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]

Probabilidad clásica

Definición

Se dice que dos variables aleatorias x_1, x_2 con pdf ρ_1, ρ_2 son estadísticamente independiente si

$$\rho_{1,2}(x_1, x_2) = \rho_1(x_1)\rho_2(x_2). \tag{1}$$

Definición

La función característica de $\rho(x)$ se define a través de la transformada de Fourier

$$\phi(t) = \left\langle e^{itx} \right\rangle = \int dx \rho(x) e^{\mathrm{i}tx} \tag{2}$$

De esta manera la independencia entre las variables aleatorias implica que $\phi_{1,2}(t_1, t_2) = \phi_1(t_1)\phi_2(t_2)$.

Probabilidad clásica

Asimismo, si consideramos la función generadora de cumulantes (varianza, *skewness*, curtosis), es decir, el logaritmo de la función característica $h(t) = log\phi(t)$, observamos que es aditiva

$$h_{1,2}(t_1, t_2) = h_1(t_1) + h_2(t_2) \tag{3}$$

De esta manera, el problema de encontrar $p(x_1 + x_2)$ se reduce a un *algoritmo* (siempre y cuando x_1, x_2 sean independientes, caso contrario surge el concepto de cópula):

1. Calcular la función caracteristica de x_1 y x_2 a través de sus funciones de probabilidad
2. Calcular los cumulates y sumar
3. Obtener la función caracteristica de la suma via exponenciación o de manera equivalente por el teorema de convolución.
4. Aplicar la transformada inversa de fourier de la suma.

Probabilidad clásica

Ejemplo: Gaussiana

Para variables aleatorias gaussianas $\{x_i\}$ tenemos que

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}, \tag{4}$$

En este caso la función caracteristica es:

$$\phi(q) = \int_{-\infty}^{\infty} p(x) e^{iqx} dx = \cdots = e^{-(\sigma^2/2)q^2} = e^{-\gamma q^2} \quad \gamma \equiv \sigma^2/2 \tag{5}$$

Para el caso de la suma $s_2 = x_1$ y x_2 se obtiene

$$p(s_2) = p(x_1) * p(x_2) \Rightarrow \phi_2(q) = [\phi(q)]^2 = e^{-2\gamma q^2} \tag{6}$$

Aplicando la transformada inversa se encuentra que $(\sigma_2 = \sqrt{2}\sigma)$

$$p(s_2) = \frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)} e^{-x^2/2(\sqrt{2}\sigma)^2} = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-x^2/2\sigma_2^2} \tag{7}$$

Probabilidad clásica

Ejemplo: Cauchy

Para variables aleatorias x_i con distribución de Cauchy (Lorentz), la pdf está dada por

$$p(x) = \frac{\gamma}{\pi} \frac{1}{\gamma^2 + x^2} \tag{8}$$

y la función característica esta dada por

$$\phi(q) = \int_{-\infty}^{\infty} p(x) e^{iqx} dx = \cdots = e^{-\gamma |q|} \tag{9}$$

En este caso

$$\phi_2(q) = [e^{-\gamma |q|}]^2 = e^{-2\gamma |q|} \tag{10}$$

Aplicando la transformada inversa

$$p(s_2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_2(q) e^{-iqx} dq = \frac{2\gamma}{\pi} \frac{1}{4\gamma^2 + x^2} \tag{11}$$

Atractores estables de $p(s_n)$ Notebook

En general,

$$p(s_n) = p(x_1) * p(x_2) \dots p(x_n) \quad \Rightarrow \quad \phi_n(q) = [\phi(q)]^n \tag{12}$$

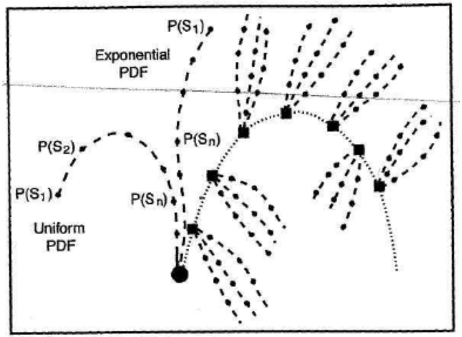


Figure (1) Mantegna, R. N., & Stanley, H. E. (1999). Introduction to econophysics: correlations and complexity in finance. Cambridge university press.

Probabilidad libre

La teoría de la probabilidad libre es un método alternativo para estudiar el comportamiento asintótico de algunas matrices aleatorias *grandes*.

Esta teoría nos proporciona una forma robusta para investigar la densidad espectral límite (LSD) de sumas o productos de matrices aleatorias con propiedades de simetría específicas.

Este es un problema importante en la práctica dado que los modelos estándar en estadística tratan con un ruido aditivo o multiplicativo.

En particular, estamos interesados en extraer la señal *verdadera* de observaciones ruidosas de la matriz de covarianza, la cual es un elemento esencial en la teoría de portafolios de Markowitz.

Antecedentes

La teoría de la probabilidad libre fue iniciada en 1985 por Dan Voiculescu para comprender clases especiales de álgebras de von Neumann, estableciendo reglas de cálculo para operadores no conmutativos.

Voiculescu y Speicher encontraron que las matrices aleatorias invariantes rotacionalmente satisfacen asintóticamente los criterios de *libertad*.

En términos generales, dos matrices grandes A y B son mutuamente libres si sus bases propias están relacionadas entre sí mediante una rotación aleatoria, es decir, cuando sus vectores propios son ortogonales.

Preliminares

Formalmente, los ingredientes que necesitaremos para definir el concepto de libertad son los siguientes:

- Una anillo $\mathcal{R}$ [1] de variables aleatorias, las cuales pueden ser no-conmutativas respecto a la multiplicación.
- Un campo de escalares que usualmente es $\mathbb{C}$: Los escalares conmutan con todo.
- Una operación *, llamada involución: el complejo conjugado, las transpuesta en matrices reales, la conjugada transpuesta en las matrices complejas.
- Una función lineal positiva $\phi(\cdot) : \mathcal{R} \to \mathbb{C}$ que satisface $\phi(AB) = \phi(BA)$ para $A, B \in \mathcal{R}$. Por positivo nos referimos a que $\phi(AA^*)$ sea real y no-negativo. Además, requerimos que $\phi(AA^*) = 0 \Rightarrow A = 0$ (*faithful*).

