An introduction to random matrices and free probability in portfolio theory: a data science approach.

Andrés García Medina

CONACyT Research Fellow at Center for Research in Mathematics, Monterrey (Mexico).
`https://sites.google.com/view/andresgm/home`

Observatory of Complex Systems at Dipartimento di Fisica e Chimica – Emilio Segrè, Palermo University. July 11-12, 2022.

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

Classical probability

Definition

Two random variables x_1, x_2 *with pdf* ρ_1, ρ_2 *are said to be statistically independent if*

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

Definition

The characteristic function of $\rho(x)$ *is defined through the Fourier transform*

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

In this way the independence between the random variables implies that $\phi_{1,2}(t_1, t_2) = \phi_1(t_1)\phi_2(t_2)$.

Classical probability

Likewise, if we consider the cumulant-generating function (variance, *skewness*, kurtosis), that is, the logarithm of the characteristic function $h(t) = log\phi(t)$, we observe that it is additive

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

In this way, the problem of finding $p(x_1 + x_2)$ is reduced to an *algorithm* (as long as x_1, x_2 are independent, otherwise the concept of a copula arises):

1. Calculate the characteristic function of x_1 and x_2 through their probability functions
2. Calculate the cumulants and sum them
3. Obtain the characteristic function of the sum via exponentiation or equivalently by the convolution theorem.
4. Apply the inverse Fourier transform of the sum.

Classical probability

Example: Gaussian

For Gaussian random variables $\{x_i\}$ we have that

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

In this case the characteristic function is:

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

For the case of the sum $s_2 = x_1$ and x_2, we obtain

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

Applying the inverse transform, we find that $(\sigma_2 = \sqrt{2}\sigma)$

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

Classical probability

Example: Cauchy

For random variables x_i with Cauchy (Lorentz) distribution, the pdf is given by

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

and the characteristic function is given by

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

In this case

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

Applying the inverse

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

Stable attractors $p(s_n)$ Notebook

In general,

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

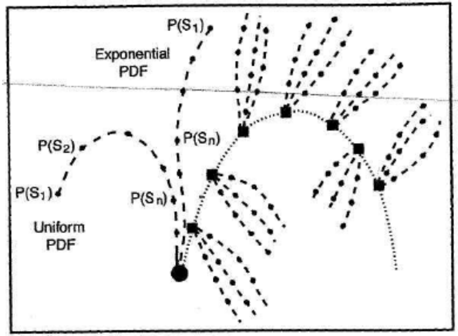


Figure (1) Mantegna, R. N., & Stanley, H. E. (1999). Introduction to econophysics: correlations and complexity in finance. Cambridge university press.

Overview of multivariate statistics

In general, suppose we have two independent Wishart matrices $A \sim W_p(n_1, \mathrm{I})$ and $B \sim W_p(n_2, \mathrm{I})$, with degrees of freedom $n_1, n_2 \geq p$. This situation is known as the double Wishart scenario.

Of central interest in multivariate analysis is to find the roots $\lambda_i, i = 1, \dots, p$ of the generalized eigenvector problem constructed from A and B

$$det[\lambda(A + B) - A] = 0 \tag{13}$$

Essentially all classical multivariate techniques involve a proper decomposition that reduces to some form of the above equation [1]

[1] Johnstone, I. M. (2006). High dimensional statistical inference and random matrices. arXiv preprint math/0611589.

Overview of multivariate statistics

In fact, it is possible to classify almost all the multivariate models that appear in any classic textbook in statistics on the following table

Double Wishart

- Canonical Correlation Analysis (CCA)
- Multivariate Analysis of Variance
- Multivariate Regression Analysis
- Discriminant Analysis
- Covariance Matrix Equality Test

Single Wishart

- Principal Component Analysis (PCA)
- Factor Analysis
- Multidimensional Scaling

This table emphasizes the importance of finding the distribution of the roots, which is a basic element of inference to the use of these methods in practice.

Overview of multivariate statistics

The joint distribution of eigenvalues for CCA and PCA has been known since 1939. The results were found almost simultaneously by 5 major statisticians around the world (Fisher, Girshick, Hsu, Mood, Roy)[2]:

$$\rho(\lambda_1, \ldots, \lambda_p) = c \prod_i w^{1/2}(\lambda_i) \prod_{i<j} (\lambda_i - \lambda_j), \quad \lambda_i \geq \cdots \geq \lambda_p, \tag{14}$$

where

$$w(\lambda) = \begin{cases} \lambda^{n-p-1} e^{-\lambda} & \text{single Wishart} \\ \lambda^{n_1-p-1}(1-\lambda)^{n_2-p-1} & \text{double Wishart} \end{cases} \tag{15}$$

$$c = \begin{cases} \frac{2^{-pn/2}\pi^{p^2/2}}{\Gamma_p(p/2)\Gamma_p(n/2)} & \text{single Wishart} \\ \frac{\pi^{p^2/2}\Gamma_p((n_1+n_2)/2)}{\Gamma_p(p/2)\Gamma_p(n_1/2)\Gamma_p(n_2/2)} & \text{double Wishart} \end{cases} \tag{16}$$

[2] Anderson, Muirhead, Mardia

Overview of multivariate statistics

Standard models in statistics deal with additive or multiplicative noise.

The original method for deriving the semicircle law and the Marchenko-Pastur law is based on an analogous procedure for finding the density of a gas of charged particles (Coulomb gas).

Unless one can write exactly the probability density function (PDF) of the perturbed matrix entries, the Coulomb gas analogy is not directly useful.

The approach does not allow investigation of the spectrum of a matrix perturbed by some noise source.

Free Probability

Free probability theory is an alternative method to study the asymptotic behavior of some *large* random matrices.

This theory provides us with a robust way to investigate the limiting spectral density (LSD) of sums or products of random matrices with specific symmetry properties.

This is an important problem in practice since the standard models in statistics deal with additive or multiplicative noise.

In particular, we are interested in extracting the *true* signal of noisy observations from the covariance matrix, which is an essential element in Markowitz portfolio theory.

