Lectures 2 – 3 : Wigner's semicircle law

Notes prepared by: M. Koyama

As we set up last week, let $M_n = [X_{ij}]_{i,j=1}^n$ be a symmetric $n \times n$ matrix with Random entries such that

- $X_{i.j} = X_{j,i}$
- $X_{i,j}$s are *iid* for all $i < j$, and X_{jj} are iid for all j with

$$E[X_{ij}^2] = 1, \quad E[X_ij] = 0$$

- All moments exists for each entries.

We considered the eigenvector of this random matrix;

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

which turns out to be random elements depending continuously on M_n;

Lemma 1. *If $\mathcal{H}_n$ is a topological space of $n \times n$ matrix with topology derived from the usual metric on product Lebesgue measurable space, then $\lambda_i(\mathcal{H})$ is a continuous function on $\mathcal{H}_n$.*

Proof. Let $H = [h_{ij}]_{i,j=1}^n$ be an element in $\mathcal{H}_n$. We know that

$$\|H\|_k = \sqrt[k]{Tr(H^k))} = \sqrt[k]{\sum \lambda_i^k}$$

So for example, $\|H\|_2 = \sqrt{\sum_i \lambda_i^2}$ Note that therefore $\|H\|_2 \geq max(\lambda_n, -\lambda_1)$. Our goal is to obtain λ in terms of H. So it is good if we can say

$$\lim_{k\to\infty} \|H\|_k \to \lambda_n$$

because λ_n dominates all the other eigen vectors, maybe except λ_1. Clearly, this logic might not work because of the presence of negative eigen values including λ_1. To fix this problem we may just shift the matrix by $\|H\|$. In particular, we can claim

$$\lim_{k\to\infty} \sqrt[k]{Tr((H + \|H\|I)^k)} \to \lambda_n + \|H\|$$

To be more precise,

$$\lambda_n(H) + \|H\| \leq \sqrt[k]{Tr((H + \|H\|I)^k)} \tag{1}$$
$$\leq \sqrt[k]{n}(\lambda_n(H) + \|H\|) \tag{2}$$
$$\leq \lambda_n(H) + \|H\|. \tag{3}$$

Having obtained $\lambda_n ... \lambda_k$, we can inductively obtain the λ_{k-1} by simply taking the limit of

$$\sqrt[k]{Tr((H + \|H\|I)^k) - \sum_{i=1}^{k} (\lambda_n(H) + \|H\|)^k}$$

□

This allows us to induce the random measure

$$\nu_n = \frac{1}{n} \sum \delta_{\frac{\lambda_i}{\sqrt{n}}}.$$

The Wigner's semicircle law claims that this ν_n has a nice distributional limit.

Theorem 2.

$$\frac{1}{n} \sum \delta_{\frac{\lambda_i}{\sqrt{n}}} \Rightarrow \nu$$

where $\frac{\nu}{dx} = \frac{1}{2\pi}\sqrt{4 - x^2} 1_{(|x| \leq 2)}$.

We will use Borrel Cantelli lemma and Carleman's condition for moment problem to show this fact. Consider the random variable

$$X_{n,k} = \int x^k d\nu_n.$$

We will show

1. $$\mathbf{EX_{n,k}} \rightarrow \mathbf{c_k} = \int \mathbf{x^k d}\nu \tag{4}$$

2. $$\mathbf{Var(X_{n,k})} \leq \frac{\mathbf{c_k}}{\mathbf{n^2}} \tag{5}$$

How do they help? Suppose these two statements are true. Then we can use Borrel Cantelli lemma to show that

$$P(|X_{n,k} - EX_{n,k}| > \frac{1}{\sqrt[4]{n}}) \leq E(X_{n,k} - EX_{n,k})^2 \sqrt{n} \tag{6}$$
$$= Var(X_{n,k})\sqrt{n} \tag{7}$$
$$= O(1/n^{3/2}) \tag{8}$$

Thus $P(|X_{n,k} - EX_{n,k}| > \frac{1}{\sqrt[4]{n}}\ i.o) = 0$ and $|X_{n,k} - EX_n| < \frac{1}{\sqrt[4]{n}}$ for some large n almost surely. If this is the case, then ν_n can be shown to be tight because this means $X_{n,k}$ is bounded by some constant C and hence by Chebyshev

$$\nu_n(\{x : |x| > m\}) < \frac{C}{m^k}.$$

We can therefore choose a converging subsequence $\nu_{n(j)}$ of measures that converge to ν^*. We would now like to show that any of these subsequencial limits ν^* of converging subsequences equals to ν. In this way, we can establish that any subsequence $\nu_{n(k)}$ has further subsequence that converges to ν. This can be done if we can characterize ν by its moments, because we know that ν^*'s moments for all subsequence agrees by the claim (1). This can be done using the following useful criterion.

