X}_n$, truncate the off-diagonal entries of the matrix, and renormalize them, without changing the LSD. We will proceed with the proof by taking the following steps.

Step 1. Removing the Diagonal Elements
Let $\widetilde{\mathbf{W}}_n$ be the matrix obtained from $\mathbf{W}_n$ by replacing the diagonal elements with zero. We shall show that the two matrices are asymptotically equivalent; i.e., their LSDs are the same if one of them exists.

Let $N_n = \#\{|x_{ii}| \geq \sqrt[4]{n}\}$. 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 $\widehat{\mathbf{W}}_n$. Then, by Corollary A.41, we have

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

On the other hand, by Theorem A.43, we have

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

Therefore, to complete the proof of our assertion, it suffices to show that $N_n/n \to 0$ almost surely. Write $p_n = \mathrm{P}(|x_{11}| \geq \sqrt[4]{n}) \to 0$. By Bernstein's inequality,[1] we have, for any $\varepsilon > 0$,

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

for some positive constant $b > 0$. This completes the proof of our assertion.

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

Step 2. Truncation

For any fixed positive constant C, truncate the variables at C and write $x_{ij(C)} = x_{ij}I(|x_{ij}| \leq C)$. Define a truncated Wigner matrix $\mathbf{W}_{n(C)}$ whose diagonal elements are zero and off-diagonal elements are $\frac{1}{\sqrt{n}}x_{ij(C)}$. Then, we have the following truncation lemma.

Lemma 2.6. *Suppose that the assumptions of Theorem 2.5 are true. Truncate the off-diagonal elements of $\mathbf{X}_n$ at C, and denote the resulting matrix by $\mathbf{X}_{n(C)}$. Write $\mathbf{W}_{n(C)} = \frac{1}{\sqrt{n}}\mathbf{X}_{n(C)}$. Then, for any fixed constant C,*

$$\limsup_n L^3(F^{\mathbf{W}_n}, F^{\mathbf{W}_{n(C)}}) \leq \mathrm{E}\big(|x_{11}|^2 I(|x_{11}| > C)\big), \quad \text{a.s.} \tag{2.1.2}$$

Proof. By Corollary A.41 and the law of large numbers, we have

$$L^3(F^{\mathbf{W}_n}, F^{\mathbf{W}_{n(C)}}) \leq \frac{2}{n^2}\left(\sum_{1\leq i<j\leq n}|x_{ij}|^2 I(|x_{11}| > C)\right)$$
$$\to \mathrm{E}\big(|x_{11}|^2 I(|x_{11}| > C)\big).$$

This completes the proof of the lemma.

Note that the right-hand side of (2.1.2) can be made arbitrarily small by making C large. Therefore, in the proof of Theorem 2.5, we can assume that the entries of the matrix $\mathbf{X}_n$ are uniformly bounded.

Step 3. Centralization

Applying Theorem A.43, we have

$$\left\| F^{\mathbf{W}_{n(C)}} - F^{\mathbf{W}_{n(C)} - a\mathbf{1}\mathbf{1}'} \right\| \leq \frac{1}{n}, \tag{2.1.3}$$

[1] Bernstein's inequality states that if $X_1, \cdots, X_n$ are independent random variables with mean zero and uniformly bounded by b, then, for any $\varepsilon > 0$,
$P(|S_n| \geq \varepsilon) \leq 2\exp(-\varepsilon^2/[2(B_n^2 + b\varepsilon)])$, where $S_n = X_1 + \cdots + X_n$ and $B_n^2 = \mathrm{E}S_n^2$.

where $a = \frac{1}{\sqrt{n}}\Re(\mathrm{E}(x_{12(C)}))$. Furthermore, by Corollary A.41, we have

$$L(F^{\mathbf{W}_{n(C)}-\Re(\mathrm{E}(\mathbf{W}_{n(C)}))}, F^{\mathbf{W}_{n(C)}-a\mathbf{1}\mathbf{1}'}) \leq \frac{|\Re(\mathrm{E}(x_{12(C)}))|^2}{n} \to 0. \qquad (2.1.4)$$

This shows that we can assume that the real parts of the mean values of the off-diagonal elements are 0. In the following, we proceed to remove the imaginary part of the mean values of the off-diagonal elements.

Before we treat the imaginary part, we introduce a lemma about eigenvalues of a skew-symmetric matrix.

Lemma 2.7. *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 $\lambda_k = i\cot(\pi(2k-1)/2n)$, $k = 1, 2, \cdots, n$. The eigenvector associated with λ_k is $\mathbf{u}_k = \frac{1}{\sqrt{n}}(1, \rho_k, \cdots, \rho_k^{n-1})'$, where $\rho_k = (\lambda_k - 1)/(\lambda_k + 1) = \exp(-i\pi(2k-1)/n)$.*

Proof. We first compute the characteristic polynomial of $\mathbf{A}_n$.

$$D_n = |\lambda\mathbf{I} - \mathbf{A}_n| = \begin{vmatrix} \lambda & -1 & -1 & \cdots & -1 \\ 1 & \lambda & -1 & \cdots & -1 \\ 1 & 1 & \lambda & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & \lambda \end{vmatrix}$$
$$= \begin{vmatrix} \lambda - 1 & -(1+\lambda) & 0 & \cdots & 0 \\ 0 & \lambda - 1 & -(1+\lambda) & \cdots & 0 \\ 0 & 0 & \lambda - 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & \lambda \end{vmatrix}.$$

Expanding the above along the first row, we get the following recursive formula

$$D_n = (\lambda - 1)D_{n-1} + (1+\lambda)^{n-1},$$

with the initial value $D_1 = \lambda$. The solution is

$$\begin{aligned} D_n &= \lambda(\lambda-1)^{n-1} + (\lambda+1)(\lambda-1)^{n-2} + \cdots + (\lambda+1)^{n-1} \\ &= \frac{1}{2}\left((\lambda-1)^n + (\lambda+1)^n\right). \end{aligned}$$

Setting $D_n = 0$, we get

$$\frac{\lambda+1}{\lambda-1} = e^{i\pi(2k-1)/n}, \quad k = 1, 2, \cdots, n, \qquad (2.1.5)$$

which implies that $\lambda = i\cot(\pi(2k-1)/2n)$.