Background

Free probability theory was started in 1985 by Dan Voiculescu to understand special classes of von Neumann algebras, establishing calculus rules for non-commutative operators.

Voiculescu and Speicher found that rotationally invariant random matrices asymptotically satisfy the *freeness* criteria.

In general terms, two large matrices A and B are mutually free if their eigenbasis are related to each other by random rotation, that is, when their eigenvectors are orthogonal.

Preliminaries

Formally, the ingredients that we will need to define the concept of *freeness* are the following:

- A ring $\mathcal{R}$ [3] of random variables, which may be non-commutative with respect to multiplication.
- A field of scalars which is usually $\mathbb{C}$: Scalars commute with everything.
- An operation *, called involution: the transposes on real matrices, the complex conjugate, the conjugate transposes on complex matrices.
- A positive linear function $\phi(\cdot) : \mathcal{R} \to \mathbb{C}$ that satisfies $\phi(AB) = \phi(BA)$ for $A, B \in \mathcal{R}$. By positive we mean that $\phi(AA^*)$ is real and non-negative. Furthermore, we require that $\phi(AA^*) = 0 \Rightarrow A = 0$ (*faithful*).

For example, ϕ can be the operator $\mathbb{E}[\cdot]$ in classical probability, or the normalized trace operator $\frac{1}{N} Tr(\cdot)$ for a ring of matrices, or even the combination: $\frac{1}{N}\mathbb{E}[Tr(\cdot)]$.

[3] set equipped with two binary operations that generalizes the arithmetic operations of addition and multiplication

Preliminaries

- We will call the elements of $\mathcal{R}$ random variables and denote them by uppercase letters.
- For any $A \in \mathcal{R}$ and $k \in \mathbb{N}$ we say that $\phi(A^k)$ is the k-th moment of A and assume that $\phi(A^k)$ is finite for all k. In particular, we say that $\phi(A)$ is the mean of A and $\phi(A^2) - \phi(A)^2$ is its variance.
- We say that two elements A and B have the same distribution if their moments are the same for all orders [4]
- The ring of variables must contain the element named $\mathbb{1}$ such that $A\mathbb{1} = \mathbb{1}A = A$ for every A. This satisfies $\phi(\mathbb{1}) = 1$. Thus, adding a constant α simply shifts the mean: $\phi(A + \alpha\mathbb{1}) = \phi(A) + \alpha$.

[4] This is not always true: some distributions are not uniquely determined by their moments.

Moments of sums: commutative random variables

It is important to remember that two r.v. commute if $AB = BA \quad \forall A, B \in \mathcal{R}$.

Note that A, B are not necessarily real (complex) numbers, but can be elements of a more abstract ring.

We say that A and B are independent if $\phi(p(A)q(B)) = \phi(p(A))\phi(q(B))$ for any polynomial p, q: This condition is equivalent to moment factorization.

By the linearity property of the ϕ operator we have that $\phi(A + B) = \phi(A) + \phi(B)$

Let us assume from here on that $\phi(A) = \phi(B) = 0$, i.e., A, B have mean zero. For a variable $\tilde{A}$ with nonzero mean, we can write $A = \tilde{A} - \phi(\tilde{A})$, such that $\phi(A) = 0$.

For the second moment of the sum we have

$$\phi((A + B)^2) = \phi(A^2) + \phi(B^2) + 2\phi(AB) = \tag{17}$$

$$\phi(A^2) + \phi(B^2) + 2\phi(A)\phi(B) = \phi(A^2) + \phi(B^2), \tag{18}$$

that is, the variance is additive. The same happens for the third moment.

Moments of sums: non-commutative random variables

One of the goals is to generalize the law of addition of independent variables.

Let us now consider the variable $A + B$, where A and B are objects that do not commute like random matrices. If we calculate the first three moments of $A + B$ no problems really arise due to the definition of the operator ϕ through the trace.

The problems become interesting when we try to calculate the fourth moment of the sum

$$\phi((A+B)^4) = \phi(A^4)+4\phi(A^3B)+4\phi(A^2B^2)+2\phi(ABAB)+4\phi(AB^3)+\phi(B^4), \quad (19)$$

where we have used the cyclic property of the trace.

In the commutative case the independence of A, B implies $\phi(A^2B^2) = \phi(A^2)\phi(B^2)$ and would be enough to deal with mixed moments.

In the non-commutative case we need to calculate the term $\phi(ABAB)$, since in general $ABAB \neq A^2B^2$. So we need a new definition of independence to deal with this term. A radical solution will be to postulate that $\phi(ABAB) = 0$ whenever $\phi(A) = \phi(B) = 0$.

Free Random Variables

Definition

Given two random variables A, B we say that they are *free* if for any polynomial $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ such that

$$\phi(p_k(A)) = 0, \quad \phi(q_k(B)) = 0, \quad \forall k, \tag{20}$$

we have

$$\phi(p_1(A)q_1(B)p_2(A)q_2(B)\ldots p_n(A)q_n(B)) = 0 \tag{21}$$

A polynomial (or variable) will be said *traceless* if $\phi(p(A)) = 0$. We can noticed that $\alpha\mathbb{1}$ is free with respect to $A \in \mathcal{R}$ because $\phi(p(\alpha\mathbb{1})) = p(\alpha\mathbb{1})$, given the definition of $\mathbb{1}$. Then

$$\phi(p(\alpha\mathbb{1})) = 0 \iff p(\alpha\mathbb{1}) = 0 \tag{22}$$

Furthermore, it is easy to see that if A, B are free, then $p(A), q(B)$ are free for any polynomial p, q. By extension, $F(A)$ and $G(B)$ are also free for any functions F and G defined by their power series.