Theorem 3. *(Carleman's condition:) Suppose*

$$\sum_{k=1}^{\infty} \frac{1}{\mu_k^{1/2k}} = \infty$$

. Then there is at most one measure F such that $\int x^k dF(x) = \mu_k$ for all positive integer k. This criterion can be made stronger: in fact, the conclusion above holds if

$$\limsup \frac{\mu_{2k}^{1/2k}}{2k} = r < \infty.$$

The logic behind the proof of this claim follows from the fact that the characteristic function $E[exp(iXt)]$ charaterizes the distribution of X. We can consider the Taylor polynomial of $exp(iXt)$. If $E[exp(iXt)] = \sum \frac{(it)^k E[X^k]}{k!}$, then the moment indeed determines the characteristic function.

Let's hence check if the Carleman's condition applies to our case. Put

$$c_k = \int_{-2}^{2} \frac{1}{2\pi} x^k \sqrt{4 - x^2} dx.$$

If k is odd, then $c_k = 0$. Therefore put $k = 2n$. Then

$$\begin{aligned}
c_k &= \frac{1}{\pi} \int_{[0,2]} x^{2n} \sqrt{4 - x^2} dx & (9)\\
&= \frac{1}{\pi} \int_{[0,\pi/2]} \sin^{2n}(t) \cos^2(t) 2^{2n+2} dt & (10)\\
&= \frac{1}{\pi} \int_{[0,\pi/2]} 2^{2n+2} \left(\sin^{2n}(t) - \sin^{2n+2} t\right) dt & (11)\\
&= \frac{1}{\pi} 2^{2n+2} \frac{(2n)!}{n! 2^{2n}} \frac{\pi}{2} \left(1 - \frac{(2n+2)(2n+1)}{4(n+1)^2}\right) & (12)\\
&= \binom{2n}{n} \frac{1}{n+1} < 4^n & (13)
\end{aligned}$$

We used the fact

$$\int_{[0,\pi/2]} \sin^{2\ell}(t) dt = \frac{(2\ell)!}{(\ell!)^2 2^{2\ell}} \frac{\pi}{2}$$

Therefore $\frac{\mu_{2k}^{1/2k}}{2k} < \frac{(4^{k/2})^{1/2k}}{2k} = \frac{\sqrt[4]{4}}{2k}$ and the claim follows.

Therefore, it remains to show (1) and (2) in (0.4) and (0.5).

Proof of (1)

Let us begin with (1). We will achieve this by a way of "controlled brute force". Note that

$$E\int x^k d\nu_n = E\frac{1}{n}\sum\left(\frac{\lambda_i}{\sqrt{n}}\right)^k \tag{14}$$

$$= n^{-1-\frac{k}{2}}E(TrM_n^k) \tag{15}$$

$$= n^{-1-\frac{k}{2}}\sum E(X_{i_1,i_2}X_{i_2,i_3}X_{i_3,i_4}\dots X_{i_k,i_1}) \tag{16}$$

To organize this, whenever we have k-tuple $(i_1, i_2, \dots i_k) = I$, put

$$E(I) = E(X_{i_1,i_2}X_{i_2,i_3}X_{i_3,i_4}\dots X_{i_k,i_1}).$$

First, observe that $E(I)$ is bounded by some constant B_k. This can be seen by applying Cauchy Shwartz inequality inductively.

Let us represent each I by a directed closed path with vertices $\{1, 2, 3, \dots n\} = V(I)$ and edges $\xi(I) = \{(i_a, i_{a+1}); a = 1, ...k, i_{k+1} = i_1\}$ For example, if $I = (2, 3, 1, 2, 2, 1)$ then this will correspond to the directed adjacency matrix [1]

$$\begin{bmatrix} 0 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \tag{17}$$

Now, **skeleton** of a directed graph is a undirected graph induced by the directed graph by replacing all the multiedges by edges. For example, the skeleton of the graph above is given by the adjacency matrix [2]

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \tag{18}$$

[1] entries in the a_{ij} represents the number of edges from vertex i to vertex j.

