

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>

Departamento de Economía y Empresa, Universidad de Almería, 21-22 junio, 2022

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

- 1 Introducción
- 2 Elementos de matrices aleatorias
- 3 Simulación ley del semicírculo y Marchenko-Pastur
- 4 Teoría de Portafolios
- 5 Simulación fronteras óptimas y RMT

Motivación

► Universalidad¹



Werner Heisenberg y 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.

Antecedentes

Las matrices aparecen en muchas áreas de las ciencias, desde las matemáticas hasta la física, ciencias de la computación, biología, economía y finanzas cuantitativas.

En muchas circunstancias las matrices que encontramos al modelar un fenómeno son grandes y sin una estructura particular.

Eugene Wigner descubrió que es posible (en muchas circunstancias) reemplazar una matriz grande y compleja (pero determinista) por un elemento típico de un cierto ensemble de matrices aleatorias.

La idea de Wigner supone el concepto de *universalidad*: *la distribución estadística de valores propios no dependen de la matriz específica que representa al sistema, sino de su simetría*.

Esta idea ha sido increíblemente fructífera y ha llevado al desarrollo de un subcampo de la física matemática denominado *Teoría de Matrices Aleatorias*.

A continuación vamos a investigar los casos más simples de ensembles de matrices aleatorias.

Definición

- ¿Qué es una matriz aleatoria? Una matriz cuyos elementos son variables aleatorias
- La teoría de matrices aleatorias (RMT) reemplaza matrices deterministas por matrices aleatorias
- Cuando se trabaja con matrices muy complicadas RMT reemplaza la matriz del sistema por una aleatoria y calcula promedios (así como las propiedades estadísticas de interés).

Ejemplo

Consideremos una matriz $H_{N \times N}$ con elementos i.i.d. tal que $H_{ij} \sim N(0, 1)$, para $i, j = 1, \dots, p$. Ejemplo ($p = 3$):

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

Notamos que $H_{ij} \neq H_{ji}$.

A cada realización la llamamos muestra del ensemble.

En general, los eigenvalores de H son complejos. Para obtener valores propios reales simetrizamos nuestra matriz H . Un ejemplo de simetrización es

$$H_s = (H + H')/2 \rightarrow H_s = H'_s \quad (2)$$

GOE

En nuestro ejemplo

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

Con la convención de que $\lambda \in \{-0.8, 1.1, 1.7\}$ son reales.

¡Felicidades! ha producido su primera matriz aleatoria a partir del ensemble ortogonal gaussiano (GOE).

GUE

En el caso de una matriz hermitiana, por ejemplo

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

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

Tenemos ahora una realización del ensemble unitario gaussiano (GUE), con la característica de que las entradas son complejas, mientras que los eigenvalores son reales.

Específicamente en nuestro ejemplo $\lambda \in \{-1.79, 0.406\}$

GSE

En el caso de una matriz simpléctica (relación con cuaterniones)

$$H_{2N \times 2N} = \begin{pmatrix} X & Y \\ -Y_* & X_* \end{pmatrix} \quad (5)$$

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

Simetrizamos

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

Se tienen $2p$ eigenvalores, $\lambda \in \{\lambda_1, \lambda_1, \dots, \lambda_p, \lambda_p\}$.

Esta es la receta para construir una matriz del ensemble simplectico gaussiano (GSE).

Densidad espectral promedio

¿Cómo se calcula la forma de los histogramas de $N \times m$ eigenvalores a partir de la jpdf $\rho(\lambda_1, \dots, \lambda_N)$?

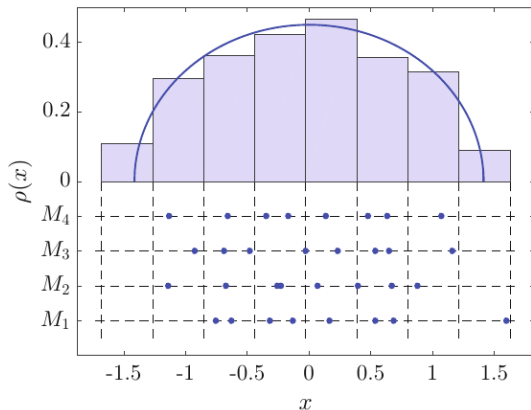


Figure (1) Source: Livan, G., Novaes, M., & Vivo, P. (2018). Introduction to random matrices theory and practice. Springer (2018).

Clasificación de ensembles de matrices aleatorias

Matrices de Wigner, Ensemble Gaussiano, Wishart-Laguerre

⇒ No free lunch

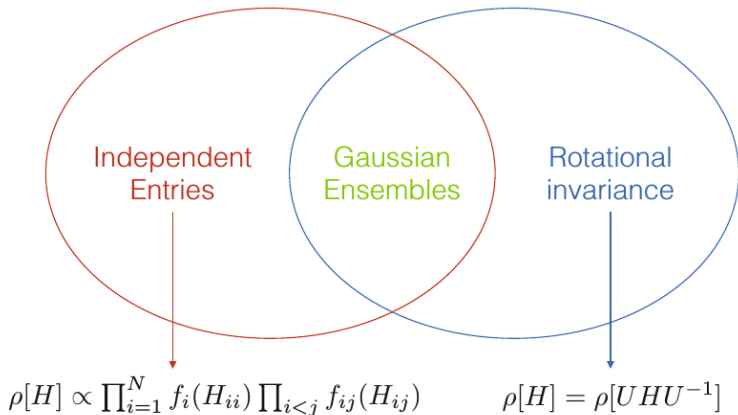


Figure (2) Source: Livan, G., Novaes, M., & Vivo, P. Introduction to random matrices theory and practice. Springer (2018).

