Solution Sheet III

Introduction to Random Matrices, IST Austria
Published November 9, 2017. *Due* November 21, 2017.
LÁSZLÓ ERDŐS, DOMINIK SCHRÖDER

1 Limiting density for XX^t random matrices

Let $X = (x_{ij})_{i\in[M],j\in[N]}$ be a (non-symmetric) random matrices with $M = M(N) \le N$, and independent, identically distributed entries

$$x_{ij} \in \mathbb{R}, \qquad \mathbf{E}\, x_{ij} = 0, \qquad \mathbf{E}\, x_{ij}^2 = 1, \qquad |x_{ij}| \le C.$$

The symmetric random matrix $H = N^{-1}XX^t \in \mathbb{R}^{M\times M}$ then has real eigenvalues $\lambda_1 \le \dots \le \lambda_M$. The goal of this exercise is to show that in the asymptotic scaling $M/N = \lambda + \mathcal{O}\left(N^{-1}\right)$ with $\lambda \in (0,1]$ the empirical spectral density $\mu_N = M^{-1}\sum_i \delta_{\lambda_i}$ of H converges weakly in probability to the absolutely continues measure $\mathrm{d}\mu = \rho\,\mathrm{d}x$ with density

$$\rho(x) = \frac{1}{2\lambda\pi x}\sqrt{(b-x)_+(x-a)_+}, \qquad a = \left(1-\sqrt{\lambda}\right)^2, \quad b = \left(1+\sqrt{\lambda}\right)^2. \tag{1}$$

Problem 1. *We define the empirical moments to be*

$$m_{k,N} := \int x^k \,\mathrm{d}\mu_N(x) = \frac{1}{MN^k}\sum_{\substack{i_1,\dots,i_k\in[M]\\ j_1,\dots,j_k\in[N]}} x_{i_1j_1}x_{i_2j_1}x_{i_2j_2}x_{i_3j_2}\dots x_{i_kj_k}x_{i_1j_k}.$$

(i) Use similar arguments to those in the moment computation for the Wigner semicircle law to show that

$$\mathbf{E}\, m_{k,N} = m_k + \mathcal{O}\left(N^{-1}\right), \qquad m_k = \sum_{2k-\text{Dyck paths}} \lambda^{u_{even}},$$

where u_{even} counts the number of upstrokes in even steps of the Dyck path.

(ii) Define the auxiliary quantity

$$m_k' := \sum_{2k-\text{Dyck paths}} \lambda^{u_{odd}}.$$

and prove the recursion

$$m_k = \sum_{j=1}^k m_{j-1}' m_{k-j}, \qquad m_k' = \lambda\sum_{j=1}^k m_{j-1}m_{k-j}', \; k\ge 1, \qquad m_0 = m_0' = 1$$

to conclude that the generating function $f(x) = \sum_{k\ge 0} m_k x^k$ of m_k satisfies the equation

$$f(x) = 1 + (1-\lambda)xf(x) + \lambda x f(x)^2.$$

(iii) Relate the Stieltjes transform of μ to f to conclude that the density of μ is indeed given by (1).

(iv) Show that $\mathbf{Var}\, m_{k,N} = \mathcal{O}\left(N^{-2}\right)$ *and conclude that $m_{k,N}$ converges almost surely to m_k for each k.* ***Hint.*** *The Borel-Cantelli Lemma might be helpful.*

(v) Conclude that almost surely μ_N converges weakly to μ. ***Hint.*** *You may assume that μ is uniquely determined by its moments and therefore convergence of moments implies weak convergence.*

Solution. (i) We compute the k-th moment of $\mathbf{E}\,\mu_N$ as

$$\mathbf{E}\int x^k\,\mathrm{d}\mu_N(x) \coloneqq \mathbf{E}\,m_{k,N} = \frac{1}{M}\,\mathbf{E}\,\mathrm{Tr}\,H^k = \frac{1}{MN^k}\sum_{\substack{i_1,\dots,i_k\in[M]\\ j_1,\dots,j_k\in[N]}}\mathbf{E}\,x_{i_1j_1}x_{i_2j_1}x_{i_2j_2}x_{i_3j_2}\dots x_{i_kj_k}x_{i_1j_k}. \tag{2}$$