[2] entries in the a_{ij} represents the number of edges between vertex i to vertex j.

Here, remark that that $E[I] = 0$ unless every edge in the skeleton is used at least twice. If, for example, an edge (i, j) happens only once, then

$$E[I] = E[X_{i,j}]E\left[\prod_{e\in\xi(I)\setminus\{(i,j)\}} X_e\right] = 0$$

This implies that if

$$\mathbf{E(I) \neq 0 \quad then \quad \xi(I) \leq \frac{k}{2}}.$$

This bound let us also put a bound on $V(I)$;

Lemma 4. *Given any graph G, denote the vertex set by $V(G)$ and edge set by $E(G)$. Then $|V(G)| \leq |E(G)| + 1$.*

Proof. To see this, first assume that G is a tree. Note that removing a leaf from the graph removes one edge and one vertex. We may continue removing leaves from the Graph until K_2(complete graph of 2 vertices) remains. Removing a leaf from K_2 results in K_1. Thus $V(G) = E(G) + 1$ in this case. For a generic graph G, we may remove edges from the graph until we obtain its spanning tree G". If we removed m edges in this process, then $V(G) = E(G') + 1 + m$ and the claim follows. □

Thus, we have

$$\mathbf{E(I) \neq 0 \quad then \quad V(I) \leq \lfloor\frac{k}{2}\rfloor + 1}.$$

We are now in position to bound the expectation of $X_{n,k}$.

Lemma 5.

$$\left|E\left[\int x^k d\nu_n\right]\right| \leq \frac{c_k}{\sqrt{n}}$$

Proof.

$$E\int x^k d\nu_n = \frac{1}{n^{k/2+1}}\sum_I E(I) \tag{19}$$

$$= \sum_{V(I)\leq\lfloor\frac{k}{2}\rfloor+1} E(I) \tag{20}$$

$$\leq \frac{B_k}{n^{1+k/2}}\left|\{I; V(I) \leq \lfloor\frac{k}{2}\rfloor + 1\}\right| \tag{21}$$

Temporarily, consider $V(I) = \ell$ for a fixed ℓ. How many ways can we choose I? Most naive bound on this number is indeed $n^\ell * \ell^k$. It turns out that this naive bound suffices. From the inequality that we obtained above, we see that if $\ell < \frac{k}{2} + 1$ then the terms with $V(I) = \ell$ will vanish in limit. Thus we can ignore the odd k all together in the limit. Let us therefore consider the case of even

k. When k is even , we see that $V(I) \leq \frac{k}{2}+1$. If the inequality is strict, again $E\int x^k d\nu_n \to 0$ in the limit. **Therefore, asymptotically, we can restrict our case to when $\mathbf{V(I)} = \frac{\mathbf{k}}{\mathbf{2}} + \mathbf{1}$ and $\xi(\mathbf{I}) \leq \frac{\mathbf{k}}{\mathbf{2}}$** . Because $V(I) \leq \xi(I)+1$, we have $\xi(I) = \frac{k}{2}$ necessarily. We are thus considering directed graphs for which the skeletons are trees, and there are exactly two edges between two adjacent vertices. This kind of directed graph is called a **double tree.** Below is a directed adjacency matrix for an example of a double tree;

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{22}$$

Now, if I is a double tree, clearly

$$E(I) = E\left(\prod_{e\in\xi(I)} X_e^2\right) = \prod_{e\in\xi(I)} E(X_e^2) = 1$$

Thus, all together, we obtain the following statement;

Proposition 6.

$$\lim_{n\to\infty} E\left(\int x^k d\nu_n\right) = \lim_{n\to\infty} \frac{1}{n^{1+\frac{k}{2}}} * \big(\textit{Number of double trees with } n \textit{ vertices}\big)$$