Clasificación de ensembles de matrices aleatorias

- **Entradas independientes:** el primer grupo de la izquierda reúne modelos matriciales cuyas entradas son variables aleatorias independientes, módulo los requisitos de simetría. Las matrices aleatorias de este tipo suelen denominarse matrices de Wigner. Ejemplos: matrices de adyacencia de gráficos aleatorios o matrices con entradas de ley de potencias independientes.
- **Invarianza rotacional:** el segundo grupo de la derecha se caracteriza por la llamada invarianza rotacional. En esencia, esta propiedad significa que dos matrices cualesquiera que estén relacionadas mediante una transformación de similitud $H^{(2)} = UH^{(1)}U'$ ocurren en el ensemble con la misma probabilidad.
- **Intersección:** ¿Qué pasa con la intersección entre las dos clases? ¡Resulta que solo contiene el conjunto gaussiano!

Trazas normalizadas y promedios de muestra

Podemos generalizar la noción de valor esperado y momentos de la probabilidad clásica hacia matrices aleatorias *grandes*.

Resulta que el análogo adecuado de valor esperado es el operador de traza normalizado definido para la matriz aleatoria A de dimensión $N \times N$ como

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

La normalización por $1/N$ es para que el operador sea finito en $N \rightarrow \infty$. Por ejemplo $\phi(\mathbb{1}) = 1$ independientemente de la dimensión.

En el caso de funciones polinomiales de la matriz A la traza de la función se puede calcular sobre los valores propios ²

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

²Propiedad cíclica de la traza

Trazas normalizadas y promedios de muestra

De aquí en adelante denotaremos $\langle \cdot \rangle$ como el promedio sobre los valores propios de una realización de A

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

En el caso de las matrices aleatorias, muchas cantidades escalares como $\phi(F(A))$ no fluctúan de muestra en muestra, o en otras palabras, tales fluctuaciones tienden a cero en límite de N grande.

A este fenómeno se le conoce en física como *self-averaging*, mientras que los matemáticos hablan de *concentration of measure*

$$\phi(F(A)) = \frac{1}{N} \mathbb{E}[Tr(F(A))] \approx \langle F(\lambda) \rangle \quad \text{para una realización de } A \quad (10)$$

Trazas normalizadas y promedios de muestra

Cuando los valores propios de una matriz aleatoria A convergen a una densidad bien definida $\rho(\lambda)$ podemos escribir

$$\phi(F(A)) = \int \rho(\lambda) F(\lambda) d\lambda \quad (11)$$

Utilizando $F(A) = A^k$ es posible definir el k -ésimo momento de una matriz aleatoria

$$m_k = \phi(A^k) \quad (12)$$

Particularmente, la raíz cuadrada de m_2 satisface los axiomas de norma y se le conoce como la norma de Frobenius de A

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

Ensemble de Wigner

Hemos definido a una matriz de Wigner como una matriz simétrica ($X=X'$) con entradas gaussianas de media cero (aunque pueden ser definidas de manera más general).

Los primeros momentos de las matriz de Wigner X están dados por

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

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

donde $\sigma_d^2, \sigma_{od}^2$ son la varianza en la diagonal y fuera de la diagonal, respectivamente.

Para respetar la invariancia rotacional y mantener el segundo momento finito elegimos $\sigma_d^2 = 2\sigma_{od}^2 = \sigma^2/N$.

De esta manera, hemos definido al ensemble ortogonal gaussiano (GOE) a través de sus dos primeros momentos.

Invariancia rotacional

Debemos recordar que para rotar a un vector v uno aplica una matriz de rotación $O : w = Ov$, donde O es una matriz ortogonal $O' = O^{-1}$ tal que $OO' = 1$.

Observaciones:

- En general O no es simétrica.
- Para rotar la base de la matrix X uno aplica $\tilde{X} = OXO'$
- Los valores propios de \tilde{X} son los mismos de X .
- Los vectores propios son $\{Ov\}$, donde $\{v\}$ son los vectores propios de X .

Un ensemble de matrices aleatorias rotacionalmente invariante es tal que la matriz OXO' es igual de probable que X misma, es decir, $OXO' \stackrel{\text{in law}}{=} X$

Invariancia rotacional

Una forma de observar la invariancia rotacional del ensemble de Wigner es a través de la densidad conjunta de sus elementos de matriz

$$P(\{X_{ij}\}) = \left(\frac{1}{2\pi\sigma_d^2}\right)^N \left(\frac{1}{2\pi\sigma_{od}^2}\right)^{N(N-1)/2} \exp \left\{ -\sum_{i=1}^N \frac{X_{ii}^2}{2\sigma_d^2} - \sum_{i<j}^N \frac{X_{ij}^2}{2\sigma_{od}^2} \right\}, \quad (16)$$

donde solo los elementos diagonales y de la diagonal superior son variables independientes.

