





Matrices aleatorias en teoría de portafolios: un enfoque desde la ciencia de datos

Andrés García Medina

Investigador Cátedras CONACyT, CIMAT-Monterrey https://sites.google.com/view/andresgm/home

Escuelas de verano del CIMAT 2022, Sede Monterrey, 6 y 7 de julio.

FOSEC SEP-INVESTIGACION BASICA-CONACY [A1-S-43514]

- Introduction
- 2 Elements of Random Matrices
- 3 Simulation: semicircle and Marchenko-Pastur laws
- Wigner's Surmise
- 5 Simulation: Wigner's Surmise
- 6 Portfolio Theory
- Simulation: optimal frontier and classical RMT technique
- Application

Motivation



Werner Heisenberg and Eugene Wigner (1928)

¹The statistical properties of the city transport in Cuernavaca (Mexico) and random matrix ensembles. J. Phys. A: Math. Gen. 33 (2000) L229–L234.

Background

Matrices appear in many areas of the sciences, from mathematics to physics, computer science, biology, economics, and quantitative finance.

In many circumstances the matrices that we encounter when modeling a phenomenon are large and without a particular structure.

Eugene Wigner discovered that it is possible (in many circumstances) to replace a large and complex (but deterministic) matrix by a typical element of a certain ensemble of random matrices.

Wigner's idea supposes the concept of *universality: the statistical distribution of eigenvalues does not depend on the specific matrix that represents the system, but on its symmetry.*

This idea has been incredibly fruitful and has led to the development of a subfield of mathematical physics called *Random Matrices Theory*.

Next we will investigate the simplest cases of ensembles of random matrices.

Definition

- What is a random matrix? An matrix whose elements are random variables
- Random Matrices Theory (RMT) replaces deterministic matrices with random matrices
- When working with very complicated matrices RMT replaces the system matrix with a random one and calculates averages (as well as the statistical properties of interest).

Example

Be a $H_{N\times N}$ matrix with i.i.d. elements such that $H_{ij} \sim N(0,1)$, para $i,j=1,\ldots,p$. Example(p=3):

$$H = \begin{pmatrix} 1.24 & 0.05 & -0.87 \\ -0.18 & 0.78 & -1.31 \\ -0.49 & -0.62 & 0.03 \end{pmatrix}$$
 (1)

We can notice $H_{ij} \neq H_{ji}$.

We call each realization a sample of the ensemble.

In general, the eigenvalues of H are complex. To obtain real eigenvalues we symmetrize our matrix H. An example of symmetrization is

$$H_s = (H + H')/2 \rightarrow H_s = H'_s \tag{2}$$

GOE

In our example

$$H = \begin{pmatrix} 1.24 & -0.065 & -0.68 \\ -0.065 & 0.78 & -0.965 \\ -0.68 & -0.965 & 0.03 \end{pmatrix}$$
(3)

With the convenience that $\lambda \in \{-0.8, 1.1, 1.7\}$ are reals

Congratulations! you have created your first random matrix from the Gaussian Orthogonal Ensemble (GOE).

GUE

In the case of a Hermitian matrix, for example

$$H_{her} = \begin{pmatrix} 0.3252 & 0.3077 + 0.2803i \\ 0.3077 - 0.2803i & -1.7115 \end{pmatrix}$$
 (4)

where $H_{her} = H_{her}^H$.

We now have a realization of the Unitary Gaussian Ensemble (GUE), with the characteristic that the inputs are complex, while the eigenvalues are real.

Specifically, in our example $\lambda \in \{-1.79, 0.406\}$

GSE

In the case of a symplectic matrix (relation to quaternions and systems with spin)

$$H_{2N\times 2N} = \begin{pmatrix} X & Y \\ -Y* & X* \end{pmatrix} \tag{5}$$

where $X, Y \in \mathbb{C}^{N \times N}$.

We symmetrize

$$H_{sc} = (H + H')/2 \tag{6}$$

It is obtained 2p eigenvalues, $\lambda \in \{\lambda_1, \lambda_1, \dots, \lambda_p, \lambda_p\}$.

Then we have a recipe to build a random matrix from the Symplectic Gaussian Ensemble (GSE).

