

Departamento de Matemáticas, Facultad de Ciencias
Universidad Autónoma de Madrid

Matrices aleatorias Modelos de Wishart y Wigner

TRABAJO DE FIN DE GRADO

Grado en Matemáticas

Autor: Alejandro Santorum Varela

Tutor: José Luis Fernández Pérez

Curso 2020-2021

Agradecimientos

En primer lugar quiero agradecer a José Luis Fernández, no solo por su entrega, guía y consejos a lo largo de este trabajo, sino también por todo lo que me ha enseñado a lo largo de estos años. También quiero mostrar mi agradecimiento a José Luis Torrea, que ha conseguido que aprenda algo de matemáticas.

Por otro lado, quiero agradecer a mis amigos, sin su ayuda y apoyo no sé donde estaría.

Y por supuesto, gracias a mi familia, por **todo**.

Resumen

La Teoría de Matrices Aleatorias estudia las propiedades de las matrices cuyas entradas son variables aleatorias. Las matrices aleatorias se aplican en diversos campos, destacando estadística, análisis matemático, teoría de números, física y hasta neurociencia.

En este trabajo nos centraremos en estudiar los modelos matriciales aleatorios de Wishart y de Wigner. Las matrices de Wishart surgen como una generalización multivariante de la distribución χ^2 y se analizarán en el capítulo 1, junto con algunas de sus propiedades más importantes. Adicionalmente, la distribución de Wishart nos permitirá formular la versión multivariante del teorema de Fisher-Cochran y definir el estadístico T^2 de Hotelling, usado para generalizar el test χ^2 univariante a varias dimensiones. Al final de este capítulo analizaremos el espectro de autovalores de las matrices de Wishart, presentando la distribución de Marchenko-Pastur.

En el capítulo 2 estudiaremos las matrices aleatorias de Wigner, las cuales conforman uno de los *ensembles* de matrices aleatorias más conocidos. El capítulo se centrará en examinar la Ley Semicircular de Wigner, que describe el espectro de autovalores de estas matrices, así como en demostrar el Teorema de Wigner. Finalmente, introduciremos la Ley de Tracy-Widom, que describe el comportamiento límite del mayor autovalor de las matrices estudiadas.

Abstract

Random Matrix Theory studies the properties of matrices whose entries are random variables. Random matrices are applied in several fields, highlighting statistics, mathematical analysis, number theory, physics and even neuroscience.

In this project we are going to study Wishart and Wigner random matrix models. Wishart matrices arise as a multivariate generalization of χ^2 distribution and it will be discussed in chapter 1, along with some important properties. Additionally, the Wishart distribution will allow us to formulate the multivariate version of Fisher-Cochran theorem and to define the Hotelling T^2 statistic, used to generalize the univariate χ^2 test to several dimensions. At the end of this chapter we will analyze eigenvalue spectrum of Wishart matrices, presenting Marchenko-Pastur distribution.

In chapter 2 we will study Wigner random matrices, which set up one of the best known random matrix ensembles. This chapter will focus on examining Wigner's Semicircular Law, which describes the eigenvalue spectrum of these matrices, as well as proving Wigner's Theorem. Finally, we will introduce Tracy-Widom Law, that describes the limiting behaviour of the largest eigenvalue of the studied matrices.

Índice general

0.1	Notación	1
1	La distribución de Wishart	3
1.1	Estimación paramétrica de la normal multivariante	3
1.2	Introducción a la distribución de Wishart	6
1.3	Propiedades de la distribución de Wishart	6
1.4	Estadístico T^2 de Hotelling	12
1.4.1	Test de una muestra T^2 de Hotelling	16
1.5	Distribución de los autovalores de una matriz de Wishart	18
1.5.1	Ley de Marchenko-Pastur	18
1.5.2	Distribución del mayor autovalor	19
2	Ley semicircular de Wigner	21
2.1	Introducción a la distribución semicircular de Wigner	21
2.1.1	Enunciados del Teorema de Wigner	23
2.2	Momentos de la Ley de Wigner	25
2.3	Grafos y caminos de Dyck	28
2.4	Demostración de la versión 3 del Teorema de Wigner	30
2.5	Demostración de la versión 2 del Teorema de Wigner	36
2.6	Demostración de la versión 1 del Teorema de Wigner	38
2.7	Ley de Tracy-Widom	39
3.1	Matrices simétricas, definidas positivas	43
3.1.1	Resultados sobre diagonalización	43
3.1.2	Funciones características de variables aleatorias con valores en matrices simétricas	44
3.1.3	Normas de matrices simétricas	45
3.2	Desigualdad de Markov	46
3.3	Sobre el ensemble ortogonal gaussiano	46

En este trabajo vamos a estudiar las propiedades de algunas matrices aleatorias. Una matriz aleatoria (por ahora cuadrada) es una matriz cuyas entradas son todas variables aleatorias, es decir, $\mathbf{X} \in \mathcal{M}_n$ es una matriz cuyas entradas $(X_{ij})_{1 \leq i, j \leq n}$ son variables aleatorias de un cierto espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$.

En el capítulo 1 estudiaremos las matrices aleatorias de Wishart, que son matrices definidas positivas con alto interés en estadística; y en el capítulo 2 analizaremos las matrices aleatorias de Wigner, que son matrices simétricas con aplicaciones en matemáticas, física, e incluso, teoría de números.

Adicionalmente, nos interesará el comportamiento de los autovalores de estas matrices aleatorias. Como para cada $\omega \in \Omega$ tenemos una instancia diferente de $\mathbf{X}(\omega)$, sus autovalores $\lambda_1(\omega), \lambda_2(\omega), \dots, \lambda_n(\omega)$ son también variables aleatorias. Buscaremos entender aspectos de las distribuciones de probabilidad de los λ_i a partir de la distribución de los X_{ij} .

0.1. Notación

Primero vamos a indicar la notación que será utilizada. Decimos que $A \in \mathcal{M}_n$, si A es una matriz cuadrada $n \times n$, y $A \in \mathcal{M}_{n \times m}$ si A es una matriz $n \times m$. Adicionalmente, diremos que $A \in \mathcal{S}_n$ si A es una matriz simétrica $n \times n$, y $A \in \mathcal{P}_n$ si A es una matriz (simétrica) definida positiva $n \times n$.

Además, si $A \in \mathcal{M}_n$, entonces $A = (a_{ij})_{1 \leq i, j \leq n}$, donde los a_{ij} son los elementos que forman la matriz A .

CAPÍTULO 1

La distribución de Wishart

Índice

1.1	Estimación paramétrica de la normal multivariante	3
1.2	Introducción a la distribución de Wishart	6
1.3	Propiedades de la distribución de Wishart	6
1.4	Estadístico T^2 de Hotelling	12
1.4.1	Test de una muestra T^2 de Hotelling	16
1.5	Distribución de los autovalores de una matriz de Wishart	18
1.5.1	Ley de Marchenko-Pastur	18
1.5.2	Distribución del mayor autovalor	19

En este capítulo haremos una breve introducción a la estimación de la normal multivariante como en el capítulo 3 de [11], y después nos centraremos en la distribución de Wishart y en el estadístico T^2 de Hotelling ([12] y [7]). Finalmente estudiaremos el espectro de autovalores de las matrices de Wishart tal y como se describe en [8].

1.1. Estimación paramétrica de la normal multivariante

Supongamos que $\mathbf{X} = [\mathbb{X}_1, \dots, \mathbb{X}_n]$ es una muestra aleatoria de tamaño n de una población que sigue una distribución $N_p(\mu, \Sigma)$. Nuestra vida sería muy fácil si los parámetros μ y Σ fueran conocidos, pero, en la mayoría de los casos no disponemos de esa información. ¿Qué podemos hacer entonces? Los siguientes resultados intentan responder a esto.

Teorema 1.1.1 (Estimadores insesgados). *Sea $\bar{\mathbb{X}} = \frac{1}{n} \sum_{i=1}^n \mathbb{X}_i$ la media muestral y $\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})^T$ la cuasivarianza muestral de la población $\mathbf{X} = [\mathbb{X}_1, \dots, \mathbb{X}_n]$. Entonces $\bar{\mathbb{X}}$ es un estimador insesgado de μ , y \mathbf{S} es un estimador insesgado de Σ .*

Demostración. Hay que probar que $\mathbf{E}(\bar{\mathbb{X}}) = \mu$ y que $\mathbf{E}(\mathbf{S}) = \Sigma$. En el primer caso, $\mathbf{E}(\bar{\mathbb{X}}) = \mathbf{E}(\frac{1}{n} \sum_{i=1}^n \mathbb{X}_i) = \frac{1}{n} \sum_{i=1}^n \mathbf{E}(\mathbb{X}_i) = \frac{1}{n} n\mu = \mu$.

En cuanto a la cuasivarianza muestral, $(n-1)\mathbf{S} = \sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})^T = \sum_{i=1}^n \mathbb{X}_i \mathbb{X}_i^T - n\bar{\mathbb{X}}\bar{\mathbb{X}}^T$, lo que implica que $\mathbf{E}[(n-1)\mathbf{S}] = \sum_{i=1}^n \mathbf{E}(\mathbb{X}_i \mathbb{X}_i^T) - n\mathbf{E}(\bar{\mathbb{X}}\bar{\mathbb{X}}^T)$.

Como cálculo intermedio, vemos que $\text{Cov}(\mathbb{X}_i) = \Sigma = \mathbf{E}[(\mathbb{X}_i - \mu)(\mathbb{X}_i - \mu)^T] = \mathbf{E}(\mathbb{X}_i \mathbb{X}_i^T) - \mu\mu^T$, y despejando obtenemos que $\mathbf{E}(\mathbb{X}_i \mathbb{X}_i^T) = \Sigma + \mu\mu^T$.

Por otro lado, $\text{Cov}(\bar{\mathbb{X}}) = \mathbf{E}[(\frac{1}{n} \sum_{i=1}^n (\mathbb{X}_i - \mu))(\frac{1}{n} \sum_{i=1}^n (\mathbb{X}_i - \mu)^T)] = \frac{1}{n^2} n\Sigma = \frac{1}{n}\Sigma$. Pero además, $\text{Cov}(\bar{\mathbb{X}}) = \mathbf{E}[(\bar{\mathbb{X}} - \mu)(\bar{\mathbb{X}} - \mu)^T] = \mathbf{E}(\bar{\mathbb{X}}\bar{\mathbb{X}}^T) - \mu\mu^T$. Por lo que, igualando $\text{Cov}(\bar{\mathbb{X}})$ y despejando, obtenemos que $\mathbf{E}(\bar{\mathbb{X}}\bar{\mathbb{X}}^T) = \frac{1}{n}\Sigma + \mu\mu^T$.

Sustituyendo los dos resultados intermedios en la expresión que teníamos de $\mathbf{E}[(n-1)S]$ llegamos a que $\mathbf{E}[(n-1)S] = \sum_{i=1}^n (\Sigma + \mu\mu^T) - n(\frac{1}{n}\Sigma - \mu\mu^T) = (n-1)\Sigma$. Se sigue entonces que $\mathbf{E}(S) = \Sigma$. \square

Teorema 1.1.2 (Estimadores de máxima verosimilitud). *Sea $\mathbf{X} = [\mathbb{X}_1, \dots, \mathbb{X}_n]$ una muestra aleatoria de tamaño n de una población que sigue una distribución $N_p(\mu, \Sigma)$, con μ y Σ desconocidos. Entonces los estimadores de máxima verosimilitud de μ y Σ son $\hat{\mu}_{EMV} = \bar{\mathbb{X}}$ y $\hat{\Sigma}_{EMV} = \frac{n-1}{n} \mathbf{S} := \mathbf{S}_n = \frac{1}{n} \sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})^T$.*

Demostración. La función de máxima verosimilitud es una función de los parámetros μ y Σ dada la muestra aleatoria \mathbf{X} definida como

$$L(\mu, \Sigma; \mathbf{X}) = \prod_{i=1}^n f(\mathbb{X}_i; \mu, \Sigma). \quad (1.1)$$

La parte derecha de la igualdad es evaluada sustituyendo los vectores $\mathbb{X}_1, \dots, \mathbb{X}_n$ uno por uno en la función de densidad de probabilidad de $N_p(\mu, \Sigma)$ y realizando el producto

$$\prod_{i=1}^n f(\mathbb{X}_i; \mu, \Sigma) = (2\pi)^{-\frac{np}{2}} \det(\Sigma)^{-n/2} \exp \left(-\frac{1}{2} \sum_{i=1}^n (\mathbb{X}_i - \mu)^T \Sigma^{-1} (\mathbb{X}_i - \mu) \right). \quad (1.2)$$