Por ejemplo, ϕ puede ser el operador $\mathbb{E}[\cdot]$ en probabilidad clásica, o el operador de traza normalizada $1/NTr(\cdot)$ para un anillo de matrices, o incluso la combinación: $\frac{1}{N}\mathbb{E}[Tr(\cdot)]$.

[1] conjunto equipado con dos operaciones binarias que generaliza las operaciones aritmeticas de adición y multiplicación

Preliminares

- Los elementos de $\mathcal{R}$ los llamaremos variables aleatorias y las denotaremos por letras mayúsculas.
- Para cualquier $A \in \mathcal{R}$ y $k \in \mathbb{N}$ decimos que $\phi(A^k)$ es el k-ésimo momento de A y asumimos que $\phi(A^k)$ es finito para todo k. En particular, decimos que $\phi(A)$ es la media de A y $\phi(A^2) - \phi(A)^2$ es su varianza.
- Decimos que dos elementos A y B tienen la misma distribución si sus momentos son iguales para todos los ordenes [2]
- El anillo de variables debe contener al elemento llamado $\mathbb{1}$ tal que $A\mathbb{1} = \mathbb{1}A = A$ para cada A. Este satisface $\phi(\mathbb{1}) = 1$. Así, añadir una constante α simplemente desplaza a la media:$\phi(A + \alpha\mathbb{1}) = \phi(A) + \alpha$.

[2] Esto no siempre se cumple: algunas distribuciones no son determinadas de manera única por sus momentos.

Momentos de suma de v.a. que conmutan

Es importante recordar que dos v.a. conmutan si $AB = BA \quad \forall A, B \in \mathcal{R}$.

Note que A, B no son necesariamente números reales (complejos), sino que pueden ser elementos de un anillo más abstracto.

Decimos que A y B son independientes si $\phi(p(A)q(B)) = \phi(p(A))\phi(q(B))$ para cualquier polinomio p, q: Esta condición es equivalente a la factorización de momentos.

Por la propiedad de linealidad del operador ϕ tenemos que $\phi(A + B) = \phi(A) + \phi(B)$

Asumamos de aquí en adelante que $\phi(A) = \phi(B) = 0$, i.e., A, B tienen media cero. Para una variable $\tilde{A}$ con media diferente de cero, podemos escribir $A = \tilde{A} - \phi(\tilde{A})$, tal que $\phi(A) = 0$.

Para el segundo momento de la suma tenemos

$$\phi((A + B)^2) = \phi(A^2) + \phi(B^2) + 2\phi(AB) = \tag{13}$$

$$\phi(A^2) + \phi(B^2) + 2\phi(A)\phi(B) = \phi(A^2) + \phi(B^2), \tag{14}$$

es decir, la varianza es aditiva. Lo mismo sucede para el tercer momento.

Momentos de suma de v.a. que no conmutan

Una de las metas es generalizar la ley de adición de variables independientes.

Consideremos ahora la variable $A+B$, donde A y B son objetos que no conmutan como las matrices aleatorias. Si calculamos los tres primeros momentos de $A+B$ no surgen problemas realmente debido a la definición del operador ϕ a través de la traza.

Los problemas se vuelven interesantes cuando intentamos calcular el cuarto momento de la suma

$$\phi((A+B)^4) = \phi(A^4)+4\phi(A^3B)+4\phi(A^2B^2)+2\phi(ABAB)+4\phi(AB^3)+\phi(B^4), \quad (15)$$

donde hemos usado la propiedad cíclica de la traza.

En el caso conmutativo la independencia de A, B implica $\phi(A^2B^2) = \phi(A^2)\phi(B^2)$ y sería suficiente para lidiar con los momentos mixtos.

En el caso no conmutativo necesitamos calcular el termino $\phi(ABAB)$, dado que en general $ABAB \neq A^2B^2$. Por lo que necesitamos una nueva definición de independencia para tratar este término. Una solución radical será postular que $\phi(ABAB) = 0$ siempre que $\phi(A) = \phi(B) = 0$.

Libertad

Dada dos variables aleatorias A, B decimos que son *libres* si para cualquier polinomio $p_1, \ldots, p_n$ y $q_1, \ldots, q_n$ tal que

$$\phi(p_k(A)) = 0, \quad \phi(q_k(B)) = 0, \quad \forall k, \tag{16}$$

tenemos

$$\phi(p_1(A)q_1(B)p_2(A)q_2(B)\ldots p_n(A)q_n(B)) = 0 \tag{17}$$

Se dirá que un polinomio (o variable) es de traza cero si $\phi(p(A)) = 0$. Noté que $\alpha\mathbb{1}$ es libre respecto a $A \in \mathcal{R}$ debido $\phi(p(\alpha\mathbb{1})) = p(\alpha\mathbb{1})$, dada la definición de $\mathbb{1}$. Entonces

$$\phi(p(\alpha\mathbb{1})) = 0 \iff p(\alpha\mathbb{1}) = 0 \tag{18}$$

Además, es fácil ver que si A, B son libres, entonces $p(A), q(B)$ son libres para cualquier polinomio p, q. Por extensión, $F(A)$ y $G(B)$ son también libres para cualquier función F y G definida por sus series de potencia.

Momentos de suma de v.a. libres

Asumiendo que A, B son libres con $\phi(A) = \phi(B) = 0$ es posible calcular la suma libre $A + B$. El segundo momento está dado

$$\phi((A+B)^2) = \phi(A^2) + \phi(B^2) + 2\phi(AB) = \phi(A^2) + \phi(B^2) \tag{19}$$

Para momentos de tercer orden y superiores el truco consiste en sumar y restar cantidades en cada término

$$\phi((A+B)^3) = \phi(A^3) + \phi(B^3) + 3\phi(A^2B) + 3\phi(AB^2) = \tag{20}$$
$$\phi(A^3) + \phi(B^3) + 3\phi((A^2 - \phi(A^2))B) + 3\phi(A^2)\phi(B) + \tag{21}$$
$$3\phi(A(B^2 - \phi(B^2))) + 3\phi(A)\phi(B^2) = \phi(A^3) + \phi(B^3) \tag{22}$$