Así, con la elección $\sigma_d^2 = 2\sigma_{od}^2 = \sigma^2/N$ tenemos

$$P(\{X_{ij}\}) \propto \exp \left\{ -\frac{N}{4\sigma^2} \text{Tr} X^2 \right\} \quad (17)$$

Bajo el cambio de variable $X \rightarrow \tilde{X} = OXO'$ el argumento de la exponencial es invariante.

En general, una matriz , será rotacionalmente invariante cuando su densidad conjunta de probabilidad de sus elementos pueda ser escrita como $P(\{M_{ij}\}) \propto \exp\{-N\text{Tr}V(M)\}$, donde $V(\cdot)$ es una función arbitraria.

Resolvente y transformada de Stieltjes

Dada una matrix real simétrica A de dimensión $N \times N$, su resolvente está dado por ³

$$G_A(z) = (z\mathbb{1} - A)^{-1}, \quad (18)$$

donde z es una variable compleja definida fuera del dominio de los valores propios de A . Entonces, la transformada de Stieltjes de A está dada por ([pizarrón](#))

$$g_N^A(z) = \frac{1}{N} \text{Tr}(G_A(z)) = \frac{1}{N} \sum_{k=1}^N \frac{1}{z - \lambda_k}, \quad (19)$$

donde λ_k son los valores propios de A .

Por otro lado, para una matriz aleatoria A podemos definir su distribución espectral empírica (ESD) llamada también densidad de valores propios muestrales

$$\rho_N(\lambda) = \frac{1}{N} \sum_{k=1}^N \delta(\lambda - \lambda_k), \quad (20)$$

donde $\delta(x)$ es la función delta de Dirac.

³El formalismo del resolvente es una técnica para aplicar conceptos de análisis complejos al estudio del espectro de ciertos operadores.

Resolvente y transformada de Stieltjes

De esta manera, la transformada de Stieltjes se puede escribir como

$$g_N(z) = \int_{-\infty}^{\infty} \frac{\rho_N(\lambda)}{z - \lambda} d\lambda \quad (21)$$

Notese que $g_N(z)$ se comporta bien para cualquier $z \notin \{\lambda_k : 1 \leq k \leq N\}$. En particular, está bien definida en ∞ :

$$\begin{aligned} g_N(z) &= \int \frac{\rho_N(\lambda)}{z - \lambda} d\lambda = \frac{1}{z} \int \frac{\rho_N(\lambda)}{1 - \lambda/z} d\lambda \\ &= \frac{1}{z} \int \rho_N(\lambda) \left(\sum_{k=0}^{\infty} \left(\frac{\lambda}{z} \right)^k \right) d\lambda \quad (\text{serie geométrica}) \\ &= \frac{1}{z} \sum_{k=0}^{\infty} \int \rho(\lambda) \left(\frac{\lambda}{z} \right)^k d\lambda = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \frac{1}{N} \text{Tr}(A^k), \quad \frac{1}{N} \text{Tr}(A^0) = 1. \end{aligned} \quad (22)$$

Resolvente y transformada de Stieltjes

Los casos de interés son las matrices aleatorias A tal que para valores grandes de N las trazas normalizadas de las potencias de A convergen a sus valores esperados

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr}(A^k) = \phi(A^k) \quad (23)$$

Se busca que para valores grandes de z , la función $g_A(z)$ converge a un límite determinista $g(z) = \lim_{N \rightarrow \infty} \mathbb{E}[g_N(z)]$, equivalentemente

$$g(z) = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \phi(A^k). \quad (24)$$

Así, $g(z)$ es la función generadora de momentos de A .

En el nivel de rigor de este curso, el conocimiento de todos los momentos de A es equivalente al conocimiento de la densidad de valores propios de A .

Resolvente y transformada de Stieltjes

De la misma manera, en el límite $N \rightarrow \infty$ los polos de la ESD se fusionan, por lo que

$$\frac{1}{N} \sum_{k=1}^N \delta(\lambda - \lambda_k) \sim \rho(\lambda) \quad (25)$$

Así, la densidad $\rho(\lambda)$ tiene un soporte extendido. Tenemos entonces

$$g(z) = \int_{\text{supp}[\rho]} \frac{\rho(\lambda) d\lambda}{z - \lambda} \quad (26)$$