Nos centramos ahora en el desarrollo de

$$\begin{aligned} \sum_{i=1}^n (\mathbb{X}_i - \mu)^T \Sigma^{-1} (\mathbb{X}_i - \mu) &= \sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}} + \bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\mathbb{X}_i - \bar{\mathbb{X}} + \bar{\mathbb{X}} - \mu) = \\ &= \sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})^T \Sigma^{-1} (\mathbb{X}_i - \bar{\mathbb{X}}) + n(\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) + 2 \sum_{i=1}^n (\mathbb{X}_i - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu). \end{aligned} \quad (1.3)$$

Notemos que

$$\sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) = \left(\sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}}) \right)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) = n \underbrace{(\bar{\mathbb{X}}^T - \bar{\mathbb{X}}^T)}_{=0} \Sigma^{-1} (\bar{\mathbb{X}} - \mu) = 0.$$

Por lo tanto

$$\sum_{i=1}^n (\mathbb{X}_i - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) = \sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})^T \Sigma^{-1} (\mathbb{X}_i - \bar{\mathbb{X}}) + n(\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu).$$

Si hacemos un cálculo mental rápido podemos ver que el sumatorio es un escalar, por lo que podemos escribir

$$\sum_{i=1}^n (\mathbb{X}_i - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) = \text{tr} \left(\sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})^T \Sigma^{-1} (\mathbb{X}_i - \bar{\mathbb{X}}) \right) + n(\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu).$$

Usando que la traza de la suma es igual a la suma de las trazas, y la propiedad de la traza $\text{tr}(AB) = \text{tr}(BA)$ para $A \in \mathcal{M}_{m_1 \times m_2}$, $B \in \mathcal{M}_{m_2 \times m_1}$:

$$\begin{aligned} \sum_{i=1}^n (\mathbb{X}_i - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) &= \sum_{i=1}^n \text{tr} ((\mathbb{X}_i - \bar{\mathbb{X}})^T \Sigma^{-1} (\mathbb{X}_i - \bar{\mathbb{X}})) + n(\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) = \\ &= \sum_{i=1}^n \text{tr} (\Sigma^{-1} (\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})^T) + n(\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) = \\ &= \text{tr} \left[\Sigma^{-1} \left(\sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})^T \right) \right] + n(\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) = \\ &= \text{tr} [\Sigma^{-1} ((n-1)\mathbf{S})] + n(\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu). \end{aligned} \quad (1.4)$$

Sustituyendo (1.4) en (1.2) llegamos a una expresión equivalente para la función de máxima verosimilitud

$$\prod_{i=1}^n f(\mathbb{X}_i; \mu, \Sigma) = (2\pi)^{-\frac{np}{2}} \det(\Sigma)^{-n/2} \exp \left(-\frac{n-1}{2} \text{tr}(\Sigma^{-1} \mathbf{S}) - \frac{n}{2} (\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) \right). \quad (1.5)$$

Notemos que para una matriz de covarianzas Σ fijada, la expresión (1.5) es maximizada con respecto a μ si $\exp(-\frac{n}{2} (\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu))$ es maximizado, es decir, si $(\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu)$ es minimizado, ya que es la única parte de la expresión (1.5) en la que interviene μ . Como $\Sigma \in \mathbf{P}_p \implies \Sigma^{-1} \in \mathbf{P}_p$, por lo que $(\bar{\mathbb{X}} - \mu)^T \Sigma^{-1} (\bar{\mathbb{X}} - \mu) \geq 0$, y se anula si y solo si $\bar{\mathbb{X}} = \mu$. Por consiguiente, $\hat{\mu} = \bar{\mathbb{X}}$ maximizaría (1.5) con respecto a μ para Σ fijada, y entonces llegamos a que $\hat{\mu}_{EMV} = \bar{\mathbb{X}}$.

A continuación llamamos $\mathbf{A} = (n-1)\mathbf{S}$, y nos centramos en maximizar

$$L(\hat{\mu}, \Sigma; \mathbf{X}) = \prod_{i=1}^n f(\mathbb{X}_i; \hat{\mu}, \Sigma) = (2\pi)^{-\frac{np}{2}} \det(\Sigma)^{-n/2} \exp \left(-\frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{A}) \right). \quad (1.6)$$

Maximizar $L(\hat{\mu}, \Sigma; \mathbf{X})$ es equivalente a maximizar

$$\log L(\hat{\mu}, \Sigma; \mathbf{X}) = -\frac{np}{2} \log(2\pi) + \frac{n}{2} \log(\det(\Sigma^{-1})) - \frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{A}), \quad (1.7)$$

que a su vez es equivalente a maximizar $\frac{n}{2} \log(\det(\Sigma^{-1})) - \frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{A})$, que es la parte que depende de Σ . Fijémonos además que

$$\frac{n}{2} \log(\det(\Sigma^{-1})) - \frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{A}) = \frac{n}{2} \log(\det(\Sigma^{-1} \mathbf{A})) - \frac{n}{2} \log(\det(\mathbf{A})) - \frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{A}).$$

Igual que antes, eliminamos los términos que no dependan de Σ e intentamos maximizar $\frac{n}{2} \log(\det(\Sigma^{-1} \mathbf{A})) - \frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{A})$.

$\mathbf{A} \in \mathcal{P}_p$ porque $\mathbf{A} = (n-1)\mathbf{S}$ y \mathbf{S} es la varianza muestral de una normal p -dimensional, por lo que es definida positiva con probabilidad 1. Por consiguiente, $\Sigma^{-1} \mathbf{A} \in \mathcal{P}_p$ y es ortogonalizable. Denotemos como λ_j los autovalores de $\Sigma^{-1} \mathbf{A}$ $j = 1, \dots, p$.

$$\begin{aligned} \frac{n}{2} \log(\det(\Sigma^{-1} \mathbf{A})) - \frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{A}) &= \frac{n}{2} \log \left(\prod_{j=1}^p \lambda_j \right) - \frac{1}{2} \sum_{j=1}^p \lambda_j \\ &= \frac{n}{2} \sum_{j=1}^p \log(\lambda_j) - \frac{1}{2} \sum_{j=1}^p \lambda_j = \frac{1}{2} \sum_{j=1}^p (n \log(\lambda_j) - \lambda_j). \end{aligned} \quad (1.8)$$

Despejamos los λ_j que maximizan lo anterior derivando e igualando a cero

$$\frac{d}{d\lambda_j} (1.8) = \frac{1}{2} \left(\sum_{j=1}^p \frac{n}{\lambda_j} - p \right) = 0 \iff \sum_{j=1}^p \frac{n}{\lambda_j} = p \iff \frac{n}{\lambda_j} = 1 \iff \lambda_j = n \quad \forall j = 1, \dots, p. \quad (1.9)$$

Para terminar el argumento, notemos que los λ_j son también los autovalores de $\Sigma^{-1} \mathbf{A} = \Sigma^{-1} \mathbf{A}^{1/2} \mathbf{A}^{1/2} = \mathbf{A}^{1/2} \Sigma^{-1} \mathbf{A}^{1/2}$, donde hemos usado que $\mathbf{A}^{1/2}$ y Σ^{-1} conmutan al ser diagonalizables ($\mathbf{A}^{1/2}, \Sigma^{-1} \in \mathcal{P}_p$). Esto nos permite escribir $\mathbf{A}^{1/2} \Sigma^{-1} \mathbf{A}^{1/2} = O^{-1} n I_p O$ para cierta matriz ortogonal O formada por los autovectores de $\mathbf{A}^{1/2} \Sigma^{-1} \mathbf{A}^{1/2}$. Como O es ortogonal, $O^{-1} n I_p O = n I_p \implies \Sigma^{-1} = n \mathbf{A}^{-1} \implies \Sigma = \frac{1}{n} \mathbf{A} = \frac{n-1}{n} \mathbf{S} \implies \hat{\Sigma}_{EMV} = \frac{n-1}{n} \mathbf{S} = \mathbf{S}_n$.

□

1.2. Introducción a la distribución de Wishart

La distribución de Wishart surge como una generalización matricial de la distribución χ^2 . Si X_1, X_2, \dots, X_n son variables aleatorias independientes con $X_i \sim N(0, 1)$ entonces $X_1^2 + X_2^2 + \dots + X_n^2 \sim \chi_n^2$, donde n indica que la distribución χ^2 tiene n grados de libertad. Cuando los \mathbb{X}_i son vectores aleatorios en lugar de variables aleatorias reales, digamos $\mathbb{X}_i \sim N_p(0, I_p)$ con valores en \mathbb{R}^p , una posible generalización de la suma de cuadrados anterior es formar la matriz $p \times p$ semidefinida positiva $\mathbf{M} = \sum_{i=1}^n \mathbb{X}_i \mathbb{X}_i^T$. Esencialmente, esta representación de \mathbf{M} es usada para definir la distribución estándar de Wishart.

De forma más general, la **distribución de Wishart** es una familia de distribuciones de probabilidad definidas sobre las matrices simétricas semidefinidas positivas. Está caracterizada por dos parámetros básicos: $p \in \mathbb{N}$ y $\Sigma \in \mathcal{P}_p$. Sean $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n$ vectores aleatorios independientes de una normal p -dimensional con matriz de covarianzas $\Sigma \in \mathcal{P}_p$, i.e. $\mathbb{X}_i \sim N_p(0, \Sigma) \forall i \in \{1, \dots, n\}$, y sea $\mathbf{X} \in \mathcal{M}_{p \times n}$ cuyas columnas son los vectores \mathbb{X}_i , $\mathbf{X} = [\mathbb{X}_1, \dots, \mathbb{X}_n]$, entonces se dice que la matriz \mathbf{M} , con $\mathbf{M} = \mathbf{X}\mathbf{X}^T = \sum_{i=1}^n \mathbb{X}_i \mathbb{X}_i^T$ sigue la distribución de Wishart y lo denotamos como $\mathbf{M} \sim W_p(n, \Sigma)$.

$$\begin{aligned} \mathbf{M} = \mathbf{X}\mathbf{X}^T &= \sum_{i=1}^n \mathbb{X}_i \mathbb{X}_i^T = \sum_{i=1}^n \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix} (X_{i1} \ X_{i2} \ \cdots \ X_{ip}) = \\ &= \sum_{i=1}^n \begin{pmatrix} X_{i1}X_{i1} & X_{i1}X_{i2} & \cdots & X_{i1}X_{ip} \\ X_{i2}X_{i1} & X_{i2}X_{i2} & \cdots & X_{i2}X_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ X_{ip}X_{i1} & X_{ip}X_{i2} & \cdots & X_{ip}X_{ip} \end{pmatrix} = \sum_{i=1}^n (X_{ij}X_{ik})_{1 \leq j, k \leq p} \end{aligned} \quad (1.10)$$

1.3. Propiedades de la distribución de Wishart

Presentamos a continuación algunas propiedades importantes de la distribución de Wishart.

1) $\mathbf{M} \sim W_p(n, \Sigma)$ es siempre semidefinida positiva.

Demostración. Sea $0 \neq b \in \mathbb{R}^p$, entonces $b^T \mathbf{M} b = b^T \mathbf{X}\mathbf{X}^T b = (\mathbf{X}^T b)^T (\mathbf{X}^T b) = (\mathbf{X}^T b)^2 \geq 0$. \square

2) Para $\mathbf{M} \sim W_p(n, \Sigma)$ y $B \in \mathcal{M}_{p \times m}$, tenemos

$$B^T \mathbf{M} B \sim W_m(n, B^T \Sigma B). \quad (1.11)$$

Demostración. Como $\mathbf{M} \sim W_p(n, \Sigma)$, entonces $\mathbf{M} = \mathbf{X}\mathbf{X}^T = \sum_{i=1}^n \mathbb{X}_i \mathbb{X}_i^T$ para ciertos vectores aleatorios $\mathbb{X}_i \sim N_p(0, \Sigma)$. Ahora multiplicamos por B^T y B y desarrollamos \mathbf{M} usando la definición de la distribución de Wishart: $B^T \mathbf{M} B = B^T \mathbf{X}\mathbf{X}^T B = (\mathbf{X}^T B)^T (\mathbf{X}^T B) = \mathbf{Y}\mathbf{Y}^T = \sum_{i=1}^n \mathbb{Y}_i \mathbb{Y}_i^T$ donde cada \mathbb{Y}_i es un vector aleatorio tal que $\mathbb{Y}_i \sim N_m(0, B^T \Sigma B)$. Finalmente, por la definición de la distribución de Wishart llegamos a que $B^T \mathbf{M} B = \mathbf{Y}\mathbf{Y}^T \sim W_m(n, B^T \Sigma B)$. \square

3) Para $\mathbf{M} \sim W_p(n, \Sigma)$ con Σ matriz $p \times p$ de covarianzas definida positiva, tenemos que

$$\Sigma^{-\frac{1}{2}} \mathbf{M} \Sigma^{-\frac{1}{2}} \sim W_p(n, I_p). \quad (1.12)$$

Para la demostración de esta propiedad será útil el siguiente lema. En él se hace referencia a la matriz $\Sigma^{-\frac{1}{2}}$. Como por definición $\Sigma \in \mathcal{P}_p$, entonces es ortogonalizable tal que $\Sigma = O^{-1} \Lambda O$ con O matriz ortogonal formada por los autovectores de Σ y Λ matriz diagonal con sus autovalores. Por lo tanto, cuando hablamos de $\Sigma^{-\frac{1}{2}}$, hablamos de $O^{-1} \Lambda^{-\frac{1}{2}} O$, donde $\Lambda^{-\frac{1}{2}}$ es la matriz diagonal de la raíz cuadrada positiva de los inversos de los autovalores de Σ . Para mayor nivel de detalle visitar (3.1.6) en el Apéndice.

