Lectures 6 – 7 : Marchenko-Pastur Law

Notes prepared by: A. Ganguly

We will now turn our attention to rectangular matrices. Let

\[ X = (X_1, X_2, \ldots, X_n) \in \mathbb{R}^{p \times n} \]

where \( X_{ij} \) are iid, \( E(X_{ij}) = 0 \), \( E(X_{ij}^2) = 1 \) and \( p = p(n) \).

Define

\[ S_n = \frac{1}{n}XX^T \in \mathbb{R}^{p \times p} \]

and let

\[ \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p \]

denote the eigenvalues of the matrix \( S_n \).

Define the random spectral measure by

\[ \mu_n = \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i} \]

We are now ready to state the Marchenko-Pastur law

**Theorem 1.** Let \( S_n, \mu_n \) be as above. Assume that \( p/n \xrightarrow{n \to \infty} y \in (0, 1] \). Then we have

\[ \mu_n(\cdot, \omega) \Rightarrow \mu \quad \text{a.s} \]

where \( \mu \) is a deterministic measure whose density is given by

\[ \frac{d\mu}{dx} = \frac{1}{2\pi xy} \sqrt{(b - x)(x - a)}1_{(a \leq x \leq b)} \tag{1} \]

Here \( a \) and \( b \) are functions of \( y \) given by

\[ a(y) = (1 - \sqrt{y})^2, \quad b(y) = (1 + \sqrt{y})^2 \]

**Remark 2.** If \( y > 1 \) then since \( \text{rank}(S) = p \wedge n \) we will have roughly \( n(y - 1) \) zero eigenvalues. Since \( \mu_n = \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i} \) we see that there will be a mass of \( (1 - y^{-1}) \) at 0 in the limiting measure. Since the nonzero eigenvalues of \( XX^T \) and \( X^TX \) are same we can say that in this case the limiting distribution is

\[ (1 - y^{-1})\delta_0 + \mu \]

where \( \mu \) satisfies (1)

**Remark 3.** Observe that if \( y = 1 \), then \( a = 0, b = 4 \), and thus

\[ \frac{d\mu}{dx} = \frac{1}{2\pi x} \sqrt{(4 - x)x}1_{(0 \leq x \leq 4)} \]

In this case \( \mu \) is the image of semicircle distribution under the mapping \( x \to x^2 \)
Proof. : We now begin the proof of Marchenko Pastur Law. Since the support of $\mu$ is compact, $\mu$ is uniquely determined by its moments. So as in the Wigners case it is enough to show

$$\int x^k \, d\mu_n \to \int x^k \, d\mu$$

Again following Wigner’s case, Borel Cantelli lemma says it is enough to show the following

1. $E \int x^k \, d\mu_n \to \int x^k \, d\mu$

2. $\text{Var}(\int x^k \, d\mu_n) \leq \frac{C_k}{n^2}$

Computation of the second integral in 1 will show that

$$\int x^k \, d\mu = \sum_{r=0}^{k-1} \frac{y^r}{r+1} \binom{k}{r} \binom{k-1}{r}$$

Now notice that

$$E \int x^k \, d\mu_n = \frac{1}{p} E\left(\sum_{i=1}^{p} \lambda_i^p\right) = \frac{1}{p} E[\text{Tr}(XX^T/n)^k]$$

$$= \frac{1}{pn^k} E\left[\sum_{I,J} X_{i_1j_1}X_{i_2j_2}X_{i_3j_3} \ldots X_{i_kj_k}X_{i_1j_1}\right] \equiv \frac{1}{pn^k} \sum_{I,J} E(I, J)$$

where $I \in [p]^k$ and $J \in [n]^k$.

Now this corresponds to a directed loop on a bipartite graph. For example if $k = 4$ then for typical $\{i_1, i_2, i_3, i_4\}$ and $\{j_1, j_2, j_3, j_4\}$ we have the following picture.

As in the Wigner’s case we see that each edge must appear at least twice, otherwise $E(I, J) = 0$. Now we have $2k$ steps in the directed loop. Thus we see that we have at most $k$ edges in the
skeleton, hence at most \( k + 1 \) vertices in the skeleton.