Por último, para el cuarto momento tenemos [3]:

$$\phi((A+B)^4) = \phi(A^4) + \phi(B^4) + 4\phi(A^2)\phi(B^2) \tag{23}$$

[3] para variables conmutativas: $\phi((A+B)^4) = \phi(A^4) + \phi(B^4) + 6\phi(A^2)\phi(B^2)$

Matrices aleatorias y libertad

Se dijo que dos variables A y B son libres si para cualquier conjunto de polinomios $p_1, \ldots, p_n$ y $q_1, \ldots, q_n$ la siguiente igualdad se cumple

$$\phi(p_1(A)q_1(b)p_2(A)q_2(B)\ldots p_n(A)q_n(B)) = 0 \tag{24}$$

- La conexión con matrices consiste en considerar A y B como matrices simétricas grandes y $\phi(M) := \frac{1}{N}Tr(M)$.
- Las matrices A, B pueden ser diagonalizadas por $U\Delta U', V\Delta V'$, respectivamente.
- Un polinomio sin traza $p_i(A)$ puede ser diagonalizado como $U\Delta_i U'$, donde U es la misma matriz ortogonal para A y $\Delta_i = p_i(\Delta)$ es una matriz diagonal sin traza, y lo mismo para $q_i(B)$.

Entonces, la condición de libertad se convierte en

$$\phi(\Delta_1 O \Delta_1' O' \Delta_2 O \Delta_2' O' \ldots \Delta_n O \Delta_n' O') = 0, \tag{25}$$

donde se ha introducido $O = U'V$ como la matriz ortogonal de cambio de base que rota a los vectores propios A hacia B.

Matrices aleatorias y libertad

En el límite $N \to \infty$ la expresión (25) se cumple al promediar sobre el grupo ortogonal O y siempre que Δ_i y Δ'_i sean de traza cero.

Se espera también que en el límite $N \to \infty$ la ecuación (25) se autopromedie de tal manera que una sola matriz O se comporte como el promedio sobre todas las matrices.

De esta manera, dos matrices simétricas grandes cuyas bases se rotan aleatoriamente una respecto a la otra son esencialmente libres.

Por ejemplo, las matrices de Wigner X y las matrices de Wishart W son rotacionalmente invariantes, lo que implica que sus eigenvectores son matrices ortogonales aleatorias.

Se concluye que para N grande, ambas X y W son libres respecto a cualquier matriz independiente de ellas, en particular son libres respecto a matrices deterministas.

Nota: Para demostrar la expresión (25) se requiere integrar sobre el grupo ortogonal, para lo cual se puede usar la función de Weingarten y las reglas de contracción de Wick (transformalo a un problema combinatorio).

Matrices aleatorias y libertad

De manera más general, la libertad permite el cálculo de momentos mixtos de suma y productos de matrices a partir del conocimiento de los momentos de A y B, similar a la independencia clásica en la teoría de probabilidades.

Un ejemplo típico de pares de matrices libres es cuando A es una matriz fija y B una matriz aleatoria que pertenece a un conjunto rotacionalmente invariante, es decir, $B = O\Lambda O'$, donde Λ es diagonal y O se distribuye de acuerdo a la medida de Haar sobre el grupo ortogonal, en el límite donde N es infinitamente grande.

Una propiedad importante de las matrices aleatorias es que sus vectores propios se distribuyen uniformemente en el grupo $O(N)$, por lo que cada vector está distribuido en la esfera unitaria S^{N-1}. Por lo que cuando N es grande es poco probable que algún vector propio de A se traslape con un vector propio de B.

El cálculo de momentos mixtos es crucial para derivar relaciones útiles en problemas de estimación en estadística multivariada de alta dimensionalidad.

Matrices aleatorias y libertad

Hemos introducido el concepto de *libertad* el cual puede ser resumido en la siguiente definición intuitiva

Definición

Dos matrices grandes son libres si sus bases propias están relacionadas por una rotación aleatoria. Consideramos una matriz grande si sus momentos calculados bajo probabilidad libre son válidos hasta una corrección de orden $\mathcal{O}(1/N)$.

Transformada $\mathcal{R}$

La transformada $\mathcal{R}$ se define como

$$\mathcal{R}_A(g) := \mathcal{B}_A(g) - \frac{1}{g} \tag{26}$$

donde $\mathcal{B}_A(g)$ es la transformada inversa de Stieltjes definida a través de su serie de potencia (función generadora de momentos) que satisface $g_A(\mathcal{B}_A(g)) = g$, a todos los ordenes.

En general, la transformada $\mathcal{R}$ es aditiva para variables aleatorias libres

$$\mathcal{R}_{A+B}(g) = \mathcal{R}_A(g) + \mathcal{R}_B(g) \tag{27}$$

Para ver esto recordemos que la transformada de Stieltjes se puede escribir como

$$g_A(z) = \phi(\left[(z - A)^{-1}\right]) = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \phi(A^k), \tag{28}$$

Considere un escalar fijo g, por construcción

$$\phi(g\mathbb{1}) = g = g_A(\mathcal{B}_A(g)) = \phi\left[(\mathcal{B}_A(g) - A)^{-1}\right]. \tag{29}$$

Transformada $\mathcal{R}$

El argumento de $\phi(\cdot)$ en la izquierda y derecha tienen el mismo significado, pero son diferentes en general, denotemos su diferencia como gX_A

$$gX_A := (z_A - A)^{-1} - g\mathbb{1}, \tag{30}$$

donde $z_A = \mathcal{B}_A(g)$, y por definición $\phi(gX_A) = 0$.

Al invertir la relación anterior

$$A - z_A = -\frac{1}{g}(1 + X_A)^{-1}. \tag{31}$$

Ahora, considere otra variables B libre en relación a A, para la cual se encuentra una relación similar

$$B - z_B = -\frac{1}{g}(1 + X_B)^{-1}. \tag{32}$$

Siendo X_A y X_B libres puesto que son funciones de A y B.

Sumando las expresiones (31) y (32) se llega al siguiente resultado (▸ ver derivación)

$$\mathcal{B}_{A+B} = z_A + z_B - g^{-1} \Rightarrow \mathcal{R}_{A+B} = \mathcal{R}_A + \mathcal{R}_B \tag{33}$$

La transformada $\mathcal{S}$

De manera análoga se puede definir la transformada $\mathcal{S}$ para el producto de variables libres

$$S_{AB}(t) = S_A(t)S_B(t), \tag{34}$$

para A y B libres.