Lema 1.3.1. Si $\Sigma \in \mathcal{P}_n$, entonces $\Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} = I_n$.

Demostración. Como $\Sigma \in \mathcal{P}_n$, es ortogonalizable, es decir, podemos escribir $\Sigma = O^{-1} \Lambda O$ para cierta matriz O ortogonal e invertible cuyas columnas son los autovectores de Σ , y cierta matriz Λ diagonal formada por los autovalores de Σ . Usando esto y el hecho de que $\Sigma^{-\frac{1}{2}} = (O^{-1} \Lambda O)^{-\frac{1}{2}} = O^{-1} \Lambda^{-\frac{1}{2}} O$:

$$\Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} = O^{-1} \Lambda^{-\frac{1}{2}} O O^{-1} \Lambda O O^{-1} \Lambda^{-\frac{1}{2}} O = O^{-1} \Lambda^{-\frac{1}{2}} \Lambda \Lambda^{-\frac{1}{2}} O.$$

Finalmente

$$O^{-1} \Lambda^{-\frac{1}{2}} \Lambda \Lambda^{-\frac{1}{2}} O = O^{-1} I_n O = O^{-1} O = O^T O = I_n,$$

donde en el último paso hemos usado la definición de matriz ortogonal $O^{-1} = O^T$.

□

Probamos ahora la propiedad (1.12).

Demostración. $\Sigma \in \mathcal{P}_p$ por definición, por lo que $\Sigma = \Sigma^T$. Esto, junto la propiedad (1.11) anteriormente demostrada podemos operar de la siguiente forma: $(\Sigma^T)^{-\frac{1}{2}} \mathbf{M} \Sigma^{-\frac{1}{2}} = \Sigma^{-\frac{1}{2}} \mathbf{M} \Sigma^{-\frac{1}{2}} \sim W_p(n, \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}})$.

Para terminar solo nos queda ver que $\Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} = I_p$, pero esto se sigue directamente del lema (1.3.1). □

4) (*Función de densidad de probabilidad*): Para $n \geq p$, la función de densidad de probabilidad de $\mathbf{M} \sim W_p(n, \Sigma)$ es:

$$f(\mathbf{M}) = \frac{1}{2^{np/2} \det(\Sigma)^{n/2} \prod_{k=1}^p \Gamma((n-k+1)/2)} \det(\mathbf{M})^{(n-p-1)/2} \exp\left(-\frac{1}{2} \text{traza}(\Sigma^{-1} \mathbf{M})\right). \quad (1.13)$$

Se puede encontrar una prueba concisa de la derivación de $f(\mathbf{M})$ en [6].

Adicionalmente, recalando el hecho de que la distribución de Wishart es una extensión multivariante de la distribución χ^2 , podemos analizar el caso en que $p = 1$ y $\sigma^2 = 1$, por lo que $\mathbf{M} \sim W_1(n, 1)$ es un real positivo (digamos x) y entonces su función de densidad de probabilidad es

$$f(x) = \frac{1}{2^{n/2} \Gamma(n/2)} x^{(n/2)-1} \exp(-x/2), \quad (1.14)$$

que coincide con la función de densidad de probabilidad de χ_n^2 . Para σ^2 genérico podemos ver que $\sigma^2 \chi_n^2 \sim W_1(n, \sigma^2) \implies \chi_n^2 \sim \frac{1}{\sigma^2} W_1(n, \sigma^2)$.

5) Si $\mathbf{M} \sim W_p(n, \Sigma)$ y $a \in \mathbb{R}^p$ con $a^T \mathbf{M} a \neq 0$, entonces

$$\frac{a^T \mathbf{M} a}{a^T \Sigma a} \sim \chi_n^2. \quad (1.15)$$

Demostración. Usando la propiedad (1.11) con $B = a$ tenemos que $a^T \mathbf{M} a \sim W_1(n, a^T \Sigma a)$. Recordando que la distribución de Wishart con $p = 1$ es una χ^2 , junto con el resultado (1.14), llegamos a que

$$\chi_n^2 \sim \frac{1}{a^T \Sigma a} a^T \mathbf{M} a = \frac{a^T \mathbf{M} a}{a^T \Sigma a}.$$

□

6) (*Función característica*): La noción de función característica puede ser extendida a matrices aleatorias. Sea $\mathbf{A} \in \mathcal{M}_p$, entonces la función característica de \mathbf{A} evaluada en una matriz $\mathbf{T} \in \mathcal{S}_p$ está definida como

$$\varphi_{\mathbf{A}}(\mathbf{T}) = \mathbf{E}[\exp(\text{traza}(\mathbf{A}\mathbf{T}))]. \quad (1.16)$$

Si ahora $\mathbf{M} \sim W_p(n, \Sigma)$, entonces

$$\varphi_{\mathbf{M}}(\mathbf{T}) = \det(I_p - 2i\Sigma\mathbf{T})^{-n/2}. \quad (1.17)$$

Demostración. La prueba de esta propiedad la podemos encontrar en el Apéndice en (3.13), donde hablamos de las funciones características de variables aleatorias con valores en matrices simétricas. □

7) Si \mathbf{M}_i son independientes $W_p(n_i, \Sigma)$ para $i \in \{1, \dots, k\}$, entonces tenemos que

$$\sum_{i=1}^k \mathbf{M}_i \sim W_p(n, \Sigma), \quad (1.18)$$

donde $n = n_1 + n_2 + \dots + n_k$.

Demostración. Basta demostrarlo para el caso $k = 2$ porque el caso general cae por si solo con un argumento de inducción.

Para $k = 2$ tenemos $\mathbf{M}_1 \sim W_p(n_1, \Sigma)$ y $\mathbf{M}_2 \sim W_p(n_2, \Sigma)$. Recordemos que $\mathbf{M}_1 = \mathbf{X}\mathbf{X}^T$ con $\mathbf{X} = [\mathbb{X}_1, \dots, \mathbb{X}_{n_1}]$, y $\mathbf{M}_2 = \mathbf{Y}\mathbf{Y}^T$ con $\mathbf{Y} = [\mathbb{Y}_1, \dots, \mathbb{Y}_{n_2}]$, donde los vectores $\mathbb{X}_i \sim N_p(0, \Sigma)$ $\forall i \in \{1, 2, \dots, n_1\}$ y los $\mathbb{Y}_j \sim N_p(0, \Sigma)$ $\forall j \in \{1, 2, \dots, n_2\}$. Si ahora calculamos $\mathbf{M}_1 + \mathbf{M}_2$ observamos:

$$\mathbf{M}_1 + \mathbf{M}_2 = \mathbf{X}\mathbf{X}^T + \mathbf{Y}\mathbf{Y}^T = \sum_{i=1}^{n_1} (X_{ij}X_{ik})_{1 \leq j, k \leq p} + \sum_{i=1}^{n_2} (Y_{ij}Y_{ik})_{1 \leq j, k \leq p} \quad (1.19)$$

donde en el último paso hemos usado el resultado (1.10). Ahora podemos juntar los dos sumatorios de (1.19) debido a que ambas matrices son del mismo tamaño ($\mathbf{M}_1, \mathbf{M}_2 \in \mathcal{M}_p$):

$$\mathbf{M}_1 + \mathbf{M}_2 = \left(\sum_{i=1}^{n_1} X_{ij}X_{ik} + \sum_{i=1}^{n_2} Y_{ij}Y_{ik} \right)_{1 \leq j, k \leq p} \quad (1.20)$$

Por construcción de la distribución de Wishart, \mathbb{X}_i y \mathbb{Y}_j son independientes $\forall i \in \{1, \dots, n_1\}$ $\forall j \in \{1, \dots, n_2\}$. Juntando la expresión (1.20) con la (1.10) y la definición de distribución de Wishart llegamos a que $\mathbf{M}_1 + \mathbf{M}_2 \sim W_p(n_1 + n_2, \Sigma)$.

Por último, por un argumento de inducción, podemos suponer cierta la propiedad para el caso $k - 1$, y entonces $\mathbf{M}' = \sum_{i=1}^{k-1} \mathbf{M}_i \sim W_p(n', \Sigma)$ con $n' = n_1 + n_2 + \dots + n_{k-1}$. Finalmente, $\mathbf{M}' + \mathbf{M}_k$ es $W_p(n' + n_k, \Sigma)$ por el caso base $k = 2$, llegando a que $\mathbf{M}' + \mathbf{M}_k = \sum_{i=1}^k \mathbf{M}_i \sim W_p(n, \Sigma)$ con $n = n_1 + n_2 + \dots + n_k$. □

8) Para $\mathbf{M} \sim W_p(n, \Sigma)$,

$$\mathbf{E}(\mathbf{M}) = n\Sigma. \quad (1.21)$$

Demostración. Teniendo en cuenta la definición de distribución de Wishart y la linealidad de la media con respecto a la suma

$$\mathbf{E}(\mathbf{M}) = \mathbf{E} \left(\sum_{i=1}^n (X_{ij} X_{ik})_{1 \leq j, k \leq p} \right) = \sum_{i=1}^n \mathbf{E} \left((X_{ij} X_{ik})_{1 \leq j, k \leq p} \right) = \sum_{i=1}^n \Sigma = n\Sigma.$$

□

9) Si \mathbf{M}_1 y \mathbf{M}_2 son independientes y satisfacen $\mathbf{M}_1 + \mathbf{M}_2 = \mathbf{M} \sim W_p(n, \Sigma)$, con $\mathbf{M}_1 \sim W_p(n_1, \Sigma)$, entonces

$$\mathbf{M}_2 \sim W_p(n - n_1, \Sigma). \quad (1.22)$$

Para la demostración de esta propiedad serán útiles los dos siguientes lemas.

Lema 1.3.2. Si $Z \sim N(0, \sigma^2)$, $X \sim N(0, \tau^2)$ y $Z = X + Y$ con X e Y independientes, entonces $Y \sim N(0, \sigma^2 - \tau^2)$.

Demostración. Recordemos que la función característica de una normal estándar es $\varphi_{N(0,1)}(t) = e^{-t^2/2}$. Por consiguiente, la función característica de una variable aleatoria $W \sim N(0, \eta^2)$ es $\varphi_W(t) = \mathbf{E}(e^{-iWt}) = \mathbf{E}(e^{-\frac{iWt\eta}{\eta}}) = \mathbf{E}(e^{-iN(0,1)t\eta}) = e^{-\frac{\eta^2 t^2}{2}}$. Como $Z = X + Y$, entonces $\varphi_Z(t) = \varphi_X(t)\varphi_Y(t)$. Usando que $Z \sim N(0, \sigma^2) \implies \varphi_Z(t) = e^{-\frac{\sigma^2 t^2}{2}}$, y que $X \sim N(0, \tau^2) \implies \varphi_X(t) = e^{-\frac{\tau^2 t^2}{2}}$, podemos despejar $\varphi_Y(t)$

$$\varphi_Z(t) = \varphi_X(t)\varphi_Y(t) \implies \varphi_Y(t) = \frac{e^{-\frac{\sigma^2 t^2}{2}}}{e^{-\frac{\tau^2 t^2}{2}}} = e^{-\frac{t^2(\sigma^2 - \tau^2)}{2}}.$$

Sabemos que si las funciones características de dos variables aleatorias coinciden, entonces tienen la misma distribución. Por lo tanto, podemos sostener que $Y \sim N(0, \sigma^2 - \tau^2)$. □

Lema 1.3.3. Si $Z \sim \chi_k^2$, $X \sim \chi_q^2$ y $Z = X + Y$ con X e Y independientes, entonces $Y \sim \chi_{k-q}^2$.