Moments of sums: free random variables

Assuming that A, B are free with $\phi(A) = \phi(B) = 0$ it is possible to calculate the free sum $A + B$. The second moment is given

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

For moments of third order and higher, the trick is to add and subtract quantities in each term.

$$\phi((A+B)^3) = \phi(A^3) + \phi(B^3) + 3\phi(A^2B) + 3\phi(AB^2) = \tag{24}$$

$$\phi(A^3) + \phi(B^3) + 3\phi((A^2 - \phi(A^2))B) + 3\phi(A^2)\phi(B) + \tag{25}$$

$$3\phi(A(B^2 - \phi(B^2))) + 3\phi(A)\phi(B^2) = \phi(A^3) + \phi(B^3) \tag{26}$$

Finally, for the fourth moment we have[5]:

$$\phi((A+B)^4) = \phi(A^4) + \phi(B^4) + 4\phi(A^2)\phi(B^2) \tag{27}$$

[5] for commutative variables: $\phi((A+B)^4) = \phi(A^4) + \phi(B^4) + 6\phi(A^2)\phi(B^2)$

Random Matrices and Freeness

Two variables A and B were said to be free if for any set of polynomials $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ the following equality holds

$$\phi(p_1(A)q_1(b)p_2(A)q_2(B)\ldots p_n(A)q_n(B)) = 0 \tag{28}$$

- The connection with matrices consists of considering A and B as large symmetric matrices and $\phi(M) := \frac{1}{N}Tr(M)$.
- The matrices A, B can be diagonalized by $U\Delta U', V\Delta V'$, respectively.
- A traceless polynomial $p_i(A)$ can be diagonalized as $U\Delta_i U'$, where U is the same orthogonal matrix for A and $\Delta_i = p_i(\Delta)$ is a traceless diagonal matrix, and the same for $q_i(B)$.

Then the freedom condition becomes

$$\phi(\Delta_1 O\Delta_1' O'\Delta_2 O\Delta_2' O' \ldots \Delta_n O\Delta_n' O') = 0, \tag{29}$$

where we introduc $O = U'V$ as the orthogonal matrix of change of basis that rotates the eigenvectors A towards B.

Random Matrices and Freeness

In the limit $N \to \infty$ the expression (29) is true when averaging over the orthogonal group O and provided that Δ_i and Δ'_i are traceless

It is also expected that in the limit $N \to \infty$ the equation (29) is self-averaging such that a single matrix O behaves as the average over all matrices.

In this way, two large symmetric matrices whose basis are randomly rotated with respect to each other are essentially free.

For example, the Wigner matrices X and the Wishart matrices W are rotationally invariant, which implies that their eigenvectors are orthogonal random matrices.

It follows that for N large, both X and W are free with respect to any matrix independent of them, in particular they are free with respect to deterministic matrices.

Note: To prove the expression (29) it is required to integrate over the orthogonal group, for which the Weingarten function and the Wick contraction rules can be used (transform it to a combinatorial problem).

Random Matrices and Freeness

More generally, freedom allows the computation of mixed moments of sum and products of matrices from knowledge of the moments of A and B, similar to classical independence in probability theory.

A typical example of pairs of free matrices is when A is a fixed matrix and B is a random matrix belonging to a rotationally invariant set, that is, $B = O\Lambda O'$, where Λ is diagonal and O is distributed according to the Haar measure over the orthogonal group, in the limit where $N \to \infty$

An important property of random matrices is that their eigenvectors are uniformly distributed over the group $O(N)$, so each vector is distributed over the unit sphere S^{N-1}. Thus when N is large it is unlikely that any eigenvector of A will overlap with an eigenvector of B.

The computation of mixed moments is crucial for deriving useful relationships in estimation problems in high-dimensional multivariate statistics.

Random matrices and Freeness

We have introduced the concept of *Freeness* which can be summarized in the following intuitive definition

Definition

Two *large* matrices are free if their proper basis are related by random rotation. We consider a matrix *large* if its moments calculated under free probability are valid up to a correction of order $\mathcal{O}(1/N)$.

$\mathcal{R}$-Transform

The $\mathcal{R}$-transform is defined as

$$\mathcal{R}_A(g) := \mathcal{B}_A(g) - \frac{1}{g} \tag{30}$$

where $\mathcal{B}_A(g)$ is the inverse Stieltjes transform defined through its power series (momentum generating function) that satisfies $g_A(\mathcal{B}_A(g)) = g$, at all orders.

In general, the $\mathcal{R}$-transform is additive for free random variables

$$\mathcal{R}_{A+B}(g) = \mathcal{R}_A(g) + \mathcal{R}_B(g) \tag{31}$$

To see this, let us recall that the Stieltjes transform can be written as

$$g_A(z) = \phi(\left[(z-A)^{-1}\right]) = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \phi(A^k), \tag{32}$$

Consider a fixed scalar g, by construction

$$\phi(g\mathbb{1}) = g = g_A(\mathcal{B}_A(g)) = \phi\left[(\mathcal{B}_A(g) - A)^{-1}\right]. \tag{33}$$

$\mathcal{R}$-Transform

The argument of $\phi(\cdot)$ on the left and right have the same meaning, but are different in general, let's denote their difference as gX_A

$$gX_A := (z_A - A)^{-1} - g\mathbb{1}, \tag{34}$$

where $z_A = \mathcal{B}_A(g)$, and by definition $\phi(gX_A) = 0$.

By reversing the above relationship

$$A - z_A = -\frac{1}{g}(1 + X_A)^{-1}. \tag{35}$$

Now, consider another free variable B relative to A, for which a similar relationship is found

$$B - z_B = -\frac{1}{g}(1 + X_B)^{-1}. \tag{36}$$

Being X_A and X_B free since they are functions of A and B.

Adding the expressions (35) and (36) we arrive at the following result (▸ Blackboard)

$$\mathcal{B}_{A+B} = z_A + z_B - g^{-1} \Rightarrow \mathcal{R}_{A+B} = \mathcal{R}_A + \mathcal{R}_B \tag{37}$$

$\mathcal{S}$-Transform

Analogously, the $\mathcal{S}$-transform can be defined for the product of free variables

$$S_{AB}(t) = S_A(t)S_B(t), \tag{38}$$

assuming A and B free.