Comparing the two sides of the equation $\mathbf{A}_n\mathbf{u}_k = \lambda_k\mathbf{u}_k$, we obtain

$$-u_{k,1} - \cdots - u_{k,\ell-1} + u_{k,\ell+1} + \cdots + u_{k,n} = \lambda_k u_{k,\ell}$$

for $\ell = 1, 2, \cdots, n$. Thus, subtracting the equations for $\ell + 1$ from that for ℓ, we get

$$u_{k,\ell} + u_{k,\ell+1} = \lambda_k (u_{k,\ell} - u_{k,\ell+1}),$$

which implies that

$$\frac{u_{k,\ell+1}}{u_{k,\ell}} = \frac{\lambda_k - 1}{\lambda_k + 1} = e^{-i\pi(2k-1)/n} := \rho_k.$$

Therefore, one can choose $u_{k,\ell} = \rho_k^{\ell-1}/\sqrt{n}$.

The proof of the lemma is complete.

Write $b = \mathrm{E}\Im(x_{12(C)})$. Then, $\mathrm{E}\Im(\mathbf{W}_{n(C)}) = ib\mathbf{A}_n$. By Lemma 2.7, the eigenvalues of the matrix $i\Im(\mathrm{E}(\mathbf{W}_{n(C)})) = ib\mathbf{A}_n$ are $ib\lambda_k = -n^{-1/2}b\cot(\pi(2k-1)/2n)$, $k = 1, \cdots, n$. If the spectral decomposition of $\mathbf{A}_n$ is $\mathbf{U}_n\mathbf{D}_n\mathbf{U}_n^*$, then we rewrite $i\Im(\mathrm{E}(\mathbf{W}_{n(C)})) = \mathbf{B}_1 + \mathbf{B}_2$, where $\mathbf{B}_j = -\frac{1}{\sqrt{n}} b\mathbf{U}_n\mathbf{D}_{nj}\mathbf{U}_n^*$, $j = 1, 2$, where $\mathbf{U}_n$ is a unitary matrix, $\mathbf{D}_n = \mathrm{diag}[\lambda_1, \cdots, \lambda_n]$, and

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

For any $n \times n$ Hermitian matrix $\mathbf{C}$, by Corollary A.41, we have

$$\begin{aligned} L^3(F^{\mathbf{C}}, F^{\mathbf{C}-\mathbf{B}_1}) &\le \frac{1}{n^2} \sum_{n^{3/4} \le k \le n - n^{3/4}} \cot^2(\pi(2k-1)/2n) \\ &< \frac{2}{n \sin^2(n^{-1/4}\pi)} \to 0 \end{aligned} \tag{2.1.6}$$

and, by Theorem A.43,

$$\|F^{\mathbf{C}} - F^{\mathbf{C}-\mathbf{B}_2}\| \le \frac{2n^{3/4}}{n} \to 0. \tag{2.1.7}$$

Summing up estimations (2.1.3)–(2.1.7), we established the following centralization lemma.

Lemma 2.8. *Under the conditions assumed in Lemma 2.6, we have*

$$L(F^{\mathbf{W}_{n(C)}}, F^{\mathbf{W}_{n(C)} - \mathrm{E}(\mathbf{W}_{n(C)})}) = o(1). \tag{2.1.8}$$

Step 4. Rescaling

Write $\sigma^2(C) = \mathrm{Var}(x_{11(C)})$, and define $\widetilde{\mathbf{W}}_n = \sigma^{-1}(C)\big(\mathbf{W}_{n(C)} - \mathrm{E}(\mathbf{W}_{n(C)})\big)$. Note that the off-diagonal entries of $\sqrt{n}\widetilde{\mathbf{W}}_n$ are $\widehat{x}_{kj} = \sigma^{-1}(C)\big(x_{kj(C)} - \mathrm{E}(x_{kj(C)})\big)$.

Applying Corollary A.41, we obtain

$$L^3(F^{\widetilde{\mathbf{W}}_n}, F^{\mathbf{W}_{n(C)}-\mathrm{E}(\mathbf{W}_{n(C)})}) \leq \frac{2(\sigma(C)-1)^2}{n^2\sigma^2(C)} \sum_{1\leq i<j\leq n} |x_{kj(C)} - \mathrm{E}(x_{kj(C)})|^2$$
$$\to (\sigma(C)-1)^2, \quad \text{a.s.} \tag{2.1.9}$$

Note that $(\sigma(C)-1)^2$ can be made arbitrarily small if C is large. Combining (2.1.9) with Lemmas 2.6 and 2.8, to prove the semicircular law, we may assume that the entries of $\mathbf{X}$ are bounded by C, having mean zero and variance 1. Also, we may assume the diagonal elements are zero.

Step 5. Proof of the Semicircular Law
We will prove Theorem 2.5 by the moment method. For simplicity, we still use $\mathbf{W}_n$ and x_{ij} to denote the Wigner matrix and basic variables after truncation, centralization, and rescaling.