Demostración. Recordemos que la función característica de χ_m^2 es $\varphi_{\chi_m^2}(t) = (1 - 2it)^{-m/2}$. Como $Z = X + Y$, entonces $\varphi_Z(t) = \varphi_X(t)\varphi_Y(t)$. Usando que $Z \sim \chi_k^2 \implies \varphi_Z(t) = (1 - 2it)^{-k/2}$, y que $X \sim \chi_q^2 \implies \varphi_X(t) = (1 - 2it)^{-q/2}$, podemos despejar $\varphi_Y(t)$

$$\varphi_Z(t) = \varphi_X(t)\varphi_Y(t) \implies \varphi_Y(t) = \frac{(1 - 2it)^{-k/2}}{(1 - 2it)^{-q/2}} = (1 - 2it)^{-(k-q)/2}.$$

Sabemos que si las funciones características de dos variables aleatorias coinciden, entonces tienen la misma distribución. Por lo tanto, podemos sostener que $Y \sim \chi_{k-q}^2$. □

La idea de la prueba de la propiedad (1.22) es la misma que la de los lemas anteriores.

Demostración. Gracias a la propiedad (1.17) sabemos que la función característica de $W_p(m, \Sigma)$ es $\varphi_{\mathbf{M}}(\mathbf{T}) = \det(I_p - 2i\Sigma\mathbf{T})^{-m/2}$.

Como $\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2$, entonces $\varphi_{\mathbf{M}}(\mathbf{T}) = \varphi_{\mathbf{M}_1}(\mathbf{T})\varphi_{\mathbf{M}_2}(\mathbf{T})$. Usando que $\mathbf{M} \sim W_p(n, \Sigma) \implies$

$\varphi_{\mathbf{M}}(\mathbf{T}) = \det(I_p - 2i\Sigma\mathbf{T})^{-n/2}$, y que $\mathbf{M}_1 \sim W_p(n_1, \Sigma) \implies \varphi_{\mathbf{M}_1}(\mathbf{T}) = \det(I_p - 2i\Sigma\mathbf{T})^{-n_1/2}$, podemos despejar $\varphi_{\mathbf{M}_2}(\mathbf{T})$

$$\varphi_{\mathbf{M}}(\mathbf{T}) = \varphi_{\mathbf{M}_1}(\mathbf{T})\varphi_{\mathbf{M}_2}(\mathbf{T}) \implies \varphi_{\mathbf{M}_2}(\mathbf{T}) = \frac{\det(I_p - 2i\Sigma\mathbf{T})^{-n/2}}{\det(I_p - 2i\Sigma\mathbf{T})^{-n_1/2}} = \det(I_p - 2i\Sigma\mathbf{T})^{-(n-n_1)/2}.$$

Sabemos que si las funciones características de dos matrices aleatorias coinciden, entonces tienen la misma distribución. Por lo tanto, podemos sostener que $\mathbf{M}_2 \sim W_p(n - n_1, \Sigma)$. \square

10) Sea $\mathbf{M} \sim W_p(n, I_p)$,

$$\text{Las entradas de la diagonal de } \mathbf{M} \text{ son } \chi_n^2. \text{ Adicionalmente, } \text{tr}(\mathbf{M}) \sim \chi_{mp}^2. \quad (1.23)$$

Demostración. Desarrollando \mathbf{M} como en (1.10):

$$\mathbf{M} = \mathbf{X}\mathbf{X}^T = \sum_{i=1}^n \mathbb{X}_i \mathbb{X}_i^T = \sum_{i=1}^n (X_{ij}X_{ik})_{1 \leq j, k \leq p}$$

Las entradas de la diagonal se obtienen cuando $j = k$, es decir, de la forma $\sum_{i=1}^n X_{ij}X_{ij} = \sum_{i=1}^n X_{ij}^2$ con $X_{ij} \sim N(0, 1)$, por lo que cada entrada de la diagonal es una χ_n^2 .

Por otro lado,

$$\text{tr}(\mathbf{M}) = \text{tr} \left(\sum_{i=1}^n \mathbb{X}_i \mathbb{X}_i^T \right) = \sum_{i=1}^n \text{tr}(\mathbb{X}_i \mathbb{X}_i^T) = \sum_{i=1}^n \text{tr}(\mathbb{X}_i^T \mathbb{X}_i) = \sum_{i=1}^n \mathbb{X}_i^T \mathbb{X}_i.$$

Recordando que $\mathbb{X}_i \sim N_p(0, I_p)$, notamos que $\mathbb{X}_i^T \mathbb{X}_i \sim \chi_p^2$, por lo que la traza de \mathbf{M} es la suma de n copias independientes de una χ_p^2 , y entonces tenemos que $\text{tr}(\mathbf{M}) \sim \chi_{np}^2$. \square

11) Sea $\mathbf{M} \sim W_p(n, \Sigma)$. Si $n \geq p$, entonces

$$\mathbf{M} \in \mathcal{P}_p \text{ casi seguramente.} \quad (1.24)$$

Demostración. Sabemos que \mathbf{M} es siempre semidefinida positiva. El extra para asegurar que es definida positiva radica en ver que si $n \geq p$, entonces \mathbf{M} es invertible casi seguro, ya que si es invertible su determinante es no nulo y por consiguiente, todos sus autovalores son positivos.

Veamos primero que $\mathbf{P}[\det(\mathbf{M}) = 0] = 0$.

Lema 1.3.4. Si $\mathbf{M} \sim W_p(n, \Sigma)$ con Σ definida positiva, entonces \mathbf{M} es invertible casi seguro cuando $n \geq p$.

Demostración. Lo probaremos usando un argumento de inducción sobre n . Si $n = 1$, $p = 1$ porque estábamos trabajando bajo la suposición de que $n \geq p \geq 1$. Entonces \mathbf{M} es XX^T con $X \sim N(0, \sigma^2)$, es decir, $\mathbf{M} = X^2 = \chi_1^2$. Por lo tanto, $\mathbf{P}(|\mathbf{M}| = 0) = \mathbf{P}(\chi_1^2 = 0) = 0$ porque la función de densidad de χ^2 es continua.

Ahora, para $n > 1$ asumimos que el resultado es cierto para $n - 1$. Escribimos \mathbf{M} de la siguiente forma

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix},$$

donde $\mathbf{M}_{22} \in \mathcal{M}_{n-1}$. Notemos que por hipótesis de inducción $\det(\mathbf{M}_{22}) \neq 0$ casi seguro, y por lo tanto invertible. Usando la fórmula del determinante de Schur¹ obtenemos que $\det(\mathbf{M}) = \det(\mathbf{M}_{22})\det(\mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21}) = \det(\mathbf{M}_{22})(\mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21})$. Sabiendo esto podemos continuar desarrollando $\mathbf{P}(|\mathbf{M}| = 0)$

$$\begin{aligned} \mathbf{P}(|\mathbf{M}| = 0) &\stackrel{(i)}{=} \mathbf{P}(|\mathbf{M}| = 0, |\mathbf{M}_{22}| \neq 0) + \mathbf{P}(|\mathbf{M}| = 0, |\mathbf{M}_{22}| = 0) = \\ &= \mathbf{P}(|\mathbf{M}| = 0, |\mathbf{M}_{22}| \neq 0) \stackrel{(ii)}{=} \mathbf{P}(\mathbf{M}_{11} = \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21}, |\mathbf{M}_{22}| \neq 0) \stackrel{(iii)}{=} \\ &\stackrel{(iii)}{=} \mathbb{E}[\mathbb{P}(\mathbf{M}_{11} = \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21}, |\mathbf{M}_{22}| \neq 0 \mid \mathbf{M}_{12}, \mathbf{M}_{22}, \mathbf{M}_{21})] \stackrel{(iv)}{=} \mathbb{E}(0) = 0. \end{aligned} \quad (1.25)$$

Donde en (i) hemos usado la ley de la probabilidad total, en (ii) la fórmula de Schur y en (iii) la ley de la esperanza total. Finalmente, en (iv) hemos tenido en cuenta que \mathbf{M}_{12} , \mathbf{M}_{22} y \mathbf{M}_{21} están fijos, por lo que $\mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21}$ es un cierto número y, análogamente al caso $n = 1$, la probabilidad que $\mathbf{M}_{11} \sim \chi_n^2$ sea igual a un cierto número es cero ya que la función de densidad de χ^2 es continua. Por consiguiente, el resultado se satisface por inducción. \square

Ahora continuamos probando el resultado principal. Notemos que $\text{rang}(\mathbf{M}) = \text{rang}(\mathbf{X}\mathbf{X}^T) = \text{rang}(\mathbf{X})$. Además, podemos escribir $\Sigma = \mathbf{L}\mathbf{L}^T$ usando la descomposición de Cholesky, y entonces tenemos $\mathbf{Y} = \mathbf{X}(\mathbf{L}^{-1})^T$, donde las columnas \mathbb{Y}_i de \mathbf{Y} son $N_p(0, I_p)$ y $\text{rang}(\mathbf{Y}) = \text{rang}(\mathbf{X})$. Finalmente, tenemos que $\text{rang}(\mathbf{M}) = \text{rang}(\mathbf{Y}) \geq \text{rang}(\mathbb{Y}_1^f, \mathbb{Y}_2^f, \dots, \mathbb{Y}_p^f) = p$ (c.s.) donde \mathbb{Y}_i^f son las filas de la matriz \mathbf{Y} . La última igualdad se debe a nuestro lema intermedio (1.3.4). Como $\text{rang}(\mathbf{M}) = p$ casi seguro, entonces es invertible casi seguro, y por lo tanto, definida positiva casi seguro si $n \geq p$. \square

12) Sea \mathbf{M}_n una sucesión de matrices de Wishart, $\mathbf{M}_n \sim W_p(n, \Sigma)$. Entonces

$$\mathbf{M}_n/n \rightarrow \Sigma \text{ en probabilidad.} \quad (1.26)$$

El siguiente resultado nos será de ayuda para la demostración, debido a que muestra un camino para probar convergencia en distribución multivariante usando resultados de convergencia en distribución univariantes.

Teorema 1.3.5 (Teorema de Cramér-Wold). *Sea \mathbb{X}_n una sucesión de vectores aleatorios en \mathbb{R}^k . Entonces \mathbb{X}_n converge en distribución a \mathbb{X} cuando $n \rightarrow \infty$ si y solo si $a^T \mathbb{X}_n$ converge en distribución a $a^T \mathbb{X}$ cuando $n \rightarrow \infty$ para todo $a \in \mathbb{R}^k$*

Demostración. Si \mathbb{X}_n converge en distribución a \mathbb{X} cuando $n \rightarrow \infty$, entonces la proyección mediante $a \in \mathbb{R}^k$ también lo hará: $a^T \mathbb{X}_n$ convergerá en distribución a $a^T \mathbb{X}$ cuando $n \rightarrow \infty$. En cuanto a la otra implicación, usando el teorema de continuidad de Levy, convergencia en distribución ocurre si las funciones características convergen. Esto se cumple porque para todo $a \in \mathbb{R}^k$,

$$\varphi_{\mathbb{X}_n}(a) = \mathbf{E}[\exp(ia^T \mathbb{X}_n)] = \varphi_{a^T \mathbb{X}_n}(1) \rightarrow \varphi_{a^T \mathbb{X}}(1) = \varphi_{\mathbb{X}}(a).$$

\square

Ahora proseguimos con la prueba de la propiedad (1.26).

Demostración de la propiedad 12. Recordemos de nuevo el desarrollo (1.10)

$$\mathbf{M}_n/n = \frac{1}{n} \sum_{i=1}^n (X_{ij}X_{ik})_{1 \leq j, k \leq p}$$

¹Para mayor información visitar (3.1.8) en el Apéndice

La expresión anterior permite observar que lo que está ocurriendo coordenada a coordenada es el promedio de las muestras $X_{ij}X_{ik}$ para $j, k \in \{1, \dots, p\}$. La Ley débil de los Grandes Números nos dice que la media muestral tiende en probabilidad al valor esperando cuando $n \rightarrow \infty$.

Por otro lado, el teorema de Cramér-Wold (1.3.5) nos permite utilizar los resultados univariantes a nuestro problema actual, por lo que, junto a la Ley débil de los Grandes Números y la propiedad (1.21) de las matrices de Wishart, podemos asegurar que $\mathbf{M}_n/n \rightarrow \Sigma$ en distribución.

Convergencia en distribución a una constante implica convergencia en probabilidad a esa misma constante², por lo que $\mathbf{M}_n/n \rightarrow \Sigma$ en probabilidad, recordando que Σ es una matriz constante una vez fijada la distribución $N(0, \Sigma)$ de los vectores que conforman la matriz de Wishart. \square

1.4. Estadístico T^2 de Hotelling

El **estadístico T^2 de Hotelling** (Hotelling 1931 [7]) juega un papel similar en el análisis multivariante al estadístico t de Student en el análisis estadístico univariante.

La aplicación del estadístico T^2 de Hotelling es de gran importancia a la hora de comprobar la hipótesis nula sobre la media de una distribución normal multidimensional cuando la matriz de covarianzas es desconocida.

