MC2 Reflection 3

Gabe Orosan-Weine

June 2023

1 Random Matrix Theory

Random matrix theory is mainly concerned with properties of matrices which have values sampled from a probability distribution. Since being introduced by Wishart in 1928, random matrices have found applications in many other fields of science. Random matrices can model the relative locations of gas particles in physics, and are used in signal processing to analyze the correlation matrix for received inputs.

One of the most common types (called ensembles) of random matrices studied is the Gaussian Orthogonal Ensemble (GOE). The GOE consists of real symmetric matrices with values samples from a Gaussian (normal) distribution. GOE matrices have been used in biology to model networks of yeast protein-protein interactions and metabolic networks (Zhong et al., 2006). One of the most famous and significant results in random matrix theory is the Semicircle Law for Gaussian random matrices, first proved by Eugene Wigner. We begin by introducing semicircles.

Definition 1.1 (Semicircle Distribution)**.** We define the standard semicircle distribution as the distribution on $[-2, 2]$ given by

$$P(x) = \frac{1}{2\pi}\sqrt{4 - x^2}.$$

Definition 1.2 (Catalan Numbers)**.** The Catalan Numbers are defined as

$$C_k = \frac{1}{k+1}\binom{2k}{k},\ k \geq 0.$$

Theorem 1.1. The moments of the semicircle distribution $P(x)$ are given by

$$\frac{1}{2\pi}\int_{-2}^{2} x^n \sqrt{4 - x^2}\, dx = \begin{cases} 0 & \text{if n is odd,} \\ C_k & \text{if } n = 2k \text{ is even.} \end{cases}$$

Proof. If n is odd, then the function being integrated is odd; $f(-x) = -f(x)$. This means that the integral from -2 to 0 is the additive inverse of the integral from 0 to 2 so the result for the integral from -2 to 2 is 0. If n is even then we can set $x = 2\sin(\theta)$ to get the $n = 2k$-th moment:

$$\frac{1}{2\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}\,dx = \frac{1}{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 2^{2k+2}\sin^{2k}\theta\,\cos^2\theta\,d\theta.$$

Bringing out the constant and making the substitution $\cos^2\theta = 1 - \sin^2\theta$ we have

$$\frac{2^{2k+1}}{\pi}(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k}\theta\ d\theta - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k+2}\theta\,d\theta).$$

By the reduction formula,

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k+2}\theta\,d\theta) = \frac{2k-1}{2k}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k}\theta\,d\theta = \pi\prod_{l=1}^{k}\frac{2l-1}{2l}.$$

Thus the 2k-th moment is equal to

$$2^{2k+1}(\prod_{l=1}^{k}\frac{2l-1}{2l} - \prod_{l=1}^{k+1}\frac{2l-1}{2l}) = 2^{2k+1}\frac{1}{2(k+1)}\prod_{l=1}^{k}\frac{2l-1}{2l} = \frac{1}{k+1}2^{2k}\prod_{l=1}^{k}\frac{2l-1}{2l}\prod_{l=1}^{k}\frac{2l}{2l}.$$

By noticing that the factors of 2 in the denominator of the products will cancel with 2^{2k} and rearranging terms in the products we get

$$\frac{1}{k+1}(\prod_{l=1}^{k}\frac{1}{l})^2\prod_{l=1}^{k}2l(2l-1) = \frac{1}{k+1}\frac{1}{k!^2}2k! = \frac{1}{k+1}\binom{2k}{k}.$$

This proves that for even n, $E[X^n] = E[X^{2k}] = C_k$ for $k \geq 0$. □

Definition 1.3 (Random Symmetric Matrix)**.** A random symmetric matrix A of dimension N from distribution D is determined as follows: A_{ij} are sampled independently from D for $i \leq j$ (that is, the main diagonal and all values the lie above it), and the other values are determined by symmetry ($A_{ji} = A_{ij}$).

Since the moments of the semicircle distribution are Catalan Numbers, we show that the distribution of eigenvalues of random matrices approaches a semicircle distribution by showing that the expected value of moments of the eigenvalues also turns out to be the Catalan Numbers.

Definition 1.4 (Dirac delta function)**.** $\delta(x)$ is defined to be the Dirac delta function given by

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{if } x \neq 0. \end{cases}$$

Theorem 1.2 (Semicircle Law)**.** Let D be a distribution having finite moments of all orders such that if X is chosen from distribution D then $E(X) =$

$0, E(X^2) = 1$, and $E(X^k)$ is finite for all $k \geq 3$. Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the eigenvalues for real-valued random symmetric matrices A of dimension N sampled from D. The expected value of the k-th moment of the distribution

$$\mu_{A,N}(x) = \frac{1}{N}\sum_{j=1}^{N}\delta(x - \frac{\lambda_j}{\sqrt{N}})$$

is equal to C_k. That is, the expected value of the moment given by

$$U(x^k) = \int_{-\infty}^{\infty} x^k \mu_{A,N}(x)\,dx$$

is equal to the Catalan Number C_k;

$$E[U(x^k)] = C_k.$$

Proof. The distribution

$$\mu_{A,N}(x) = \frac{1}{N}\sum_{j=1}^{N}\delta(x - \frac{\lambda_j}{\sqrt{N}})$$

has nonzero values only when $x = \frac{\lambda_j}{\sqrt{N}}$, so

$$U(x^k) = \int_{-\infty}^{\infty} x^k \mu_{A,N}(x)\,dx = \frac{1}{N}\sum_{j=1}^{N}(\frac{\lambda_j}{\sqrt{N}})^k.$$

