Math 606, Summer 2023[1]: Introduction to Random Matrices; version of June 27, 2023

Notations.

(1) $1_{\{\cdot\}}$ — the indicator function for the event $\{\cdot\}$.
(2) $\mathcal{N}(0,1)$ — standard Gaussian random variable.
(3) $\mathfrak{i}$ — the imaginary unit: $\mathfrak{i}^2 = -1$.
(4) $\mathfrak{j}, \mathfrak{k}$ — the additional imaginary units for quaternions.
(5) $A^\top$ — the transpose of the matrix A.
(6) A^* — the transpose of the matrix A and complex (or quaternion) conjugation of the entries.
(7) I_n — the identity matrix of size n.
(8) $\mathrm{Tr}(A)$ — trace of the (square) matrix A.

Abbreviations

(1) CLT — Central Limit Theorem;
(2) ECM — Empirical Correlation Matrix;
(3) ESD — Empirical Spectral Distribution: empirical probability measure generated by the eigenvalues of a (random) matrix;
(4) GOE — Gaussian orthogonal ensemble ($\beta = 1$);
(5) GSE — Gaussian Symplectic Ensemble ($\beta = 4$);
(6) GUE — Gaussian unitary ensemble ($\beta = 2$);
(7) RMT — Random Matrix Theory;
(8) SVD — Singular Value Decomposition.

Ideas for Homework

From the Book.

(1) Produce your version of Fig. 1.1.
(2) Derive (2.17) from (2.15).
(3) Produce your version of Fig. 3.1.
(4) Derive (3.9) from (1.7).
(5) Confirm (6.6).
(6) Confirm (6.21).
(7) Confirm the second equality in (8.1).
(8) Confirm the second equality in (8.23).
(9) Confirm (9.10) and (9.11).
(10) Confirm (10.21).
(11) Produce your version of Fig. 12.1 for $N = 5$.

Theoretical Exercises

(1) Confirm that quaternions $\mathbb{H}$ form a four-dimensional associative division algebra over the field of real numbers.
(2) Confirm that the equation $q^2 + 1 = 0$ has infinitely many solutions over $\mathbb{H}$.
(3) Confirm that if p is a quaternion with zero real part and q is a unit quaternion, then qpq^{-1} has zero real part.
(4) Derive Stirling's formula using Laplace's method.
(5) Evaluate the integral $\int_0^{+\infty} e^{\mathfrak{i}x^p}\, dx$, $p > 1$.
(6) Confirm that the (a) determinant of an odd-dimensional skew-symmetric matrix is zero; (b) every eigenvalue of a skew-symmetric matrix has real part equal to zero.

[1]Sergey Lototsky, USC

(7) Confirm that, for two square matrices A, B, the trace of the product is the dot product of the vectors $\mathrm{vec}(A^\top)$ and $\mathrm{vec}(B)$.
(8) Confirm that both the usual and the Kronecker products of two stochastic matrices are stochastic matrices [that is, non-negative entries, with all row sums equal to 1].
(9) Confirm that $\mathrm{Tr}(A \otimes B) = \mathrm{Tr}(A)\mathrm{Tr}(B)$.
(10) Confirm that, for an n-by-n matrix $A = (a_{ij})$, the norm $\|A\| = \max_{i,j} |a_{ij}|$ is not sub-multiplicative, but $n\|A\|$ is.
(11) Give an example of two real Gaussian random variables such that their sum is not Gaussian.
(12) Give an example of two complex (jointly) Gaussian random variables that are uncorrelated but dependent.
(13) Given two independent complex standard Gaussian random variables U and V and a real number r, compute

$$\mathbb{E}\frac{U}{U+rV}.$$

(14) Given iid standard Gaussian random variables X and Y and a real number r, compute

$$\mathbb{E}\frac{X+\mathfrak{i}rY}{X+\mathfrak{i}Y}.$$

(15) Let $X_1, \ldots, X_n$ be iid standard Gaussian random variables. Confirm that the distribution of the random vector

$$\boldsymbol{Y} = \frac{(X_1, \ldots, X_n)}{\sqrt{X_1^2 + \ldots + X_n^2}}$$

is uniform on the unit sphere in $\mathbb{R}^n$.
(16) Let $U_1, \ldots, U_n$ be iid uniform on $[0,1]$ and let $U_{(1)} < U_{(2)} < \ldots < U_{(n)}$ be the corresponding order statistics. Setting $U_{(n+1)} = 1$, confirm that

$$\lim_{n\to\infty} \mathbb{P}\Big(n^2 \min_{1\le k\le n} (U_{(k+1)} - U_{(k)}) > x\Big) = e^{-x},\ x > 0,$$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{n}{\ln n} \max_{1\le k\le n} (U_{(k+1)} - U_{(k)}) - 1\right| > \varepsilon\right) = 0,\ \varepsilon > 0.$$