The semicircular distribution satisfies the Riesz condition. Therefore it is enough to show that the moments of the spectral distribution converge to the corresponding moments of the semicircular distribution almost surely. The k-th moment of the ESD of $\mathbf{W}_n$ is

$$\begin{aligned}\beta_k(\mathbf{W}_n) = \beta_k(F^{\mathbf{W}_n}) &= \int x^k \, dF^{\mathbf{W}_n}(x) \\ &= \frac{1}{n}\sum_{i=1}^n \lambda_i^k = \frac{1}{n}\mathrm{tr}(\mathbf{W}_n^k) = \frac{1}{n^{1+\frac{k}{2}}}\mathrm{tr}(\mathbf{X}_n^k) \\ &= \frac{1}{n^{1+\frac{k}{2}}}\sum_{\mathbf{i}} X(\mathbf{i}),\end{aligned} \tag{2.1.10}$$

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

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

(1) $E[\beta_k(\mathbf{W}_n)]$ converges to the k-th moment β_k of the semicircular distribution, which are $\beta_{2m-1} = 0$ and $\beta_{2m} = (2m)!/m!(m+1)!$ given in Lemma 2.1.
(2) For each fixed k, $\sum_n \mathrm{Var}[\beta_k(\mathbf{W}_n)] < \infty$.

The Proof of (1); i.e., $\mathrm{E}[\beta_k(\mathbf{W}_n)] \to \beta_k$.

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

For each vector $\mathbf{i}$, construct a graph $G(\mathbf{i})$ as in Subsection 2.1.2. To specify the graph, we rewrite $X(\mathbf{i}) = X(G(\mathbf{i}))$. The summation is taken over all sequences $\mathbf{i} = (i_1, i_2, \cdots, i_k) \in \{1, 2, \cdots, n\}^k$.

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

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

where

$$S_j = n^{-1-k/2} \sum_{\Gamma(k,t)\in C_j} \sum_{G(\mathbf{i})\in\Gamma(k,t)} \mathrm{E}[XG(\mathbf{i})],$$

in which the summation $\sum_{\Gamma(k,t)\in C_j}$ is taken on all canonical $\Gamma(k,t)$-graphs in category j and the summation $\sum_{G(\mathbf{i})\in\Gamma(k,t)}$ is taken on all isomorphic graphs for a given canonical graph.

By definition of the categories and by the assumptions on the entries of the random matrices, 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 2.2, and $t \le (k+1)/2$ by Lemma 2.3, 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 with an edge of opposite direction, each term in S_1 is $(\mathrm{E}|x_{12}|^2)^m = 1$. So, by Lemma 2.4,

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

Assertion (1) is then proved.

The proof of (2). We only need to show that $\mathrm{Var}(\beta_k(\mathbf{W}_n))$ is summable for all fixed k. We have

$$\begin{aligned} &\mathrm{Var}(\beta_k(\mathbf{W}_n)) = E[|\beta_k(\mathbf{W}_n)|^2] - |E[\beta_k(\mathbf{W}_n)]|^2 \\ &= \frac{1}{n^{2+k}} \sum{}^* \{\mathrm{E}[X(\mathbf{i})X(\mathbf{j})] - \mathrm{E}[X(\mathbf{i})]\mathrm{E}[X(\mathbf{j})]\}, \end{aligned} \tag{2.1.11}$$

where $\mathbf{i} = (i_1, \cdots, i_k)$, $\mathbf{j} = (j_1, \cdots, j_k)$, and $\sum^*$ is taken over all possibilities for $\mathbf{i}$, $\mathbf{j} \in \{1, \cdots, n\}^k$. Here, the reader should notice that $\beta_k(\mathbf{W}_n)$ is real and hence the second equality in the above is meaningful, although the variables $X(\mathbf{i})$ and $X(\mathbf{j})$ are complex.

Using $\mathbf{i}$ and $\mathbf{j}$, one can construct two graphs $G(\mathbf{i})$ and $G(\mathbf{j})$, as in the proof of (1). If there are no coincident edges between $G(\mathbf{i})$ and $G(\mathbf{j})$, then $X(\mathbf{i})$ is independent of $X(\mathbf{j})$, and thus the corresponding term in the sum is 0. If the combined graph $G = G(\mathbf{i}) \cup G(\mathbf{j})$ has a single edge, then $\mathrm{E}[X(\mathbf{i})X(\mathbf{j})] = \mathrm{E}[X(\mathbf{i})]\mathrm{E}[X(\mathbf{j})] = 0$, and hence the corresponding term in (2.1.11) is also 0.

Now, suppose that G contains no single edges and the graph of noncoincident edges has a cycle. Then the noncoincident vertices of G are not more than k. If 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, and thus the number of noncoincident vertices is not larger than k. Also, each term in (2.1.11) is not larger than $2C^{2k}n^{-2-k}$. Consequently, we can conclude that

$$\mathrm{Var}(\beta_k(\mathbf{W}_n)) \leq K_k C^{2k} n^{-2}, \tag{2.1.12}$$

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

The proof of Theorem 2.5 is then complete.

2.2 Generalizations to the Non-iid Case

Sometimes, it is of practical interest to consider the case where, for each n, the entries above or on the diagonal of $\mathbf{W}_n$ are independent complex random variables with mean zero and variance σ^2 (for simplicity we assume $\sigma = 1$ in the following), but may depend on n. For this case, we present the following theorem.

Theorem 2.9. *Suppose that $\mathbf{W}_n = \frac{1}{\sqrt{n}}\mathbf{X}_n$ is a Wigner matrix and the entries above or on the diagonal 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 zero and variance 1 and satisfy the condition that, for any constant $\eta > 0$,*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{jk} \mathrm{E}|x_{jk}^{(n)}|^2 I(|x_{jk}^{(n)}| \geq \eta\sqrt{n}) = 0. \tag{2.2.1}$$

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

Remark 2.10. In Girko's book [121], it is stated that condition (2.2.1) is necessary and sufficient for the conclusion of Theorem 2.9.