Our proof of (1) is therefore simplified to the counting of the number of double trees with n vertices. To answer this, first let us answer the following question; "If I fix a shape of a tree, just how many double trees of that shape exist?" We may achieve this by making a bijection between the shape of a double tree and a random walk on $\mathbb{N}$ beginning from 0 and returning in exactly k-step. For example, suppose that a double tree is given by the directed adjacency matrix above; then fixing the vertex 1 as the starting point of the walk, the shape of this double tree corresponds to the random walk $(0,1,2,1,2,1,0)$ (the kth entry is the distance of the walker from the vertex 1 at kth step). Counting this way, we will show next week that the shape of a double tree with $\frac{k}{2}$ edges are given by

$$\binom{k}{\frac{k}{2}} \frac{1}{k+1}.$$

Now, given a fixed shape, the number of double trees of that shape is given by

$$\underbrace{\binom{n}{\frac{k}{2}+1}}_{\text{choosing the vertices}} \underbrace{\left(\frac{k}{2}+1\right)!}_{\text{permutation}}$$

Thus at last, we obtain that

$$\lim_{n\to\infty} \int x^k d\nu_n = \lim_{n\to\infty} \frac{1}{n^{k/2}+1} \binom{k}{\frac{k}{2}} \frac{1}{k+1} \frac{1}{\frac{k}{2}+1} n(n-1)\cdots\left(n-\frac{k}{2}\right) \tag{23}$$

$$= \binom{k}{\frac{k}{2}} \frac{1}{\frac{k}{2}+1} = \binom{2n}{n} \frac{1}{n+1} \tag{24}$$

and the claim follows. □

Lectures 4 – 5 : Wigner's semicircle law

Notes prepared by: H. Lin

Following from last week, we let k is an even number, say, $k = 2l$. we first briefly show how to derive the number of paths from $(0,0)$ to $(k,0)$ while not allowed to go below the $x-$axis.

For any path from $(0,0)$ to $(2l,0)$ which intersects with $y = -1$, we can let $(a,-1)$ be the last intersection, and reflect the part after $(a,-1)$ with respect to $y = -1$, and get a path from $(0,0)$ to $(2l,-2)$. On the other hand, given a path from $(0,0)$ to $(2l,-2)$, we can reflect similarly and obtain a path $(0,0)$ to $(2l,0)$. Since a path from $(0,0)$ to $(2l,-2)$ takes $l-1$ steps upward, and $l+1$ steps downward, we have $\binom{2l}{l-1}$ such paths. Therefore we have $\binom{2l}{l-1}$ paths from $(0,0)$ to $(2l,0)$ which intersects with $y = -1$. Now we can claim that there are

$$\binom{2l}{l} - \binom{2l}{l-1} = \binom{2l}{l}\frac{1}{l+1}$$

paths from $(0,0)$ to $(k,0)$ without going below the x-axis.

Proof of (2)

We follow the notation used in the proof of (1). Let $I = (i_1, i_2, ..., i_n) \in [n]^k$ be a $k-$tuple, and write X_I for the product of the entries $X_{i_1 i_2} X_{i_2 i_3} ... X_{i_k i_1}$. From (16), we have

$$\int x^k d\nu_n = n^{-1-\frac{k}{2}} \sum_{I \in [n]^k} X_I.$$

So

$$Var(\int x^k d\nu_n) = n^{-2-k} \sum_{I,J \in [n]^k} cov(X_I, X_J). \tag{25}$$

Now we again represent I and J as closed directed path as in the proof of (1), and give the following facts:

1. If there are no common "edges" in I and J, then X_I and X_J are independent and hence $cov(X_I, X_J) = 0$.

2. If there is an edge which only appears once in I or J, say, the edge $i_1 i_2$, then $X_{i_1 i_2}$ is independent of rest of the terms in X_I and $X_I X_J$. Since $EX_{i_1 i_2} = 0$, we know $EX_I = EX_I X_J = 0$, so $cov(X_I, X_J) = EX_I X_J - EX_I EX_J = 0$.

Let $m = |V(I \cup J)|$ be the size of the set corresponding to the union of I and J. For a given m, we at most have Cn^m ways to select the $k-$tuples I and J, where is C is a constant independent of n.

If $m \leq k$, then we see the contribution of these term in (25) is of order $\frac{1}{n^2}$.

Now we consider the terms with $m \geq k+1$. From the two facts we just mentioned, I and J give a connected graph, with each edge used at least twice, and hence in the skeleton of the graph of $I \cup J$ we only have at most k edges. However, $m \geq k+1$, in this situation we know we must have $m = k+1$ and $I \cup J$ is actually a double tree.

By erasing vertices not belonging to I, we can see that I is also a double tree, and so is J. We now look at a "common edge" in both I and J, it appears twice in I and twice in J, and thus four times in $I \cup J$, which contradicts with the observation that $I \cup J$ is a double tree.