La cual es la transformada de Stieltjes de la medida límite $\rho(\lambda)$

Densidad espectral límite

Si conocemos la transformada de Stieltjes para un ensemble dado es posible conocer su densidad espectral límite

$$\begin{aligned}
 g(x - i\epsilon) &= \int \frac{\rho(x') d\lambda}{z - x'} = \int dx' \frac{\rho(x')}{x - i\epsilon - x'} \left(\frac{x + i\epsilon - x'}{x + i\epsilon - x'} \right) \\
 &= \int dx' \frac{\rho(x')(x - x')}{(x - x')^2 + \epsilon^2} + i \int dx' \rho(x') \frac{\epsilon}{(x - x')^2 + \epsilon^2}.
 \end{aligned} \tag{27}$$

Tomemos el límite $\epsilon \rightarrow 0^+$ ([► Pizarron](#))

$$\lim_{\epsilon \rightarrow 0^+} g(x - i\epsilon) = Pr \left[\int dx' \frac{\rho(x')}{x - x'} \right] + i\pi\rho(x) \tag{28}$$

Por lo que podemos escribir

$$\rho(x) = \frac{1}{\pi} \lim_{\epsilon \rightarrow 0^+} Im\{g(x - i\epsilon)\} \tag{29}$$

Esto implica que si podemos calcular la transformada de Stieltjes en el plano complejo es posible entonces obtener la densidad espectral límite.

Transformada de Stieltjes del ensemble de Wigner

Para encontrar la transformada de Stieltjes del ensemble de Wigner podemos usar el método de la cavidad o *self-consistent equation*.

Este método consiste en encontrar una relación entre la transformada de Stieltjes de una matriz de Wigner de tamaño N y una de tamaño $N - 1$.

En el límite de N grande, ambas transformadas deben converger al mismo límite, con lo que obtenemos una ecuación auto-consistente que puede ser resuelta relativamente fácil.

Entonces, deseamos calcular $g_N^X(z)$ cuando X es una matriz de Wigner con $X_{ij} \sim N(0, \sigma^2/N)$ y $X_{ii} \sim N(0, 2\sigma^2/N)$. En el límite de N grande, g_N^X converge a una función bien definida $g(z)$.

Para esto necesitamos revisar la fórmula del complemento de Schur.

Transformada de Stieltjes del ensemble de Wigner

El complemento de Schur relaciona los bloques de la inversa de una matriz con la inversa de los bloques de la matriz original.

Resultado

Sea una matriz M invertible la cual dividimos en cuatro bloques

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \quad y \quad M^{-1} = Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \quad (30)$$

donde

$[M_{11}] = n \times n$, $[M_{12}] = n \times (N - n)$, $[M_{21}] = (N - n) \times n$, $[M_{22}] = (N - n) \times (N - n)$, y M_{22} es invertible. Además, el entero n puede tomar cualquier valor entre 1 y $N - 1$. Entonces, el bloque superior izquierdo de Q ($n \times n$) está dado por

$$Q_{11}^{-1} = M_{11} - M_{12}(M_{22})^{-1}M_{21}, \quad (31)$$

donde el lado derecho se le conoce como el complemento de Schur del bloque M_{22} de la matriz M

Transformada de Stieltjes del ensemble de Wigner

Usando la formula del complemento de Schur podemos calcular los elementos de $M = z\mathbb{1} - X$. Así $M^{-1} = G_X$, y

$$\frac{1}{(G_X)_{11}} = M_{11} - \sum_{k,l=2}^N M_{1k} (M_{22})_{kl}^{-1} M_{l1}, \quad (32)$$

donde M_{22} es la submatriz de tamaño $(N-1) \times (N-1)$ de M tal que el primer renglón y columna se ha removido.

Se argumenta que para N grande el lado derecho está dominado por sus valor esperado con fluctuaciones pequeñas del orden $\mathcal{O}(1/\sqrt{N})$. Por lo que sólo se calculará su valor esperado, aunque es posible estimar también sus fluctuaciones ⁴.

⁴las variables gaussianas están completamente determinadas por su media y varianza

Transformada de Stieltjes del ensemble de Wigner

Notese que

- $\mathbb{E}[M_{11}] = z$.
- Las entradas de M_{22} son independientes de $M_{1i} = -X_{1i}$

Así, podemos tomar la esperanza parcial sobre los elementos $\{X_{1i}\}$ y obtener ⁵

$$\mathbb{E}_{X_{1i}} \left[M_{1i} (M_{22})_{ij}^{-1} M_{j1} \right] = \frac{\sigma^2}{N} (M_{22})_{ii}^{-1} \delta_{ij} \quad (33)$$

Entonces

$$\mathbb{E}_{X_{1i}} \left[\sum_{k,l=2}^N M_{1k} (M_{22})_{kl}^{-1} M_{l1} \right] = \frac{\sigma^2}{N} \text{Tr}((M_{22})^{-1}) \quad (34)$$

Además

- $\frac{1}{(N-1)} \text{Tr}((M_{22})^{-1})$ es la transformada de Stieltjes de una matriz de Wigner de tamaño $N-1$ y varianza $\sigma^2(N-1)/N$
- En el límite de N grande, la transformada de Stieltjes debe ser independiente del tamaño de la matriz, por lo que la diferencia entre N y $N-1$ es despreciable.