2.2.1 Proof of Theorem 2.9

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 positions. Instead, we shall truncate them at $\eta_n\sqrt{n}$ for some sequence $\eta_n \downarrow 0$.

Step 1. Truncation

Note that Corollary A.41 may not be applicable in proving the almost sure asymptotic equivalence between the ESD of the original matrix and that of the truncated one, as was done in the last section. In this case, we shall use the rank inequality (see Theorem A.43) to truncate the variables.

Note that condition (2.2.1) is equivalent to: for any $\eta > 0$,

$$\lim_{n\to\infty} \frac{1}{\eta^2 n^2} \sum_{jk} \mathrm{E}|x_{jk}^{(n)}|^2 I(|x_{jk}^{(n)}| \geq \eta\sqrt{n}) = 0. \tag{2.2.2}$$

Thus, one can select a sequence $\eta_n \downarrow 0$ such that (2.2.2) remains true when η is replaced by η_n. Define $\widetilde{\mathbf{W}}_n = \frac{1}{\sqrt{n}} n(x_{ij}^{(n)} I(|x_{ij}^{(n)}| \leq \eta_n\sqrt{n})$. By using Theorem A.43, one has

$$\begin{aligned}\|F^{\mathbf{W}_n} - F^{\widetilde{\mathbf{W}}_n}\| &\leq \frac{1}{n}\mathrm{rank}(\mathbf{W}_n - \mathbf{W}_{n(\eta_n\sqrt{n})}) \\ &\leq \frac{2}{n} \sum_{1\leq i\leq j\leq n} I(|x_{ij}^{(n)}| \geq \eta_n\sqrt{n}). \end{aligned}\tag{2.2.3}$$

By condition (2.2.2), we have

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

and

$$\begin{aligned}&\mathrm{Var}\left(\frac{1}{n}\sum_{1\leq i\leq j\leq n} I(|x_{ij}^{(n)}| \geq \eta_n\sqrt{n})\right) \\ &\leq \frac{4}{\eta_n^2 n^3}\sum_{jk} \mathrm{E}|x_{ij}^{(n)}|^2 I(|x_{ij}^{(n)}| \geq \eta_n\sqrt{n}) = o(1/n).\end{aligned}$$

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

$$\mathrm{P}\left(\frac{1}{n}\sum_{1\leq i\leq j\leq n} I(|x_{ij}^{(n)}| \geq \eta_n\sqrt{n}) \geq \varepsilon\right) \leq 2e^{-\varepsilon n}, \tag{2.2.4}$$

which is summable. Thus, by (2.2.3) and (2.2.4), to prove that with probability one $F^{\mathbf{W}_n}$ converges to the semicircular law, it suffices to show that with probability one $F^{\widetilde{\mathbf{W}}_n}$ converges to the semicircular law.

Step 2. Removing diagonal elements

Let $\widehat{\mathbf{W}}_n$ be the matrix $\widetilde{\mathbf{W}}_n$ with diagonal elements replaced by 0. Then, by Corollary A.41, we have

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

Step 3. Centralization

By Corollary A.41, it follows that

$$\begin{aligned}
&L^3\left(F^{\widehat{\mathbf{W}}_n}, F^{\widehat{\mathbf{W}}_n - \mathrm{E}\widehat{\mathbf{W}}_n}\right)\\
&\leq \frac{1}{n^2}\sum_{i\neq j}|\mathrm{E}(x_{ij}^{(n)} I(|x_{ij}^{(n)}| \leq \eta_n\sqrt{n}))|^2\\
&\leq \frac{1}{n^3\eta_n^2}\sum_{ij}\mathrm{E}|x_{jk}^{(n)}|^2 I(|x_{jk}^{(n)}| \geq \eta_n\sqrt{n}) \to 0. \qquad (2.2.5)
\end{aligned}$$

Step 4. Rescaling

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

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

$\sigma_{ij}^2 = \mathrm{E}|x_{ij}^{(n)} I(|x_{ij}^{(n)}| \leq \eta_n\sqrt{n}) - \mathrm{E}(x_{ij}^{(n)} I(|x_{ij}^{(n)}| \leq \eta_n\sqrt{n}))|^2$ and δ_{ij} is Kronecker's delta.

By Corollary A.41, it follows that

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

Note that

$$\begin{aligned}
&\mathrm{E}\left(\frac{1}{n^2}\sum_{i\neq j}(1-\delta_{ij}^{-1})^2|x_{ij}^{(n)} I(|x_{ij}^{(n)}| \leq \eta_n\sqrt{n}) - \mathrm{E}(x_{ij}^{(n)} I(|x_{ij}^{(n)}| \leq \eta_n\sqrt{n}))|^2\right)\\
&\leq \frac{1}{n^2\eta_n^2}\sum_{ij}(1-\sigma_{ij})^2
\end{aligned}$$

$$\leq \frac{1}{n^2\eta_n^2}\sum_{ij}(1-\sigma_{ij}^2)$$
$$\leq \frac{1}{n^2\eta_n^2}\sum_{ij}[\mathrm{E}|x_{jk}^{(n)}|^2 I(|x_{jk}^{(n)}|\geq \eta_n\sqrt{n})+\mathrm{E}^2|x_{jk}^{(n)}|I(|x_{jk}^{(n)}|\geq \eta_n\sqrt{n})]\to 0.$$