Para definir la transformada $\mathcal{S}$ se necesita introducir la transformada $\mathcal{T}$

$$\mathcal{T}_A(z) = \phi\left[(1 - z^{-1}A)^{-1}\right] - 1 = zg_A(z) - 1 = \sum_{k=1}^{\infty} \frac{\phi(A^k)}{z^k} \tag{35}$$

Otra forma de definir la transformada $\mathcal{S}$ es

$$S_A(t) := \frac{t+1}{t\mathcal{Z}_A(t)}, \tag{36}$$

donde $\mathcal{Z}_A(t)$ es la inversa de la $\mathcal{T}$.

Aplicando a matrices aleatorias es mejor trabajar con el producto de la forma $A^{1/2}BA^{1/2}$ puesto que AB no es necesariamente simétrico.

La transformada R y S

En suma, A y OBO' son libres cuando O es una matriz de rotación aleatoria (en límite de dimensión grande).

Cuando A y B son libres su transformada $\mathcal{R}$ y $\mathcal{S}$ nos permiten su adición y multiplicación:

$$\mathcal{R}_{A+B}(g) = \mathcal{R}_A(g) + \mathcal{R}_B(g) \tag{37}$$

$$\mathcal{S}_{AB}(t) = \mathcal{S}_A(t)\mathcal{S}_B(t) \tag{38}$$

En general AB no es una matriz simétrica, de hecho la transformada $S_{AB}(t)$ se relaciona con los valores propios de la matrix $\sqrt{A}B\sqrt{A}$, que son los mismos de $\sqrt{B}A\sqrt{B}$ cuando A y B son positivas semidefinidas (valores propios $\lambda \geq 0$).

La transformada R y S

La transformada $\mathcal{R}$ y $\mathcal{S}$ se pueden definir a través de las siguientes relaciones:

$$g_A(z) = \phi\left[(z-A)^{-1}\right] \quad \text{(Transformada Stieltjes de A)} \tag{39}$$

$$\mathcal{R}_A(g) = \mathcal{B}_A(g) - \frac{1}{g} \quad (\mathcal{B} \text{ Inversa de } g_A) \tag{40}$$

$$\mathcal{T}_A(z) = zg_A(z) - 1 \quad (\text{Transformada } \mathcal{T}) \tag{41}$$

$$\mathcal{S}_A(t) = \frac{t+1}{t\mathcal{Z}_A(t)} \quad , \text{Si} \quad \phi(A) \neq 0 \quad (\mathcal{Z} \text{ inversa de } \mathcal{T}) \tag{42}$$

La transformada R y S: Propiedades

- Multiplicación por un escalar:

$$\mathcal{R}_{\alpha A}(g) = \alpha \mathcal{R}_A(\alpha g), \quad \mathcal{S}_{\alpha A}(t) = \alpha^{-1}\mathcal{S}_A(t) \tag{43}$$

- Desplazamiento [4]

$$\mathcal{R}_{A+\alpha I}(g) = \alpha + \mathcal{R}_A(g) \tag{44}$$

- Inversión matricial

$$\mathcal{S}_{A^{-1}}(t) = \frac{1}{\mathcal{S}_A(-t-1)} \tag{45}$$

  Sin embargo, no existe una formula simple equivalente para $\mathcal{S}$
- Relación entre $\mathcal{R}$ y $\mathcal{S}$

$$\mathcal{S}_A(t) = \frac{1}{\mathcal{R}_A(t\mathcal{S}_A(t))}, \quad \mathcal{R}_A(g) = \frac{1}{\mathcal{S}_A(g\mathcal{R}_A(g))} \tag{46}$$

- Cuando $A = I$

$$g_I(z) = \frac{1}{z-1} \quad \mathcal{T}_I = \frac{1}{z-1} \tag{47}$$

$$\mathcal{R}_I(g) = 1 \quad \mathcal{S}_I(t) = 1 \tag{48}$$

[4] no existe una formula simple equivalente para $\mathcal{S}$

La transformada $\mathcal{R}$ y $\mathcal{S}$: Serie de Taylor

Las transformadas $\mathcal{R}$ y $\mathcal{T}$ tienen la siguiente expansión de Taylor para argumentos pequeños:

$$\mathcal{R}_A(g) = \kappa_1 + \kappa_2 g + \kappa_3 g^2 + \ldots, \tag{49}$$

$$\mathcal{S}_A(t) = \frac{1}{\kappa_1} - \frac{\kappa_2}{\kappa_1^3} t + \frac{2\kappa_2^2 - \kappa_1\kappa_3}{\kappa_1^5} t^2 + \ldots, \tag{50}$$

donde κ_n son los cumulantes libres de A:

$$\kappa_1 = \phi(A) \tag{51}$$

$$\kappa_2 = \phi(A^2) \tag{52}$$

$$\kappa_3 = \phi(A^3) - 3\phi(A)\phi(A^2) + 2\phi^3(A) \tag{53}$$

$$\vdots \tag{54}$$

Procedimiento para encontrar la densidad de valores propios: Suma

La transformada $\mathcal{R}$ nos brinda un método sistemático para obtener el espectro de la suma C de dos matrix independientes A y B, donde al menos una de ellas es rotacionalmente invariante:

1. Encuentre $g_A(z)$ y $g_B(z)$
2. Invierta $g_A(z)$ y $g_B(z)$ para obtener la transformada *Blue* $\mathcal{B}_A(g)$ y $\mathcal{B}_B(g)$ y por ende $\mathcal{R}_A(g)$ y $\mathcal{R}_B(g)$.
3. Así, $\mathcal{R}_C(g) = \mathcal{R}_A(g) + \mathcal{R}_B(g)$, donde $\mathcal{B}_C(g) = \mathcal{R}_C(g) + \frac{1}{g}$
4. Calcule la inversa para obtener $g_C(z)$
5. Utilice la relación para la densidad

$$\rho_C(\lambda) = \frac{1}{\pi} \lim_{\epsilon \to 0^+} Im\{g_C(\lambda - i\epsilon)\} \tag{55}$$