Average spectral density

How do you calculate the shape of the histograms of $N \times m$ eigenvalues from the jpdf $\rho(\lambda_1, \ldots, \lambda_N)$?

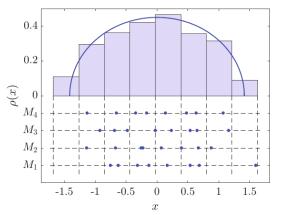


Figure (1) Source: Livan, G., Novaes, M., & Vivo, P. Introduction to random matrices theory and practice. Springer (2018).

Classification of Random Matrices Ensembles

⇒ No free lunch

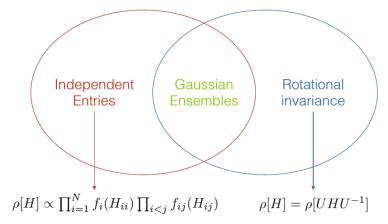


Figure (2) Livan, G., Novaes, M., & Vivo, P. Introduction to random matrices theory and practice. Springer (2018).

Classification of Random Matrices Ensembles

- Independent Entries: the first group on the left brings together matrix
 models whose inputs are independent random variables, modulo the
 symmetry requirements. Random matrices of this type are often called
 Wigner matrices. Examples: adjacency matrices of random graphs or
 matrices with independent power-law entries.
- **Rotational Invariance**: the second group from the right is characterized by the so-called rotational invariance. In essence, this property means that any two matrices that are related by a similarity transformation: $H^{(2)} = UH^{(1)}U'$ occur in the ensemble with equal probability.
- Intersección: What about the intersection between the two classes? It turns out that it only contains the Gaussian Ensembles!

Normalized traces and sample averages

We can generalize the notion of expected value and moments of classical probability to *large* random matrices.

It turns out that the proper analog of expected value is the normalized trace operator defined for the random matrix A of dimension $N \times N$ as

$$\phi(A) = \frac{1}{N} \mathbb{E}[TrA] \tag{7}$$

The normalization by 1/N is to have a finite operator when $N \to \infty$. For example $\phi(\mathbb{1}) = 1$ independent of the dimension.

In the case of polynomial functions of the matrix \boldsymbol{A} the trace of the function can be calculated on the eigenvalues 2

$$\frac{1}{N}Tr(F(A)) = \frac{1}{N}\sum_{k=1}^{N}F(\lambda_k)$$
 (8)

²Cyclic trace property

Normalized traces and sample averages

From here on we will denote $\langle \cdot \rangle$ as the average over the eigenvalues of a realization of A

$$\langle F(\lambda) \rangle := \frac{1}{N} \sum_{k=1}^{N} F(\lambda_k)$$
 (9)

In the case of random matrices, many scalar quantities like $\phi(F(A))$ do not fluctuate from sample to sample, or in other words, such fluctuations tend to zero in the large N limit.

This phenomenon is known in physics as *self-averaging*, while mathematicians speak of *concetration of measure*

$$\phi(F(A)) = \frac{1}{N} \mathbb{E}[Tr(F(A))] \approx \frac{1}{N} Tr(F(A)) = \langle F(\lambda) \rangle, \tag{10}$$

for a single realization of A.

Normalized traces and sample averages

When the eigenvalues of a random matrix A converge to a well-defined density $\rho(\lambda)$ we can write

$$\phi(F(A)) = \int \rho(\lambda)F(\lambda)d\lambda \tag{11}$$

Using $F(A) = A^k$ it is possible to define the k-th moment of a random matrix

$$m_k = \phi(A^k) \tag{12}$$

The first moment is simply the normalized trace of A, while $m_2 = \frac{1}{N} \sum_{ij} A_{ij}^2$ is the normalized sum of squares of all items. In particular, the square root of m_2 satisfies the norm axioms and is known as the Frobenius norm of A

$$||A||_F := \sqrt{m_2}$$
 (13)

Wigner Matrices

We have defined a Wigner matrix as a symmetric matrix (X=X') with Gaussian entries of mean zero (although they can be defined more generally).