In other words, the iid uniforms fill in the interval in a highly non-uniform way: the smallest gap is of order $1/n^2$ and the largest gap is of order $\ln n/n$.
(17) Confirm that the even moments of the semi-circle law are Catalan numbers.
(18) Consider the `Stieltjes-Wigert weight function`

$$\varphi(x) = \frac{1}{\sqrt{\pi}} e^{-(\ln x)^2},\ x > 0.$$

(a) Confirm that

$$s_n = \int_0^{+\infty} x^n \varphi(x)dx = e^{(n+1)^2/4},\ n = 0, 1, 2, \ldots.$$

(b) Confirm that

$$\int_0^{+\infty} x^n \sin(2\pi \ln x)\, \varphi(x)dx = 0$$

for every $n = 0, 1, 2, \ldots$ and so the polynomials are not complete in $L_2\big((0, +\infty), \varphi(x)dx\big)$.
(c) Confirm that the integral

$$\int_0^{+\infty} e^{ax}\varphi(x)dx$$

diverges for every $a > 0$, but the integral

$$\int_0^{+\infty} \frac{\ln \varphi(x)}{1+x^2}\, dx$$

and the series

$$\sum_{k\geq 1} \frac{1}{\sqrt[2k]{s_{2k}}}$$

both converge.

(19) Let U be a complex-valued random variable with uniform distribution in the unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$. Confirm that (a) both real and imaginary parts of U follow a semi-circle law; (b) the absolute values of the real and imaginary parts of U follow a quarter-circle law; (c) identify the distribution of $|U|$.

(20) Compute the Stieltjes transform of the Cauchy distribution with pdf $1/(\pi(x^2+1))$. The answer is $1/(z \pm \mathfrak{i})$, depending on the sign of the imaginary part of z.

(21) Identify the main shapes of the Marchenko-Pastur distribution and sketch the graphs of the corresponding densities.

(22) Confirm that the distribution of a GOE/GUE/GSE matrix is invariant under orthogonal/unitary/symplectic transformation.

(23) Let A be a square matrix with iid standard Gaussian entries. Write the joint distribution of the entries of the matrix $(A - A^{\top})/2$.

(24) State and solve the corresponding version of the previous problem for matrices with complex and quaternion entries.

(25) `The Stieltjes transform of the semi-circle law.` Confirm that, for complex numbers z with positive imaginary part,

$$\frac{1}{2\pi}\int_{-2}^{2} \frac{\sqrt{4-x^2}}{x-z}\,dx = \frac{-z+\sqrt{z^2-4}}{2},$$

where the complex square root on the right has positive imaginary part. A possible way to proceed: $x = 2\cos t$; integration in t from 0 to π can be extended to full circle; then get a complex integral in ζ over the unit circle using $\zeta = e^{\mathrm{i}t}$, $1/\zeta = e^{-\mathrm{i}t}$, as well as Euler's formula for $\sin t$ and $\cos t$; the integrand (a rational function of ζ) has a second-order pole at the center of the circle and another (simple) somewhere inside the unit circle; residue integration completes the process.

(26) `The Stieltjes transform of the Marchenko-Pastur law` leads to a complex integral over a unit circle with a rational integrand having *five* simple poles inside. The computations are outlined in the proof of Lemma 3.11 in the book

Bai, Z. D. and Silverstein, J. W. Spectral Analysis of Large Dimensional Random Matrices, Second edition, Springer, 2010.

See if you can fill in the details.

(27) Denote by δ_a the point mass at the point $a \in \mathbb{R}$, and define probability measures μ_n on $\mathbb{R}$ by

$$\mu_n = \left(1 - \frac{1}{n}\right)\delta_0 + \frac{1}{n}\,\delta_n, \quad n = 1, 2, \ldots.$$

Confirm that each μ_n has finite moments of every order and, as $n \to \infty$, the sequence $\{\mu_n,\ n \geq 1\}$ converges weakly to δ_0, but there is no convergence of moments. [The point of this exercise is that (a) under some conditions, convergence of moments implies weak convergence, (b) under some *other* conditions, weak convergence implies convergence of moments, but (c) *in general*, the two types of convergence are different.]

Computer Exercises

(1) Confirm the semi-circle law for the Gaussian Orthogonal Ensemble. How will the picture change if you replace Gaussian random variables with Cauchy, using scale parameter instead of standard deviation?