To define the $\mathcal{S}$-transform you need to introduce the transform $\mathcal{T}$

$$\mathcal{T}_A(z) = \phi\left[(1 - z^{-1}A)^{-1}\right] - 1 = zg_A(z) - 1 = \sum_{k=1}^{\infty} \frac{\phi(A^k)}{z^k} \tag{39}$$

Then, the $\mathcal{S}$-transform is given by

$$S_A(t) := \frac{t+1}{t\mathcal{Z}_A(t)}, \tag{40}$$

where $\mathcal{Z}_A(t)$ is the inverse of $\mathcal{T}$.

Applying to random matrices it is better to work with the product of the form $A^{1/2}BA^{1/2}$ since AB is not necessarily symmetric.

$\mathcal{R}$- and $\mathcal{S}$-Transforms

To summarize, A and OBO' are free when O is a random rotation matrix (in the large dimension limit).

When A and B are free, their transform $\mathcal{R}$ and $\mathcal{S}$ allow us to sum and multiply them:

$$\mathcal{R}_{A+B}(g) = \mathcal{R}_A(g) + \mathcal{R}_B(g) \tag{41}$$

$$\mathcal{S}_{AB}(t) = \mathcal{S}_A(t)\mathcal{S}_B(t) \tag{42}$$

In general AB is not a symmetric matrix, in fact the transform $S_{AB}(t)$ is related to the eigenvalues of the matrix $\sqrt{A}B\sqrt{A}$, which are the same of $\sqrt{B}A\sqrt{B}$ when A and B are positive semidefinite (eigenvalues $\lambda \geq 0$).

$\mathcal{R}$- and $\mathcal{S}$-Transforms

The transform $\mathcal{R}$ and $\mathcal{S}$ can be defined through the following relations:

$$g_A(z) = \phi\left[(z - A)^{-1}\right] \quad \text{(Stieltjes transform of A)} \tag{43}$$

$$\mathcal{R}_A(g) = \mathcal{B}_A(g) - \frac{1}{g} \quad (\mathcal{B} \text{ Inverse of } g_A) \tag{44}$$

$$\mathcal{T}_A(z) = z g_A(z) - 1 \quad (\mathcal{T}\text{-transform}) \tag{45}$$

$$\mathcal{S}_A(t) = \frac{t+1}{t\mathcal{Z}_A(t)} \quad , \text{Si} \quad \phi(A) \neq 0 \quad (\mathcal{Z} \text{ inverse of } \mathcal{T}) \tag{46}$$

$\mathcal{R}$- and $\mathcal{S}$-Transforms: properties

- Multiplication by a scalar:

$$\mathcal{R}_{\alpha A}(g) = \alpha \mathcal{R}_A(\alpha g), \quad \mathcal{S}_{\alpha A}(t) = \alpha^{-1} \mathcal{S}_A(t) \tag{47}$$

- Translation [6]

$$\mathcal{R}_{A+\alpha I}(g) = \alpha + \mathcal{R}_A(g) \tag{48}$$

- Matrix inversion

$$\mathcal{S}_{A^{-1}}(t) = \frac{1}{\mathcal{S}_A(-t-1)} \tag{49}$$

- The two transforms are related by the following equivalent identities

$$\mathcal{S}_A(t) = \frac{1}{\mathcal{R}_A(t\mathcal{S}_A(t))}, \quad \mathcal{R}_A(g) = \frac{1}{\mathcal{S}_A(g\mathcal{R}_A(g))} \tag{50}$$

- When $A = I$

$$g_I(z) = \frac{1}{z-1} \quad \mathcal{T}_I = \frac{1}{z-1} \tag{51}$$

$$\mathcal{R}_I(g) = 1 \quad \mathcal{S}_I(t) = 1 \tag{52}$$

[6] there is no equivalent simple formula for $\mathcal{S}$

$\mathcal{R}$- and $\mathcal{S}$-Transforms: Taylor expansion

The transforms $\mathcal{R}$ and $\mathcal{S}$ have the following Taylor expansion for small arguments:

$$\mathcal{R}_A(g) = \kappa_1 + \kappa_2 g + \kappa_3 g^2 + \ldots, \tag{53}$$

$$\mathcal{S}_A(t) = \frac{1}{\kappa_1} - \frac{\kappa_2}{\kappa_1^3} t + \frac{2\kappa_2^2 - \kappa_1\kappa_3}{\kappa_1^5} t^2 + \ldots, \tag{54}$$

where κ_n are the free cumulants of A:

$$\kappa_1 = \phi(A) \tag{55}$$

$$\kappa_2 = \phi(A^2) \tag{56}$$

$$\kappa_3 = \phi(A^3) - 3\phi(A)\phi(A^2) + 2\phi^3(A) \tag{57}$$

(58)

Procedure for finding the density of eigenvalues: sums

The $\mathcal{R}$-transform gives us a systematic method to obtain the spectrum of the sum C of two free matrices A and B, where at least one of them is rotationally invariant:

1. Find $g_A(z)$ and $g_B(z)$
2. Invert $g_A(z)$ and $g_B(z)$ to obtain the *Blue* transform $\mathcal{B}_A(g)$ and $\mathcal{B}_B(g)$ and then $\mathcal{R}_A(g)$ y $\mathcal{R}_B(g)$.
3. Thus, $\mathcal{R}_C(g) = \mathcal{R}_A(g) + \mathcal{R}_B(g)$, where $\mathcal{B}_C(g) = \mathcal{R}_C(g) + \frac{1}{g}$
4. Compute the inverse and get $g_C(z)$
5. Use the relation for the density

$$\rho_C(\lambda) = \frac{1}{\pi} \lim_{\epsilon \to 0^+} Im\{g_C(\lambda - i\epsilon)\} \tag{59}$$