Next assume that number of vertices = \( m \leq k \). Let \( m = a + b \) where \( a = \# \) of \( I \) vertices and \( b = \# \) of \( J \) vertices. Then the total number of ways choosing \( a \) \( I \) vertices and \( b \) \( J \) vertices \( \leq C p^a n^b \), where \( C \) is a constant independent of \( n \). The contribution of these terms in the expectation \( \leq C' p^a n^b / p n^k \to 0 \) as \( n \to \infty \).

Thus we need to look at loops which have exactly \( k + 1 \) vertices and \( k \) edges. These are exactly the double trees.

Reshuffle them to get the following structure.

![Diagram of double trees](image)

Start with an \( I \) vertex. Vertices that can be reached in even steps are the \( I \) vertices, the rest are the \( J \) vertices.

Next we ask the question: How many double trees are there for a given shape? Here by the shape of a tree we mean the vertices numbered in order of appearance. For example

\[
2 \ 3 \ 4 \ 5 \ 4 \ 6 \ 7 \ 6 \ 8 \ 6 \ 4 \ 3 \ 9 \ 3 \ 10 \ 11 \ 10 \ 3 \ 2, \quad 12 \ 13 \ 14 \ 15 \ 16 \ 17 \ 16 \ 18 \ 16 \ 14 \ 13 \ 19 \ 13 \ 6 \ 7 \ 6 \ 13 \ 12
\]

will give us the same shape, because after renumbering in order of appearance both will give us the following double tree

\[
1 \ 2 \ 3 \ 4 \ 3 \ 5 \ 6 \ 5 \ 7 \ 5 \ 3 \ 2 \ 8 \ 2 \ 9 \ 10 \ 9 \ 2 \ 1
\]

and all of them look like the figure above. Thats is we have to choose \( r + 1 \) \( I \) vertices from \([p]\) and \( k - r \) \( J \) vertices from \([n]\). This can be done in \( P(p, r + 1)P(n, k - r) \) where \( P(n, k) = n(n-1)\ldots(n-k+1) \) is permutation of \( k \) objects from \( n \) distinct objects. Notice that

\[
P(p, r + 1)P(n, k - r) = np^k y_n^r (1 + O(n^{-1})), \quad \text{where} \quad y_n = p/n
\]

Thus

\[
E(\int x^k \, d\mu_n) = \frac{1}{p n^k} \sum_{I,J} E(I, J) = \sum_{r=0}^{k-1} y_n^r (1 + O(n^{-1})) \times \#\{\text{double tree shapes with } r+1 \ I \text{ and } k-r \ J \text{ vertices}\}
\]
Since \( y_n \to y \), as \( n \to \infty \), it's clear that all we need now is to show that

\[
\# \{ \text{double tree shapes with } r + 1 \ I \text{ and } k - r \ J \text{ vertices} \} = \frac{1}{r + 1} \binom{k}{r} \binom{k - 1}{r}
\]

Towards this end we try to correspond each double tree shape with the following type of path/sequence of \( 2k \) steps.

1. If \( i \) is odd then \( s_i \in \{-1, 0\} \)
2. If \( i \) is even then \( s_i \in \{0, 1\} \), \( s_{2k} = 0 \)
3. For any \( l = 1, 2, \ldots, 2k \), we have \( \sum_{i=1}^{l} s_i \geq 0 \). That is the path is never below 0.
4. \( \# \{ i : s_i = 1 \} = \# \{ i : s_i = -1 \} = r \). That is there are exactly \( r \) upsteps and \( r \) downsteps
5. \( \sum_{i=1}^{2k} s_i = 0 \). That is we return to 0 at the end.

Given any such sequence \( \{s_i\}_{i=1}^{2k} \), clearly we can construct a tree as following:

- Suppose \( i \) is odd. If \( s_i = -1 \) then we go down the double tree, if \( s_i = 0 \) then we go up from an \( I \) vertex but we will return
- Suppose now \( i \) is even. If \( s_i = 1 \) then we go one step up in the double tree. If \( s_i = 0 \) then we go one step down

Next given a double tree shape we construct such a sequence \( \{s_i\}_{i=1}^{2k} \). First for each \( I \) vertex we mark the first edge leading to it and the last edge leaving it. After marking the previous double will look like the following. The circled vertices are the \( I \) vertices.