Definición 1.4.1 (Estadístico T^2 de Hotelling). *Supongamos que \mathbb{X} es un vector aleatorio y \mathbf{M} una matriz aleatoria tales que \mathbb{X} y \mathbf{M} son independientes y $\mathbb{X} \sim N_p(\mu, \Sigma)$ y $\mathbf{M} \sim W_p(n, \Sigma)$, entonces*

$$T_p^2(n) = n(\mathbb{X} - \mu)^T \mathbf{M}^{-1}(\mathbb{X} - \mu), \quad (1.27)$$

es conocido como el estadístico T^2 de Hotelling, el cual es una variable aleatoria que toma valores en los reales.

Si $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n$ son i.i.d. $N_p(\mu, \Sigma)$, entonces denotamos la media muestral y la cuasivarianza muestral como

$$\begin{aligned} \bar{\mathbb{X}} &= \frac{1}{n} \sum_{i=1}^n \mathbb{X}_i \\ \mathbf{S} &= \frac{1}{n-1} \sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})^T. \end{aligned}$$

Teorema 1.4.2 (versión multivariante de Fisher-Cochran). *$\bar{\mathbb{X}}$ y \mathbf{S} son independientes con*

$$\begin{aligned} \sqrt{n}(\bar{\mathbb{X}} - \mu) &\sim N_p(0, \Sigma), \\ (n-1)\mathbf{S} &\sim W_p(n-1, \Sigma). \end{aligned} \quad (1.28)$$

Antes de mostrar la prueba del teorema, observemos que efectivamente para el caso $p = 1$ tenemos la versión univariante del teorema de Fisher-Cochran. Si $p = 1$, X es una variable

²El recíproco, por lo general, no es cierto

aleatoria $N(\mu, \sigma^2)$ y tenemos que $\bar{X} \sim N(\mu, \sigma^2/n)$, por lo que $\sqrt{n}(X - \mu) \sim N(0, \sigma^2)$. Adicionalmente, la χ^2 juega el papel unidimensional de la distribución de Wishart, por lo que $(n-1)S/\sigma^2 \sim \chi^2_{n-1}$. Finalmente, Fisher-Cochran asegura también que \bar{X} y S son independientes.

Demostración. Primero notemos que podemos suponer que $\mathbb{X}_i \sim N_p(0, I_p)$. Consideremos el vector aleatorio $\mathbb{Z} = \Sigma^{-\frac{1}{2}}(\mathbb{X} - \mu) \sim N_p(0, \Sigma)$ y la muestra $\mathbb{Z}_1, \dots, \mathbb{Z}_n$, que son copias independientes de \mathbb{Z} . Obsérvese que la media muestral de \mathbb{Z} es

$$\bar{\mathbb{Z}} = \frac{1}{n} \sum_{i=1}^n \mathbb{Z}_i = \frac{1}{n} \sum_{i=1}^n \Sigma^{-\frac{1}{2}}(\mathbb{X}_i - \mu) = \Sigma^{-\frac{1}{2}}(\bar{\mathbb{X}} - \mu) \implies \bar{\mathbb{Z}} = \Sigma^{-\frac{1}{2}}(\bar{\mathbb{X}} - \mu), \quad (1.29)$$

y la cuasivarianza muestral de \mathbb{Z}

$$\begin{aligned} (n-1)\mathbf{S}_{\mathbb{Z}} &= \sum_{i=1}^n (\mathbb{Z}_i - \bar{\mathbb{Z}})(\mathbb{Z}_i - \bar{\mathbb{Z}})^T = \sum_{i=1}^n \left(\Sigma^{-\frac{1}{2}}(\mathbb{X}_i - \mu - \bar{\mathbb{X}} + \mu) \right) \left(\Sigma^{-\frac{1}{2}}(\mathbb{X}_i - \mu - \bar{\mathbb{X}} + \mu) \right)^T = \\ &= \Sigma^{-\frac{1}{2}} \left[\sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})^T \right] \Sigma^{-\frac{1}{2}} = \Sigma^{-\frac{1}{2}}(n-1)\mathbf{S}_{\mathbb{X}}\Sigma^{-\frac{1}{2}} \Rightarrow (n-1)\mathbf{S}_{\mathbb{Z}} = (n-1)\Sigma^{-\frac{1}{2}}\mathbf{S}_{\mathbb{X}}\Sigma^{-\frac{1}{2}}. \end{aligned} \quad (1.30)$$

La independencia de $\bar{\mathbb{Z}}$ y $\mathbf{S}_{\mathbb{Z}}$ nos daría la independencia de $\bar{\mathbb{X}}$ y $\mathbf{S}_{\mathbb{X}}$. El que $\sqrt{n}\bar{\mathbb{Z}}$ siguiera una $N_p(0, I_p)$ nos llevaría a que $\sqrt{n}\bar{\mathbb{X}} \sim N_p(\mu, \Sigma)$, es decir, $\sqrt{n}(\bar{\mathbb{X}} - \mu) \sim N_p(0, \Sigma)$; y si $(n-1)\mathbf{S}_{\mathbb{Z}} \sim W_p(n-1, I_p)$, entonces $(n-1)\mathbf{S}_{\mathbb{X}} \sim W_p(n-1, \Sigma^{\frac{1}{2}}I_p\Sigma^{\frac{1}{2}}) = W_p(n-1, \Sigma)$ por la propiedad (1.11) de Wishart. Así que, como afirmábamos, basta probar el teorema para $\mu = 0$ y $\Sigma = I_p$.

Digamos entonces que $\mathbb{X}_1, \dots, \mathbb{X}_n$ es una muestra i.i.d. con distribución $N_p(0, I_p)$. Consideremos ahora unos nuevos vectores aleatorios $\mathbb{Y}_1, \dots, \mathbb{Y}_n$ provenientes de una combinación lineal de los \mathbb{X}_i 's por una matriz ortogonal \mathbf{D} que satisface

$$\mathbf{D} = [d_1, d_2, \dots, d_n], \quad d_1 = \frac{1}{\sqrt{n}}\mathbf{1}_n, \quad \mathbf{D}\mathbf{D}^T = \mathbf{D}^T\mathbf{D} = I_n,$$

$$\mathbb{Y}_j = \sum_{i=1}^n d_{ji}\mathbb{X}_i$$

donde d_1, \dots, d_n son vectores columna n -dimensionales.

En primer lugar, veamos que $\mathbb{Y}_j \sim N_p(0, I_p) \quad \forall j \in \{1, \dots, n\}$. Para probarlo vamos a comprobar que

$$\mathbf{E}(\mathbb{Y}_j) = 0 \quad \forall j \in \{1, \dots, n\}$$

$$\mathbf{E}(\mathbb{Y}_j \mathbb{Y}_k^T) = \begin{cases} I_p & \text{si } j = k \\ 0 & \text{si } j \neq k \end{cases} \quad (1.31)$$

Vemos primero que $\mathbf{E}(\mathbb{Y}_j) = 0$,

$$\mathbf{E}(\mathbb{Y}_j) = \mathbf{E} \left(\sum_{i=1}^n d_{ji}\mathbb{X}_i \right) \stackrel{*}{=} \sum_{i=1}^n \mathbf{E}(d_{ji}\mathbb{X}_i) \stackrel{**}{=} \sum_{i=1}^n d_{ji} \underbrace{\mathbf{E}(\mathbb{X}_i)}_{=0} = 0,$$

donde en $*$ y en $**$ hemos usado la linealidad de la esperanza.

Continuamos probando la afirmación desarrollando $\mathbf{E}(\mathbb{Y}_j \mathbb{Y}_k^T)$:

- Caso 1: $j = k$:

$\mathbf{E}(\mathbb{Y}_j \mathbb{Y}_j^T) = \mathbf{E} \left[\left(\sum_{i=1}^n d_{ji} \mathbb{X}_i \right) \left(\sum_{l=1}^n d_{jl} \mathbb{X}_l \right)^T \right] = \mathbf{E} \left(\sum_{i=1}^n \sum_{l=1}^n d_{ji} d_{jl} \mathbb{X}_i \mathbb{X}_l^T \right) = \sum_{i=1}^n \sum_{l=1}^n d_{ji} d_{jl} \mathbf{E}(\mathbb{X}_i \mathbb{X}_l^T)$
 Recordemos que \mathbf{D} era una matriz ortogonal, por lo que $d_{ji} d_{jl} = 0$ cuando $i \neq l$, y $d_{ji} d_{jl} = 1$ cuando $i = l$. Con esto podemos reducir los dos sumatorios a uno solo:

$$\sum_{i=1}^n \sum_{l=1}^n d_{ji} d_{jl} \mathbf{E}(\mathbb{X}_i \mathbb{X}_l^T) = \sum_{i=1}^n d_{ji}^2 \mathbf{E}(\mathbb{X}_i \mathbb{X}_i^T) = \sum_{i=1}^n d_{ji}^2 (\mathbf{E}(\mathbb{X}_{ir} \mathbb{X}_{is}^T))_{1 \leq r, s \leq p},$$

donde en el último paso hemos usado que la esperanza de una matriz es la matriz de las esperanzas de sus elementos. Resulta ahora que si $r = s$, entonces $\mathbf{E}(\mathbb{X}_{ir} \mathbb{X}_{ir}^T) = 1$; y si $r \neq s$, entonces $\mathbf{E}(\mathbb{X}_{ir}) = 0$, ya que la matriz de covarianzas de \mathbb{X} es I_p . Por lo tanto,

$$\mathbf{E}(\mathbb{Y}_j \mathbb{Y}_j^T) = \sum_{i=1}^n d_{ji}^2 I_p = I_p \underbrace{\sum_{i=1}^n d_{ji}^2}_{=1} = I_p.$$

- Caso 2: $j \neq k$:

$$\mathbf{E}(\mathbb{Y}_j \mathbb{Y}_k^T) = \sum_{i=1}^n \sum_{l=1}^n \underbrace{d_{ji} d_{kl}}_{=0} \mathbf{E}(\mathbb{X}_i \mathbb{X}_l^T) = 0,$$

donde hemos usado que \mathbf{D} es ortogonal.

Con esto probamos la afirmación (1.31), y entonces $\mathbb{Y}_j \sim N_p(0, I_p) \forall j \in \{1, \dots, n\}$. Fijémonos además que $\mathbf{E}(\mathbb{Y}_{ji} \mathbb{Y}_{qk})_{1 \leq i, k \leq p} = 0$ para $j \neq q$, y como $\mathbb{Y}_{ji}, \mathbb{Y}_{qk}$ para $1 \leq i, k \leq p$ son combinación lineal de normales estándar, entonces son independientes.

Notemos ahora que $\mathbb{Y}_1 = \sum_{i=1}^n d_{1i} \mathbb{X}_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{X}_i = \frac{n}{\sqrt{n}} \bar{\mathbb{X}} = \sqrt{n} \bar{\mathbb{X}}$, teniendo en cuenta que $d_1 = \frac{1}{\sqrt{n}} \mathbf{1}_n$. Por consiguiente, $\bar{\mathbb{X}} = \frac{1}{\sqrt{n}} \mathbb{Y}_1$, y como $\mathbb{Y}_1 \sim N_p(0, I_p)$, entonces $\sqrt{n} \bar{\mathbb{X}} \sim N_p(0, I_p)$, que era el primer punto que queríamos probar.

Por otro lado, si calculamos $\sum_{j=1}^n \mathbb{Y}_j \mathbb{Y}_j^T$ obtenemos el siguiente resultado

$$\sum_{j=1}^n \mathbb{Y}_j \mathbb{Y}_j^T = \sum_{j=1}^n \left[\left(\sum_{i=1}^n d_{ji} \mathbb{X}_i \right) \left(\sum_{k=1}^n d_{jk} \mathbb{X}_k^T \right) \right] = \sum_{i,k=1}^n \mathbb{X}_i \mathbb{X}_k^T \sum_{j=1}^n \underbrace{d_{ji} d_{jk}}_{\substack{1 \text{ si } i=k \\ 0 \text{ si } i \neq k}} = \sum_{i=1}^n \mathbb{X}_i \mathbb{X}_i^T. \quad (1.32)$$

Además, notemos que

$$n \bar{\mathbb{X}} \bar{\mathbb{X}}^T = n \frac{1}{\sqrt{n}} \mathbb{Y}_1 \frac{1}{\sqrt{n}} \mathbb{Y}_1^T = \mathbb{Y}_1 \mathbb{Y}_1^T. \quad (1.33)$$

Finalmente, es importante darse cuenta de la siguiente descomposición

$$\sum_{i=1}^n (\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})^T = \sum_{i=1}^n \mathbb{X}_i \mathbb{X}_i^T - n \bar{\mathbb{X}} \bar{\mathbb{X}}^T. \quad (1.34)$$

Reuniendo los resultados (1.32), (1.33) y (1.34), obtenemos

$$(n-1) \mathbf{S}_{\mathbb{X}} = \sum_{i=1}^n \mathbb{X}_i \mathbb{X}_i^T - n \bar{\mathbb{X}} \bar{\mathbb{X}}^T = \sum_{i=1}^n \mathbb{Y}_i \mathbb{Y}_i^T - \mathbb{Y}_1 \mathbb{Y}_1^T \implies (n-1) \mathbf{S}_{\mathbb{X}} = \sum_{i=2}^n \mathbb{Y}_i \mathbb{Y}_i^T.$$

En consecuencia, y de un solo golpe, hemos probado la independencia entre $\bar{\mathbb{X}}$ y $(n-1)\mathbf{S}_{\mathbb{X}}$, y que $(n-1)\mathbf{S}_{\mathbb{X}} \sim W_p(n-1, I_p)$.