Procedimiento para encontrar la densidad de valores propios: Producto

En el caso multiplicativo $C = A^{1/2} B A^{1/2}$ la receta es similar:

1. Encuentre $\mathcal{T}_A(z)$ y $\mathcal{T}_B(z)$
2. Invierta $\mathcal{T}_A(z)$ y $\mathcal{T}_B(z)$ para obtener $\mathcal{Z}_A(t)$ y $\mathcal{Z}_B(t)$, y consecuentemente $\mathcal{S}_A(t)$ y $\mathcal{S}_B(t)$
3. Así, $\mathcal{S}_C(t) = \mathcal{S}_B(t)\mathcal{S}_A(t)$, de donde $\mathcal{Z}_C(t) = \frac{t+1}{\mathcal{S}_C(t)t}$
4. Calcule la inversa para obtener $\mathcal{T}_C(z)$
5. Utilice la relación para la densidad en $g_C(z) = (\mathcal{T}_C(z) + 1)/z$, o de manera equivalente:

$$\rho_C(\lambda) = \frac{1}{\pi\lambda} \lim_{\epsilon \to 0^+} Im\{\mathcal{T}_C(\lambda - i\epsilon)\} \tag{56}$$

Analogía probabilidad clásica y libre

Probabilidad Clásica	Probabilidad libre
Función característica $\phi(t) \equiv \mathbb{E}[e^{itx}]$	Función de Stieltjes: $g_M(z) = \phi[(z\mathbb{1} - M)^{-1}]$ $= \frac{1}{N}\mathbb{E}\left(Tr\left[(z\mathbb{1} - M)^{-1}\right]\right)$
Independencia	Libertad (freeness)
Suma de r.v. El logaritmo es aditivo $h(t) = log\phi(t)$ $h_{1,2}(t_1, t_2) = h_1(t_1) + h_2(t_2)$	Suma de f.r.v. La función R es aditiva $R_H(g) = B_H(g) - 1/g$ $\mathcal{R}_{A+B}(g) = \mathcal{R}_A(g) + \mathcal{R}_B(g)$
Multiplicación de r.v. Se reduce al problema aditivo Mapeo exponencial $e^{itx_1}e^{itx_2} = e^{it(x_1+x_2)}$	Multiplicación de f.r.v. La transformación S $\mathcal{S}_C(t) = \frac{t+1}{t\mathcal{Z}_C(t)}$ $\mathcal{S}_{AB}(t) = \mathcal{S}_A(t)\mathcal{S}_B(t)$

Ejemplo: GOE + Wishart

Suponga que desea calcular la densidad espectral promedio de la suma de matrices aleatorias grandes ($N \to \infty$) pertenecientes a dos ensembles diferentes.

- Consideremos las matrices aleatorias A y B pertenecientes a ensembles caracterizados por las transformadas de Stieltjes $g_A(z)$ y $g_B(z)$, respectivamente.
- La transformada $\mathcal{R}$ de $M = A + B$ esta dado por

$$\mathcal{R}_M(g) = \mathcal{R}_A(g) + \mathcal{R}_B(g) \tag{57}$$

- Podemos escribir un algoritmo para obtener la densidad de valores propios de la matriz M a través de la transformada $\mathcal{R}$.

Ejemplo: GOE + Wishart

En particular, consideremos la combinación de una matriz X del GOE y una matriz W del ensemble de Wishart:

$$M = rX + (1-r)W, \quad \text{donde} \quad r \in [0,1] \tag{58}$$

Se puede verificar que las transformadas $\mathcal{R}$ de los resolventes de cada ensemble son (pizarrón)

$$\begin{aligned} \mathcal{R}_H(g) &= \frac{g}{2} \\ \mathcal{R}_W(g) &= \frac{1}{1-qg} \end{aligned} \tag{59}$$

donde $q = p/n$. Ahora, utilizando la propiedad de escalamiento $\mathcal{R}_{\alpha A}(g) = \alpha \mathcal{R}_A(\alpha g)$, se obtiene

$$\begin{aligned} \mathcal{R}_M(g) &= r\mathcal{R}_X(rg) + (1-r)\mathcal{R}_W((1-r)g) \\ &= \frac{r^2}{2}g + \frac{(1-r)}{1-q(1-r)g}. \end{aligned} \tag{60}$$

Ejemplo: GOE + Wishart

Dado que $\mathcal{B}_M(g_M(z)) = z$, donde $\mathcal{B}_M(g) = \mathcal{R}_M(g) + 1/g$,

se llega a una ecuación para $g_M(z)$:

$$z = \frac{r^2}{2} g_M(z) + \frac{(1-r)}{1 - q(1-r)g_M(z)} + \frac{1}{g_M(z)}. \tag{61}$$

En general, esta ecuación de tercer grado tiene una solución real, y dos soluciones complejas conjugadas.

Finalmente, la densidad de valores propios de M se calcula con

$$\rho(\lambda) = \frac{1}{\pi} \lim_{\varepsilon \to 0^+} \operatorname{Im} g_M(\lambda - \mathrm{i}\varepsilon) \tag{62}$$

Ejemplo: GOE + Wishart

Simbólicamente

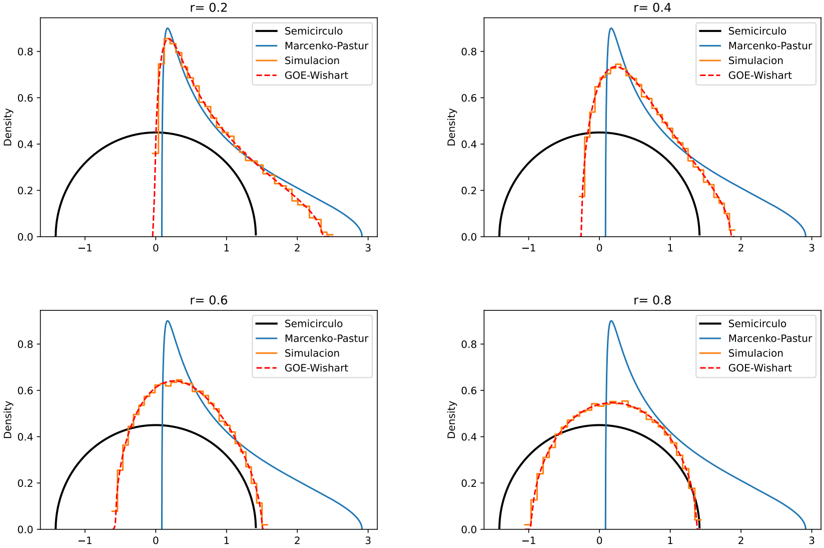
```
sol = sym.solveset((w**2)*(G/2) + (1-w)/(1-q*(1-w)*G) + 1/G - z, G)
```