(2) Repeat the previous exercise for GUE and GSE.

(3) Confirm the circle law in the standard Gaussian case. How will the picture change if you replace Gaussian with Cauchy?
(4) Write a program generating a random orthogonal matrix.
(5) Generate a sample trajectory of the Dyson Brownian motion with $N \geq 5$.

Problems

(1) There are infinitely many non-isomorphic commutative, non-associative, finite-dimensional division algebras over the field of real numbers; they all have dimension 2. One example is the complex numbers with "multiplication" defined by

$$(a+b\mathfrak{i})\circ(x+y\mathfrak{i}) = (a-b\mathfrak{i})(x-y\mathfrak{i}) \equiv (a+b\mathfrak{i})^*\circ(x+y\mathfrak{i})^* \equiv \big((a+b\mathfrak{i})\circ(x+y\mathfrak{i})\big)^* = (ax-by)-(ay+bx)\mathfrak{i}.$$

Construct another example.
(2) Investigate the asymptotic of the Airy function $\mathrm{Ai}(x)$ as $x \to +\infty$.
(3) Investigate the distribution of eigenvalues of a random matrix that is (a) zero-trace (b) skew-symmetric (c) skew-symmetric and has trace zero.
(4) Let κ be the condition number of a 2-by-2 GOE matrix. Determine the values of $r \in \mathbb{R}$ for which $\mathbb{E}\kappa^r < \infty$. If this is too hard (or too easy), consider a 2-by-2 matrix with iid uniform on $[-1,1]$ entries instead.

Main facts of general interest.

(1) GOE$=\{(A+A^\top)/2 : A=(a_{ij}) \in \mathbb{R}^n,\ a_{ij}$ iid $\mathcal{N}(0,1)\}$: symmetric matrices with Gaussian entries. To achieve certain normalization, one can divide by numbers other than 2. For example, $\{(A+A^\top)/\sqrt{2n}\}$ gives, for the limit of the distribution of eigenvalues, the (more traditional) semi-circle law with density $(2\pi)^{-1}\sqrt{4-x^2}\, I(|x|\leq 2)$.
(2) GUE$=\Big\{\big(A+A^\top+\mathfrak{i}(B-B^\top)\big)/2, A=(a_{ij}) \in \mathbb{R}^n,\ B=(b_{ij}) \in \mathbb{R}^n,\ a_{ij}, b_{ij}$ iid $\mathcal{N}(0,1)\Big\}$: Hermitian matrices with Gaussian entries. Similarly, normalizations can vary. In particular, a rather different normalization comes from considering Hermitian matrices with independent zero-mean Gaussian entries that are *real* standard (variance one) on the diagonal and *complex* standard off diagonal (that is, real and imaginary pars are iid with mean zero and variance 1/2).
(3) GSE$=\Big\{\big(A+A^*\big)/2$, where A is a matrix with iid *quaternion*-valued Gaussian random variables, and A^* is quaternion conjugate of A.
(4) A somewhat unified approach to Gaussian ensembles comes from considering square n-by-n matrices $H=H^\top$ $(\beta=1)$ or $H=H^*$ $(\beta=2,4)$ with iid standard normal on the diagonal and iid *β-dimensional normal* with mean zero and covariance $I_\beta/2$ above the diagonal; $\beta=1$ corresponds to GOE, $\beta=2$ corresponds to GUE, and $\beta=4$ corresponds to GSE. Then the joint pdf of the upper-triangular part of the matrix entries is

$$p_\beta(H) = (2\pi)^{-n/2}\ \pi^{-\beta\frac{n(n-1)}{4}} e^{-\frac{1}{2}\mathrm{Tr}(H^2)}.$$

The corresponding joint pdf for the (real) eigenvalues is

$$\rho_\beta(\lambda_1,\ldots,\lambda_n) = \frac{1}{n!}\,\frac{\big(\Gamma(\beta/2)\big)^n}{\prod_{k=1}^n \Gamma(k\beta/2)} \left(\prod_{k<\ell} |\lambda_\ell-\lambda_k|^\beta\right) \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\sum_{k=1}^n \lambda_k^2}.$$

(5) `Wigner` matrix $W=(\xi_{ij},\ i,j=1,\ldots,n)$ with $\xi_{ij}=\xi_{ji}$, independent $\xi_{ij},\ j\geq i$, and identically distributed ξ_{ij} for $j>i$ and for $j=i$ (but the distributions of the diagonal and off-diagonal elements are usually different). It is typically assumed that each ξ_{ij} has zero mean and finite moment of some order bigger than or equal to 2.
(6) `Wishart`[2] matrix is a matrix W of the form $W=MM^*$, where M is a *rectangular* matrix with iid mean zero-variance one entries. The `Wishart ensemble` corresponds to Gaussian

[2]John Wishart (1898–1956) was a Scottish statistician with special interests in agriculture

M with real/complex/quaternion entries ($\beta = 1, 2, 4$). Sometimes, the name of Laguerre[3] is also included, especially when $\beta = 2$.