Graphically this corresponds to a looped bipartite directed graph on the independent vertex sets $i_1,\dots,i_k$ and $j_1,\dots,j_k$, which in the sequel we shall call i- and j-vertices. By independence it follows that each of the $2k$ edges occurs at least twice and therefore the skeleton of this graph has at most k edges, hence at most $k+1$ vertices. If the graph has $n_i+n_j=n<k+1$ vertices with n_i type i and n_j type j ones, then those terms in (2) can be estimated by $M^{a-1}N^{b-k}C^{2k}\lesssim N^{m-1-k}\le N^{-1}$. Thus it suffices to estimate those loops with exactly k edges and $k+1$ vertices which are exactly graphs whose skeleton is a tree.

Due to the bipartite structure of the graph it follows that there are $1\le n_i+1\le k$ type-i and $1\le n_j=k-n_i\le k$ type-j vertices. The number of index assignments to such a tree is given by $M^{n_i+1}N^{k-n_i}+\mathcal{O}\left(N^k\right)$ and thus

$$\mathbf{E}\int x^k\,\mathrm{d}\mu_N(x) = \sum_{n_i=0}^{k-1}\lambda^{n_i}\times\#\Big\{\text{backtracking trees with } n_i+1 \text{ type-}i \text{ and } n_j=k-n_i \text{ type-}j \text{ vertices}\Big\}+\mathcal{O}\left(N^{-1}\right).$$

As shown in the lecture the backtracking trees with $k+1$ vertices are in one-to-one correspondence to Dyck paths of length $2k$. We now note that in that correspondence n_i is the number of up-strokes in even steps of the Dyck path and $n_j=k-n_i$ is the number of up-strokes in odd steps of the Dyck path. Indeed, even steps in the Dyck path correspond to steps to type-i vertices and up-strokes correspond to the first visit of a previously unexplored vertex. Moreover, the very first (type-i) vertex is not reached via an upstroke and therefore the n_i+1 type-i vertices correspond to n_i upstrokes in even steps. Therefore we have that

$$\mathbf{E}\,m_{k,N}=m_k+\mathcal{O}\left(N^{-1}\right),\qquad m_k=\sum_{2k-\text{Dyck paths}}\lambda^{u_{\text{even}}},$$

where u_{even} counts the number of even upstrokes.

(ii) We now want to derive a recursion for m_k. To that end we first introduce a complementary quantity counting the odd upstrokes of a path,

$$m_k' \coloneqq \sum_{2k-\text{Dyck paths}}\lambda^{u_{\text{odd}}}.$$

For each Dyck path there is a first index $1\le j\le k$ such that the path returns to zero in the $2j$-th step. Then the subpaths from $(1,1)$ to $(1,2j-1)$ and from $(2j,0)$ to $(2k,0)$ are again Dyck paths of length $2(j-1)$ and $2(k-j)$, where the number of even upstrokes of the original path is the number of odd upstrokes in the first subpath plus the number of even upstrokes in the second subpath. Conversely the number of odd up-strokes in the original path is 1 plus the number of even upstrokes in the first subpath plus the number of odd up-strokes in the second subpath. Consequently we have the recursion formulas

$$m_k=\sum_{j=1}^k m_{j-1}'m_{k-j},\qquad m_k'=\lambda\sum_{j=1}^k m_{j-1}m_{k-j}'=\lambda\sum_{j=1}^k m_{k-j}m_{j-1}'=\lambda m_k,\ k\ge 1,\qquad m_0=m_0'=1.$$

It now follows that

$$m_k=(1-\lambda)m_{k-1}+\lambda\sum_{j=1}^k m_{j-1}m_{k-j}.$$

This implies that the generating function $f(x)=\sum_{k\ge0}m_kx^k$ satisfies the equation

$$f(x)=\sum_{k\ge0}m_kx^k=1+(1-\lambda)x\sum_{k\ge1}m_{k-1}x^{k-1}+\lambda x\sum_{k\ge1}\sum_{j=1}^k m_{j-1}x^{j-1}m_{k-j}x^{k-j}=1+(1-\lambda)xf(x)+\lambda xf(x)^2.$$