Also, we have[2]

$$\mathrm{E}\left|\frac{1}{n^2}\sum_{i\neq j}(1-\delta_{ij}^{-1})^2\left|x_{ij}^{(n)}I(|x_{ij}^{(n)}|\leq \eta_n\sqrt{n})-\mathrm{E}(x_{ij}^{(n)}I(|x_{ij}^{(n)}|\leq \eta_n\sqrt{n}))\right|^2\right|^4$$
$$\leq \frac{C}{n^8}\left[\sum_{i\neq j}\mathrm{E}|x_{ij}^{(n)}|^8 I(|x_{ij}^{(n)}|\leq \eta_n\sqrt{n})+\left(\sum_{i\neq j}\mathrm{E}|x_{ij}^{(n)}|^4 I(|x_{ij}^{(n)}|\leq \eta_n\sqrt{n})\right)^2\right]$$
$$\leq Cn^{-2}[n^{-1}\eta_n^6+\eta_n^4],$$

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

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

Step 5. Proof by MCT

Up to here, we have proved that we may truncate, centralize, and rescale the entries of the Wigner matrix at $\eta_n\sqrt{n}$ and remove the diagonal elements without changing the LSD. These four steps are almost the same as those we followed for the iid case.

Now, 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}, We assume:

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

Similar to what we did in the last section, in order to prove Theorem 2.9, we need to show that:

(1) $E[\beta_k(\mathbf{W}_n)]$ converges to the k-th moment β_k of the semicircular distribution.
(2) For each fixed k, $\sum_n \mathrm{E}|\beta_k(\mathbf{W}_n)-\mathrm{E}(\beta_k(\mathbf{W}_n))|^4<\infty$.

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

[2] Here we use the elementary inequality $\mathrm{E}|\sum X_i|^{2k}\leq C_k(\sum \mathrm{E}|X_i|^{2k}+(\sum \mathrm{E}|X_i|^2)^k)$ for some constant C_k if the X_i's are independent with zero means.

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

where $X(G(\mathbf{i})) = x_{i_1,i_2} x_{i_2,i_3} \cdots, x_{i_k,i_1}$, and $G(\mathbf{i})$ is the graph defined by $\mathbf{i}$.

By the same method for the iid case, we split $\mathrm{E}[\beta_k(\mathbf{W}_n)]$ into three sums according to the categories of graphs. We know that the terms in S_2 are all zero, that is, $S_2 = 0$.

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 at least one noncoincident edge with multiplicity greater than 2 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 multiplicities $\nu_1, \cdots, \nu_\ell$ among which at least one is greater than or equal to 3. Note that the multiplicities are subject to $\nu_1 + \cdots + \nu_\ell = k$. Also, each term in S_{31} is bounded by

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

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

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

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

To estimate S_{32}, we note that the $\Gamma_3(k,t)$-graph contains exactly $k/2$ noncoincident edges, each with multiplicity 2 (thus k must be even). Then each term of 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

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

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

The proof of (2)

Unlike in the proof of (2.1.11), the almost sure convergence cannot follow by estimating the second moment of $\beta_k(\mathbf{W}_n)$. We need to estimate its fourth moment as

$$\begin{aligned} &\mathrm{E}(\beta_k(\mathbf{W}_n) - \mathrm{E}(\beta_k(\mathbf{W}_n)))^4 \\ &= n^{-4-2k} \sum_{\mathbf{i}_j, j=1,2,3,4} \mathrm{E} \prod_{j=1}^{4} \left(X[\mathbf{i}_j] - \mathrm{E}(X[\mathbf{i}_j]) \right), \end{aligned} \tag{2.2.6}$$

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 $G_j = G(\mathbf{i}_j)$.

Obviously, if, for some j, $G(\mathbf{i}_j)$ does not have any edges coincident with edges of the other three graphs, then the term in (2.2.6) equals 0 by independence. Also, if $G = \bigcup_{j=1}^4 G_j$ has a single edge, the term in (2.2.6) equals 0 by centralization.

Now, let us estimate the nonzero terms in (2.2.6). Assume that G has ℓ noncoincident edges with multiplicities $\nu_1, \cdots, \nu_\ell$, subject to the constraint $\nu_1 + \cdots + \nu_\ell = 4k$. Then, the term corresponding to G is bounded by

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

Since the graph of noncoincident edges of G can have at most two pieces of connected subgraphs, the number of noncoincident vertices of G is not greater than $\ell + 2$. If $\ell = 2k$, then $\nu_1 = \cdots = \nu_\ell = 2$. Therefore, there is at least one noncoincident edge consisting of edges from two different subgraphs and hence there must be a cycle in the graph of noncoincident edges of G. Therefore,

$$\begin{aligned} &\mathrm{E}(\beta_k(\mathbf{W}_n) - \mathrm{E}(\beta_k(\mathbf{W}_n)))^4 \\ &\le C_k n^{-2k-4} \left[\sum_{\ell < 2k} n^{\ell+2} (\eta_n^2 n)^{2k-\ell} + n^{2k+1} \right] \le C_k \eta_n n^{-2}, \end{aligned}$$

which is summable, and thus (2) is proved. Consequently, the proof of Theorem 2.9 is complete.

2.3 Semicircular Law by the Stieltjes Transform

As an illustration of the use of Stieltjes transforms, in this section we shall present a proof of Theorem 2.9 using them.