Monte Carlo

```
for i in range(m):
    H = np.sqrt(1/p)*np.random.normal(loc =0, scale =1, size=(p, p))
    Hs = (H + np.transpose(H))/2
    X = np.random.normal(loc =0, scale =1, size=(p,n))
    W = np.dot(X,np.transpose(X))
    Hs_W = r*Hs + (1-r)*W/n
    w,v = LA.eigh(Hs_W)
```

Ejemplo: GOE + Wishart

Notebook



Ejemplo: Correlaciones espaciales

La matriz de covarianza muestral se define como

$$E = \frac{1}{n} HH', \tag{63}$$

donde H es una matriz de datos rectangular de dimensión $n \times p$. Sí las p series de tiempo son estacionarias se espera que para $n >> p$, E converja a la matriz de covarianza poblacional Σ.

Consideremos el caso donde H son observaciones gaussianas multivariadas provenientes de $\mathcal{N}(0, \Sigma)$. En este caso E esta dada por una matriz de Wishart general con matriz de covarianza Σ (▸ pizarron)

$$E = \Sigma^{1/2} W_q \Sigma \tag{64}$$

Ejemplo: Correlaciones espaciales

Usando probabilidad libre es posible inferir la densidad espectral poblacional (*verdadera*) $\rho_\Sigma(\lambda)$ a partir del espectro de la matriz muestral E

$$g_E(z) = \int \frac{\rho_\Sigma(\mu) d\mu}{z - \mu(1 - q + qzg_E(z))} \tag{65}$$

Esta expresión resulta muy útil en la teoría portafolios para modelar las correlaciones cruzadas en la matriz de covarianza.

El límite de alta dimensionalidad

En el límite de matrices grandes y con algunos supuestos sobre la estructura g, podemos precisar las desigualdades del riesgo dentro y fuera de muestra a través de probabilidad libre.

En particular, podemos vincular el riesgo verdadero y el riesgo realizado utilizando la ecuación de Marčenko-Pastur y la teoría de la probabilidad libre.

Supongamos por simplicidad que

$$g = N_p(0, I_p), \tag{66}$$

Aunque nuestros argumentos serán válidos para cualquier vector g cuya dirección sea independiente de Σ o E, siempre que $g'g = p$, es decir, cada componente de g es de orden unitario.

Enfatizamos que estos supuestos no son necesariamente realistas (los predictores pueden estar sesgados a lo largo de los componentes principales de Σ) pero nos permiten cuantificar con mayor precisión la relación entre el riesgo dentro/verdadero/fuera de la muestra.

El límite de alta dimensionalidad

Sea M una matriz definida positiva independiente del vector g. Entonces, tenemos en el límite grande de p,

$$\frac{g'Mg}{p} = \frac{1}{p}Tr[g'gM] \underset{freeness}{=} \frac{g'g}{p}\phi(M) \tag{67}$$

donde recordamos que ϕ es el operador normalizado de la traza

$$\phi(M) = \frac{1}{p}Tr(M) \tag{68}$$

y es el primer momento de $\rho(M)$.

Lo anterior implica que podemos usar el álgebra de momentos proveniente de probabilidad libre siempre que $p \to \infty$

De esta manera, dado que $g'g = p$ tenemos

$$\frac{g'Mg}{p} - \phi(M) \underset{p\to\infty}{\to} 0 \tag{69}$$

El límite de alta dimensionalidad

Fijemos ahora $M = \{E^{-1}, \Sigma^{-1}\}$ y usemos la relación anterior en las ecs. del riesgo dentro de muestra, verdadero, y fuera de muestra:

$$\mathcal{R}_{in}^2 \to \frac{\mathcal{G}^2}{p\phi(E^{-1})} \tag{70}$$

$$\mathcal{R}_{true}^2 \to \frac{\mathcal{G}^2}{p\phi(\Sigma^{-1})} \tag{71}$$

$$\mathcal{R}_{out}^2 \to \frac{\mathcal{G}^2\phi(E^{-1}\Sigma E^{-1})}{p\phi^2(E^{-1})} \tag{72}$$

$$\tag{73}$$

Para $q < 1$ ($q = p/n$) se ha demostrado en el régimen de alta dimensionalidad que (ver correlaciones espaciales)

$$\phi(\Sigma^{-1}) = (1-q)\phi(E^{-1}) \tag{74}$$

Como resultado cuando $p \to \infty$

$$\mathcal{R}_{in}^2 = (1-q)\mathcal{R}_{true}^2 \tag{75}$$

El límite de alta dimensionalidad

Así, para cualquier $q \in (0,1)$ vemos que el riesgo dentro de muestra asociado a los pesos w_E nos dan una sobre estimación (optimista). Afortunadamente, es posible cuantificar esta sobre estimación en el límite de alta dimensionalidad.

Buscamos ahora una relación similar para el riesgo fuera de muestra.