⁵iid, media cero, y varianza fuera de la diagonal definen a un matriz de Wigner

Transformada de Stieltjes del ensemble de Wigner

Por lo anterior tenemos que

$$\mathbb{E} \left[\frac{1}{N} \text{Tr} ((M_{22})^{-1}) \right] \rightarrow g(z) \quad (35)$$

Así $1/(G_X)_{11}$ tiende a un número determinista con fluctuaciones despreciables. Entonces, en el límite de N grande

$$\mathbb{E} \left[\frac{1}{(G_X)_{11}} \right] = \frac{1}{\mathbb{E} [(G_X)_{11}]} \quad (36)$$

De la invariancia rotacional de X y consecuentemente de G_X , todos los elementos diagonales de G_X deben tener el mismo valor esperado

$$\mathbb{E} [(G_X)_{11}] = \frac{1}{N} \mathbb{E} [\text{Tr}(G_X)] = \mathbb{E}[g_N] \rightarrow g \quad (37)$$

Transformada de Stieltjes del ensemble de Wigner

Considerando todos los argumentos anteriores se obtiene la siguiente expresión cuadrática para $g(z)$:

$$\frac{1}{g(z)} = z - \sigma^2 g(z) \quad (38)$$

Con un poco de álgebra extra podemos obtener finalmente la ley del semicírculo (► pizarrón) ...

$$\rho(x) = \frac{1}{\pi} \sqrt{2 - x^2} \quad (39)$$

Ley del semicírculo

Teorema

Suponga que $H = (A + A')/2$ es una matriz $N \times N$, donde los elementos $A_{i,j}$ son variables aleatorias reales i.i.d. $\sim \mathcal{N}(0, 1)$. Entonces, cuando $N \rightarrow \infty$ la función de densidad espectral de H converge (a.s.) a la *ley del semicírculo de Wigner*:

$$\rho(x) = \frac{1}{\pi} \sqrt{2 - x^2} \quad (40)$$

Transformada de Stieltjes del ensemble de Wishart

De manera equivalente se puede encontrar una ecuación cuadrática en $g(z)$ para el ensemble de Wishart

$$\frac{1}{g(z)} = z - 1 + q - qzg(z) \quad (41)$$

de donde se obtiene la densidad de valores propios

$$\rho(\lambda) = \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{2\pi q\lambda}, \quad (42)$$

donde

$$\lambda_{\min}^{\max} = (1 \pm \sqrt{q})^2, \quad q = p/n \quad (43)$$

Ley de Marcenko-Pastur

Teorema

Sea X una matriz de dimensiones $p \times n$, donde los elementos $X_{i,j}$ are i.i.d. $\mathcal{N}(0, 1)$. Entonces, cuando $p, n \rightarrow \infty$, tal que $\frac{p}{n} \rightarrow q \in (0, \infty)$, la densidad espectral de eigenvalores de la matriz de Wishart $W = \frac{1}{T}XX'$ converge (a.s.) a la *ley de Marcenko-Pastur*^a

$$\rho(\lambda) = \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{2\pi q\lambda}, \quad (44)$$

donde

$$\lambda_{\min}^{\max} = (1 \pm \sqrt{q})^2. \quad (45)$$

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

Ejemplo: ley de semicírculo

Se genera $m = 10000$ matrices para $N = 8$, y se producen los histogramas normalizados de la muestra completa de $m \times N$ valores propios.

Para obtener los histogramas normalizados, se deben escalar por el factor $\frac{1}{\sqrt{\beta p}}$, donde $\beta = 1, 2, 4$, para el caso *GOE*, *GUE*, y *GSE*; respectivamente.

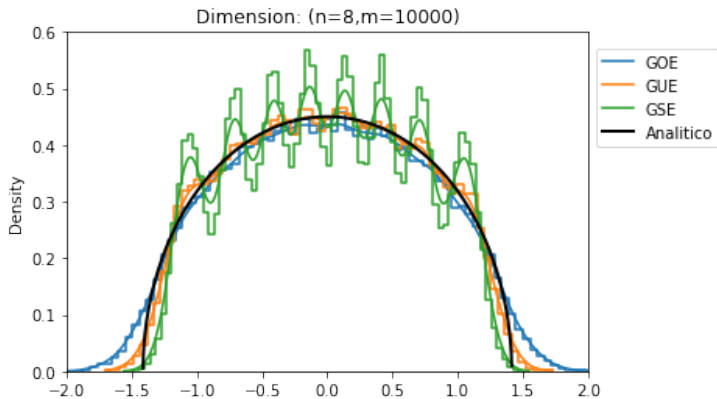
Se compara con el resultado analítico $\rho(x) = \frac{1}{\pi} \sqrt{2 - x^2}$, llamada *ley del semicírculo de Wigner*. Se tienen los siguiente resultados analíticos para las cotas del soporte de las distribuciones

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

Ejemplo: ley de semicírculo

[▶ Notebook](#)

