Assignment 2

Random Matrix Theory in Data Science and Statistics

(EN.553.796, Fall 2024)

Assigned: October 7, 2024   Due: 12pm EST, October 23, 2024

Solve Problem 1, and any three out of the five remaining problems. If you solve more, we will grade the first three solutions (past Problem 1) that you include. 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 try to avoid writing excessive explanation that is not contributing to our understanding 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 some ideas. In any case, write roughly one paragraph here doing 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 (8-10 minutes) and a short write-up about whatever you choose here.

Problem 2 (Robustness of semicircle limit theorems)**.** We showed in class that, if ν has mean zero, variance 1, and all moments finite, then $\boldsymbol{W}^{(d)} \sim \mathsf{Wig}(d, \nu)$ (see the lecture notes for this notation) have $\mathsf{esd}(\frac{1}{\sqrt{d}}\boldsymbol{W}^{(d)})$ converging weakly in probability to μ_{SC}. That is, for any

$f : \mathbb{R} \to \mathbb{R}$ smooth and of compact support, $\frac{1}{d}\sum_{i=1}^d f(\lambda_i(\frac{1}{\sqrt{d}}\boldsymbol{W}^{(d)})) \to \int f d\mu_{\mathsf{SC}}$ in probability (we considered more general f, but only worry about convergence in probability for these f for this problem). In this problem, you will probe what conditions on ν are really necessary for what kinds of limit theorems.

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 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$,

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

3. Prove that Wigner's semicircle limit theorem (convergence in probability of averages of smooth and compactly supported functions, as stated above) holds only under the assumption that ν has mean 0 and variance 1.

   (**Hint:** Define a version of $\boldsymbol{W} = \boldsymbol{W}^{(d)}$ where entries W_{ij} are replaced with the centered *truncations* $W_{ij}\mathbb{1}\{|W_{ij}| \le C\} - \mathbb{E}[W_{ij}\mathbb{1}\{|W_{ij}| \le C\}]$ for a large C and use the limit theorem from class, as cited above, 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) does require further moment assumptions on ν.

   (**Hint:** Prove and use that $\|\boldsymbol{W}\| \ge \max_{i,j\in[d]} |W_{ij}|$. As I mentioned in class, the Wigner edge limit theorem *does* hold provided that the fourth moment of ν is finite, so your choice must not have that property.)

Problem 3 (Eigenvector perturbation bound)**.** Write $\boldsymbol{v}_1(\boldsymbol{X})$ for the unit-norm eigenvector of $\lambda_1(\boldsymbol{X})$ for $\boldsymbol{X} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$. Whenever this notation is used below, you may assume that $\lambda_1(\boldsymbol{X})$ occurs with multiplicity 1 as an eigenvalue of $\boldsymbol{X}$.

Suppose $\boldsymbol{M} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$, and $\boldsymbol{\Delta}$ has the same dimensions as $\boldsymbol{M}$ with $\|\boldsymbol{\Delta}\| < \lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M})$ (the matrix norm without a subscript always denotes the operator norm). You will show the perturbation inequality

$$\langle \boldsymbol{v}_1(\boldsymbol{M}), \boldsymbol{v}_1(\boldsymbol{M}+\boldsymbol{\Delta})\rangle^2 \geq 1 - \left(\frac{\|\boldsymbol{\Delta}\|}{\lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M}) - \|\boldsymbol{\Delta}\|}\right)^2.$$

Follow these steps, where we abbreviate $\boldsymbol{v} := \boldsymbol{v}_1(\boldsymbol{M})$ and $\widetilde{\boldsymbol{v}} := \boldsymbol{v}_1(\boldsymbol{M}+\boldsymbol{\Delta})$.

1. Show that $\lambda_1(\boldsymbol{M}) - \lambda_i(\boldsymbol{M}+\boldsymbol{\Delta}) \geq \lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M}) - \|\boldsymbol{\Delta}\|$ for all $i \geq 2$.

   (HINT: You may use the Courant-Fischer min-max theorem. Look it up and take a minute to internalize it if you are not familiar with this.)

2. Using Part 1, show that $\|\boldsymbol{\Delta}\boldsymbol{v}\| \geq (\lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M}) - \|\boldsymbol{\Delta}\|) \cdot \|(\boldsymbol{I} - \widetilde{\boldsymbol{v}}\widetilde{\boldsymbol{v}}^\top)\boldsymbol{v}\|$.

   (HINT: Expand $\boldsymbol{v}$ in the orthonormal basis of eigenvectors of $\boldsymbol{M}+\boldsymbol{\Delta}$.)