The first few moments of the Wigner matrix X are given by

$$\phi(X) = \frac{1}{N} \mathbb{E}[TrX] = \frac{1}{N} Tr \mathbb{E}[X] = 0$$
 (14)

$$\phi(X^2) = \frac{1}{N} \mathbb{E}[TrXX'] = \frac{1}{N} \mathbb{E}\left[\sum_{ij=1}^{N} X_{ij}^2\right] = \frac{1}{N} [N(N-1)\sigma_{od}^2 + N\sigma_d^2]$$
 (15)

where $\sigma_d^2, \sigma_{od}^2$ are the diagonal and off-diagonal varianza, respectively.

To satisfy rotational invariance and keep the second finite moment is necessary to set

$$\sigma_d^2 = 2\sigma_{od}^2 = 2\sigma^2/N. \tag{16}$$

The above equations is another way to define the Orthogonal Gaussian Ensemble (GOE).

Rotational Invariance

We must remember that to rotate a vector v one applies a rotation matrix O: w = Ov, where O is an orthogonal matrix $O' = O^{-1}$ such that OO' = 1.

Observations:

- In general O is not symmetric.
- To rotate the base of X apply $\tilde{X} = OXO'$
- The eigenvalues of \tilde{X} are the same as those of X.
- The corresponding eigenvectors are $\{Ov\}$, where $\{v\}$ are the eigenvectors of X.

A rotationally invariant ensemble of random matrices is such that the matrix OXO' is just as likely as X itself, that is, $OXO' \stackrel{\text{in law}}{=} X$

Rotational Invariance

One way to observe the rotational invariance of the Wigner ensemble is through the joint density of its matrix elements

$$P(\{X_{ij}\}) = \left(\frac{1}{2\pi\sigma_{od}^2}\right)^N \left(\frac{1}{2\pi\sigma_{od}^2}\right)^{N(N-1)/2} \exp\left\{-\sum_{i=1}^N \frac{X_{ii}^2}{2\sigma_{od}^2} - \sum_{i< j}^N \frac{X_{ij}^2}{2\sigma_{od}^2}\right\}, \quad (17)$$

where only the diagonal and upper diagonal elements are independent variables.

Thus, with the election $\sigma_d^2=2\sigma_{od}^2=2\sigma^2/N$ we have (PBlackboard)

$$P(\lbrace X_{ij}\rbrace) \propto \exp\left\{-\frac{N}{4\sigma^2} Tr X^2\right\} \tag{18}$$

Then, under the change of variable $X \to \tilde{X} = OXO'$ the argument of the exponential is invariant.

In general, a matrix will be rotationally invariant when its joint probability density of its elements can be written as $P(\{M_{ij}\}) \propto \exp\{-NTrV(M)\}$, where $V(\cdot)$ is an arbitrary function.

Given a symmetric real matrix A of dimension $N \times N$, its resolvent is given by ³

$$G_A(z) = (z\mathbb{1} - A)^{-1},$$
 (19)

where z is a complex variable defined outside the domain of the eigenvalues of A. Then the Stieltjes transform of A is given by (blackboard)

$$g_N^A(z) = \frac{1}{N} Tr(G_A(z)) = \frac{1}{N} \sum_{k=1}^N \frac{1}{z - \lambda_k},$$
 (20)

where λ_k are the eigenvalues of A.

On the other hand, for a random matrix A we can define its empirical spectral distribution (ESD) also called sample eigenvalue density

$$\rho_N(\lambda) = \frac{1}{N} \sum_{k=1}^{N} \delta(\lambda - \lambda_k), \tag{21}$$

where $\delta(x)$ is the Dirac delta function.

³The resolvent formalism is a technique for applying complex analysis concepts to the study of the spectrum of certain operators.