Lo primero es debido a que $\bar{\mathbb{X}}$ solo depende de \mathbb{Y}_1 y $(n-1)\mathbf{S}_{\mathbb{X}}$ depende de $\mathbb{Y}_2, \dots, \mathbb{Y}_n$. Como $\mathbb{Y}_1, \mathbb{Y}_2, \dots, \mathbb{Y}_n$ eran independientes, podemos asegurar que $\bar{\mathbb{X}}$ y $(n-1)\mathbf{S}_{\mathbb{X}}$ también lo son. Finalmente, habíamos visto que $\mathbb{Y}_j \sim N_p(0, I_p) \forall j \in \{1, \dots, n\}$, por lo que si $(n-1)\mathbf{S}_{\mathbb{X}} = \sum_{i=2}^n \mathbb{Y}_i \mathbb{Y}_i^T$, por la definición de matriz de Wishart, $(n-1)\mathbf{S}_{\mathbb{X}} \sim W_p(n-1, I_p)$. \square

Corolario 1.4.3. Si $\bar{\mathbb{X}}$ es la media y \mathbf{S} la cuasivarianza de una muestra de tamaño n de una $N_p(\mu, \Sigma)$, entonces

$$n(\bar{\mathbb{X}} - \mu)^T \mathbf{S}^{-1} (\bar{\mathbb{X}} - \mu) \sim T_p^2(n-1)$$

Demostración. Por el teorema (1.4.2) sabemos que $\sqrt{n}(\bar{\mathbb{X}} - \mu) \sim N_p(0, \Sigma)$ y que $(n-1)\mathbf{S} \sim W_p(n-1, \Sigma)$. Sustituyendo esto directamente en la definición (1.4.1) obtenemos

$$(n-1) [\sqrt{n}(\bar{\mathbb{X}} - \mu)]^T (n-1)^{-1} \mathbf{S}^{-1} [\sqrt{n}(\bar{\mathbb{X}} - \mu)] = n(\bar{\mathbb{X}} - \mu)^T \mathbf{S}^{-1} (\bar{\mathbb{X}} - \mu) \sim T_p^2(n-1).$$

\square

Teorema 1.4.4. Si $n \geq p$, entonces

$$\frac{n-p+1}{np} T_p^2(n) \sim F_{p, n-p+1}. \quad (1.35)$$

En particular, si $p = 1$ entonces $T_1^2(n) \sim F_{1, n}$.

Demostración. Sea $\mathbb{X} \sim N_p(\mu, \Sigma)$ y $\mathbf{M} \sim W_p(n, \Sigma)$. Por la definición (1.4.1) sabemos que $n(\mathbb{X} - \mu)^T \mathbf{M}^{-1} (\mathbb{X} - \mu) \sim T_p^2(n)$. Observar que \mathbf{M} es invertible cuando $n \geq p$ con probabilidad 1 al ser definida positiva.

Por otro lado, por definición de la distribución de Wishart, Σ es también definida positiva y podemos calcular Σ^{-1} . Llamando $\mathbb{D} = (\mathbb{X} - \mu)^T$ y multiplicando y dividiendo por $\mathbb{D}^T \Sigma^{-1} \mathbb{D}$ nos queda

$$\alpha = n \frac{\mathbb{D}^T \mathbf{M}^{-1} \mathbb{D}}{\mathbb{D}^T \Sigma^{-1} \mathbb{D}} \mathbb{D}^T \Sigma^{-1} \mathbb{D} \sim T_p^2(n).$$

Gracias a la propiedad de Wishart (1.15) sabemos que $\frac{\mathbb{D}^T \Sigma^{-1} \mathbb{D}}{\mathbb{D}^T \mathbf{M}^{-1} \mathbb{D}} \sim \chi_{n-p+1}^2$. Adicionalmente, notemos que $\Sigma^{-1} \mathbb{D} = \Sigma^{-1} (\mathbb{X} - \mu) \sim N_p(0, I_p)$, y por consiguiente $\mathbb{D}^T \Sigma^{-1} \mathbb{D} = \sum_{i=1}^p \mathbb{Y}_i^2$ con $\mathbb{Y}_j \sim N_p(0, I_p) \implies \mathbb{D}^T \Sigma^{-1} \mathbb{D} \sim \chi_p^2$.

Finalmente, uniendo lo anterior, podemos expresar α como cociente de χ^2 independientes (gracias a que \mathbf{M} y \mathbb{X} eran independientes por definición y por tanto también \mathbf{M} y \mathbb{D}):

$$\alpha = n \frac{\chi_p^2}{\chi_{n-p+1}^2} \sim \frac{np}{n-p+1} F_{p, n-p+1} \implies \frac{n-p+1}{np} T_p^2(n) \sim F_{p, n-p+1}.$$

\square

Corolario 1.4.5. El estadístico T^2 de Hotelling para una muestra normal multivariante está definida como

$$T_p^2(n-1) = n(\bar{\mathbb{X}} - \mu)^T \mathbf{S}^{-1} (\bar{\mathbb{X}} - \mu),$$

y tenemos que

$$\frac{n-p}{p} \frac{n}{n-1} (\bar{\mathbb{X}} - \mu)^T \mathbf{S}^{-1} (\bar{\mathbb{X}} - \mu) \sim F_{p, n-p}.$$

Demostración. El resultado es derivado directamente del hecho de que $\mathbf{S} \sim W_p(n-1, \Sigma)$ gracias al teorema (1.4.2), y del teorema (1.4.4). \square

1.4.1. Test de una muestra T^2 de Hotelling

En el caso univariante, el test- t es usado para contrastar la hipótesis nula $H_0 : \mu = \mu_0$, es decir, contrastar la media de la población de una variable aleatoria con un valor concreto. El estadístico del test viene dado por $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$, donde el test univariante de la hipótesis nula está basado en el hecho de que $t \sim STU(n-1)$. La hipótesis nula es rechazada cuando $|t| > t_{crit}$ con $t_{crit} = t_{\{n-1; \alpha\}}$. El parámetro α es el nivel de significación.

Análogamente, la versión multivariante contrasta la media muestral de una población normal (vector aleatorio $p \times 1$) con un cierto vector, es decir, $H_0 : \mu = \mu_0$, donde μ y μ_0 son vectores. Es aquí donde interviene el estadístico T^2 de Hotelling, que estaba definido como

$$T^2 = n(\bar{\mathbb{X}} - \mu_0)^T \mathbf{S}^{-1}(\bar{\mathbb{X}} - \mu_0),$$

y el corolario (1.4.5) nos dice que

$$\frac{n-p}{p(n-1)} T^2 \sim F_{p, n-p}.$$

Ahora la hipótesis nula es rechazada cuando $F > F_{crit} := F_{\{p; n-p; \alpha\}}$.

Para ejemplificar lo anterior, vamos a analizar un caso práctico diseñado por Charles Zaiontz [15]: Una compañía de calzado evalúa los nuevos modelos de zapato con 5 diferentes criterios: estilo, confort, estabilidad, amortiguamiento y durabilidad, donde los 4 primeros criterios están en una escala del 1 al 10 y el criterio de durabilidad está evaluado del 1 al 20.

Los objetivos por cada criterio de la compañía están reflejados en la siguiente tabla.

	Estilo	Confort	Estabilidad	Amortiguamiento	Durabilidad
Objetivo (μ_0)	7	8	5	7	9
Media muestral ($\bar{\mathbb{X}}$)	5.6	7.4	5.08	5.04	12.88

Tabla 1.1: Objetivos de la compañía y media muestral

La siguiente tabla muestra las evaluaciones de 25 personas que han probado el último prototipo de calzado y que van a ser utilizadas para determinar si el modelo está listo para salir al mercado.

Muestra	Estilo	Confort	Estabilidad	Amortiguamiento	Durabilidad
1	6	8	3	5	19
2	6	7	3	4	9
3	5	7	1	4	16
4	10	9	8	4	4
5	7	9	7	6	9
6	6	6	3	9	17
7	5	8	6	7	6
8	3	7	3	6	16
9	8	8	9	3	8
10	8	6	5	3	13
11	5	9	5	4	17
12	8	8	2	3	5
13	5	8	7	5	8
14	4	9	10	2	16
15	2	9	4	10	14
16	7	5	8	6	15
17	4	8	8	2	16
18	5	10	9	3	11
19	7	7	3	7	12
20	1	5	2	7	17
21	5	6	7	7	20
22	4	3	1	2	15
23	7	9	6	6	9
24	4	5	2	4	12
25	8	9	5	7	18

Tabla 1.2: Opiniones del prototipo

En primer lugar vamos a determinar qué criterios cumplen los objetivos y cuales no. Para ello utilizaremos el test- t de Student para cada criterio (test unidimensional) dado por $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ y $t \sim STU(n-1)$. Por supuesto supondremos que la hipótesis nula es que $H_0 : \mu = \mu_0$ y la rechazaremos si $|t| > t_{crit}$ donde $t_{crit} = t_{\{n-1; \alpha\}}$ con $\alpha = 0,05$ que es el estándar de significación.

Los cálculos explícitos son los siguientes

	Estilo	Confort	Estabilidad	Amortiguamiento	Durabilidad
Objetivo (μ_0)	7	8	5	7	9
Media muestral (\bar{x})	5.6	7.4	5.08	5.04	12.88
Varianza muestral (s)	2.082	1.708	2.722	2.169	4.567
Tamaño muestra (n)	25	25	25	25	25
$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	-3.363	-1.757	0.147	-4.517	4.248
Significación (α)	0.05	0.05	0.05	0.05	0.05
t_{crit}	2.064	2.064	2.064	2.064	2.064
¿Se cumple el objetivo?	No	Sí	Sí	No	No

Tabla 1.3: Tests- t de Student univariantes

Tanto el criterio de confort como el de estabilidad son satisfechos. El estilo y el amortiguamiento están significativamente por debajo de los objetivos y, el test del criterio de durabilidad rechaza la hipótesis nula pero para un caso favorable, ya que supera de largo el objetivo especificado inicialmente en la tabla 1.1.

Ahora realicemos un test multivariante, analizando los cinco objetivos a la vez, utilizando para ello el estadístico T^2 de Hotelling. Tanto el nivel de significación ($\alpha = 0,05$) como la hipótesis nula ($\mu = \mu_0$) se mantienen intactos, pero rechazamos cuando $F > F_{crit}$ donde $F_{crit} = F_{\{p; n-p; \alpha\}}$.

Calculamos la matriz de covarianzas muestral

$$\mathbf{S} = \frac{1}{n-1} \sum (\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})^T = \begin{pmatrix} 4,333 & 0,917 & 1,533 & -0,817 & -4,467 \\ 0,917 & 2,917 & 2,425 & 0,025 & -2,408 \\ 1,533 & 2,425 & 7,41 & -1,212 & -2,573 \\ -0,817 & 0,025 & -1,212 & 4,707 & 1,963 \\ -4,467 & -2,408 & -2,573 & 1,963 & 20,86 \end{pmatrix}.$$

Con los datos anteriores, y teniendo en cuenta que $n = 25$ y $p = 5$ podemos proseguir con el test T^2 de Hotelling. Usando el corolario (1.4.5) calculamos T^2

$$T^2 = n(\bar{\mathbb{X}} - \mu_0)^T \mathbf{S}^{-1} (\bar{\mathbb{X}} - \mu_0) = 52,672$$

$$F_{p, n-p} \sim \frac{n-p}{p(n-1)T^2} = \frac{25-5}{5(25-1)} 52,672 = 8,779$$

$$F_{crit} = F_{\{p; n-p; \alpha\}} = 2,711$$

Finalmente, podemos observar que $F > F_{crit}$, por lo que rechazamos la hipótesis nula y sostenemos que existe una diferencia significativa entre la media muestral de las cinco categorías con respecto a los objetivos planteados en la tabla 1.1.