En general, la matriz de covarianza muestral E se puede escribir como $E = \Sigma^{1/2} W_q \Sigma^{1/2}$, donde W_q es una matriz con distribución $W_p(n, I_p)$ independiente de Σ. Así, cuando $p \to \infty$

$$\mathcal{R}^2_{out} \to \frac{\mathcal{G}^2 \phi(\Sigma^{-1} W_q^{-2})}{p\phi^2(E^{-1})} \tag{76}$$

Al considerar matrices grandes, W y Σ_q son asintoticamente libres, por lo que tenemos la relación

$$\phi(\Sigma^{-1} W^{-2}) = \phi(\Sigma^{-1})\phi(W^{-2}) \tag{77}$$

El límite de alta dimensionalidad

Usando la relación asintótica (74) se llega a

$$\mathcal{R}_{out}^2 = \mathcal{G}^2(1-q)^2 \frac{\phi(W_q^{-2})}{p\phi(\Sigma^{-1})} \tag{78}$$

Finalmente, se puede obtener $\phi(W^{-2}) = (1-q)^{-3}$ para $q < 1$, con lo que encontramos

$$\mathcal{R}_{out}^2 = (1-q)^{-1} \frac{\mathcal{G}^2}{p\phi(\Sigma^{-1})} = \frac{\mathcal{R}_{true}^2}{1-q} \tag{79}$$

En suma, tenemos la siguiente relación asintótica valida para una estructura general de Σ:

$$\frac{\mathcal{R}_{in}^2}{1-q} = \mathcal{R}_{true}^2 = (1-q)\mathcal{R}_{out}^2 \tag{80}$$

Simulación

Notebook

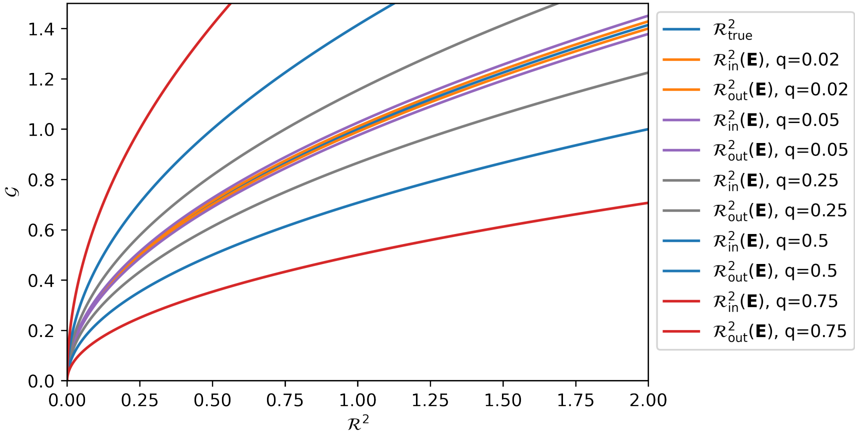


Figure (2) $q = \{1/50, 1/20, 1/4, 1/2, 3/4\}$, $r = 0.5$, $p = 100$. Se han usado unidades arbitrarias de tal manera que $\mathcal{R}^2_{true} = 1$ cuando $\mathcal{G} = 1$.

El límite de alta dimensionalidad

Observaciones:

- Si uno realiza una inversión con los pesos *ingenuos* w_E, resulta que el riesgo predicho (dentro de muestra) subestima el riesgo realizado por un factor $(1-q)^2$
- En el caso extremo $p = n$ $(q = 1)$, el riesgo dentro de muestra es igual a cero, mientras que el riesgo fuera de muestra diverge.
- En conclusión, el uso de la matriz de covarianza muestral E para el problema de optimización de Markowitz puede conducir a resultados desastrosos.
- Esto sugiere que deberíamos tener un estimador de Σ confiable para controlar el riesgo fuera de la muestra.

Estimador Rotacionalmente Invariante (RIE)

En general, dada la invariancia rotacional de Σ se puede obtener la invariancia de la probabilidad a priori (en el sentido bayesiano) $P_0(\Sigma)$

$$P_0(\Sigma) = P_0(O\Sigma O'), \tag{81}$$

donde O es una matriz de rotación arbitraria.

Asimismo, se puede obtener la probabilidad a posterior de Σ dada la matriz de covarianza muestral E

$$P(\Sigma|E) \propto (det\Sigma)^{-n/2} \exp\left[-\frac{n}{2} Tr\left(\Sigma^{-1}E\right)\right] P_0(\Sigma) \tag{82}$$

RIE

También, se puede verificar que el estimador MMSE (mínimo error cuadratico medio) de Σ se transforma en el mismo sentido que E bajo una rotación arbitraria

$$\begin{aligned}\mathbb{E}[\Sigma|OEO'] &= \int \Sigma P(\Sigma|OEO')P_0(\Sigma)d\Sigma \\ &= O\left[\int \tilde{\Sigma}P(\tilde{\Sigma}|E)P_0(\tilde{\Sigma})d\tilde{\Sigma}\right]O' \\ &= O\mathbb{E}(\Sigma|E)O',\end{aligned} \tag{83}$$

donde se uso el cambio de variable $\tilde{\Sigma} = O'\Sigma O$, y la forma explicita de $P(\Sigma|E)$

De manera general, si denotamos como $\Xi(E)$ como el estimador de Σ dado E, este estimador es rotacionalmente invariante si y sólo si

$$\Xi(OEO') = O\Xi(E)O' \tag{84}$$

Para cualquier matriz ortogonal O

RIE

- Si la matriz de covarianza muestral E se rota mediante alguna O, entonces la estimación de Σ también se debe rotar en la misma dirección.
- Esto es así por que no tenemos una suposición a priori acerca de los vectores propios de Σ.
- Los estimadores dados por la expresión (84) se les denomina *Estimadores Rotacionalmente Invariantes* (RIE)
- $\Xi(E)$ pueden ser diagonalizado en la misma base de E salvo una rotación fija Ω
- Sin embargo, no existe un candidato natural para Ω, excepto la matriz identidad I
- Podemos concluir que $\Xi(E)$ tiene los mismos vectores propios que E y escribir

$$\Xi(E) = \sum_{i=1}^{p} \xi_i v_i v_i', \tag{85}$$

donde v_i son los vectores propios de E, y ξ_i es función de los valores propios $[\lambda_j]_{j\in\{1,p\}}$ de E.

RIE óptimo

Veamos ahora como elegir las ξ_is de manera óptima y estimarlas computacionalmente a partir de los datos en el límite $p \to \infty$.

La pregunta es: ¿Cuál la elección óptima de ξ_i tal que $\Xi(E)$ es tan cercano como sea posible a Σ?

Si los vectores propios de E fueran igual a los de Σ $(v_i = u_i, \forall i)$ la solución sería trivialmente $\xi_i = \mu_i$. Pero cuando $v_i \neq u_i$ es no trivial a priori.