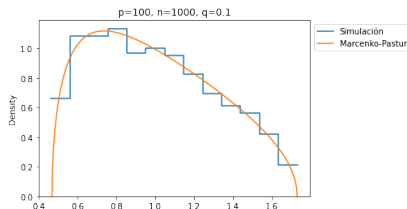
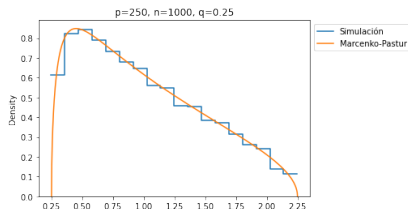
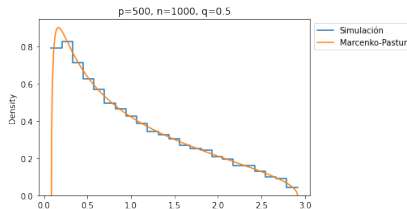
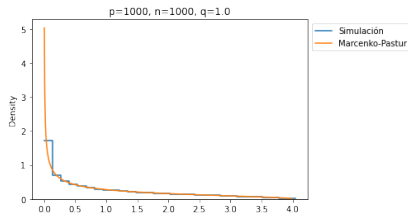
github.com/agarciam/curso-de-matrices-aleatorias-y-portafolios



Ejemplo: ley de Marcenko-Pastur

[▶ Notebook](#)

github.com/agarciam/curso-de-matrices-aleatorias-y-portafolios



Teoría de portafolios

La implementación de la teoría de portafolios no es sencilla. Requiere la estimación de la matriz de covarianza y el pronóstico del rendimiento de los activos.

Un problema es que los portafolios óptimos son muy sensibles a las entradas.

Algunos problemas de estabilidad hacen que la práctica de la optimización de portafolios de inversión sea menos atractiva que la teoría.

En este caso, se pueden emplear técnicas de regularización para atenuar estos problemas.

Este enfoque es ampliamente apoyado por Ledoit y Wolf (2003), quienes proponen combinar diferentes estimadores de matrices de covarianza para estabilizar la solución.

Teoría de portafolios

Hoy en día, uno de los enfoques más prometedores consiste en interpretar los portafolios óptimos como la solución de un problema de regresión lineal y utilizar la penalización tipo *Lasso* o *Ridge*.

Sin embargo, la regularización no es suficiente para obtener soluciones satisfactorias, por lo que los profesionales introducen algunas limitaciones en el problema de optimización.

Estas restricciones pueden interpretarse como un método de contracción.

Al imponer restricciones de peso, el administrador del portafolio cambia implícitamente la matriz de covarianza.

Este enfoque equivale entonces a tener algunas suposiciones a priori y, por tanto, se relaciona con el modelo de Black y Litterman (1992).

La frontera eficiente

- En 1952 Markowitz introdujo el concepto de la frontera eficiente. Esta fue la primera formulación matemática acerca de la optimización de portafolios.
- La idea detrás es de que los inversionistas ven el retorno esperado como una cosa buena, mientras que la varianza de los retornos como una cosa mala.
- Lo que demostro Markowitz es de que existe un conjunto de portafolios optimos, los cuales maximizan el retorno esperado, dado un valor de riesgo. A esto fue lo que llamo *frontera eficiente*.

Definiciones básicas

Considere un universo de p activos, y denotemos como $s_{i,t}$ el precio del activo i al tiempo t . Así, el retorno o rendimiento del activo i sobre una unidad de tiempo (días) se define como

$$r_{i,t} = \frac{s_{i,t} - s_{i,t-1}}{s_{i,t-1}} \quad (46)$$

donde $i = 1, \dots, p$.

y el vector de pesos del portafolio como

$$w = \{w_1, \dots, w_p\} \quad (47)$$

donde w_i es la cantidad de dinero (dolares) invertida en el activo i , el cual puede ser positivo o negativo.

Definiciones básicas

La cantidad total de capital invertido lo expresaremos como \mathcal{C} , donde usualmente $\mathcal{C} = \sum_i w_i$.

De manera más general existen los siguiente escenarios:

- $\sum_i w_i > \mathcal{C}$: Se pide prestado y se paga una una tasa libre de riesgo (interés) r_0 sobre el crédito
- $\sum_i w_i < \mathcal{C}$: Se reserva una cantidad de capital para invertir en la tasa libre de riesgo (cetes) r_0
- En general las tasas son distintas en cada caso, pero las asumiremos iguales por simplicidad

Nociones básicas

Así, el retorno del portafolio al tiempo t está dado

$$R_t = \sum_{i=1}^P w_i r_{i,t} + (\mathcal{C} - \sum_i w_i) r_0, \quad (48)$$

y definimos como *exceso de rendimiento* (sobre la tasa libre de riesgo) a

$$R_t - \mathcal{C} r_0 = \sum_i w_i (r_{i,t} - r_0) = w' e, \quad (49)$$

donde e es el vector de exceso de rendimiento (omitimos el tiempo de aquí en adelante por simplicidad).