Procedure for finding the density of eigenvalues: product

In the multiplicative case $C = A^{1/2}BA^{1/2}$ the recipe is quite similar

1. Find $\mathcal{T}_A(z)$ and $\mathcal{T}_B(z)$
2. Invert $\mathcal{T}_A(z)$ and $\mathcal{T}_B(z)$ to obtain $\mathcal{Z}_A(t)$ and $\mathcal{Z}_B(t)$. Consequently, you get $\mathcal{S}_A(t)$ and $\mathcal{S}_B(t)$
3. Thus, $\mathcal{S}_C(t) = \mathcal{S}_B(t)\mathcal{S}_A(t)$, then $\mathcal{Z}_C(t) = \frac{t+1}{\mathcal{S}_C(t)t}$
4. Compute the inverse and get $\mathcal{T}_C(z)$
5. Use the relation for the density $g_C(z) = (\mathcal{T}_C(z) + 1)/z$, equivalently

$$\rho_C(\lambda) = \frac{1}{\pi\lambda} \lim_{\epsilon \to 0^+} Im\{\mathcal{T}_C(\lambda - i\epsilon)\} \tag{60}$$

Classical and free probability analogy

Classical probability	Free probability
Characteristic function $\phi(t) \equiv \mathbb{E}[e^{itx}]$	Stieltjes function: $g_M(z) = \phi[(z\mathbb{1} - M)^{-1}]$ $= \frac{1}{N}\mathbb{E}\left(Tr\left[(z\mathbb{1} - M)^{-1}\right]\right)$
Independence	Freeness
Sum r.v. Logarithm is additive $h(t) = log\phi(t)$ $h_{1,2}(t_1, t_2) = h_1(t_1) + h_2(t_2)$	Sum f.r.v $\mathcal{R}$-transform is additive $R_H(g) = B_H(g) - 1/g$ $\mathcal{R}_{A+B}(g) = \mathcal{R}_A(g) + \mathcal{R}_B(g)$
Multiplication r.v. Exponential map $e^{itx_1}e^{itx_2} = e^{it(x_1+x_2)}$	Multiplication f.r.v. $\mathcal{S}$-transform $\mathcal{S}_{AB}(t) = \mathcal{S}_A(t)\mathcal{S}_B(t)$

Example: GOE + Wishart

Suppose you want to calculate the average spectral density of the sum of large random matrices ($N \to \infty$) belonging to two different ensembles.

- Consider the random matrices A and B belonging to ensembles characterized by the Stieltjes transforms $g_A(z)$ and $g_B(z)$, respectively.
- The $\mathcal{R}$-transform of $M = A + B$ is given by

$$\mathcal{R}_M(g) = \mathcal{R}_A(g) + \mathcal{R}_B(g) \tag{61}$$

- We can write an algorithm to obtain the density of eigenvalues of the matrix M through the transform $\mathcal{R}$.

Example: GOE + Wishart

In particular, let us consider the combination of a GOE matrix X and a Wishart ensemble matrix W:

$$M = rX + (1-r)W, \quad \text{where} \quad r \in [0,1] \tag{62}$$

It can be verified that the transforms $\mathcal{R}$ of the resolvents of each ensemble are given by (Blackboard)

$$\begin{aligned} \mathcal{R}_H(g) &= \frac{g}{2} \\ \mathcal{R}_W(g) &= \frac{1}{1-qg} \end{aligned} \tag{63}$$

where $q = p/n$. Now using the scaling property $\mathcal{R}_{\alpha A}(g) = \alpha \mathcal{R}_A(\alpha g)$, it is obtained

$$\begin{aligned} \mathcal{R}_M(g) &= r\mathcal{R}_X(rg) + (1-r)\mathcal{R}_W((1-r)g) \\ &= \frac{r^2}{2}g + \frac{(1-r)}{1-q(1-r)g}. \end{aligned} \tag{64}$$

Example: GOE + Wishart

Given that $\mathcal{B}_M(g_M(z)) = z$, where $\mathcal{B}_M(g) = \mathcal{R}_M(g) + 1/g$,

we arrive to an equation for $g_M(z)$:

$$z = \frac{r^2}{2} g_M(z) + \frac{(1-r)}{1 - q(1-r)g_M(z)} + \frac{1}{g_M(z)}. \tag{65}$$

In general, this third degree equation has one real solution, and two complex conjugate solutions.

Finally, the density of eigenvalues of M is calculated with

$$\rho(\lambda) = \frac{1}{\pi} \lim_{\varepsilon \to 0^+} \operatorname{Im} g_M(\lambda - \mathrm{i}\varepsilon) \tag{66}$$

Example: GOE + Wishart

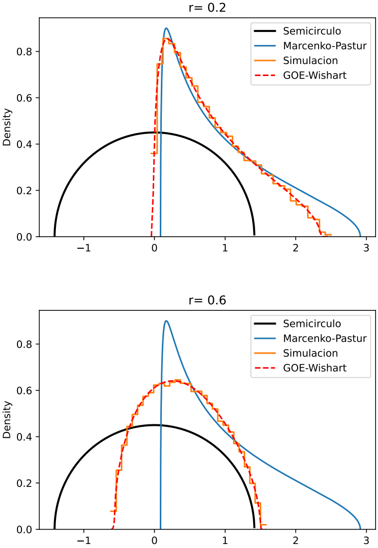
Symbolically

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

Simulation

```
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)
```

Example: GOE + Wishart ▸ Notebook



Example: Spatial correlations

The sample covariance matrix is defined as

$$E = \frac{1}{n} H H', \tag{67}$$

where H is a rectangular data matrix of dimension $n \times p$. If the p time series are stationary, it is expected that for $n >> p$, E converges to the population covariance matrix Σ.