3. Complete the proof.

Also show the following application:

4. Suppose that $\boldsymbol{W} \sim \mathsf{GOE}^{(0)}(d)$, and let $\boldsymbol{x} \in \mathbb{R}^d$ with $\|\boldsymbol{x}\| = 1$. Let $\lambda > 0$ and consider the matrix $\boldsymbol{Y} = \lambda\sqrt{d}\,\boldsymbol{x}\boldsymbol{x}^\top + \boldsymbol{W}$. Show that there is a function $f(\lambda) \in \mathbb{R}$ such that $f(\lambda) \to 1$ as $\lambda \to \infty$ and such that, for any fixed $\lambda > 0$, we have that

   $$\lim_{d\to\infty} \mathbb{P}[\langle \boldsymbol{v}_1(\boldsymbol{Y}), \boldsymbol{x}\rangle^2 \geq f(\lambda)] = 1.$$

   You may use the Wigner edge limit theorem mentioned above in Problem 2 and cited in class and the lecture notes. More colloquially, this says that the top eigenvector of $\boldsymbol{Y}$ can achieve an arbitrarily good estimate of a rank one perturbation of $\boldsymbol{W}$ of sufficiently large magnitude λ. We will see finer results in class soon.

Problem 4 (Free probability)**.** Define the 2×2 matrix

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

Let $t \sim \mathsf{Unif}([0,\pi])$ and define the random rotation matrix

$$\boldsymbol{U} := \begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix}.$$

Finally, define $\boldsymbol{X}^{(2d)} := \boldsymbol{I}_d \otimes \boldsymbol{A} \in \mathbb{R}^{2d\times 2d}$ and $\boldsymbol{Y}^{(2d)} := \boldsymbol{I}_d \otimes (\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^\top) \in \mathbb{R}^{2d\times 2d}$ random matrices.

1. Show that the sequences $\boldsymbol{X}^{(2d)}$ and $\boldsymbol{Y}^{(2d)}$ have converging empirical spectral moments (i.e., that $\lim_{d\to\infty} \frac{1}{2d}\mathbb{E}\,\mathsf{Tr}\,\boldsymbol{X}^{(2d)k}$ exists for all k and likewise for $\boldsymbol{Y}^{(2d)}$) and that the pair of sequences is asymptotically free. (View the definition of asymptotic freeness as restricted to a sequence of matrices in only even dimensions.)

   (HINT: Boil this down to a statement about the 2×2 matrices $\boldsymbol{A}$ and $\boldsymbol{U}$.)

2. To what measure must the empirical spectral distribution of $\boldsymbol{X}^{(2d)} + \boldsymbol{Y}^{(2d)}$ then converge in expected moments? Why?

   (**Hint:** You do not need to calculate an additive free convolution by hand if you paid attention in class.)

3. Show that the empirical spectral distribution of $\boldsymbol{X}^{(2d)} + \boldsymbol{Y}^{(2d)}$ almost surely consists of at most two atoms. Therefore, qualitatively, it will never resemble the measure you described in Part 2. For example, in a histogram of the eigenvalues, only at most two bins will ever be non-empty. Explain formally and precisely why this is not a contradiction to Part 2.

Problem 5 (More on Gaussian random vectors)**.** This problem is a continuation of Problem 3 from Homework 1. In the next homework, a final problem in the sequence will have you derive powerful consequences of these ideas for random matrices. For now, you will derive some more general tools.

1. Suppose $F : \mathbb{R}^d \to \mathbb{R}$ is a smooth function with $\max\{|F(\boldsymbol{x})|, \|\nabla F(\boldsymbol{x})\|_2^2, \|\nabla^2 F(\boldsymbol{x})\|_F^2\} \leq C(1 + \|\boldsymbol{x}\|)^K$ for some $C, K > 0$ and all $\boldsymbol{x} \in \mathbb{R}^d$, where $\nabla^2 F$ is the $d \times d$ Hessian matrix of second derivatives. Let $\boldsymbol{\Sigma}, \boldsymbol{\Lambda} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$ be positive semidefinite. Define $\boldsymbol{\Sigma}(t) := (1-t)\boldsymbol{\Sigma} + t\boldsymbol{\Lambda}$ for $t \in [0,1]$, and write

   $$f(t) := \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma}(t))} F(\boldsymbol{g}).$$

   That is, we are considering the value of an expectation of a general function of a Gaussian vector as the covariance matrix moves along a line in matrix space. Show that the derivative of this value is

   $$f'(t) = \frac{1}{2}\left\langle \boldsymbol{\Lambda} - \boldsymbol{\Sigma}, \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma}(t))} \nabla^2 F(\boldsymbol{g}) \right\rangle.$$

   Here, $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \mathsf{Tr}(\boldsymbol{A}\boldsymbol{B}) = \sum_{i,j} A_{ij}B_{ij}$ is the Frobenius inner product.