In this way, the Stieltjes transform can be written as

$$g_N(z) = \int_{-\infty}^{\infty} \frac{\rho_N(\lambda)}{z - \lambda} d\lambda$$
 (22)

Note that $g_N(z)$ behaves well for any $z \notin \{\lambda_k : 1 \le k \le N\}$. In particular, it is well-defined at ∞ :

$$g_{N}(z) = \int \frac{\rho_{N}(\lambda)}{z - \lambda} d\lambda = \frac{1}{z} \int \frac{\rho_{N}(\lambda)}{1 - \lambda/z} d\lambda$$

$$= \frac{1}{z} \int \rho_{N}(\lambda) \left(\sum_{k=0}^{\infty} \left(\frac{\lambda}{z} \right)^{k} \right) d\lambda \quad \text{(geometric series)}$$

$$= \frac{1}{z} \sum_{k=0}^{\infty} \int \rho(\lambda) \left(\frac{\lambda}{z} \right)^{k} d\lambda = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \frac{1}{N} Tr(A^{k}), \quad \frac{1}{N} Tr(A^{0}) = 1.$$
(23)

The cases of interest are the random matrices A such that for large values of N the normalized traces of the powers of A converge to their expected values

$$\lim_{N \to \infty} \frac{1}{N} \operatorname{Tr}(A^k) = \phi(A^k) \tag{24}$$

It is found that for large values of z, the function $g_A(z)$ converges to a deterministic limit $g(z) = \lim_{N \to \infty} \mathbb{E}[g_N(z)]$, equivalently

$$g(z) = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \phi(A^k).$$
 (25)

Thus, g(z) is the moment-generating function of A.

In random matrices it holds that the knowledge of all moments of A is equivalent to knowledge of the eigenvalue densities of A.

In the same way, in the limit $N \to \infty$ the poles of the ESD merge, so

$$\frac{1}{N} \sum_{k=1}^{N} \delta(\lambda - \lambda_k) \sim \rho(\lambda) \tag{26}$$

Thus, the density $\rho(\lambda)$ has an extended support. Tenemos entonces

$$g(z) = \int_{supp[\rho]} \frac{\rho(\lambda)d\lambda}{z - \lambda}$$
 (27)

which is the Stieltjes transform of the limit measure $\rho(\lambda)$

Limit spectral density

If we know the Stieltjes transform for a given ensemble it is possible to know its limiting spectral density

$$g(x - i\epsilon) = \int \frac{\rho(x')d\lambda}{z - x'} = \int dx' \frac{\rho(x')}{x - i\epsilon - x'} \left(\frac{x + i\epsilon - x'}{x + i\epsilon - x'}\right)$$
$$= \int dx' \frac{\rho(x')(x - x')}{(x - x')^2 + \epsilon^2} + i \int dx' \rho(x') \frac{\epsilon}{(x - x')^2 + \epsilon^2}.$$
(28)

Let us take the limit $\epsilon \to 0^+$ (blackboard)

$$\lim_{\epsilon \to 0^+} g(x - i\epsilon) = Pr\left[\int dx' \frac{\rho(x')}{x - x'}\right] + i\pi \rho(x)$$
 (29)

So we can write

$$\rho(x) = \frac{1}{\pi} \lim_{\epsilon \to 0^+} Im\{g(x - i\epsilon)\}$$
(30)

This implies that if we can calculate the Stieltjes transform in the complex plane then it is possible to obtain the limiting spectral density.

To find the Stieltjes transform of the Wigner ensemble we can use the cavity method or *self-consistent equation*.

This method consists of finding a relation between the Stiltjes transform of a Wigner matrix of size N and one of size N-1

In the large N limit, both transformations must converge to the same limit, so we get a self-consistent equation that can be solved relatively easily.

So, we want to compute $g_N^X(z)$ when X is a Wigner matrix with $X_{ij} \sim N(0, \sigma^2/N)$ and $X_{ii} \ sim N(0, 2\sigma^2/N)$. In the limit of N large, g_N^X converges to a well-defined function g(z).

For this we need to remember the Schur's complement formula.

Schur's complement relates the blocks of the inverse of a matrix to the inverse of the blocks of the original matrix.

Property

Let be an invertible M matrix which we divide into four blocks

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \quad y \quad M^{-1} = Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \tag{31}$$

where

 $[M_{11}] = n \times n$, $[M_{12}] = n \times (N-n)$, $[M_{21}] = (N-n) \times n$, $[M_{22}] = (N-n) \times (N-n)$, $y \ M_{22}$ is invertible. Also, the integer n can take any value between 1 and N-1. So, the upper left block of Q $(n \times n)$ is given by

$$Q_{11}^{-1} = M_{11} - M_{12}(M_{22})^{-1}M_{21}, (32)$$

where the right hand side is known as the Schur's complement of the block M_{22} of the matrix M.