![Diagram](image.png)

Now put \( s_i = 1 \) if the \( i \)-th edge is marked and its going up, \( s_i = -1 \) if \( i \)-th edge is marked and going down, \( s_i = 0 \) otherwise. For example the above double tree will give the following path.
We have to verify this allocation of \(-1, 0, 1\) would still make \(\{s_i\}_{i=1}^{2k}\) satisfy condition (3). Suppose if possible we have a first \(l\) such that

\[
\sum_{i=1}^{2l-1} s_i = -1, \quad \text{hence} \quad \sum_{i=1}^{2l} s_i = 0, \quad \text{and} \quad s_{2l-1} = -1
\]

Then the other part tells us that we can construct a double tree with vertices \(\{1, 2, \ldots, 2l\}\), and since \(s_{2l-1} = -1\), the second bullet from the other part says that we are not moving up to a new vertex, but going down to an old vertex in \(\{1, 2, \ldots, 2l\}\). But this destroys the double tree shape giving us a contradiction. Hence we see that if we allocate \(-1, 0, 1\) by the above rule then we indeed get a sequence \(\{s_i\}_{i=1}^{2k}\) satisfying required conditions.

Thus the set of the double tree shapes is in bijection with the set of sequence \(\{s_i\}_{i=1}^{2k}\), so all we need to do now is count such sequences \(\{s_i\}_{i=1}^{2k}\).

Since \(s_{2k} \neq +1\), not considering condition (3) for the moment we see that out of \(k - 1\) positions for \(+1\) and \(k\) positions for \(-1\) we have to choose \(r\) each. Therefore the number of such sequences is \({k-1 \choose r}{k \choose r}\).

Let’s now count the number of sequences which fail condition (3). Since those paths hit \(-1\) there exists a first \(l\), such that \(\sum_{i=1}^{2l-1} s_i = -1\) (By construction of the sequence \(s_k\) can be \(-1\) only when \(k\) is odd.) We now construct a new sequence \(\{s'_i\}_{i=1}^{2k}\) by ‘reflection’. Put

\[s'_i = s_i, \quad \text{for } i = 1, \ldots, 2l - 1, \quad s'_i = s_{2k} = 0\]

For \(l \leq i \leq k - 1\), put

\[
(s'_{2i}, s'_{2i+1}) = (1, -1) \quad \text{if} \quad (s_{2i}, s_{2i+1}) = (1, -1)
\]
\[
= (0, 0) \quad \text{if} \quad (s_{2i}, s_{2i+1}) = (0, 0)
\]
\[
= (1, 0) \quad \text{if} \quad (s_{2i}, s_{2i+1}) = (0, -1)
\]
\[
= (0, -1) \quad \text{if} \quad (s_{2i}, s_{2i+1}) = (1, -1)
\]

Clearly the set of all sequences \(\{s'_i\}_{i=1}^{2k}\), which fail condition (3) is in bijection with the set of sequences \(\{s'_i\}_{i=1}^{2k}\) but to count the number of such sequences \(\{s'_i\}_{i=1}^{2k}\) we just have to count the number of ways we can choose \(r - 1\) ‘+1’ from \(k - 1\) of them, and \(r + 1\) ‘-1’ from \(k\) of them. This can be done in \({k-1 \choose r-1}{r+1 \choose r}\). So the total number of sequences \(\{s'_i\}_{i=1}^{2k}\) which satisfies condition (1-4) is given by

\[
\binom{k-1}{r} \binom{k}{r} - \binom{k-1}{r-1} \binom{k}{r+1} = \frac{1}{r+1} \binom{k-1}{r} \binom{k}{r}
\]

This proves the fact about expectation and the proof of the variance bound is similar to that of Wigner Matrix. \(\square\)
We now move to some particular type of random matrices, namely the **Gaussian Ensembles**.