Solving for $f(x)$ we find

$$f(x)=\frac{1-x+\lambda x\pm\sqrt{(1-x+x\lambda)^2-4x\lambda}}{2x\lambda}. \tag{3}$$

(iii) This is related to the Stieltjes transform m_μ of the limiting density via

$$m_\mu(z) = \int \frac{1}{x-z} \,\mathrm{d}\mu(x) = -\frac{1}{z}\sum_{k\geq 0} \int (x/z)^k \,\mathrm{d}\mu(x) = -\frac{1}{z}\sum_{k\geq 0} m_k z^{-k} = -\frac{f(1/z)}{z}$$

and we can recover the density ρ of μ by taking the limit

$$\rho(x) = \frac{1}{\pi}\lim_{\eta\searrow 0} m_\mu(x+i\eta) = \frac{\sqrt{\left[4\lambda x - (-1+x+\lambda)^2\right]_+}}{2\lambda\pi x}.$$

Strictly speaking, one has to choose the correct branch of the complex square-root in this problem. The square-root in (3) is defined to have a branch-cut along the negative real axis such that

$$\sqrt{z} = \sqrt{\frac{|z|+\Re z}{2}} \pm i\sqrt{\frac{|z|-\Re z}{2}},$$

where the principal choice would be to choose $+$ whenever $\Im z > 0$ or $\Im z = 0, \Re z < 0$. It turns out that the correct choice is completely determined by the fact that m_μ is a Stieltjes transform. We have that

$$m_\mu(z) = \frac{z-1+\lambda \pm \sqrt{(z-1+\lambda)^2 - 4\lambda z}}{2\lambda z} = \frac{z-1+\lambda - \sqrt{(z-b)(z-a)}}{2\lambda z}, \tag{4}$$

where in the second equality we were forced to choose $-$ in the $\pm$ as for large real z, $\Im m_\mu$ has to tend to zero for all λ. For the square-root we choose the branch-cut of $\sqrt{(z-b)(z-a)}$ in such a way that the discontinuity is in $[a,b]$ and $\Im m_\mu > 0$ whenever $\Im z > 0$. Thus $\Im m_\mu(z)$ for small $\Im z > 0$ is given by

$$\Im m_\mu(z) \approx \frac{1}{2\lambda \Re z}\sqrt{\frac{|f(z)| - \Re f(z)}{2}}, \qquad f(z) = (z-b)(z-a)$$

and (1) follows since $\Re f(z) \approx -|f(z)| \approx (b-x)(x-a)$ for $x = \Re z \in (a,b)$ and $\Re f(z) \approx |f(z)|$ otherwise.

(iv) The variance of the k-th moment is given by

$$\mathbf{Var} \int x^k \,\mathrm{d}\mu_N(x) = \frac{1}{M^2 N^{2k}} \sum_{\substack{\boldsymbol{i},\boldsymbol{i}'\in[M]^k \\ \boldsymbol{j},\boldsymbol{j}'\in[N]^k}} \mathbf{Cov}\left(x_{\boldsymbol{i},\boldsymbol{j}}, x_{\boldsymbol{i}',\boldsymbol{j}'}\right),$$

where $x_{\boldsymbol{i},\boldsymbol{j}}$ are defined as in (2). We make two observations: (i) If some edge (ij) is only visited once, then the corresponding covariance vanishes, (ii) If the graphs generated by $(\boldsymbol{i},\boldsymbol{j})$ and $(\boldsymbol{i}',\boldsymbol{j}')$ have no edges in common, then the covariance vanishes by independence. Those terms with $n \leq 2k$ unique vertices are trivially of size $\mathcal{O}\left(N^{-2}\right)$. From (i) it follows that there are at most $2k$ unique edges and therefore at most $2k+2$ unique vertices since the graph has at most 2 connected components. If it has two connected component, then it gives no contribution due to (ii) and since those graphs with at most $2k$ unique vertices contribute at most $\mathcal{O}\left(N^{-2}\right)$ it only remains to control those graphs with exactly $2k+1$ unique vertices. This is only the case for trees where each edge occurs exactly twice. Since both $(\boldsymbol{i},\boldsymbol{j})$ and $(\boldsymbol{i}',\boldsymbol{j}')$ are cycles it follows that they have to visit each of its edges exactly twice which implies that each each can only be visited by either $(\boldsymbol{i},\boldsymbol{j})$ or $(\boldsymbol{i}',\boldsymbol{j}')$ but never both. According to (ii) the contribution of these graphs vanishes as well and we have that