2.3.1 Stieltjes Transform of the Semicircular Law

Let $z = u + iv$ with $v > 0$ and $s(z)$ be the Stieltjes transform of the semicircular law. Then, we have

$$s(z) = \frac{1}{2\pi\sigma^2} \int_{-2\sigma}^{2\sigma} \frac{1}{x - z} \sqrt{4\sigma^2 - x^2}\, dx.$$

Letting $x = 2\sigma \cos y$, then

$$\begin{aligned} s(z) &= \frac{2}{\pi}\int_0^{\pi} \frac{1}{2\sigma\cos y - z}\sin^2 y dy \\ &= \frac{1}{\pi}\int_0^{2\pi} \frac{1}{2\sigma\frac{e^{iy}+e^{-iy}}{2} - z}\Big(\frac{e^{iy}-e^{-iy}}{2i}\Big)^2 dy \\ &= -\frac{1}{4i\pi}\oint_{|\zeta|=1} \frac{1}{\sigma(\zeta+\zeta^{-1})-z}(\zeta-\zeta^{-1})^2\zeta^{-1}d\zeta \text{ (setting } \zeta = e^{iy}) \\ &= -\frac{1}{4i\pi}\oint_{|\zeta|=1} \frac{(\zeta^2-1)^2}{\zeta^2(\sigma\zeta^2+\sigma-z\zeta)}d\zeta. \end{aligned} \tag{2.3.1}$$

We will use the residue theorem to evaluate the integral. Note that the integrand has three poles, at $\zeta_0 = 0$, $\zeta_1 = \frac{z+\sqrt{z^2-4\sigma^2}}{2\sigma}$, and $\zeta_2 = \frac{z-\sqrt{z^2-4\sigma^2}}{2\sigma}$, where here, and throughout the book, the square root of a complex number is specified as the one with the positive imaginary part. By this convention, we have

$$\sqrt{z} = \text{sign}(\Im z)\frac{|z|+z}{\sqrt{2(|z|+\Re z)}} \tag{2.3.2}$$

or

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

and

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

This shows that the real part of $\sqrt{z}$ has the same sign as the imaginary part of z. Applying this to ζ_1 and ζ_2, we find that the real part of $\sqrt{z^2-4\sigma^2}$ has the same sign as z, which implies that $|\zeta_1| > |\zeta_2|$. Since $\zeta_1\zeta_2 = 1$, we conclude that $|\zeta_2| < 1$ and thus the two poles 0 and ζ_1 of the integrand are in the disk $|z| \le 1$. By simple calculation, we find that the residues at these two poles are

$$\frac{z}{\sigma^2} \text{ and } \frac{(\zeta_2^2-1)^2}{\sigma\zeta_2^2(\zeta_2-\zeta_1)} = \sigma^{-1}(\zeta_2-\zeta_1) = -\sigma^{-2}\sqrt{z^2-4\sigma^2}.$$

Substituting these into the integral of (2.3.1), we obtain the following lemma.

Lemma 2.11. *The Stieltjes transform for the semicircular law with scale parameter σ^2 is*

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

2.3.2 Proof of Theorem 2.9

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 Steps 1–4 in the last section. That is, we assume that:

(i) For $i \neq j$, $|x_{ij}| \leq \eta_n\sqrt{n}$ and $x_{ii} = 0$.
(ii) For all $i \neq j$, $\mathrm{E}x_{ij} = 0$, $\mathrm{E}|x_{ij}|^2 = \sigma^2$.
(iii) The variables $\{x_{ij}, i < j\}$ are independent.

For brevity, we assume $\sigma^2 = 1$ in what follows.

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

$$s_n(z) = \frac{1}{n}\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1}. \tag{2.3.3}$$

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

(i) For any fixed $z \in \mathbb{C}^+ = \{z, \Im(z) > 0\}$, $s_n(z) - \mathrm{E}s_n(z) \to 0$, a.s.
(ii) For any fixed $z \in \mathbb{C}^+$, $\mathrm{E}s_n(z) \to s(z)$, the Stieltjes transform of the semicircular law.
(iii) Outside a null set, $s_n(z) \to s(z)$ for every $z \in \mathbb{C}^+$.

Then, applying Theorem B.9, it follows that, except for this null set, $F^{\mathbf{W}_n} \to F$ weakly.

Step 1. Almost sure convergence of the random part

For the first step, we show that, for each fixed $z \in \mathbb{C}^+$,

$$s_n(z) - \mathrm{E}(s_n(z)) \to 0 \quad \text{a.s.} \tag{2.3.4}$$

We need the extended Burkholder inequality.

Lemma 2.12. *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $p > 1$,*

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

Proof. Burkholder [67] proved the lemma for a real martingale difference sequence. Now, both $\{\Re X_k\}$ and $\{\Im X_k\}$ are martingale difference sequences. Thus, we have

$$\begin{aligned}
\mathrm{E}\left|\sum X_k\right|^p &\leq C_p\left[\mathrm{E}\left|\sum \Re X_k\right|^p + \mathrm{E}\left|\sum \Im X_k\right|^p\right] \\
&\leq C_p\left[K_p\mathrm{E}\left(\sum |\Re X_k|^2\right)^{p/2} + K_p\mathrm{E}\left(\sum |\Im X_k|^2\right)^{p/2}\right] \\
&\leq 2C_pK_p\mathrm{E}\left(\sum |X_k|^2\right)^{p/2},
\end{aligned}$$

where $C_p = 2^{p-1}$. This lemma is proved.

For later use, we introduce here another inequality proved in [67].

Lemma 2.13. *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 $p \geq 2$,*

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

Similar to Lemma 2.12, Burkholder proved this lemma for the real case. Using the same technique as in the proof of Lemma 2.12, one may easily extend the Burkholder inequality to the complex case.