**Gaussian Orthogonal Ensemble (GOE):** Here we look at matrices $M_n$ of the form $M_n = [X_{i,j}]_{i,j=1}^n$ where

$$X_{i,j} = X_{j,i}, \quad X_{i,j} \overset{iid}{\sim} N(0,1), \quad i < j, \quad \text{and} \quad X_{i,i} \sim \sqrt{2}N(0,1)$$

and they are all independent.

We can construct them in following way. Take a matrix $A = [Y_{i,j}]_{i,j=1}^n$, where $Y_{i,j} \overset{iid}{\sim} N(0,1)$. Then

$$M_n = (A + A^T)/\sqrt{2}$$

is a GOE.

**Gaussian Unitary Ensemble (GUE):** These are very similar to GOE. Here we look at matrices $M_n$ of the form $M_n = [X_{i,j}]_{i,j=1}^n$ where

$$X_{i,j} = \bar{X}_{j,i}, \quad X_{i,j} \sim N(0,1/2) + iN(0,1/2), \quad i < j, \quad \text{and} \quad X_{i,i} \sim N(0,1)$$

and they are all independent. We can construct them in following way. Take a matrix $A = [Y_{i,j}]_{i,j=1}^n$, where $Y_{i,j} \overset{iid}{\sim} N(0,1/2) + iN(0,1/2)$. Then

$$M_n = (A + A^*)/\sqrt{2}$$

is GUE.

**Gaussian Symplectic Ensemble (GSE)** Define $Z$ as the following block diagonal matrix

$$Z_{2n \times 2n} = diag(A, A \ldots A), \quad \text{where}$$

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

Call a matrix $M \in \mathbb{C}^{2n \times 2n}$ symplectic if

$$Z = MZM^T$$

We next define the space of quaternions. Define the following $2 \times 2$ matrices

$$e_1 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad 1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Note that

$$e_1 \cdot e_2 = -e_2 \cdot e_1 = e_3, \quad e_1^2 = e_2^2 = e_3^2 = -1$$
The conjugation rule is as follows

\[ \bar{1} = 1, \bar{e}_2 = -e_2, \bar{e}_3 = -e_3, \bar{e}_4 = -e_4 \]

The vector space generated by \( \{e_1, e_2, e_3, 1\} \) over \( \mathbb{C} \) is called the space of quaternions.

A quaternion \( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \) is real if \( q^{(1)}, q^{(2)}, q^{(3)}, q^{(4)} \) are real where

\[
\begin{bmatrix} a & b \\ c & d \end{bmatrix} = q^{(1)} \cdot 1 + q^{(2)} \cdot e_2 + q^{(3)} \cdot e_3 + q^{(4)} \cdot e_4
\]

A random quaternion \( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \) is called real standard quaternion if its real and if

\[
q^{(1)}, q^{(2)}, q^{(3)}, q^{(4)} \sim \text{iid } N(0, 1/4)
\]

Let \( Q_{i,j} = q_{i,j}^{(1)} \cdot 1 + q_{i,j}^{(2)} \cdot e_2 + q_{i,j}^{(3)} \cdot e_3 + q_{i,j}^{(4)} \cdot e_4 \)

A GSE is defined by \( M_n = [Q_{i,j}]_{i,j=1}^n \) where for \( i < j \), \( Q_{i,j} \) are iid standard quaternions, \( Q_{i,j} = Q_{j,i} \), and on the diagonal \( i = j \) we have \( q_{i,i}^{(0)} \sim N(0, 1/2) \). We can construct such a matrix as follows. Let \( A = [Y_{i,j}]_{i,j=1}^n \) where \( Y_{i,j} \) are iid real standard quaternions. Then \( M_n = (A + A^*) / \sqrt{2} \) is GSE.

Let \( dM \) be the reference lebesgue measure, based on the determining entries. Define the density function w.r.t \( dM \) as

\[
\frac{1}{Z_{n,\beta}} \exp(-\frac{\beta}{4} \text{Tr}(M^2))
\]

Then this defines the density of GOE, GUE and GSE for \( \beta = 1, 2, 4 \) respectively.