Assignment 2

Random Matrix Theory in Data Science and Statistics

(EN.553.796, Fall 2025)

Assigned: September 29, 2025   Due: 11:59pm EST, October 16, 2025

Solve all problems. Each problem is worth an equal amount towards your grade.

Submit solutions in LaTeX. Write in complete sentences. Include and justify all steps of your arguments, but avoid writing excessive discussion that is not contributing to your solution. You are welcome to include images if you think that will help explain your solutions.

Problem 1 (Project proposal)**.** Propose a paper or collection of related papers to read, a computational experiment to perform, or an open problem to study for your final project. See the course website for ideas. Write roughly one paragraph here doing one of the following:

- If you plan to read a paper, tell us the title and author(s), read the abstract and introduction, and describe how it relates to the course, what aspect of the paper you are interested in, and what else you might have to read or do to understand it.
- If you plan to run a computational experiment, describe the experiment, give one or two relevant references that might help set your expectations for what you will find, and explain what the experiment will tell you about random matrices.
- If you plan to try working on an open problem, write down the problem, give a few references concerning it, and outline in one paragraph what approach you plan to try or what experiments you can perform.

In all cases, your final project will consist of a short presentation at the end of the class (about 10 minutes) and a short write-up about whatever you choose here.

Problem 2 (Invariance)**.** We have repeatedly seen various notions of orthogonal invariance in class. This problem will ask you to work out some of their properties.

1. Suppose $\boldsymbol{x} \in \mathbb{R}^d$ is a random vector whose law has a smooth density ρ on $\mathbb{R}^d$ and is orthogonally invariant, i.e., having $\mathsf{Law}(\boldsymbol{Q}\boldsymbol{x}) = \mathsf{Law}(\boldsymbol{x})$ for all $\boldsymbol{Q} \in \mathcal{O}(d)$. Show that the scalar $\|\boldsymbol{x}\|$ and the vector $\hat{\boldsymbol{x}} := \boldsymbol{x}/\|\boldsymbol{x}\|$ are independent random variables.

   You do not need to show it here, but you should be aware that a similar general fact (though a little tricky to state precisely) is true of random matrices: the eigen- or singular values of an orthogonally invariant matrix are independent of the eigen- or singular vectors.

2. Consider the following abstracted version of a situation we encountered in discussing the singular vectors of a Gaussian random matrix. Suppose that $\boldsymbol{L} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$ is a random matrix that always takes values having $\mathsf{rank}(\boldsymbol{L}) = 1$ and $\lambda_1(\boldsymbol{L}) = 1$, i.e., $\boldsymbol{L}$ is a rank-one projection matrix. There then exist exactly two distinct unit vectors $\boldsymbol{u}_1, \boldsymbol{u}_2$ such that $\boldsymbol{L}\boldsymbol{u}_i = \boldsymbol{u}_i$, and $\boldsymbol{u}_1 = -\boldsymbol{u}_2$. Draw $i \sim \mathsf{Unif}(\{1, 2\})$ independently of $\boldsymbol{L}$, and define $\widetilde{\boldsymbol{u}} := \boldsymbol{u}_i$.

   Suppose that $\boldsymbol{L}$ is symmetric-orthogonally invariant, i.e., has $\mathsf{Law}(\boldsymbol{Q}\boldsymbol{L}\boldsymbol{Q}^\top) = \mathsf{Law}(\boldsymbol{L})$ for all $\boldsymbol{Q} \in \mathcal{O}(d)$. Show (precisely!) that the law of $\widetilde{\boldsymbol{u}}$ is then orthogonally invariant, i.e., having $\mathsf{Law}(\boldsymbol{Q}\widetilde{\boldsymbol{u}}) = \mathsf{Law}(\widetilde{\boldsymbol{u}})$.

   (**Hint:** Make an unambiguous definition of $\boldsymbol{u}_1 = \boldsymbol{u}_1(\boldsymbol{L})$ in the above setting, breaking the symmetry between $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. Then, express $\mathbb{P}[\widetilde{\boldsymbol{u}} \in A]$ first in terms of probabilities involving $\boldsymbol{u}_1$, and then in terms of probabilities involving $\boldsymbol{L} = \boldsymbol{u}_1\boldsymbol{u}_1^\top$.)

3. Let $\boldsymbol{Q} \sim \mathsf{Haar}(\mathcal{O}(d))$. Calculate the entrywise moments

$$\mathbb{E}Q_{ij}, \quad \mathbb{E}Q_{ij}Q_{k\ell}, \quad \mathbb{E}Q_{ij}Q_{k\ell}Q_{mn}, \quad \mathbb{E}Q_{ij}Q_{k\ell}Q_{mn}Q_{rs},$$

   in terms of d and the indices $i, j, k, \ell, m, n, r, s \in [d]$.

   (**Hint:** Use that the Haar measure is invariant, so $\mathsf{Law}(\boldsymbol{A}\boldsymbol{Q}\boldsymbol{B}) = \mathsf{Law}(\boldsymbol{Q})$ for all $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{O}(d)$. Make some simple choices of $\boldsymbol{A}$ and $\boldsymbol{B}$, and also use the relation $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}_d$, expanded into polynomials of the entries of $\boldsymbol{Q}$. You should find that, for most choices of indices, these moments are zero.)