Now, we proceed to the proof of the almost sure convergence (2.3.4). Denote by $\mathrm{E}_k(\cdot)$ conditional expectation with respect to the σ-field generated by the random variables $\{x_{ij}, i, j > k\}$, with the convention that $\mathrm{E}_n s_n(z) = \mathrm{E}s_n(z)$ and $\mathrm{E}_0 s_n(z) = s_n(z)$. Then, we have

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

where, by Theorem A.5,

$$\begin{aligned}
\gamma_k &= \frac{1}{n}\left(\mathrm{E}_{k-1}\mathrm{tr}(\mathbf{W}_n - z\mathbf{I})^{-1} - \mathrm{E}_k \mathrm{tr}(\mathbf{W}_n - z\mathbf{I})^{-1}\right) \\
&= \frac{1}{n}\Big(\mathrm{E}_{k-1}[\mathrm{tr}(\mathbf{W}_n - z\mathbf{I})^{-1} - \mathrm{tr}(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}] \\
&\qquad -\mathrm{E}_k[\mathrm{tr}(\mathbf{W}_n - z\mathbf{I})^{-1} - \mathrm{tr}(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}]\Big) \\
&= \frac{1}{n}\left(\mathrm{E}_{k-1}\frac{1+\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k}{-z-\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}\right. \\
&\qquad \left. -\mathrm{E}_k\frac{1+\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k}{-z-\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}\right),
\end{aligned}$$

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.

Note that

$$\begin{aligned}
&|1+\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k| \\
&\leq 1+\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}(\mathbf{W}_k - \bar{z}\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k \\
&= v^{-1}\Im(z+\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k)
\end{aligned}$$

which implies that

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

Noting that $\{\gamma_k\}$ forms a martingale difference sequence, applying Lemma 2.12 for $p = 4$, we have

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

By the Borel-Cantelli lemma, we know that, for each fixed $z \in \mathbb{C}^+$,

$$s_n(z) - \mathrm{E}(s_n(z)) \to 0, \text{ a.s.}$$

Step 2. Convergence of the expected Stieltjes transform
By Theorem A.4, we have

$$\begin{aligned}
s_n(z) &= \frac{1}{n}\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1} \\
&= \frac{1}{n}\sum_{k=1}^{n} \frac{1}{-z - \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}.
\end{aligned} \tag{2.3.5}$$

Write $\varepsilon_k = \mathrm{E}s_n(z) - \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k$. Then we have

$$\begin{aligned}
\mathrm{E}s_n(z) &= \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\frac{1}{-z - \mathrm{E}s_n(z) + \varepsilon_k} \\
&= -\frac{1}{z + \mathrm{E}s_n(z)} + \delta_n,
\end{aligned} \tag{2.3.6}$$

where

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

Solving equation (2.3.6), we obtain two solutions:

$$\frac{1}{2}(-z + \delta_n \pm \sqrt{(z+\delta_n)^2 - 4}).$$

We show that

$$\mathrm{E}s_n(z) = \frac{1}{2}(-z + \delta_n + \sqrt{(z+\delta_n)^2 - 4}). \tag{2.3.7}$$

When fixing $\Re z$ and letting $\Im z = v \to \infty$, we have $\mathrm{E}s_n(z) \to 0$, which implies that $\delta_n \to 0$. Consequently,

$$\Im\Big(\frac{1}{2}(-z+\delta_n-\sqrt{(z+\delta_n)^2-4})\Big)\le-\frac{v-|\delta_n|}{2}\to-\infty,$$

which cannot be $\mathrm{E}s_n(z)$ since it violates the property that $\Im s_n(z)\ge 0$. Thus, assertion (2.3.7) is true when v is large. Now, we claim that assertion (2.3.7) is true for all $z\in\mathbb{C}^+$.

It is easy to see that $\mathrm{E}s_n(z)$ and $\frac{1}{2}(-z+\delta_n\pm\sqrt{(z+\delta_n)^2-4})$ are continuous functions on the upper half plane $\mathbb{C}^+$. If $\mathrm{E}s_n(z)$ takes a value on the branch $\frac{1}{2}(-z+\delta_n-\sqrt{(z+\delta_n)^2-4})$ for some z, then the two branches $\frac{1}{2}(-z+\delta_n\pm\sqrt{(z+\delta_n)^2-4})$ should cross each other at some point $z_0\in\mathbb{C}^+$. At this point, we would have $\sqrt{(z_0+\delta_n)^2-4}=0$ and hence $\mathrm{E}s_n(z_0)$ has to be one of the following:

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

However, both of the two values above have negative imaginary parts. This contradiction leads to the truth of (2.3.7).

From (2.3.7), to prove $\mathrm{E}s_n(z)\to s(z)$, it suffices to show that

$$\delta_n\to 0. \tag{2.3.8}$$

Now, rewrite

$$\begin{aligned}\delta_n&=-\frac{1}{n}\sum_{k=1}^{n}\frac{\mathrm{E}(\varepsilon_k)}{(z+\mathrm{E}s_n(z))^2}+\frac{1}{n}\sum_{k=1}^{n}\mathrm{E}\left(\frac{\varepsilon_k^2}{(z+\mathrm{E}s_n(z))^2(-z-\mathrm{E}s_n(z)+\varepsilon_k)}\right)\\&=J_1+J_2.\end{aligned}$$

By (A.1.10) and (A.1.12), we have

$$\begin{aligned}|\mathrm{E}\varepsilon_k|&=\left|\frac{1}{n}\mathrm{E}(\mathrm{tr}(\mathbf{W}_n-z\mathbf{I})^{-1}-\mathrm{tr}(\mathbf{W}_k-z\mathbf{I}_{n-1})^{-1})\right|\\&=\left|\frac{1}{n}\cdot\mathrm{E}\frac{1+\boldsymbol{\alpha}_k^*(\mathbf{W}_k-z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k}{-z-\boldsymbol{\alpha}_k^*(\mathbf{W}_k-z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}\right|\le\frac{1}{nv}.\end{aligned}$$

Note that

$$|z+\mathrm{E}s_n(z)|\ge\Im(z+\mathrm{E}s_n(z))=v+\mathrm{E}(\Im(s_n(z))\ge v.$$