Consider the case where H are multivariate Gaussian observations from $\mathcal{N}(0, \Sigma)$. In this case E is given by a general Wishart matrix with covariance matrix Σ (Blakboard)

$$E = \Sigma^{1/2} W_q \Sigma \tag{68}$$

Example: Spatial correlations

We can recognize the expression (68) as the free product of Σ and a white Wishart matrix W_q. Thus, its $\mathcal{S}$ -transformturns out to be (Blackboard)

$$\mathcal{S}_E(t) = \mathcal{S}_\Sigma(t)\mathcal{S}_{W_q}(t) = \frac{\mathcal{S}_\Sigma(t)}{1+qt} \tag{69}$$

Making use of the free product subordination relation we can rewrite (69) in terms of the $\mathcal{T}$-transform (Blackboard)

$$\mathcal{T}_E(z) = \mathcal{T}_\Sigma\left(\frac{z}{1+q\mathcal{T}_E(z)}\right) \tag{70}$$

Now, in terms of the Stieltjes transform (Blackboard)

$$zg_E(z) = Zg_\Sigma(Z), \quad \text{donde} \quad Z = \frac{z}{1-q+qzg_E(z)} \tag{71}$$

Example: Spatial correlations

The last expression leads us to infer the population spectral density (*true*) $\rho_\Sigma(\lambda)$ from the spectrum of the sample matrix E(Blackboard)

$$g_E(z) = \int \frac{\rho_\Sigma(\mu)d\mu}{z - \mu(1 - q + qzg_E(z))} \tag{72}$$

Therefore, by the use of free probability tools it is possible to infer the population spectral density (*true*) $\rho_\Sigma(\lambda)$ from the spectrum of the sample matrix E

This expression is very useful in portfolio theory to model cross-correlations in the covariance matrix.

The high dimensional limit: portfolio theory

In the limit of large matrices and with some assumptions about the g structure, we can specify the in- and out-of-sample risk inequalities through free probability.

Let's suppose for simplicity that

$$g = N_p(0, I_p), \tag{73}$$

Although our arguments will be valid for any vector g whose direction is independent of Σ or E, as long as $g'g = p$, that is, each component of g is of unitary order.

We emphasize that these assumptions are not necessarily realistic (predictors may be biased along the principal components of Σ) but allow us to more accurately quantify the relationship between in-sample/true/out-sample risk.

The high dimensional limit: portfolio theory

Let M be an independent positive definite matrix of the vector g. So, we have in the large limit of p,

$$\frac{g'Mg}{p} = \frac{1}{p}Tr[g'gM] \underset{freeness}{=} \frac{g'g}{p}\phi(M) \tag{74}$$

where we remember that ϕ is the normalized trace operator

$$\phi(M) = \frac{1}{p}Tr(M) \tag{75}$$

and the first moment of $\rho_M(\lambda)$.

This implies that we can use the algebra of free moments whenever $p \to \infty$

In this way, since $g'g = p$ we have

$$\frac{g'Mg}{p} - \phi(M) \underset{p\to\infty}{\to} 0 \tag{76}$$

The high dimensional limit: portfolio theory

Let us now set $M = \{E^{-1}, \Sigma^{-1}\}$ and use the above relation on the in-sample, true, and out-sample risk expressions:

$$\mathcal{R}_{in}^2 \to \frac{\mathcal{G}^2}{p\phi(E^{-1})} \tag{77}$$

$$\mathcal{R}_{true}^2 \to \frac{\mathcal{G}^2}{p\phi(\Sigma^{-1})} \tag{78}$$

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

$$\tag{80}$$

For $q < 1$ ($q = p/n$) it has been shown in the high dimensional regime that (see spatial correlations)

$$\phi(\Sigma^{-1}) = (1-q)\phi(E^{-1}) \tag{81}$$

As a result when $p \to \infty$

$$\mathcal{R}_{in}^2 = (1-q)\mathcal{R}_{true}^2 \tag{82}$$

The high dimensional limit: portfolio theory

Thus, for any $q \in (0,1)$ we see that the in-sample risk associated with the weights w_E gives us an overestimate (optimistic). Fortunately, it is possible to quantify this overestimation in the high-dimensional limit.

We now look for a similar relationship for the out-sample risk.

In general, the sample covariance matrix E can be written as $E = \Sigma^{1/2} W_q \Sigma^{1/2}$, where W_q is a matrix with distribution $W_p(n, I_p)$ independent of Σ. Thus, when $p \to \infty$

$$\mathcal{R}^2_{out} \to \frac{\mathcal{G}^2 \phi(\Sigma^{-1} W_q^{-2})}{p\phi^2(E^{-1})} \tag{83}$$

Considering large matrices, W and Σ_q are asymptotically free, so we have the relation

$$\phi(\Sigma^{-1} W^{-2}) = \phi(\Sigma^{-1})\phi(W^{-2}) \tag{84}$$

The high dimensional limit: portfolio theory

Using the asymptotic relation (81) we get

$$\mathcal{R}_{out}^2 = \mathcal{G}^2 (1-q)^2 \frac{\phi(W_q^{-2})}{p\phi(\Sigma^{-1})} \tag{85}$$

Finally, $\phi(W^{-2}) = (1-q)^{-3}$ can be obtained for $q < 1$, which gives us

$$\mathcal{R}_{out}^2 = (1-q)^{-1} \frac{\mathcal{G}^2}{p\phi(\Sigma^{-1})} = \frac{\mathcal{R}_{true}^2}{1-q} \tag{86}$$

In sum, we have the following valid asymptotic relation for a general structure of Σ:

$$\frac{\mathcal{R}_{in}^2}{1-q} = \mathcal{R}_{true}^2 = (1-q)\mathcal{R}_{out}^2 \tag{87}$$

Simulation ▸ Notebook

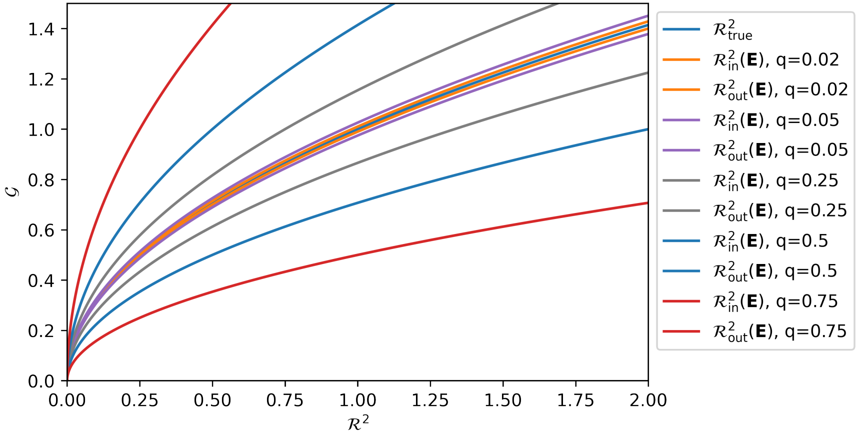


Figure (2) $q = \{1/50, 1/20, 1/4, 1/2, 3/4\}$, $r = 0.5$, $p = 100$. Arbitrary units have been used such that $\mathcal{R}^2_{true} = 1$ when $\mathcal{G} = 1$.