4. Let $\boldsymbol{U} \sim \mathsf{Haar}(\mathsf{Stief}(m, d))$ for some $1 \leq d \leq m$, and let $\boldsymbol{x} \in \mathbb{S}^{m-1} \subset \mathbb{R}^m$ be deterministic. Let $\boldsymbol{P} := \boldsymbol{U}\boldsymbol{U}^\top \in \mathbb{R}^{m\times m}_{\mathsf{sym}}$ and note that this is a projection matrix. Let $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_m)$. Show that

$$\mathsf{Law}(\|\boldsymbol{P}\boldsymbol{x}\|^2) = \mathsf{Law}\left(\frac{g_1^2 + \cdots + g_d^2}{g_1^2 + \cdots + g_d^2 + g_{d+1}^2 + \cdots + g_m^2}\right).$$

   State and prove a concentration inequality expressing that, if d and m are both large, then $\|\boldsymbol{P}\boldsymbol{x}\|^2$ is close to $\frac{d}{m}$ with high probability.

   (**Hint:** Recall that $\boldsymbol{U}$ has the same law as $\boldsymbol{Q} \sim \mathsf{Haar}(\mathcal{O}(m))$ truncated to the first d columns. Express this fact in linear algebraic terms, writing $\mathsf{Law}(\boldsymbol{U}) = \mathsf{Law}(\boldsymbol{Q}\boldsymbol{U}^{(0)})$ for a suitable deterministic $\boldsymbol{U}^{(0)}$. Plug that into the expression $\|\boldsymbol{P}\boldsymbol{x}\|^2 = \|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{x}\|^2$.)

Problem 3 (Semicircle limit theorem)**.** We showed in class that, if ν has mean 0, variance 1, and all moments finite, then $\boldsymbol{W}^{(d)} \sim \mathsf{Wig}(d, \nu)$ (see the lecture notes for this notation) have the random measures $\mu_d := \mathsf{esd}(\widehat{\boldsymbol{W}}^{(d)})$ converging in expected moments to the semicircle distribution μ_{SC}, where $\widehat{\boldsymbol{W}}^{(d)} := \frac{1}{\sqrt{d}}\boldsymbol{W}^{(d)}$. That is, for all $k \geq 1$,

$$\lim_{d\to\infty} \mathbb{E}\left[\int x^k d\mu_d(x)\right] = \int x^k d\mu_{\mathsf{SC}}(x).$$

1. Show that the above convergence, for any given $k \geq 1$, also holds in L^2. That is, show that, for all $k \geq 1$,

$$\lim_{d\to\infty} \mathsf{Var}\left[\int x^k d\mu_d(x)\right] = 0.$$

Conclude that, for any polynomial $p(x)$,

$$\int p\,d\mu_d - \mathbb{E}\left[\int p\,d\mu_d\right] \xrightarrow{\mathbb{P}} 0.$$

(**Hint:** The (random) integral in the variance is equal to $I_k = \frac{1}{d^{k/2+1}}\mathsf{Tr}(\boldsymbol{W}^{(d)^k})$. Expand $\mathsf{Var}[I_k] = \mathbb{E}[I_k^2] - (\mathbb{E}[I_k])^2$. The convergence of expected moments tells you the asymptotics of the second term. For the first term, adapt our combinatorial argument from class. When you expand out the square of the trace completely, associate each term with a graph whose edges are given by *two* closed walks (it might help to say that this graph has k "red" edges and k "blue" edges). Show that for odd k this expression is $o(1)$, while for even k its leading order term is the same as that of $(\mathbb{E}[I_k])^2$, so that we still have $\mathsf{Var}[I_k] = o(1)$.)

2. Let $N(\epsilon) := \#\{i : |\lambda_i(\widehat{\boldsymbol{W}}^{(d)})| \geq 2+\epsilon\}$, the number of eigenvalues of $\widehat{\boldsymbol{W}}^{(d)}$ that lie outside of the support of μ_{SC} by ϵ or more. Show that, for any $\ell \geq 1$,

$$N(\epsilon) \leq \frac{1}{(2+\epsilon)^{2\ell}}\mathsf{Tr}(\widehat{\boldsymbol{W}}^{(d)^{2\ell}}).$$

Conclude that

$$N(\epsilon) \xrightarrow{\mathbb{P}} 0$$

as $d \to \infty$ for any fixed $\epsilon > 0$, and therefore that also

$$\frac{1}{d}\sum_{i:|\lambda_i(\widehat{\boldsymbol{W}}^{(d)})|>2+\epsilon} |f(\lambda_i(\widehat{\boldsymbol{W}}^{(d)}))| \xrightarrow{\mathbb{P}} 0$$

for any fixed bounded function f and $\epsilon > 0$.

(**Hint:** Use Markov's inequality on $\mathbb{P}[N(\epsilon) \geq 1]$, then use the convergence of expected moments and a suitable bound on the Catalan numbers.)