Using Schur's complement formula we can calculate the elements of $M=z\mathbb{1}-X$. Así $M^{-1}=G_X$, y

$$\frac{1}{(G_X)_{11}} = M_{11} - \sum_{k,l=2}^{N} M_{1k} (M_{22})_{kl}^{-1} M_{l1}, \tag{33}$$

where M_{22} is the submatrix of size $(N-1) \times (N-1)$ of M such that the first row and column have been removed.

Assumptions: (i)
$$\phi(X) = 0$$
, (ii) $\phi(X^2) = \frac{1}{N}[N(N-1)\sigma_{od}^2 + N\sigma_d^2]$, (iii) $\sigma_d^2 = 2\sigma_{od}^2 = 2\sigma^2/N$. (iv) i.i.d. (v) $\phi(X^M) = 0$, $M > 2$

It is argued that for large N the right hand side is dominated by its expected values with small fluctuations of the order $\mathcal{O}(1/\sqrt{N})$. Therefore, only its expected value will be calculated, although it is also possible to estimate its fluctuations (assumption iii and v).

Then, we need to compute the expected value of the terms:

- $\frac{1}{(G_X)_{11}}$
- M_{11}

The second term is direct: $\mathbb{E}[M_{11}] = \mathbb{E}[z - X_{11}] = z$ (assumption i)

The entries of M_{22} are independent of $M_{1i} = -X_{1i}$ (assumption iv).

Thus, we can take the partial expectation on the elements $\{X_{1i}\}, i=2,\ldots,N;$ first and obtain (assumption i and ii)

$$\mathbb{E}_{X_{1i}}\left[M_{1i}(M_{22})_{ij}^{-1}M_{j1}\right] = (M_{22})_{ij}^{-1}\mathbb{E}_{X_{1i}}\left[M_{1i}M_{j1}\right] =$$
(34)

$$(M_{22})_{ii}^{-1}\sigma_{od}^2\delta_{ij} = \frac{\sigma^2}{N}(M_{22})_{ii}^{-1}\delta_{ij}$$
 (35)

Then

$$\mathbb{E}_{X_{1i}} \left[\sum_{k,l=2}^{N} M_{1k} (M_{22})_{kl}^{-1} M_{l1} \right] = \frac{\sigma^2}{N} \operatorname{Tr} \left((M_{22})^{-1} \right)$$
 (36)

Further $\frac{1}{(N-1)} Tr \left((M_{22})^{-1} \right)$ is the Stieltjes transform of a Wigner matrix of size

N-1 and variance $\sigma^2(N-1)/N$.

In the large N limit, the Stieltjes transform must be independent of the size of the matrix, so the difference between N and N-1 is negligible.

From the above we have

$$\mathbb{E}\left[\sum_{k,l=2}^{N} M_{1k} (M_{22})_{kl}^{-1} M_{l1}\right] = \mathbb{E}\left[\frac{\sigma^2}{N} \operatorname{Tr}\left((M_{22})^{-1}\right)\right] \to \sigma^2 g(z) \tag{37}$$

Thus $1/(G_X)_{11}$ approaches a deterministic number with negligible fluctuations.

The expectation of left side can be approximated by (Jensen inequality)

$$\mathbb{E}\left[\frac{1}{(G_X)_{11}}\right] \ge \frac{1}{\mathbb{E}\left[(G_X)_{11}\right]} \tag{38}$$

The equality holds when $N \to \infty$ (negligible variance)

Now, from the rotational invariance of X and consequently of G_X , all diagonal elements of G_X must have the same expected value

$$\mathbb{E}[(G_X)_{11}] = \frac{1}{N} (\mathbb{E}[(G_X)_{11}] + \dots + \mathbb{E}[(G_X)_{11}]) =$$
(39)

$$\frac{1}{N}\mathbb{E}[(G_X)_{11} + \dots + (G_X)_{NN}] =$$
 (40)

$$\frac{1}{N}\mathbb{E}[Tr(G_X)] = \mathbb{E}[g_N] \to g \tag{41}$$

Then, for $N \to \infty$

$$\mathbb{E}\left[\frac{1}{(G_X)_{11}}\right] \to \frac{1}{g(z)} \tag{42}$$

Considering all the previous arguments, the following quadratic expression for g(z) is finally obtained

$$\frac{1}{g(z)} = z - \sigma^2 g(z) \tag{43}$$

With a little extra algebra we can finally get the well-known semicircle law(blackboard) . . .