The high dimensional limit: portfolio theory

Observations:

- If one makes an investment with the *naive* weights w_E, it turns out that the risk predicted (in-sample) underestimates the risk taken by a factor $(1-q)^2$
- In the extreme case $p = n$ $(q = 1)$, the in-sample risk equals zero, while the out-of-sample risk diverges.
- In conclusion, the use of the sample covariance matrix E for the Markowitz optimization problem can lead to disastrous results.
- This suggests that we should have a reliable estimator of Σ to control the out-sample risk.

Rotationally Invariant Estimator (RIE)

Let's review some elements of Bayesian inference before introduce the Rotationally Invariant Estimator (RIE).

In the case of additive noise, one observes a matrix E which is the true matrix Σ plus a random matrix X that plays the role of noise

$$E = \Sigma + X \tag{88}$$

In the case of multiplicative noise, the observed matrix E has the form

$$E = \Sigma^{1/2} W \Sigma^{1/2}. \tag{89}$$

When W is a white Wishart matrix, this is the problem of sample covariance matrix with spatial correlations (cross-correlations).

We would like to know the probability of Σ given that we have observed E, i.e. compute $P(\Sigma|E)$. This is the general subject of Bayesian estimation.

Rotationally Invariant Estimator (RIE)

Suppose we have an observable variable y that we would like to infer from the observation of a related variable x.

The variables x and y can be scalars, vectors, matrices, higher dimensional objects.

Assume that we know the random process that generates y given x.

The generation process of y is encoded in a probability distribution $P(y|x)$, which is called the sampling distribution or the likelihood function.

We would like to write the inference probability $P(x|y)$ given $P(y|x)$, also called the posterior distribution.

To accomplish this we can use Bayes' rule:

$$P(x|y) = \frac{P(y|x)P_0(x)}{P(y)} \tag{90}$$

Rotationally Invariant Estimator (RIE)

To obtain the desired probability, Bayes' rule tells us that we need to know the prior distribution $P_0(x)$.

$P_0(x)$ encodes our ignorance of x. It represent our best (probabilistic) guess of x before we observe the data y.

The other distribution appearing in Bayes'rule $P(y)$ is assumed to be known, and we can simplify the rule as

$$P(x|y) = \frac{1}{Z} P(y|z) P_0(x) \quad Z = \int P(y|x) P_0(x) dx \tag{91}$$

Then, from the posterior distribution $P(x|y)$ we can build an estimator of x.

The optimal estimator depends on the problem at hand, namely, which quantity are we trying to optimize (loss function).

Rotationally Invariant Estimator (RIE)

We now apply the Bayesian estimation method to covariance matrices

$$P(\Sigma|E) \propto P(E|\Sigma)P_0(\Sigma) \tag{92}$$

We have for gaussian observations

$$P(E|\Sigma) = (det\Sigma)^{-n/2} \exp\left[-\frac{n}{2} Tr\left(\Sigma^{-1}E\right)\right] \tag{93}$$

Then, the posterior probability of Σ can be obtained given the sample covariance matrix E

$$P(\Sigma|E) \propto (det\Sigma)^{-n/2} \exp\left[-\frac{n}{2} Tr\left(\Sigma^{-1}E\right)\right] P_0(\Sigma) \tag{94}$$

In general, given the rotational invariance of Σ, we can obtain the invariance of the prior probability (in the Bayesian sense) $P_0(\Sigma)$

$$P_0(\Sigma) = P_0(O\Sigma O'), \tag{95}$$

where O is an arbitrary rotation matrix.

Likewise,

RIE

Also, it can be verified that the MMSE (minimum mean squared error) estimator of Σ transforms in the same direction as E under an arbitrary rotation

$$
\begin{aligned}
\mathbb{E}[\Sigma|OEO'] &= \int \Sigma P(\Sigma|OEO')P_0(\Sigma)d\Sigma \\
&= O\left[\int \tilde{\Sigma}P(\tilde{\Sigma}|E)P_0(\tilde{\Sigma})d\tilde{\Sigma}\right]O' \\
&= O\mathbb{E}(\Sigma|E)O',
\end{aligned} \tag{96}
$$

where the change of variable $\tilde{\Sigma} = O'\Sigma O$ was used, and the explicit form of $P(\Sigma|E)$

In general, if we denote $\Xi(E)$ as the estimator of Σ given E, this estimator is rotationally invariant if and only if

$$
\Xi(OEO') = O\Xi(E)O' \tag{97}
$$

For any orthogonal matrix O

RIE

- If the sample covariance matrix E is rotated by some O, then the estimate of Σ must also be rotated in the same direction.
- This is so because we have no prior assumption about the eigenvectors of Σ.
- The estimators given by the expression (97) are called *Rotationally Invariant Estimators* (RIE)
- $\Xi(E)$ can be diagonalized to the same base as E except for a fixed rotation Ω
- However, there is no natural candidate for Ω, except the identity matrix I
- We can conclude that $\Xi(E)$ has the same eigenvectors as E and write

$$\Xi(E) = \sum_{i=1}^{p} \xi_i v_i v_i', \tag{98}$$

where v_i are the eigenvectors of E, and ξ_i is a function of the eigenvalues $[\lambda_j]_{j \in \{1,p\}}$ of E.