La ganancia esperada (exceso de rendimiento) del portafolio se define como

$$\mathcal{G} = \mathbb{E}(w' e) = w' \mathbb{E}(e) = w' g, \quad (50)$$

donde g es el vector de ganancias (exceso de rendimiento) esperadas

Nociones básicas

El riesgo del portafolio se define tradicionalmente como la varianza de los retornos. Aquí se definirá en relación al exceso de retorno

$$\begin{aligned}
 \mathcal{R}^2 &= \text{Var}[R] = \text{Var}[w'e] = \mathbb{E}[(w'e - w'g)(w'e - w'g)'] \\
 &= \mathbb{E}[w'(e - g)(e - g)'w] \\
 &= w' \mathbb{E}[(e - g)(e - g)'] w \\
 &= w' \Sigma w
 \end{aligned}
 \tag{51}$$

Portafolios óptimos

Bajo el enfoque de Markowitz, la idea es resolver el siguiente problema de optimización cuadrática

$$\left\{ \begin{array}{l} \min_{w \in \mathbb{R}^p} \frac{1}{2} w' \Sigma w \\ \text{s.t. } w' g \geq \mathcal{G} \end{array} \right. \quad (52)$$

Este problema matemático se puede resolver fácilmente al introducir un multiplicador de lagrange γ para reescribir el sistema como un problema de optimización sin restricciones

$$\min_{w \in \mathbb{R}^p} \frac{1}{2} w' \Sigma w - \gamma w' g \quad (53)$$

Solución

Escribiendo lo anterior con una función de Lagrange

$$\mathcal{L}(w) = \frac{1}{2}w'\Sigma w - \gamma w'g \quad (54)$$

Por lo que se debe cumplir

$$\frac{d\mathcal{L}}{dw} = \Sigma w - \gamma g = 0 \quad (55)$$

Suponiendo que la inversa de Σ existe, se encuentra que el portafolio óptimo está dado por

$$\hat{w} = \gamma \Sigma^{-1}g \quad (56)$$

Por otra lado al sustituir (56) en (50) se encuentra la expresión para γ

$$\gamma = \frac{\mathcal{G}}{g' \Sigma^{-1} g} \quad (57)$$

De esta manera, la solución óptima está dada por

$$\hat{w} = \mathcal{G} \frac{\Sigma^{-1} g}{g' \Sigma^{-1} g} \quad (58)$$

Esto requiere del conocimiento de Σ y g , lo cual no se conoce *a priori*.

Dejaremos el trabajo de crear expectativas acerca del comportamiento futuro de los retornos al inversionista o analista financiero.

Aquí asumiremos que g está dado, y estimaremos Σ en el régimen de alta dimensionalidad, i.e., cuando $p, n \rightarrow \infty$ siendo $q = p/n$ constante.

Lo cual es natural plantearse en el contexto de “*Big Data*”.

Riesgo verdadero ⁶

¿Cuál es el riesgo mínimo asociando a esta asignación de activos cuando se mide el riesgo como la varianza de los retornos del portafolio?

Para responder esto sustituycamos (58) en (51). Así

$$\mathcal{R}_{true}^2 = \frac{\mathcal{G}^2}{\mathbf{g}'\Sigma^{-1}\mathbf{g}} \quad (59)$$

Sin embargo, la estrategia óptima no es alcanzable en la práctica ya que en principio no se conoce Σ .

¿Qué podemos hacer entonces? ¿Que tan distinto es el *riesgo realizado* o fuera de muestra (out-sample) del *riesgo verdadero* (true) del modelo óptimo?

⁶Cleaning large correlation matrices: Tools from Random Matrix Theory, 2016.

Riesgo predicho y riesgo realizado.

Una aproximación es aplicar la matriz de covarianza empírica E en lugar de Σ .

Sin embargo, podemos anticipar que esta estrategia sufrirá de un fuerte sesgo siempre que n no sea demasiado grande comparado con p .

A pesar de lo anterior, es fácil darse cuenta que los pesos óptimos de inversión están dados de manera análoga en este caso por

$$w_E = \mathcal{G} \frac{E^{-1}g}{g'E^{-1}g}, \quad (60)$$

Por lo que el mínimo riesgo asociado a este portafolio está dado por

$$\mathcal{R}_{in}^2 = \frac{\mathcal{G}^2}{g'E^{-1}g}, \quad (61)$$

al que se le conoce como riesgo *riesgo predicho* o dentro de muestra (in-sample), debido a que se utiliza la información disponible

Riesgo predicho y riesgo realizado.

Teorema

Si X es una variable aleatoria y ϕ es una función convexa entonces $\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$, donde $\mathbb{E}[X]$ representa el valor esperado de X

Dado el resultado anterior (desigualdad de Jensen) se tiene que

$$\mathbb{E}[g'E^{-1}g] \geq g'\mathbb{E}[E^{-1}]g = g'\Sigma^{-1}g \quad (62)$$

- Dado que E es positiva semidefinida en general su inversa también lo es e implica convexidad.
- Así, la primera desigualdad se cumple para vectores fijos g fijos.
- La igualdad se cumple dado que E es un estimador no sesgado de Σ .

De esta manera $\mathcal{R}_{in}^2 \geq \mathcal{R}_{true}^2$, por lo que el portafolio óptimo w_E subestima el riesgo verdadero de la inversión \mathcal{R}_{true}^2 .