   You may differentiate under the expectation (i.e., bring a derivative inside an expectation) without justification, but you should consider on your own time what the justification would be.

   (**Hint:** If $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda})$ independently, construct a Gaussian vector with covariance $\boldsymbol{\Sigma}(t)$ to make differentiating under the expectation easier. Then, use Homework 1, Problem 3, Part 1—you may use it even if you did not solve that problem.)

2. Show that, if F as above is also convex, and $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma})$ are independent Gaussian vectors (that is, the entries of $\boldsymbol{g}$ may be correlated with one another, and likewise for $\boldsymbol{h}$, but entries of $\boldsymbol{g}$ are independent of entries of $\boldsymbol{h}$) for any $\boldsymbol{\Sigma}, \boldsymbol{\Gamma} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$ positive semidefinite, then

   $$\mathbb{E}F(\boldsymbol{g}) \leq \mathbb{E}F(\boldsymbol{g} + \boldsymbol{h}).$$

   Informally, expectations of convex functions of Gaussians are only increased by adding noise. Show that the same also holds for $F(\boldsymbol{x}) = \max_{i\in[d]} x_i$, though it is not smooth.

   (**Hint:** $\mathsf{Law}(\boldsymbol{g} + \boldsymbol{h}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda})$ for some $\boldsymbol{\Lambda}$—write this out and use Part 1. For the last part, consider the "soft-max" function $F(\boldsymbol{x}) = \beta^{-1}\log(\sum_{i=1}^d \exp(\beta x_i))$ and take $\beta \to \infty$.)

3. Suppose that $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{h} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$ are arbitrary centered Gaussian vectors as in Part 1. Suppose that, for all $i, j \in [d]$, we have $\mathbb{E}(g_i - g_j)^2 \le \mathbb{E}(h_i - h_j)^2$. Show that

$$\mathbb{E} \max_{i \in [d]} g_i \le \mathbb{E} \max_{i \in [d]} h_i.$$

(**Hint:** Expand the condition on $\boldsymbol{g}$ and $\boldsymbol{h}$ into a condition on $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$. Again consider the soft-max function and use Part 1, explicitly calculating the Hessian.)

Problem 6 (Numerical study of random regular graphs)**.** We discussed p-regular graphs of large girth in class. In this problem, you will study p-regular graphs chosen uniformly at random numerically and observe that they share some but not all of the same properties.

1. Write code to generate a p-regular graph on d vertices (pd must be even) at random, as follows. View the d vertices as each having p "half-edges" attached to them, for a total of pd. A graph may be viewed as formed by gluing together half-edges in pairs to form full edges. As we have seen from the combinatorics of Gaussian moments, there are $(pd - 1)!!$ possible perfect matchings among pd objects. Choose such a perfect matching uniformly at random (come up with and justify a way to perform this sampling). This forms a random p-regular *multigraph* G_0 on d vertices, since it is possible that you created self-loops or parallel edges in choosing your matching. Now, perform rejection sampling: repeat the procedure until you choose a matching that yields a simple graph G. Include this part of your code in your homework submission.

   You do not need to prove it, but the resulting G is uniformly random among simple p-regular graphs on d vertices with labelled vertices.

2. Write code to estimate $f(p, d) := \mathbb{P}[G_0 \text{ is simple}]$ in the above procedure. For $p \in \{3, 4\}$, estimate $f(p) := \lim_{d \to \infty} f(p, d)$ by taking d large. That is, for each p, for a sequence of growing d, report the fraction of trials giving G_0 simple out of a large total. Plot data to illustrate the convergence of your estimate as $d \to \infty$. (Optionally, if you are very patient, you may try $p = 5$. It helps to not generate an entire perfect matching before rejecting a G_0 that is not simple.)

3. Write $\boldsymbol{A}$ for the adjacency matrix of G formed above. For $p \in \{3, 4\}$, confirm that, for large d, $\mathsf{esd}(\boldsymbol{A})$ is close to the Kesten-McKay measure with parameter 3 and 4 (respectively) as predicted in class for regular graphs of large girth. How large of d is needed? Include convincing plots.

4. Let $T = T(G)$ be the number of triangles in G. Estimate $t(p, d) := \mathbb{E}T(G)$ for $p = 3$ and a sequence of growing d. Does our reasoning from class apply to the random G? Try to identify what $\mathsf{Law}(T)$ converges to as $d \to \infty$. Include numerical evidence for your prediction of any kind you think is reasonable—histograms, experimental estimates of moments, etc.

   (**Hint:** Compute $T(G)$ with matrix algebra, not for loops.)