1.5. Distribución de los autovalores de una matriz de Wishart

1.5.1. Ley de Marchenko-Pastur

Sea $\mathbf{M} \sim W_p(n, \Sigma)$ matriz aleatoria y denotemos como $\lambda_i(\mathbf{M})$ $i \in \{1, \dots, p\}$ a los autovalores (aleatorios) de \mathbf{M} .

La distribución de Marchenko-Pastur o **Ley de Marchenko-Pastur** (1967) describe el comportamiento asintótico de los autovalores de una matriz de Wishart. Si el radio $p/n \rightarrow \lambda \in (0, \infty)$ cuando $p, n \rightarrow \infty$, entonces la distribución espectral empírica (DEE) de $\mathbf{M} \in \mathcal{M}_p$, que es una función de densidad discreta, converge débilmente con probabilidad 1 a una función determinista y continua conocida como la Ley de Marchenko-Pastur con parámetro λ .

Si $\lambda \in (0, 1]$, entonces con probabilidad 1 lo siguiente se cumple para cada $x \in \mathbb{R}$

$$F^{\mathbf{M}}(x) := \frac{1}{p} \#\{1 \leq i \leq p : \lambda_i(\mathbf{M}) \leq x\} \xrightarrow[p/n \rightarrow \lambda]{n, p \rightarrow \infty} \int_{-\infty}^x f_\lambda(t) dt, \quad (1.36)$$

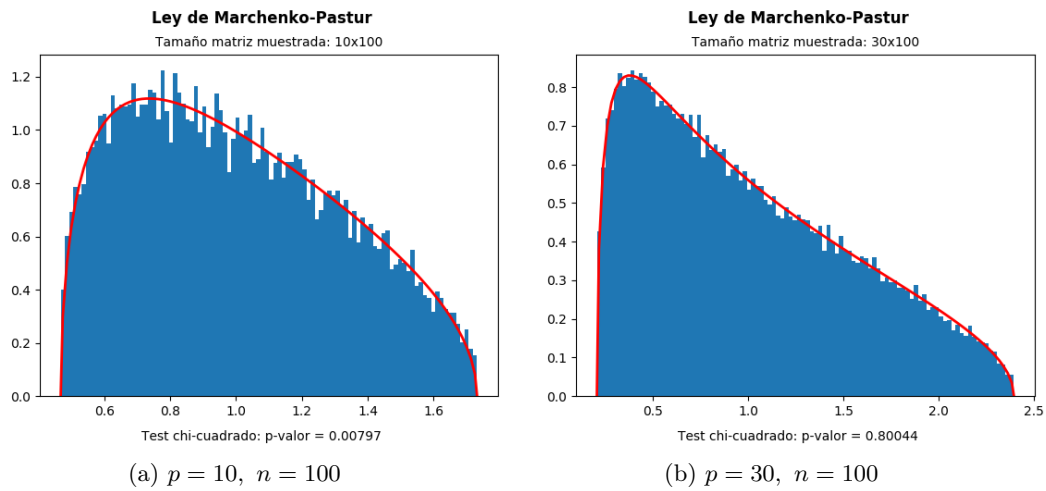
donde $f_\lambda(t)$ es la función de densidad de probabilidad de Marchenko-Pastur

$$f_\lambda(x) = \frac{1}{2\pi\sigma^2} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{\lambda x}, \text{ donde } \lambda_\pm = \sigma^2(1 \pm \sqrt{\lambda})^2, \quad (1.37)$$

y σ^2 es la varianza de las entradas de \mathbf{X} , que conforman la matriz de Wishart $\mathbf{M} = \mathbf{X}\mathbf{X}^T$.

Un resultado similar se cumple si $\lambda > 1$, pero en este caso la distribución límite tiene un punto de masa adicional de $1 - \frac{1}{\lambda}$ en el origen.

La derivación de la función de densidad de probabilidad anterior, así como pruebas de la Ley de Marchenko-Pastur, mediante el método de momentos o usando la transformada de Stieltjes, se pueden encontrar en el capítulo 3 de [5], o en el capítulo 3 de [2].



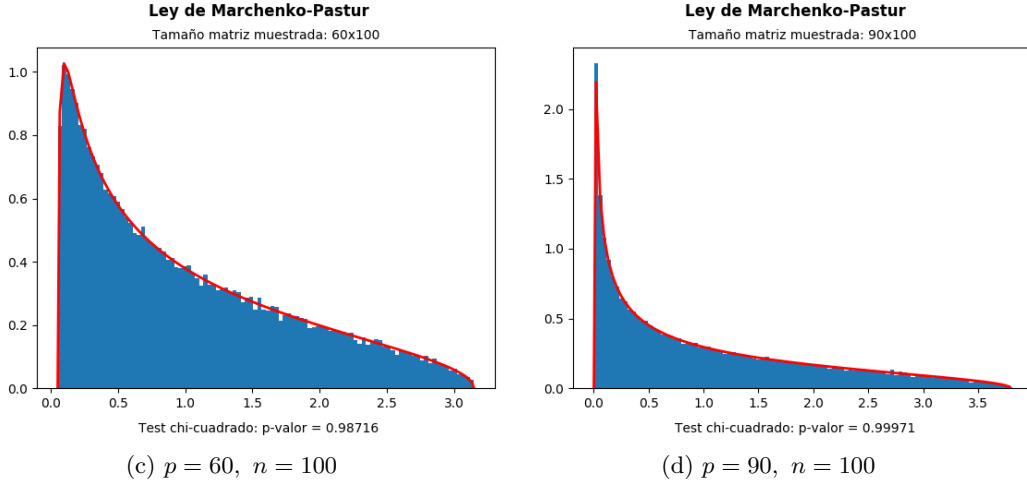


Figura 1.1: Simulación y gráfica de la FDP de Marchenko-Pastur para diferentes $\lambda \in (0, 1)$

1.5.2. Distribución del mayor autovalor

Cuando se analiza la distribución empírica espectral de cualquier matriz, uno siempre se pregunta por los casos extremos, en este caso por la distribución del mayor autovalor de una matriz de Wishart. Denotemos a este último como λ_{\max} .

En la sección 2.7 estudiaremos la **Ley de Tracy-Widom** de orden 1, F_1 , que es la distribución del mayor autovalor de una matriz Gaussiana $n \times n$ simétrica con entradas en los reales.

Como una matriz de Wishart está generada por una matriz \mathbf{X} , cuyos vectores columna son $\mathbb{X}_i \sim N_p(0, \Sigma)$, no es de extrañar que el mayor autovalor de una matriz de Wishart siga una distribución similar a la ley de Tracy-Widom.

Teorema 1.5.1. Sea $\mathbf{M} \sim W_p(n, \Sigma)$, para n y p suficientemente grandes y $n/p \rightarrow \lambda \in (0, 1]$ tenemos que

$$\frac{\lambda_{\max} - \mu_{np}}{\sigma_{np}} \xrightarrow{d} F_1, \text{ donde} \quad (1.38)$$

$$\mu_{np} = (\sqrt{n-1} + \sqrt{p})^2, \quad \sigma_{np} = (\sqrt{n-1} + \sqrt{p}) \left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}} \right)^{1/3}$$

El teorema está formulado para la situación (habitual) en que $n \geq p$. No obstante, se aplica equivalentemente cuando $n < p$, es decir cuando $\lambda > 1$, simplemente intercambiando los papeles de n y p en las expresiones de μ_{np} y σ_{np} mostradas arriba.

Por lo tanto, escribiremos $TW_1(n, p)$ para la ley de $\mu_{np} + \sigma_{np}F_1$, con la cual usaremos para aproximar la distribución de λ_{\max} .

La demostración del teorema (1.5.1), y más información sobre la relación entre la ley de Tracy-Widom y la distribución del mayor autovalor de una matriz de Wishart, se puede encontrar en la fuente [8].

CAPÍTULO 2

Ley semicircular de Wigner

Índice

2.1	Introducción a la distribución semicircular de Wigner	21
2.1.1	Enunciados del Teorema de Wigner	23
2.2	Momentos de la Ley de Wigner	25
2.3	Grafos y caminos de Dyck	28
2.4	Demostración de la versión 3 del Teorema de Wigner	30
2.5	Demostración de la versión 2 del Teorema de Wigner	36
2.6	Demostración de la versión 1 del Teorema de Wigner	38
2.7	Ley de Tracy-Widom	39
3.1	Matrices simétricas, definidas positivas	43
3.1.1	Resultados sobre diagonalización	43
3.1.2	Funciones características de variables aleatorias con valores en matrices simétricas	44
3.1.3	Normas de matrices simétricas	45
3.2	Desigualdad de Markov	46
3.3	Sobre el ensemble ortogonal gaussiano	46

2.1. Introducción a la distribución semicircular de Wigner

La clase de matrices aleatorias que vamos a estudiar en este capítulo son las **matrices de Wigner**. Nos vamos a ayudar de las fuentes [4] y [9], así como de los artículos originales [13] y [14] del autor de estos resultados y premio Nobel de física en 1963, Eugene Wigner.

Existen diferentes *ensembles* de estas matrices, pero todos ellos cumplen unas ciertas condiciones.

Definición 2.1.1. *Decimos que una matriz $\mathbf{Y}_n \in \mathcal{S}_n$ pertenece a un ensemble de matrices de Wigner si las siguientes condiciones se cumplen.*

- *Todas las entradas son independientes, es decir, $\{Y_{ij}\}_{1 \leq i, j \leq n}$ son independientes.*
- *Las entradas de la diagonal $\{Y_{ii}\}_{1 \leq i \leq n}$ están idénticamente distribuidas, y las entradas fuera de la diagonal $\{Y_{ij}\}_{1 \leq i < j \leq n}$ están idénticamente distribuidas.*

- Su segundo momento no centrado está acotado, es decir, $\mathbf{E}(Y_{ij}^2) < \infty$ para todo i, j . Podemos simplificar esto usando la hipótesis anterior, por lo que pedimos que

$$\beta_2 = \max\{\mathbf{E}(Y_{11}^2), \mathbf{E}(Y_{12}^2)\} < \infty.$$

También será de ayuda suponer que todos los momentos están acotados, es decir, $\beta_k = \max\{\mathbf{E}(Y_{11}^k), \mathbf{E}(Y_{12}^k)\} < \infty$ para todo k entero positivo.

Es importante notar que las matrices de un ensemble de Wigner no tienen por qué ser definidas positivas; y si \mathbf{Y}_n pertenece a un ensemble de Wigner, $-\mathbf{Y}_n$ también.

La hipótesis de que todos los momentos β_k están acotados para todo k entero positivo se puede relajar a solo suponer que β_2 es finito. Podremos encontrar este argumento con todo detalle en el capítulo 5 de [9].

Como $\mathbf{Y}_n \in \mathcal{S}_n$, ésta tiene n autovalores reales, los cuales los podemos ordenar de forma creciente $\lambda_1(\mathbf{Y}_n) \leq \dots \leq \lambda_n(\mathbf{Y}_n)$.

Por otro lado, como vamos a analizar el comportamiento de los autovalores, es natural añadir un factor de escalado con el objetivo de que exista un comportamiento límite de los mismos. $\mathbf{X}_n = \alpha_n \mathbf{Y}_n$.

Por suerte, podemos estudiar la estabilidad de los autovalores sin la referencia explícita de estos, utilizando para ello la norma de Frobenius de una matriz simétrica, la cual se define como en (3.1.10) en el Apéndice:

$$\frac{1}{n} \|\mathbf{X}\|_2^2 = \frac{1}{n} \sum_{i=1}^n \lambda_i(\mathbf{X}_n)^2 = \frac{1}{n} \sum_{1 \leq i, j \leq n} X_{ij}^2.$$

Utilizando esto, podemos calcular el valor esperado de esta norma:

$$\begin{aligned} \mathbf{E} \left[\frac{1}{n} \|\mathbf{X}_n\|_2^2 \right] &= \frac{1}{n} \mathbf{E} [\|\mathbf{X}_n\|_2^2] = \frac{1}{n} \alpha_n^2 \sum_{1 \leq i, j \leq n} \mathbf{E}(Y_{ij}^2) = \\ &= \frac{\alpha_n^2}{n} \sum_{i=1}^n \mathbf{E}(Y_{ii}^2) + \frac{2\alpha_n^2}{n} \sum_{1 \leq i, j \leq n} \mathbf{E}(Y_{ij}^2) = \alpha_n^2 \cdot \mathbf{E}(Y_{11}^2) + (n-1)\alpha_n^2 \cdot \mathbf{E}(Y_{12}^2). \end{aligned}$$

Si $\mathbf{E}(Y_{12}^2) = 0$ entonces $\frac{1}{n} \mathbf{E} [\|\mathbf{X}_n\|_2^2] = \alpha_n^2 \mathbf{E}(Y_{11}^2) < \infty$ por hipótesis. Nos bastaría escoger $\alpha_n = 1$.