En este caso nos interesa minimizar el siguiente error de mínimos cuadrados

$$Tr\left[\Xi(E) - \Sigma\right]^2 = \sum_{i=1}^{p} v_i'(\Xi(E) - \Sigma)^2 v_i = \sum_{i=1}^{p}\left(\xi_i^2 - 2\xi_i v_i'\Sigma v_i + v_i'\Sigma^2 v_i\right) \tag{86}$$

Si notamos que el tercer término es independiente de las ξ's al minimizar sobre ξ_k obtenemos la siguiente expresión

$$\xi_k = v_k'\Sigma v_k \tag{87}$$

El *milagro* de alta dimensión

El resultado anterior da la apariencia de absurdo: asumimos que no conocemos Σ y queremos obtener el mejor estimador de Σ dado E, y encontramos una ecuación para ξ que no podemos resolver a menos que conozcamos Σ.

A la expresión (87) se le conoce como estimador *oráculo* dado que requiere de información que no conocemos.

El estimador RIE solo depende de las transformaciones de la matriz de covarianza muestral E (observable), tales como $G_E(z)$, $T_E(z)$, y de las transformadas R y S del ruido.

Sin embargo, es necesario conocer el comportamiento de las transformadas en el límite asintótico para el eje real, precisamente donde las transformadas discretas $G_E^p(z)$ y $T_E^p(z)$ no convergen.

Afortunadamente, en el limite de $p \to \infty$ es posible calcular el valor óptimo de ξ partiendo directamente de los datos, sin necesidad de conocer Σ.

Solución de Ledoit& Peché (2011)

En el caso particular de correlaciones espaciales es posible utilizar el resultado (clase anterior)

$$g_E(z) = \int \frac{\rho_\Sigma(\mu) d\mu}{z - \mu(1 - q + qzg_E(z))} \tag{88}$$

para inferir la densidad espectral poblacional (*verdadera*) $\rho_\Sigma(\lambda)$ a partir del espectro de la matriz muestral E.

Esta relación la utilizaron Ledoit & Peche(2011) para demostrar que el RIE óptimo al modelar correlaciones espaciales está dado por la expresión [5]:

$$\xi(\lambda) = \frac{\lambda}{|1 - q + q\lambda g_E(\lambda - i\epsilon)|}\Bigg|_{\epsilon \to 0^+} \tag{89}$$

De esta manera, el estimador RIE de E esta dado por

$$\Xi(E) = \sum_{i=1}^{p} \xi_i(\lambda) v_i v_i', \tag{90}$$

[5] Los detalles escapan a los objetivos de este curso

Ejemplo: S&P 500

- Se consideraron retornos de las empresas que componen el S& P500.
- El periodo de estudio comprendió del 1992-01-06 al 2021-10-15 a una frecuencia diaria para un total de $t = 7503$ días de transacción.
- Se incluyeron los mercados que tuvieron menos del 10% de días faltantes.
- De esta manera se estudio un conjunto de $p = 250$ series financieras imputados con splines de orden tres.

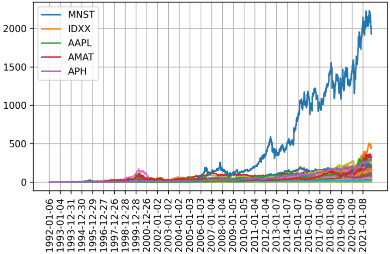
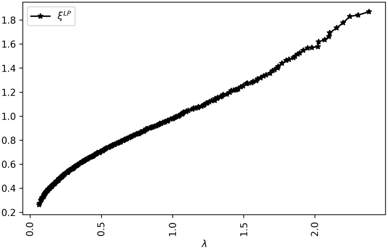


Figure (3) Retornos acumulados

Ejemplo: S&P 500



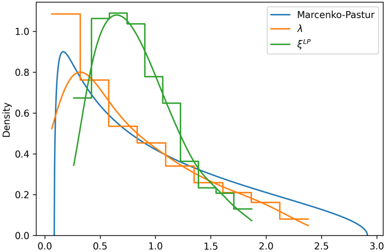


Figure (4) valores propios estimados para primera ventana de tiempo (a) contracción no-lineal (b) densidad

Ejemplo: S&P 500

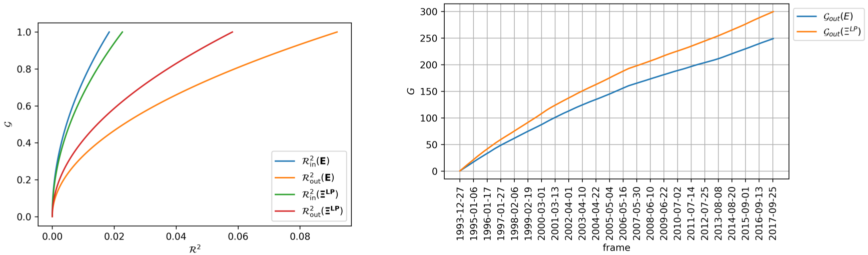


Figure (5) (a) frontera óptima para primera ventana de tiempo (b) ganancia acumulada a través del periodo de estudio.

Conclusiones

- La teoría de matrices aleatorias surge en la física estadística, pero en la actualidad es de gran relevancia en la ciencia de datos y en especial dentro de la estadística de alta dimensionalidad.
- La probabilidad libre permite trabajar con estructuras más complejas dentro de la teoría de matrices aleatorias.
- Cuando el número de activos es del mismo orden que el número de observaciones el riesgo de inversión está fuertemente sesgado.
- La familia de estimadores RIE nos permiten estimar correctamente la matriz de covarianza muestral.
- Se están desarrollando aplicaciones en finanzas en contextos de alta dimension o *big data*.

Proyectos

Modelación en Finanzas y Econometría desde el paradigma de la Econofísica (A1-S-43514).

- Rotationally invariant estimators on portfolio optimization to unveil financial states. En colaboración con Rodrigo Macías Páez - CIMAT-MTY (`http://dx.doi.org/10.2139/ssrn.4126928`).
- Reducción de Sesgo en Matrices de Covarianza y sus Efectos en Portafolios de Alta Dimensionalidad. Benito Rodríguez Camejo. Maestría en Cómputo Estadístico.
- Métodos de agrupamiento jerárquico para diversificación de portafolios y asignación de capital. José Antonio Duarte Mendieta. Maestría en Cómputo Estadístico.

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. (2018). 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.