For $k = 0$, the expected value is

$$E[U(x^0)] = \frac{1}{N}\sum_{j=1}^{N} 1 = 1.$$

For $k = 1$, because the sum of eigenvalues is equal to the trace of a matrix and expected value of terms in the matrix is 0, we have

$$E[U(x^1)] = E\left[\frac{1}{N}\sum_{j=1}^{N}\frac{\lambda_j}{\sqrt{N}}\right] = \frac{1}{N^{3/2}}E\left[\sum_{j=1}^{N}\lambda_j\right] = \frac{1}{N^{3/2}}E\left[\sum_{j=1}^{N}A_{jj}\right] = 0.$$

For $k = 2$, squaring an eigenvalue for A gives you an eigenvalue for A^2, so by the same reasoning used above we get

$$E[U(x^2)] = E\left[\frac{1}{N}\sum_{j=1}^{N}(\frac{\lambda_j}{\sqrt{N}})^2\right] = \frac{1}{N^2}E\left[\sum_{j=1}^{N}\lambda_j^2\right] = \frac{1}{N^2}E\left[\sum_{j=1}^{N}(A^2)_{jj}\right].$$

Noting that since A is symmetric,

$$(A^2)_{jj} = \sum_{k=1}^{N}A_{jk}A_{kj} = \sum_{k=1}^{N}(A_{jk})^2$$

so we have

$$\frac{1}{N^2} E\left[\sum_{j=1}^{N}\sum_{k=1}^{N}(A_{jk})^2\right] = \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N} E\left[(A_{jk})^2\right].$$

We defined the values in A as coming from a distribution as with $E[X^2] = 1$ so we get

$$\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N} 1 = 1.$$

Isserlis' theorem (which relates the product of normal random variables to their covariance) together with combinatorial arguments can be used to deduce that for even $n = 2k > 2$,

$$E[U(x^{2k})] = \frac{1}{k+1}\binom{2k}{k} = C_k$$

and for odd $n = 2k+1$, $E[U(x^{2k+1})] = 0$. This equates the expected value of the moments with the moments of the semicircle distribution. We refer the reader to Wolf, 2021 for a full proof. □

The method of moments is useful in many areas of statistics because it allows you to characterize a distribution by values that tend to be relatively easy to compute. Partially because the nonlinearities in machine learning algorithms preclude the application of traditional proof techniques like the method of moments, the uses of random matrix theory to deep learning have been limited. However, a recent paper showed how the pointwise nonlinearities imposed by prevalent algorithms can be accounted for in the method of moments. Their results suggest a class of activation functions which lead to eigenvalues of the data covariance matrix being constant in distribution, which could decrease training time (Pennington et al., 2017).

One of fundamental questions problems in random matrix theory is universality: given a matrix with values sampled from a certain probability distribution, is there a law which governs the distribution of properties of the matrix like eigenvalues as the size of the matrix goes to infinity? While the answer is known for certain classes of random matrices, as we have established with Wigner's Semicircle law, this remains an active field of research for other classes of matrices. Tools of random matrix theory developed to study this problem of universality often end up being useful for understanding complex dynamics in other fields such as physics, biology, and machine learning. Overall, random matrix theory is an exciting field of research that has the potential to give us insight into deep problems across various disciplines.

2 References

Jiang, Tianchong. “Wigner’s Semicircle Law for Gaussian Random Matrices”. University of Chicago, 2021.
http://math.uchicago.edu/~may/REU2021/REUPapers/Jiang,Tianchong.pdf

Liu, Yi-Kai. “Statistical Behavior of the Eigenvalues of Random Matrices”. Princeton University, 2001.
https://web.math.princeton.edu/mathlab/projects/ranmatrices/yl/randmtx.PDF

Luo, F., Zhong, J., Yang, Y., Scheuermann, R. H., & Zhou, J. (2006). Application of random matrix theory to biological networks. Physics Letters A, 357(6), 420-423.
https://arxiv.org/pdf/q-bio/0503035.pdf

Pennington, Jeffrey, and Pratik Worah. “Nonlinear random matrix theory for deep learning.” Advances in neural information processing systems 30 (2017).
https://proceedings.neurips.cc/paper_files/paper/2017/file/0f3d014eead934bbdbacb62a01dc4831-
pdf

Wolf, Vanessa, “Random Matrix Theory: A Combinatorial Proof of Wigner’s Semicircle Law” (2021). Scripps Senior Theses. 1683.
https://scholarship.claremont.edu/scripps_theses/1683