3. Adapt the argument from Part 2 to show that the latter conclusion also holds for polynomials (even though they are not bounded): show that for any polynomial p,

$$\frac{1}{d}\sum_{i:|\lambda_i(\widehat{\boldsymbol{W}}^{(d)})|>2+\epsilon} |p(\lambda_i(\widehat{\boldsymbol{W}}^{(d)}))| \xrightarrow{\mathbb{P}} 0.$$

(**Hint:** Reduce to the case of monomials $p(x) = x^k$. Then, bound $\frac{1}{d}\sum_{i:|\lambda_i|>2+\epsilon}|\lambda_i|^k \leq \frac{1}{d(2+\epsilon)^{2\ell-k}}\sum_{i=1}^d \lambda_i^{2\ell}$ for $2\ell > k$. Apply Markov's inequality again, use convergence of expected moments, and take ℓ growing.)

4. Following the sketch from class, put together Parts 1, 2, and 3 to prove a version of the semicircle limit theorem with weak convergence in probability: show that, for any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$,

$$\int f\,d\mu_d \xrightarrow{\mathbb{P}} \int f\,d\mu_{\mathsf{SC}}.$$

You may use the Weierstrass approximation theorem (stated in the lecture notes).

Problem 4 (Robustness of semicircle limit theorems). Now you will probe the assumptions needed on the result of Problem 3. Recall that, in proving convergence of expected moments in class (which Problem 3 uses heavily), we relied on all moments of ν being finite. In this problem, you will show that that assumption can be relaxed considerably. To make your life a little easier, we will work only with smooth and compactly supported test functions f; the arguments are not hard to adapt to general bounded continuous functions.

1. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$. Show the perturbation inequality

$$\min_{\pi \text{ permutation of } [d]} \sum_{i=1}^{d} (\lambda_i(\boldsymbol{A}) - \lambda_{\pi(i)}(\boldsymbol{B}))^2 \le \|\boldsymbol{A} - \boldsymbol{B}\|_F^2.$$

   You may use the *Birkhoff–von Neumann theorem*, which states that the set of doubly stochastic $d \times d$ matrices (i.e., $\boldsymbol{P} \in \mathbb{R}^{d\times d}$ such that $P_{ij} \ge 0$ for all $i, j \in [d]$, $\sum_j P_{ij} = 1$ for all $i \in [d]$, and $\sum_i P_{ij} = 1$ for all $j \in [d]$) is the convex hull of the set of the $d \times d$ permutation matrices (those $\boldsymbol{P}$ with exactly one 1 in each row and each column and all other entries 0, of which there are $d!$). You may also use that a linear function over a convex polytope is maximized at one of the vertices.

   (**Hint:** Write the spectral decomposition of $\boldsymbol{A}$ and $\boldsymbol{B}$. Consider the matrix of $\langle \boldsymbol{u}_i, \boldsymbol{v}_j\rangle^2$ for $\boldsymbol{u}_i$ eigenvectors of $\boldsymbol{A}$ and $\boldsymbol{v}_j$ eigenvectors of $\boldsymbol{B}$.)

2. Let f be a smooth and compactly supported function. Show that there is a constant $K = K(f)$ depending only on f such that, for any $d \ge 1$ and $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$,

$$\left| \frac{1}{d}\sum_{i=1}^{d} f(\lambda_i(\boldsymbol{A})) - \frac{1}{d}\sum_{i=1}^{d} f(\lambda_i(\boldsymbol{B})) \right| \le \frac{K}{\sqrt{d}}\|\boldsymbol{A} - \boldsymbol{B}\|_F.$$

3. Prove that, if ν has mean 0 and variance 1 (but other moments not necessarily finite), then for any smooth and compactly supported f we have

$$\int f d\mu_d \xrightarrow{\mathbb{P}} \int f d\mu_{\mathsf{SC}}.$$

   (**Hint:** Define a version of $\boldsymbol{W} = \boldsymbol{W}^{(d)}$ where entries W_{ij} are replaced with the centered *truncations* $(T_{ij} - \mathbb{E}[T_{ij}])/\sqrt{\mathsf{Var}[T_{ij}]}$ for $T_{ij} := W_{ij}\mathbb{1}\{|W_{ij}| \le C\}$. Take a large constant C and use the result of Problem 3 on this matrix.)

4. Find a choice of ν that has mean 0 and variance 1 but such that, for $\boldsymbol{W}^{(d)} \sim \mathsf{Wig}(d, \nu)$, we have $\lim_{d\to\infty} \mathbb{P}[\|\frac{1}{\sqrt{d}}\boldsymbol{W}^{(d)}\| \ge C_d] = 1$ for some diverging sequence $C_d \to \infty$. Consequently, the Wigner edge or norm limit theorem (the statement that $\|\frac{1}{\sqrt{d}}\boldsymbol{W}^{(d)}\| \to 2$ in probability) that we will mention in class does require further moment assumptions on ν.

   (**Hint:** Prove and use that $\|\boldsymbol{W}\| \ge \max_{i,j\in[d]} |W_{ij}|$. Try to make ν as heavy-tailed as possible provided that its mean and variance must be finite.)