$$\mathbf{Var}\, m_{k,N} = \mathcal{O}\left(N^{-2}\right).$$

It follows from Chebyshev's inequality that

$$\mathbf{P}\left(|m_{k,N} - \mathbf{E}\, m_{k,N}| > \epsilon\right) \leq \frac{\mathbf{Var}\, m_{k,N}}{\epsilon^2}$$

and therefore from the Borel-Cantelli Lemma that $m_{k,N}$ for each k converges almost surely to m_k as $N\to\infty$. Consequently by taking the union of countably many events of probability 1 it follows that on a set of probability 1, $m_{k,N}$ converges to m_k for all k. Because μ is uniquely determined by its moments it follows that on this set of probability 1, μ_N converges weakly to μ.

(v) Using Carleman's continuity condition we can check that μ is uniquely determined by its moments and therefore on a set of probability 1, μ_N converges weakly to μ.

$\square$

Problem 2. *Use Problem 1 to show that for square X as in Problem 1, the singular value distribution of $N^{-1/2}X$ converges weakly to a quarter-circle distribution with density*

$$\frac{\sqrt{(4-s^2)_+}}{\pi}\mathbf{1}_{s\geq 0}\,\mathrm{d}s.$$

Solution. For square i.i.d. random matrices X (where $\lambda = 1$) the distribution of the eigenvalues of $N^{-1}XX^t$ simplifies to

$$\frac{\sqrt{(4-x)_+ x_+}}{2\pi x}\,\mathrm{d}x$$

and we can conclude after a change of variables $x = s^2$, $\mathrm{d}x = 2s\,\mathrm{d}s$ for $s \geq 0$ that the singular values $s_1, \ldots, s_N$ of $N^{-1/2}X$ are approximately distributed in a quarter circle distribution

$$\frac{\sqrt{(4-s^2)_+}}{\pi}\mathbf{1}_{s\geq 0}\,\mathrm{d}s.$$

$\square$

2 Operator norm of Wigner matrices

In the lecture it was proved that the empirical spectral density $\mu_N = N^{-1}\sum_i \delta_{\lambda_i}$ concentrated in the eigenvalues $\lambda_1 \leq \cdots \leq \lambda_N$ of a Wigner matrix H converges weakly in probability to a semi-circular distribution $\mathrm{d}\mu_{sc} = \rho_{sc}(x)\,\mathrm{d}x$ with density

$$\rho_{sc}(x) = \frac{\sqrt{(4-x^2)_+}}{2\pi}.$$

This might suggest that λ_N converges to 2 in probability. This is indeed the case but requires an extra argument. The lower bound, however, follows directly from the weak convergence in probability:

Problem 3. *Prove that for each $\epsilon > 0$,*

$$\lim_{N\to\infty} \mathbf{P}(\lambda_N < 2-\epsilon) = 0. \tag{5}$$

Solution. Choose any continuous test function f supported in $[2-\epsilon, 2]$ such that $\int f\rho_{sc}\,\mathrm{d}x = 1$. Then we have that

$$\mathbf{P}(\lambda_N < 2-\epsilon) \leq \mathbf{P}\left(\int f(x)\,\mathrm{d}\mu_N(x) = 0\right) \leq \mathbf{P}\left(\left|\int f(x)\,\mathrm{d}(\mu_N - \mu_{sc})(x)\right| \geq \frac{1}{2}\right)$$

and the right hand side converges to 0 because μ_N converges to μ_{sc} weakly in probability. $\square$

For the upper bound we impose additional assumptions on the growth of moments. Specifically, we will assume that

$$\sup_{ij} \mathbf{E}\left|\sqrt{N}h_{ij}\right|^k \leq \mu_k \leq k^{Ck} \tag{6}$$

for some constant C.