Optimal RIE

Let us now see how to choose the ξ_is optimally and estimate them computationally from the data in the limit $p \to \infty$.

The question is: What is the optimal choice of ξ_i such that $\Xi(E)$ is as close as possible to Σ?

If the eigenvectors of E were equal to those of Σ ($v_i = u_i, \forall i$) the solution would trivially be $\xi_i = \mu_i$. But when $v_i \neq u_i$ the solution is non-trivial a priori.

In this case we are interested in minimizing the following least squares error [7]

$$Tr\left[\Xi(E) - \Sigma\right]^2 = \sum_{i=1}^{p} v_i'(\Xi(E) - \Sigma)^2 v_i = \sum_{i=1}^{p} \left(\xi_i^2 - 2\xi_i v_i' \Sigma v_i + v_i' \Sigma^2 v_i\right) \quad (99)$$

If we notice that the third term is independent of the ξ's when minimizing over ξ_k we obtain the following expression

$$\xi_k = v_k' \Sigma v_k \quad (100)$$

[7] It is equivalent to minimizing with respect to the Frobenius norm

The high dimensional *miracle*

The previous result gives the appearance of absurd: we assume that we do not know Σ and we want to obtain the best estimator of Σ given E, and we find an equation for ξ that we can not solve unless we know Σ.

The expression (100) is known as an *oracle* estimator since it requires information that we do not know.

Fortunately, in the limit of $p \to \infty$ it is possible to calculate the optimal value of ξ starting directly from the data, without needing to know Σ.

Ledoit& Peché (2011) solution

In the particular case of spatial correlations it is possible to use the result

$$g_E(z) = \int \frac{\rho_\Sigma(\mu) d\mu}{z - \mu(1 - q + qzg_E(z))} \tag{101}$$

to infer the population spectral density (*true*) $\rho_\Sigma(\lambda)$ from the spectrum of the sample matrix E.

This relationship was used by Ledoit & Peche(2011) to show that the optimal RIE when modeling spatial correlations is given by the expression [8]:

$$\xi(\lambda) = \left. \frac{\lambda}{|1 - q + q\lambda g_E(\lambda - i\epsilon)|} \right|_{\epsilon \to 0^+} \tag{102}$$

In this way, the RIE estimator of E is given by

$$\Xi(E) = \sum_{i=1}^{p} \xi_i(\lambda) v_i v_i', \tag{103}$$

[8] yet derivation and numerical implementation is not trivial

Projects

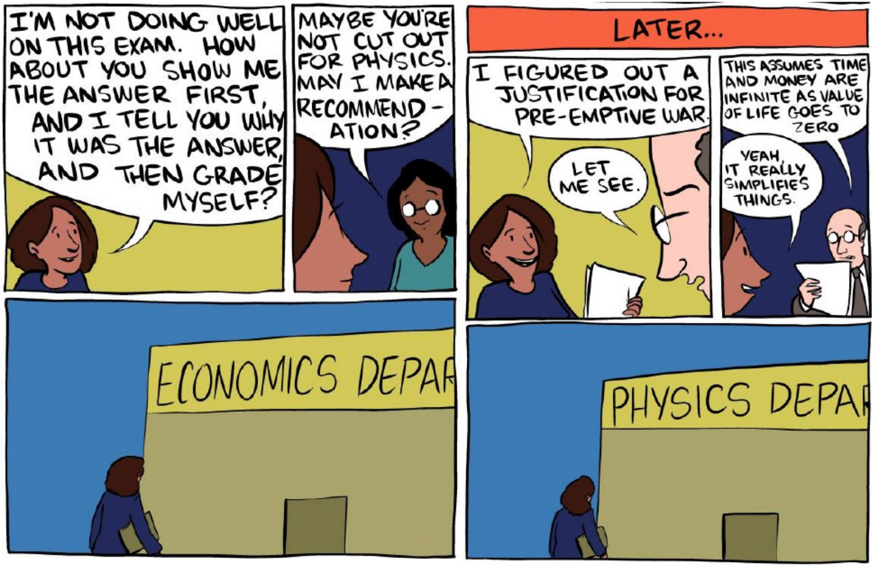
Modeling in Finance and Econometrics from the Econophysics paradigm (A1-S-43514).

- Rotationally invariant estimators on portfolio optimization to unveil financial states. In collaboration with Rodrigo Macías Páez - CIMAT-MTY (`http://dx.doi.org/10.2139/ssrn.4126928`).
- Bias Reduction in Covariance Matrices and their Effects on high-dimensional portfolios. Benito Rodriguez Camejo. Master in Statistical Computing.
- Hierarchical clustering methods for portfolio diversification and capital allocation. Jose Antonio Duarte Mendieta. Master in Statistical Computing.

Conclusions

- The theory of random matrices arises in statistical physics, but is currently of great relevance in data science and especially on high-dimensional statistics.
- Free probability allows to work with more complex structures within the theory of random matrices.
- When the number of assets is of the same order as the number of observations, the investment risk is strongly biased.
- The family of RIE allow us to correctly estimate the sample covariance matrix.
- Applications are being developed in finance from the approach of data science.

Thank you!


I'M NOT DOING WELL ON THIS EXAM. HOW ABOUT YOU SHOW ME THE ANSWER FIRST, AND I TELL YOU WHY IT WAS THE ANSWER, AND THEN GRADE MYSELF?
MAYBE YOU'RE NOT CUT OUT FOR PHYSICS. MAY I MAKE A RECOMMEND-ATION?
ECONOMICS DEPAR
LATER...
I FIGURED OUT A JUSTIFICATION FOR PRE-EMPTIVE WAR.
LET ME SEE.
THIS ASSUMES TIME AND MONEY ARE INFINITE AS VALUE OF LIFE GOES TO ZERO
YEAH, IT REALLY SIMPLIFIES THINGS.
PHYSICS DEPAR