(7) `Spectrum of a real random matrix:` an informal summary. Let A be a rectangular p-by-n, $p \le n$, matrix with entries that are iid real random variables, each with mean zero, variance one, and *finite fourth moment*.[4] Then, for large n,
   - The *singular* values of A are mostly in the interval $\sqrt{n} \pm \sqrt{p}$ and approximately distributed according to the (square root of the) Marchenko-Pastur law;
   - In the square case $p = n$, the (complex) eigenvalues of A are mostly spread out uniformly around the disk $|z| \le \sqrt{n}$ in the complex plane

(8) `The Circle Laws.` Let A be an n-by-n matrix with real or complex iid entries, *each having zero mean and unit variance.* In the complex case, the real and imaginary parts of each entry are iid with mean zero and variance 1/2. Then
   - The empirical distribution of the (complex) *eigenvalues* of the matrix $A/\sqrt{n}$ converges to the uniform distribution in the unit disk of the complex plane: the (full) `circle law`, which started in 1960-s with the work of Ginibre[5] for the Gaussian case, continued in the 1980-s by V. L. Girko[6] for more general cases, and fully established by T. Tao and V. H. Vu in 2010.
   - The empirical distribution of the (real) *eigenvalues* of the matrix $\frac{A+A^*}{2\sqrt{n}}$ converges to the distribution with pdf equal to $(1/\pi)\sqrt{2-x^2}\, 1_{|x|\le\sqrt{2}}$: the `semi-circle law`, which goes back to the 1955 paper by Wigner.[7]
   - The empirical distribution of the *singular values* of the matrix $A/\sqrt{n}$ converges to the distribution with pdf equal to $(1/\pi)\sqrt{4-x^2}\, 1_{0<x<2}$: the `quarter-circle law`, which is a (very) particular case of the `Marchenko-Pastur law`.[8]

   Recall that, given a finite collection of points $x_1, \dots, x_n$ in a measurable space $(\mathbf{X}, \mathcal{X})$, the corresponding `empirical distribution` is a measure L_n on $(\mathbf{X}, \mathcal{X})$ defined by $L_n(G) = \frac{1}{n}\sum_{k=1}^{n} 1_{x_k \in G}$, $G \in \mathcal{X}$. If the points are random, then L_n is a random measure, and the (weak) convergence to the corresponding limit can be studied in distribution, in probability, or with probability one. For a random square matrix, the empirical distribution of the corresponding eigenvalues is sometimes referred to as the `empirical spectral distribution` (ESD).

(9) The `Bai-Yin Theorems`[9] can refer to various bounds on singular values or, more generally, matrix norms. The original paper contains the following result: If $S_n = n^{-1}XX^\top$, with a p-by-n matrix X having iid real entries with mean zero, variance one, and *finite fourth moment*, and $\lim_{n\to\infty} p/n = y \in (0, 1]$, then, with probability one,

$$\lim_{n\to\infty} \lambda_{\min}(S_n) = (1-\sqrt{y})^2, \ \lim_{n\to\infty} \lambda_{\max}(S_n) = (1+\sqrt{y})^2.$$

   Equivalently, with probability one, the *singular values* of X satisfy

$$\sigma_{\min}(X) = \sqrt{n} - \sqrt{p} + o(\sqrt{p}), \ \sigma_{\max}(X) = \sqrt{n} + \sqrt{p} + o(\sqrt{p}), \ n, p \to \infty.$$

   The result is consistent with the *quarter-circle law*: If the case $p = n$, then $y = 1$ and the singular values of $n^{-1/2}X$ are the *square roots* of the eigenvalues of S_n; the theorem

[3] Edmond Nicolas Laguerre (1834–1886) was a French mathematician, most famous in connection with the corresponding polynomials.

[4] Sometimes, finite fourth moment is necessary; sometimes, the assumption simplifies the proof and can be relaxed.

[5] Jean Ginibre (1938–2020), French. Also, the "G" in the FKG inequality.