To sum up, all terms in (25) satisfy $m \leq k$ and are $O(\frac{1}{n^2})$, which proves (2).

Before proceeding to the next theorem, we prove the following lemma:

Lemma 7. (**Hoffman – Wielandt**) *Let A and B be two $n \times n$ symmetric (or Hermitian) matrices with eigenvalues $\lambda_1(A) \leq \lambda_2(A) \leq ... \leq \lambda_n(A)$, and $\lambda_1(B) \leq \lambda_2(B) \leq ... \leq \lambda_n(B)$, then*

$$\sum_{i=1}^{n} |\lambda_i(A) - \lambda_i(B)| \leq Tr[(A-B)^2] \tag{26}$$

Proof. Since $\sum_i \lambda_i^2(A) = Tr(A^2)$, we only need to prove

$$Tr(AB) \leq \sum_{i=1}^{n} \lambda_i(A)\lambda_i(B). \tag{27}$$

Because A and B are symmetric, we can write

$$A = UD_AU^T \quad and \quad B = VD_BV^T$$

for some diagonal matrices D_A, D_B and orthogonal matrices U, V.

Then let $W = [w_{ij}]_{i,j=1}^n = U^T V$, we get

$$\begin{aligned} Tr(AB) &= Tr(UD_A U^T V D_B V^T) \\ &= Tr(D_A U^T V D_B V^T U) \\ &= Tr(D_A W D_B W^T) \\ &= \sum_{1\le i,j\le n} \lambda_i(A)\lambda_j(B) w_{ij}^2 \end{aligned} \tag{28}$$

So now we try to maximize $\sum_{i,j} \lambda_i(A)\lambda_j(B) v_{ij}$ with the constraints that $v_{ij} \ge 0, \sum_{i=1}^n v_{ij} = 1$ for $j = 1, ..., n$, and $\sum_{j=1}^n v_{ij} = 1$ for $i = 1, ..., n$.

Suppose $v_{11} < 1$, then there must exist i and j such that $v_{i1} > 0$ and $v_{1j} > 0$. Let $v = \min\{v_{i1}, v_{j1}\}$. Then define $v'_{11} = v_{11} + v$, $v'_{1j} = v_{1j} - v$, $v'_{i1} = v_{i1} - v$, and $v'_{ij} = v_{ij} + v$.

Since

$$\begin{aligned} &\lambda_1(A)\lambda_1(B)(v'_{11} - v_{11}) + \lambda_1(A)\lambda_j(B)(v'_{1j} - v_{1j}) \\ &\quad + \lambda_i(A)\lambda_1(B)(v'_{i1} - v_{i1}) + \lambda_i(A)\lambda_j(B)(v'_{ij} - v_{ij}) \\ =&v(\lambda_1(A) - \lambda_i(A))(\lambda_1(B) - \lambda_j(B)) \\ \ge&0, \end{aligned} \tag{29}$$

we see that if we repeat the same argument, we maximize $\sum_{i,j} \lambda_i(A)\lambda_j(B) v_{ij}$ when all $v_{ij} = 0$ for $i \ne j$ and $v_{ii} = 1$ for $i = 1, 2, ..., n$. Therefore (27) is proved and we conclude the proof of the lemma.

□

Now we look at a more generalized version of Wigner's theorem without assuming finiteness of higher moments:

Theorem 8. *Let $M_n = [X_{ij}]_{i,j=1}^n$ be a symmetric $n \times n$ matrix with Random entries such that*

- *X_{ij} are i.i.d., with $EX_{ij} = 0$ and $EX_{ij}^2 = 1$ for all $i < j$.*
- *X_{ii} are i.i.d., with $EX_{ii} = 0$ and EX_{ii}^2 is finite for $1 \le i \le n$.*

Let ν_n and ν be defined as before, then we have

$$\nu_n \Rightarrow \nu \tag{30}$$

Proof. Fix $C > 0$, for $i \neq j$ define

$$\sigma^2(C) = Var\big(X_{ij}1_{(|X_{ij}|\leq C)}\big)$$

and for all i and j define

$$X_{ij}^C = \frac{X_{ij}1_{(|X_{ij}|\leq C)} - EX_{ij}1_{(|X_{ij}|\leq C)}}{\sigma(C)}$$