Riesgo predicho y riesgo realizado.

Al riesgo realizado se le conoce como fuera de muestra (out-sample) ya que utiliza datos nuevos una vez optimizados los pesos.

Denotemos por \tilde{E} a la matriz muestral de covarianza fuera de muestra. En este caso, el riesgo asociado está dado por la expresión

$$\mathcal{R}_{out}^2 = w_E' \tilde{E} w_E \quad (63)$$

Se puede asumir que el ruido dentro de muestra (contenido en w_E) es independiente al proveniente fuera de muestra, entonces w_E y \tilde{E} no estarán correlacionados ⁷, obteniendo para $p \rightarrow \infty$

$$\mathcal{R}_{out}^2 = w_E' \tilde{E} w_E \approx w_E' \Sigma w_E = \frac{g^2 g' E^{-1} \Sigma E^{-1} g}{(g' E^{-1} g)^2} \quad (64)$$

⁷S. Pafka, I. Kondor, Noisy covariance matrices and portfolio optimization II, Physica A 319 (2003) 487–494.

Riesgo predicho y riesgo realizado.

Además, dada la optimalidad de $\tilde{w} = w_{\Sigma}$ sabemos que

$$w'_{\Sigma} \Sigma w_{\Sigma} \leq w'_E \Sigma w_E \quad (65)$$

Por lo que se llega a la siguiente desigualdad general

$$\mathcal{R}_{in}^2 \leq \mathcal{R}_{true}^2 \leq \mathcal{R}_{out}^2 \quad (66)$$

En suma, w_E es sobre optimista y puede llevar a resultados desastrosos en la práctica.

Esta conclusión es válida incluso para otras medidas de riesgo como el *expected shortfall* mencionada al inicio.

Desigualdad frontera eficiente [▶ Notebook](#)

La siguiente figura muestra la frontera óptima para el proceso verdadero, dentro de muestra, y fuera de muestra en el caso en caso de una matriz de Wishart W_q .

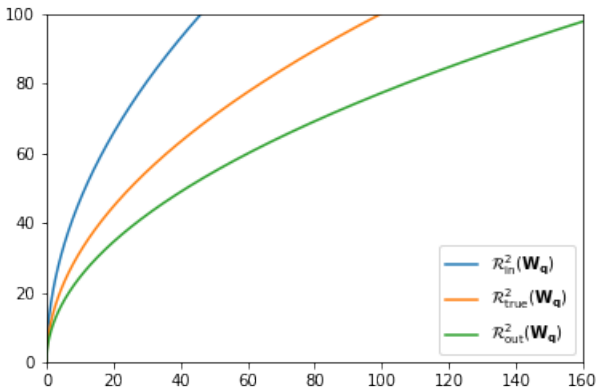


Figure (3) $X \sim (0, \sigma^2 I)$, $q = 1/2$, $p = 100$, $\sigma^2 = 0.2$ y $g = 1$

Técnica RMT de Bouchaud et. al.

Procedimiento para filtrar matriz de covarianza propuesta por Bouchaud et. al. (2002):

- $\hat{\lambda}_k^B = \begin{cases} \bar{\lambda} & \text{if } \lambda_k < (1 + \sqrt{q})^2 \\ \lambda_k & \text{otro caso} \end{cases}$
- $Q^B = V \hat{\Lambda}^B V'$
- $\Xi_{ij}^B = \frac{q_{ij}^B}{\sqrt{q_{ii}^B q_{jj}^B}}$

Desigualdad frontera eficiente aplicando filtro RMT

[▶ Notebook](#)

github.com/agarciam/curso-de-matrices-aleatorias-y-portafolios

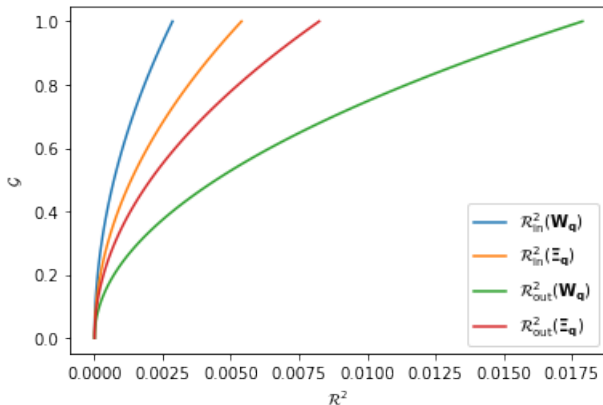


Figure (4) Sistema con 10 señales artificiales, $q = 1/2$, $p = 100$

Referencias

- ① Potters, M., & Bouchaud, J. P. (2020). A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists. Cambridge University Press.
- ② Livan, G., Novaes, M., & Vivo, P. Introduction to random matrices theory and practice. Springer (2018).
- ③ Mehta, M. L. (2004). Random matrices. Elsevier.
- ④ Mingo, J. A., & Speicher, R. (2017). Free probability and random matrices (Vol. 35). New York: Springer.