[6] Vyacheslav Leonidovich Girko (b. 1946); `http://www.general-statistical-analysis.girko.freewebspace.com/`

[7] Eugene Paul "E. P." Wigner (1902–1995) was born in Hungary (Budapest), educated in Germany (TU Berlin), became US citizen in 1937, and received Nobel Prize in Physics in 1963.

[8] Vladimir Alexandrovich Marchenko (b. 1922); Leonid Andreevich Pastur (b. 1937) was a Ph.D. student of Marchenko.

[9] Zhidong Bai (b. 1943); Yong Quan Yin (1930–2020); the original paper is *Limit of the smallest eigenvalue of a large dimensional sample covariance matrix*, Annals of Probability, Vol. 21, No. 3, pp. 1275–1294, 1993.

asserts that, in the limit, all these singular values will be in the interval $[0, 2]$. Some further directions include results such as

$$\sqrt{n} - \sqrt{p} \le \mathbb{E}\sigma_{\min}(X) \le \ \mathbb{E}\sigma_{\max}(X) \le \sqrt{n} + \sqrt{p}$$

for *Gaussian* X, or [there exist numbers a, b such that, for every $t > 0$]

$$\mathbb{P}\big(\sqrt{n} - a\sqrt{p} - t \le \sigma_{\min}(X) \le \ \sigma_{\max}(X) \le \sqrt{n} + a\sqrt{p} + t\big) \ge 1 - e^{-bt^2}$$

for *sub-Gaussian* X.

(10) The `Marchenko-Pastur theorem` describes the limit of the ESD for a Wishart matrix $W_n = n^{-1}AA^{\top}$, with a p-by-n matrix A having iid real entries with mean zero, variance one, and $\lim_{n\to\infty} p/n = y \in (0, +\infty)$; the convergence is weak, with probability one, and the distribution in the limit is known as the Marchenko-Pastur law or Marchenko-Pastur distribution.[10] This distribution is usually written as

$$\max\left\{\left(1 - \frac{1}{y}\right), 0\right\} \cdot \delta_0 + \frac{\sqrt{(b-x)(x-a)}}{2\pi c x} 1_{a<x<b}\, dx,$$

with $a = (1 - \sqrt{y})^2$, $b = (1 + \sqrt{y})^2$, and δ_0 denoting the point mass at zero: if $y > 1$, then, in the limit, $n < p$ so that the matrix W_n does not have full rank and has $p - n$ zero eigenvalues, leading to the point mass at zero. The extreme case $y = 0$ ($n/p \to \infty$) corresponds to a consistent estimator of the (identity) covariance matrix for the components of the first column of the matrix A, and, after a CLT-type normalization, leads to the *semi-circle law*:[11] as $n/p \to \infty$, the empirical cdf for the eigenvalues of

$$\sqrt{\frac{n}{p}}\Big(W_n - I_p\Big)$$

converges to the semi-circle law with radius 2. Note that $W_n \in \mathbb{R}^{p\times p}$ for every n; assumptions about the matrix A and the strong law of large numbers imply $\lim_{n\to\infty} W_n = I_p$ with probability one. The extreme case $y = +\infty$ corresponds to a "non-identifiable model" and might require further attention.

(11) If the entries of a Gaussian (O/U/S) ensemble become Brownian motions in such a way that, at time t the distribution of the matrix is $t^{1/2}$ times the corresponding standard ensemble, then the term `Dyson's Brownian motion` typically refers to the process describing the resulting time evolution of the eigenvalues; occasionally, the same term can refer to the matrix itself.

(12) From complex analysis.
   (a) The original `Cauchy transform` is for measures on the unit circle; the `Stieltjes transform` does something similar for measures on the real line, and with an opposite sign.
   (b) `Helffer-Sjöstrand` identity is an analog of the Cauchy integral formula for (certain extensions of real) functions that are not analytic.
   (c) The `Sokhotski-Plemelj` theorem investigates the Cauchy integral formula when the point where the function is evaluated is approaching the path of integration.
   (d) `Painlevé` classified 1-st and 2-nd order ODEs for which a pole is the only possible singularity of the solution that can depend on the initial condition.

[10] For a short proof under even more general conditions, see P. Yaskov, *A short proof of the Marchenko–Pastur theorem*, C. R. Acad. Sci. Paris, Ser. I, vol. 354, pp. 319–322, 2016. The proof uses the Stieltjes transform.

[11] Z. D. Bai, Y. Q. Yin, *Convergence to the semicircle law*, Annals of Probability, Vol. 16, No. 2, pp. 863–875, 1988; the proof is using truncation and the method of moments, under an additional assumption that the fourth moment of A_{11} is finite.