Let $\tilde{M}_n = [X_{ij}^C]_{i,j=1}^n$, and define the corresponding $\tilde{\lambda}_i$ and $\tilde{\nu}_n$ as before, then we see all entries have bounded support and thus $\tilde{M}_n$ satisfy all conditions of theorem 2 (we actually didn't use the condition $EX_{ii}^2 = 1$ in the proof of theorem 2, we only need finiteness), so

$$\tilde{\nu}_n \Rightarrow \nu \quad a.s. \tag{31}$$

From the definition of X_{ij}^C, we have

$$X_{ij} - X_{ij}^C = \frac{1}{\sigma(C)}\big(X_{ij}1_{(|X_{ij}\geq C|)} - EX_{ij}1_{(|X_{ij}\geq C|)}\big) + \big(1 - \frac{1}{\sigma(C)}\big)X_{ij}$$

By lemma (7) and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \Big|\frac{\lambda_i}{\sqrt{n}} - \frac{\tilde{\lambda}_i}{\sqrt{n}}\Big| \\
\leq &\frac{1}{n^2}Tr[(M_n - \tilde{M}_n)^2] \\
= &\frac{1}{n^2}\sum_{i,j}(X_{ij} - X_{ij}^C)^2 \\
\leq &\frac{2}{n^2}\frac{1}{\sigma(C)^2}\sum_{i,j}\big(X_{ij}1_{(|X_{ij}\geq C|)} - EX_{ij}1_{(|X_{ij}\geq C|)}\big)^2 + \frac{2}{n^2}\big(1 - \frac{1}{\sigma(C)}\big)^2\sum_{i,j}X_{ij}^2 \\
&= \frac{(1-\sigma(C)^2)}{\sigma(C)^2}O(1) + \big(1 - \frac{1}{\sigma(C)}\big)^2 O(1) \quad as\ n \to \infty
\end{aligned} \tag{32}$$

The last step comes from the observation that as $n \to \infty$,

$$\frac{1}{n^2}\sum_{i,j}\big(X_{ij}1_{(|X_{ij}\geq C|)} - EX_{ij}1_{(|X_{ij}\geq C|)}\big)^2 \to Var\big(X_{ij}1_{(|X_{ij}\geq C|)} - EX_{ij}1_{(|X_{ij}\geq C|)}\big) = 1 - \sigma(C)^2$$

and

$$\frac{1}{n^2}\sum_{i,j}X_{ij}^2 \to VarX_{ij} = 1$$

Now we look at any bounded and Lipschitz continuous function $f(x)$. There exists a constant $K > 0$ such that for all x and y,

$$|f(x) - f(y)| \le K|x - y|.$$

Hence

$$\begin{aligned}
&|\int f d\tilde{\nu}_n - \int f d\nu_n| \\
\le& \frac{1}{n}\sum_i |f(\frac{\tilde{\lambda}_i}{\sqrt{n}}) - f(\frac{\lambda_i}{\sqrt{n}})| \\
\le& K\sqrt{\frac{1}{n}\sum_i (\frac{\tilde{\lambda}_i}{\sqrt{n}} - \frac{\lambda_i}{\sqrt{n}})^2} \\
\le& K'\sqrt{\frac{(1-\sigma(C)^2)}{\sigma(C)^2} + \big(1 - \frac{1}{\sigma(C)}\big)^2}
\end{aligned} \tag{33}$$

for some constant K'.

So

$$\begin{aligned}
&\limsup_{n\to\infty} |\int f d\nu_n - \int f d\nu| \\
\le& \limsup_{n\to\infty} |\int f d\tilde{\nu}_n - \int f d\nu| + \limsup_{n\to\infty} |\int f d\nu_n - \int f d\tilde{\nu}_n| \\
\le& K'\sqrt{\frac{(1-\sigma(C)^2)}{\sigma(C)^2} + \big(1 - \frac{1}{\sigma(C)}\big)^2}.
\end{aligned} \tag{34}$$

This holds for any $C > 0$, so we let C go to infinity, and obtain

$$\lim_{n\to\infty} |\int f d\nu_n - \int f d\nu|) = 0 \quad a.s. \tag{35}$$

Take an arbitrary function that is Lipschitz continuous with the following conditions

1. $f(x) = 1$ for$|x| \le 2$ and $x = 0$ for $|x| \ge 3$.
2. $0 \le f(x) \le 1$ for all real number x.