Therefore, for any fixed $z\in\mathbb{C}^+$,

$$|J_1|\le\frac{1}{nv^3}\to 0.$$

On the other hand, we have

$$\begin{aligned}&|-z-\mathrm{E}s_n(z)+\varepsilon_k|=|-z-\boldsymbol{\alpha}_k^*(\mathbf{W}_k-z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k|\\&\ge\Im(z+\boldsymbol{\alpha}_k^*(\mathbf{W}_k-z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k)\end{aligned}$$

$$= v(1 + \boldsymbol{\alpha}_k^*((\mathbf{W}_k - z\mathbf{I}_{n-1})(\mathbf{W}_k - \bar{z}\mathbf{I}_{n-1}))^{-1}\boldsymbol{\alpha}_k) \geq v.$$

To prove $J_2 \to 0$, it is sufficient to show that

$$\max_k \mathrm{E}|\varepsilon_k|^2 \to 0.$$

Write $(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1} = (b_{ij})_{i,j \leq n-1}$. We then have

$$\begin{aligned}
\mathrm{E}|\varepsilon_k - \mathrm{E}\varepsilon_k|^2 &= \mathrm{E}|\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k - \frac{1}{n}\mathrm{Etr}((\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1})|^2 \\
&= \mathrm{E}|\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k - \frac{1}{n}\mathrm{tr}((\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1})|^2 \\
&\quad + \mathrm{E}\left|\frac{1}{n}\mathrm{tr}((\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}) - \frac{1}{n}\mathrm{Etr}((\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1})\right|^2.
\end{aligned}$$

By elementary calculations, we have

$$\begin{aligned}
&\mathrm{E}|\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k - \frac{1}{n}\mathrm{tr}((\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1})|^2 \\
&= \frac{1}{n^2}\left[\sum_{ij\neq k}[\mathrm{E}|b_{ij}|^2\mathrm{E}|x_{ik}|^2\mathrm{E}|x_{jk}|^2 + \mathrm{E}b_{ij}^2\mathrm{E}x_{ik}^2\mathrm{E}x_{jk}^2] + \sum_{i\neq k}\mathrm{E}|b_{ii}^2|(\mathrm{E}|x_{ik}^4| - 1)\right] \\
&\leq \frac{2}{n^2}\sum_{ij}\mathrm{E}|b_{ij}|^2 + \frac{\eta_n^2}{n}\sum_{i\neq k}\mathrm{E}|b_{ii}|^2 \\
&= \frac{2}{n^2}\mathrm{Etr}((\mathbf{W}_k - z\mathbf{I}_{n-1})(\mathbf{W}_k - \bar{z}\mathbf{I}_{n-1}))^{-1} + \frac{\eta_n^2}{n}\sum_{i\neq k}\mathrm{E}|b_{ii}|^2 \\
&\leq \frac{2}{nv^2} + \eta_n^2 \to 0. \qquad (2.3.9)
\end{aligned}$$

By Theorem A.5, one can prove that

$$\mathrm{E}\left|\frac{1}{n}\mathrm{tr}((\mathbf{W}_n - z\mathbf{I}_{n-1})^{-1}) - \frac{1}{n}\mathrm{Etr}((\mathbf{W}_n - z\mathbf{I}_{n-1})^{-1})\right|^2 \leq 1/n^2v^2.$$

Then, the assertion $J_2 \to 0$ follows from the estimates above and the fact that

$$\mathrm{E}|\varepsilon_n|^2 = \mathrm{E}|\varepsilon_n - \mathrm{E}\varepsilon_n|^2 + |\mathrm{E}\varepsilon_n|^2.$$

The proof of the mean convergence is complete.

Step 3. Completion of the proof of Theorem 2.9
In this step, we need Vitali's convergence theorem.

Lemma 2.14. *Let $f_1, f_2, \cdots$ be analytic in D, a connected open set of $\mathbb{C}$, satisfying $|f_n(z)| \leq M$ for every n and z in D, and $f_n(z)$ 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 $z \in D$. Moreover, on any set bounded by a contour interior to D, the convergence is uniform and $\{f_n'(z)\}$ is uniformly bounded.*

Proof. The conclusions on $\{f_n\}$ are from Vitali's convergence theorem (see Titchmarsh [275], p. 168). Those on $\{f_n'\}$ follow from the dominated convergence theorem (d.c.t.) and the identity

$$f_n'(z) = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{f_n(w)}{(w-z)^2} dw,$$

where $\mathcal{C}$ is a contour in D and enclosing z. The proof of the lemma is complete.

By Steps 1 and 2, for any fixed $z \in \mathbb{C}^+$, we have

$$s_n(z) \to s(z), \quad \text{a.s.},$$

where $s(z)$ is the Stieltjes transform of the standard semicircular law. That is, for each $z \in \mathbb{C}^+$, there exists a null set N_z (i.e., $P(N_z) = 0$) such that

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

Now, let $\mathbb{C}_0^+ = \{z_m\}$ be a dense subset of $\mathbb{C}^+$ (e.g., all z of rational real and imaginary parts) and let $N = \cup N_{z_m}$. Then

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

Let $\mathbb{C}_m^+ = \{z \in \mathbb{C}^+, \Im z > 1/m, \ |z| \le m\}$. When $z \in \mathbb{C}_m^+$, we have $|s_n(z)| \le m$. Applying Lemma 2.14, we have

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

Since the convergence above holds for every m, we conclude that

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

Applying Theorem B.9, we conclude that

$$F^{\mathbf{W}_n} \xrightarrow{w} F, \ \text{a.s.}$$

Springer
http://www.springer.com/978-1-4419-0660-1

Spectral Analysis of Large Dimensional Random
Matrices
Bai, Z.; Silverstein, J.W.
2010, XVI, 552 p., Hardcover
ISBN: 978-1-4419-0660-1