$$\rho(x) = \frac{1}{\pi} \sqrt{2 - x^2} \tag{44}$$

Semicircle law

Teorema

Suppose H=(A+A')/2 is a $N\times N$ matrix, where the elements $A_{i,j}$ are real random variables i.i.d. $\sim \mathcal{N}(0,1)$. Then, when $N\to\infty$ the spectral density function of H converges (a.s.) to Wigner's semicircle law:

$$\rho(x) = \frac{1}{\pi} \sqrt{2 - x^2} \tag{45}$$

Equivalently, a quadratic equation in g(z) can be found for the Wishart ensemble

$$\frac{1}{g(z)} = z - 1 + q - qzg(z) \tag{46}$$

from where we get the density of eigenvalues

$$\rho(\lambda) = \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{2\pi q \lambda},\tag{47}$$

where

$$\lambda_{min}^{max} = (1 \pm \sqrt{q})^2, \quad q = p/n \tag{48}$$

Marchenko-Pastur law

Teorema

Let X be a matrix of dimensions $p \times n$, where the elements $X_{i,j}$ are i.i.d. $\mathcal{N}(0,1)$. Then, when $p, n \to \infty$, such that $\frac{p}{n} \to q \in (0,\infty)$, the eingenvalue spectral density of the Wishart matrix $W = \frac{1}{T}XX'$ converges (a.s.) to the Marcenko-Pastur law a

$$\rho(\lambda) = \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{2\pi q \lambda},\tag{49}$$

where

$$\lambda_{min}^{max} = (1 \pm \sqrt{q})^2. \tag{50}$$

^aMarchenko VA, Pastur LA. Distribution of eigenvalues for some sets of random matrices. Sb. Math. 1967;114(4):507-36.

Example: semicircle law

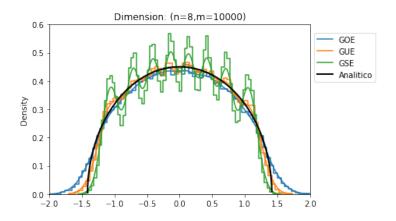
We generate m=10000 matrices for N=8, and the normalized histograms of the full sample of $m \times N$ eigenvalues are computed.

To obtain the normalized histograms, they must be scaled by the factor $\frac{1}{\sqrt{\beta N}}$, where $\beta=1,2,4$, for the case *GOE*, *GUE*, and *GSE*; respectively.

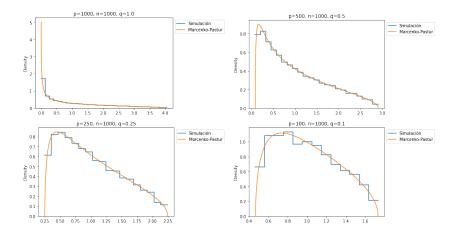
The simulations are compared with the analytical result $\rho(x) = \frac{1}{\pi}\sqrt{2-x^2}$, called Wigner's semicircle law. The following analytical results are obtained for the bounds of the support of the distributions

- $\pm\sqrt{2N}$ (GOE)
- $\pm \sqrt{4N}$ (GUE)
- $\pm\sqrt{8N}$ (GSE)

Example: semicircle law Notebook



Example: Marchenko-Pastur law Notebook



Exercises and homework

- Exercise: Reproduce the density for the GUE and GSE cases
- Exercise: The Marchenko-Pastur law is valid for $N(0, \sigma^2)$ processes. In this case the density is divided by σ^2 and the bounds are extended by the same factor. Simulate processes with $\sigma^2 = 0.1, 2, 10$. What do you observe?
- Homework: read Krbálek, M., & Seba, P. (2000). The statistical properties of the city transport in Cuernavaca (Mexico) and random matrix ensembles. Journal of Physics A: Mathematical and General, 33(26), L229.
- Material: github.com/agarciam/ curso-de-matrices-aleatorias-y-portafolios