Immediately from (35) we have

$$\lim_{n\to\infty} \int f d\nu_n = \int f d\nu$$

Recall that ν is a measure with support $[-2,2]$, and we can also get $\int f d\nu = 1$. Since we let $0 \le f(x) \le 1$ have support $[-3,3]$, it follows that $\nu_n([-3,3]) \to 1$ as n approaches infinity. Now define a new measure $\bar{\nu}_n(A) = \nu_n(A \cap [-3,3])$.

Now we claim $\bar{\nu}_n \Rightarrow \nu \quad a.s.$. It suffices to show that

$$\int x^k d\bar{\nu}_n \to \int x^k d\nu$$

for all $k \in Z_+$. This is clear from previous argument because $\int x^k d\bar{\nu}_n = \int x^k 1_{(|x|\le 3)} d\nu_n$, and $x^k 1_{(|x|\le 3)}$ is bounded and Lipschitz continuous.

So for any bounded and continuous function $f(x)$,

$$\begin{aligned}
&\lim_{n\to\infty} |\int f d\nu_n - \int f d\nu| \\
\le &\lim_{n\to\infty} |\int f d\nu_n - \int f d\bar{\nu}_n| + \lim_{n\to\infty} |\int f d\nu - \int f d\bar{\nu}_n| \\
\le &\|f\|_\infty \lim_{n\to\infty} \nu_n(\mathbb{R}\backslash[-3,3]) + 0 \\
= &0
\end{aligned} \tag{36}$$

This concludes the proof of the theorem.

□

Remark 9. When we proved (35) for Lipschitz functions, we could get that for each t, the corresponding characteristic function $c_n(t) = \int e^{itx} d\nu_n$ converges to $c(t) = \int e^{itx} d\nu$ almost surely. Another fact is that there is a countable selection of bounded Lipschitz funcitons on $\mathbb{R}$ which determines the convergence in distribution. So in this way we can prove the theorem.

We can relax the conditions even further and give the following two theorems:

Theorem 10. *For each $n \in \mathbb{Z}_+$, let $M_n = [X_{ij}^{(n)}]_{i,j=1}^n$ be a symmetric $n \times n$ matrix with Random entries such that*

- $X_{ij}^{(n)}$ *are independent with mean zero and variance 1.*
- $\sup_{i,j,n} E|X_{ij}^{(n)}|^4 < C$ *for some constant C.*

If we define ν_n and ν as before, then

$$\nu_n \Rightarrow \nu \quad a.s.$$

The second condition could also be replaced by some sort of uniform integrability of variance.

Theorem 11. *For each $n \in \mathbb{Z}_+$, let $M_n = [X_{ij}^{(n)}]_{i,j=1}^n$ be a symmetric $n \times n$ matrix. Assume the matrix EM_n has rank $r(n)$, with $\lim_{n\to\infty} \frac{r(n)}{n} = 0$. If also assuming $VarX_{ij}^{(n)} = 1$ and*

$$\sup_{i,j,n} E|X_{ij}^{(n)} - EX_{ij}^{(n)}|^4 < \infty,$$

then for any bounded and continuous function $f(x)$,

$$\int f d\nu_n = \int f d\nu$$

With some tightness on ν_n the above conditions imply $\nu_n \Rightarrow \nu$. Note that when all entries have the same mean, then $r(n) = 1$, which gives a special case of the theorem.

Another note is that if A is symmetric matrix, then the eigenvalues of A and $A + \lambda ee^T$ are interlaced for $\lambda \in \mathbb{R}, e \in \mathbb{R}^n$.

For matrices with complex entries, we have the following theorem:

Theorem 12. *$M_n = [X_{ij}]_{i,j=1}^n$ be an $n \times n$ matrix with Random entries such that*

- $X_{ij} = \overline{X_{ji}}$
- *$X_{ij}s$ are i.i.d for all $i < j$, and X_{ii} are i.i.d for all i. For all $1 \le i, j \le n$,*

  $$E|X_{ij}|^2 = 1, \qquad E[X_ij] = 0$$

- *All moments exists for each entry.*

Define ν_n and ν as before, then

$$\nu_n \Rightarrow \nu.$$

This is analogue of theorem 2, and the proof is very similar, with the only difference $X_{ij} = \overline{X_{ji}}$, so we will have $X_{ij}X_{ji} = |X_{ij}|^2$ in our computation.