Problem 4. *Prove that for each $\epsilon > 0$,*

$$\lim_{N\to\infty} \mathbf{P}(\lambda_N > 2+\epsilon) = 0. \tag{7}$$

Hint. *In the lecture it was proved that for each fixed k, the empirical moments $m_{k,N} = N^{-1}\,\mathbf{E}\,\mathrm{Tr}\,H^k$ converge to $C_{k/2}$ for even k and 0 for odd k. This was achieved by counting the so called back-tracking graphs which form the leading term in the expansion*

$$N^{-1}\,\mathbf{E}\,\mathrm{Tr}\,H^k = N^{-1}\sum_{i_1,\ldots,i_k\in[n]} \mathbf{E}\,h_{i_1 i_2}\ldots h_{i_k i_1}. \tag{8}$$

Using a more careful counting of the remaining graphs (i.e., all graphs in which each edge occurs at least twice which are not backtracking trees) this can be improved to also cover slowly growing k. One can show[1] *that the number $N_{k,j}$ of cycles $i_1, \ldots, i_k, i_1$ which visit j unique vertices and pass along each edge at least twice is bounded by*

$$N_{k,j} \leq 2^k k^{3(k-2j+2)} N^j, \quad j < \frac{k}{2} + 1, \qquad N_{k,\lfloor k/2+1 \rfloor} \leq C_{k/2} N^{\lfloor k/2+1 \rfloor} \leq 2^k N^{\lfloor k/2+1 \rfloor}. \tag{9}$$

Use the combinatorial bound (9) *in* (8) *to prove that $N^{-1} \,\mathbf{E}\, \mathrm{Tr}\, H^k \leq C' 2^k$ for $k \sim \log^2 N$.*

Solution. Denote the number of edges which are visited twice by l. Then it follows from the monotonicity of moments that

$$|\mathbf{E}\, h_{i_1 i_2} \ldots h_{i_k i_1}| \leq N^{-k/2} \mu_{k-2l}.$$

Since all other edges are visited at least three times we have $k \geq 3(j-1-l) + 2l = 3(j-1) - l$ and it follows that $k - 2l \leq 6(k/2 + 1 - j)$. Thus we have that

$$N^{-1} \,\mathbf{E}\, \mathrm{Tr}\, H^k = \mathcal{O}\left(\frac{1}{N^{k/2+1}} \sum_{1 \leq j \leq k/2+1} N_{k,j} \mu_{6(k/2+1-j)} \right).$$

and we compute

$$\frac{1}{N^{k/2+1}} \sum_{1 \leq j \leq k/2+1} N_{k,j} \mu_{6(k/2+1-j)} \leq 2^k \sum_{1 \leq j \leq k/2+1} \left(\frac{k^6 [6(k/2+1-j)]^{6C}}{N} \right)^{k/2+1-j} \leq 2^k \sum_{0 \leq j \leq k/2} \left(\frac{k^{C'}}{N} \right)^j$$

for some constant C'. We now choose $k = k(N)$ such that $k^{C'}/N \to 0$ but $k/\log N \to \infty$ as $N \to \infty$ (for example $k = \log^2 N$ would work). Then it holds that

$$\mathbf{P}(\lambda_N > 2 + \epsilon) \leq \mathbf{P}\left(\mathrm{Tr}\, H^k > (2+\epsilon)^k \right) \leq \frac{\mathbf{E}\, \mathrm{Tr}\, H^k}{(2+\epsilon)^k} \leq 2 \frac{N 2^k}{(2+\epsilon)^k} \to 0.$$

$\square$

[1] For those interested this combinatorial argument with a slightly less precise bound is spelled out in the proof of Theorem 12 of `https://terrytao.wordpress.com/2010/01/09/254a-notes-3-the-operator-norm-of-a-random-matrix`