Por el contrario, un caso menos aburrido, si $\mathbf{E}(Y_{12}^2) > 0$ tenemos que

$$\frac{1}{n} \mathbf{E} [\|\mathbf{X}_n\|_2^2] \rightarrow (n-1)\alpha_n^2 \cdot \mathbf{E}(Y_{12}^2) \text{ cuando } n \rightarrow \infty.$$

Por tanto, para que $\frac{1}{n} \mathbf{E} [\|\mathbf{X}_n\|_2^2]$ sea una sucesión acotada, tenemos que escoger $\alpha_n \sim n^{-1/2}$.

Podemos ahora mostrar la definición de las matrices de Wigner.

Definición 2.1.2 (Matriz de Wigner). Sea \mathbf{Y}_n como en la definición (2.1.1), con $\beta_2 < \infty$ y $\mathbf{E}(Y_{12}^2) > 0$. Entonces la matriz $\mathbf{X}_n = n^{-1/2} \mathbf{Y}_n$ es una matriz de Wigner. Notemos que \mathbf{X}_n es un objeto aleatorio, determinado por n muestras independientes de la distribución de las entradas de la diagonal, y $n(n-1)/2$ copias independientes de la distribución de las entradas fuera de la diagonal.

El cálculo anterior muestra que si \mathbf{X}_n es una matriz de Wigner, entonces la esperanza de la norma de Frobenius converge (cuando $n \rightarrow \infty$) al segundo momento de las entradas fuera de la diagonal.

Existe una distribución de probabilidad universal μ_{sc} tal que la densidad de los autovalores de cualquier matriz de Wigner (con segundo momento ρ) converge a ella. Esta distribución límite es conocida como la Ley Semicircular de Wigner.

Definición 2.1.3 (Ley Semicircular de Wigner). *La distribución de probabilidad μ_{sc} con parámetro ρ se conoce como la Ley Semicircular de Wigner:*

$$\mu_{sc} = \frac{1}{2\pi\rho} \sqrt{4\rho - x^2} \mathbf{1}_{|x| \leq 2\sqrt{\rho}} dx.$$

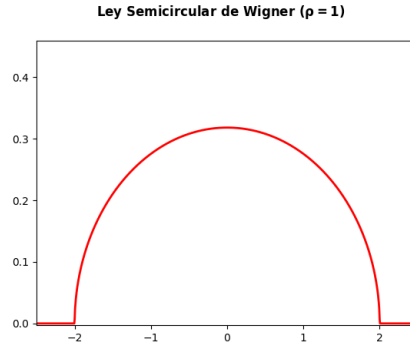


Figura 2.1: Ley Semicircular de Wigner con $\rho = 1$

Los siguientes resultados son diferentes versiones del Teorema de Wigner. Demostrar la Ley de Wigner y comprobar la equivalencia de las distintas versiones será el objetivo de nuestro trabajo a partir de ahora.

2.1.1. Enunciados del Teorema de Wigner

Teorema 2.1.4 (1ª versión Teorema de Wigner). *Sea $\mathbf{X}_n = n^{-1/2}\mathbf{Y}_n$ una matriz de Wigner con $\mathbf{E}(Y_{ij}) = 0 \ \forall i, j \in \{1, \dots, n\}$ y $\mathbf{E}(Y_{12}^2) = \rho$. Fijemos un intervalo $I \subset \mathbb{R}$ y definimos la sucesión de variables aleatorias*

$$E_n(I) = \frac{\#(\{\lambda_1(\mathbf{X}_n), \dots, \lambda_n(\mathbf{X}_n)\} \cap I)}{n}.$$

Entonces $E_n(I) \rightarrow \mu_{sc}(I)$ en probabilidad cuando $n \rightarrow \infty$.

La observación clave que Wigner hizo es que uno puede estudiar el comportamiento de las variables aleatorias $E_n(I)$ sin la necesidad de calcular los autovalores de \mathbf{X}_n . Esto se logra reinterpretando el teorema en términos de una medida aleatoria: la distribución empírica espectral (DEE). La DEE de una matriz aleatoria \mathbf{X} se define como

$$\mu_{\mathbf{X}}(x) = \frac{1}{n} \# \{i \leq n : \lambda_i(\mathbf{X}) \leq x\} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(\mathbf{X})}(x).$$

Por lo tanto, $\mu_{\mathbf{X}_n}$ es una medida de probabilidad aleatoria, ya que para cada \mathbf{X}_n la medida cambia al variar sus autovalores. Observemos que, dada una función continua $f \in C(\mathbb{R})$, la integral $\int f d\mu_{\mathbf{X}_n}$ es la variable aleatoria

$$\int f d\mu_{\mathbf{X}_n} = \frac{1}{n} \sum_{j=1}^n f(\lambda_j(\mathbf{X}_n)).$$

Notar que las variables aleatorias $E_n(I)$ en el teorema (2.1.4) se obtienen tomando $f = \mathbf{1}_I$, ya que $E_n(I) = \int \mathbf{1}_I d\mu_{\mathbf{X}_n} = \mu_{\mathbf{X}_n}(I)$. No olvidar que $\mathbf{1}_I$ no es una función continua, pero ésta es fácil de aproximar por funciones continuas.

Con esto en mente, llegamos a la segunda versión del teorema de Wigner.

Teorema 2.1.5 (2ª versión Teorema de Wigner). *Sea $\mathbf{X}_n = n^{-1/2}\mathbf{Y}_n$ una matriz de Wigner con $\mathbf{E}(Y_{ij}) = 0 \ \forall i, j \in \{1, \dots, n\}$ y $\mathbf{E}(Y_{12}^2) = \rho$. Entonces para $f \in C_b(\mathbb{R})$ y $\epsilon > 0$:*

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\left| \int f d\mu_{\mathbf{X}_n} - \int f d\mu_{sc} \right| > \epsilon \right) = 0,$$

donde $C_b(\mathbb{R})$ denota al conjunto de funciones continuas y acotadas sobre \mathbb{R} .

Con esta formulación, podemos usar el teorema espectral para eliminar la aparición explícita de los autovalores de la medida $\mu_{\mathbf{X}_n}$. Como las matrices de Wigner son simétricas, podemos diagonalizar $\mathbf{X}_n = O_n^T \Lambda_n O_n$, y entonces

$$\begin{aligned} \int f d\mu_{\mathbf{X}_n} &= \frac{1}{n} \sum_{j=1}^n f(\lambda_j(\mathbf{X}_n)) = \frac{1}{n} \sum_{j=1}^n f([\Lambda_n]_{jj}) = \\ &= \frac{1}{n} \sum_{j=1}^n \text{traza}(f(\Lambda_n)) = \frac{1}{n} \text{traza}(O_n^T f(\Lambda_n) O_n) = \frac{1}{n} \text{traza}(f(\mathbf{X}_n)). \end{aligned} \quad (2.1)$$

La última igualdad es consecuencia del teorema espectral (3.1.1) del Apéndice.

Si ahora tomamos f como un polinomio, llegamos a la tercera versión del Teorema de Wigner. Notar que un polinomio no constante nunca está acotado, pero probaremos utilizando el teorema de aproximación de Weierstrass que ambas versiones son finalmente equivalentes.

Teorema 2.1.6 (3ª versión Teorema de Wigner). *Sea $\mathbf{X}_n = n^{-1/2}\mathbf{Y}_n$ una matriz de Wigner con $\mathbf{E}(Y_{ij}) = 0 \ \forall i, j \in \{1, \dots, n\}$ y $\mathbf{E}(Y_{12}^2) = \rho$. Sea f un polinomio. Entonces la sucesión de variables aleatorias $\int f d\mu_{\mathbf{X}_n}$ converge a $\int f d\mu_{sc}$ en probabilidad cuando $n \rightarrow \infty$. Equivalentemente, para un $k \in \mathbb{N}$ y un $\epsilon > 0$ fijados,*

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\left| \frac{1}{n} \text{traza}(\mathbf{X}_n^k) - \int x^k d\mu_{sc}(dx) \right| > \epsilon \right) = 0.$$

Antes de proceder con la demostración, notemos ciertas características de la Ley Semicircular de Wigner. Primero, y lo más obvio, es que es conocida como la Ley del semicírculo o Ley Semicircular debido a que la función de densidad de probabilidad μ_{sc} toma la apariencia de un semicírculo centrado en el $(0, 0)$ y de radio $R = 2\sqrt{\rho}$.

Esto lo podemos ejemplificar simulando el *ensemble* de matrices de Wigner más conocido: el *Ensemble* Ortogonal Gaussiano (GOE)¹. Las entradas de la diagonal son $N(0, 2)$, y las entradas fuera de la diagonal $N(0, 1)$. Por tanto $\mathbf{E}(Y_{12}^2) = \rho = 1$ y

$$\mu_{sc} = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbf{1}_{|x| \leq 2} \quad (2.2)$$

Para probar la versión del teorema (2.1.6) necesitaremos conocer los momentos de μ_{sc} y de $\mu_{\mathbf{X}_n}$. En la sección (2.2) será donde expondremos los momentos de μ_{sc} , y en la sección (2.3) introduciremos ciertas nociones de teoría de grafos para codificar eficientemente los momentos de $\mu_{\mathbf{X}_n}$.

¹Para saber más sobre él, y por qué se le denomina 'ortogonal', visitar 3.3

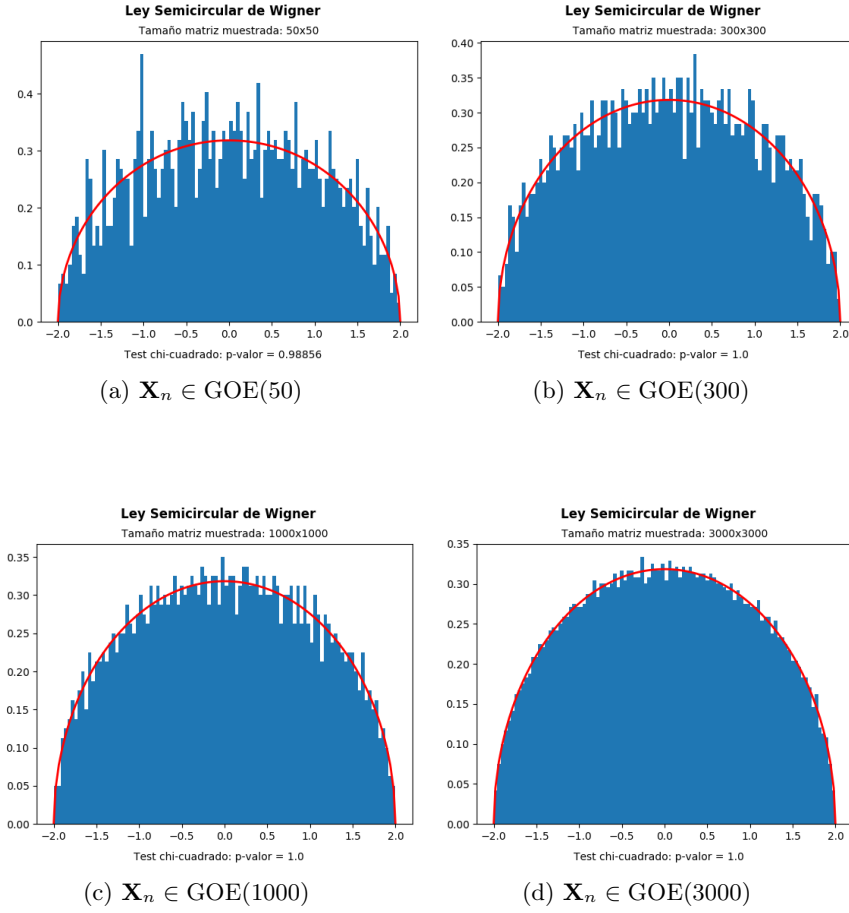


Figura 2.2: Simulación de la FDP de la Ley Semicircular de Wigner para $\mathbf{X}_n \in \text{GOE}(n)$

2.2. Momentos de la Ley de Wigner

En esta sección ilustraremos unos resultados previos a la demostración de la Ley Semicircular de Wigner que nos serán de importante utilidad.

Definición 2.2.1 (Número de Catalan).

$$C_m = \frac{1}{m+1} \binom{2m}{m} = \frac{(2m)!}{(m+1)!m!}, \text{ para } m \geq 0.$$

Los primeros números de Catalan son $C_0 = 1$, $C_1 = 1$, $C_2 = 2$, $C_3 = 5$. Una manera de argumentar cómo de relevante es esta sucesión es ver que es la entrada más larga en La Enciclopedia electrónica de secuencias de enteros (OEIS): <https://oeis.org/search?q=catalan+number&language=english&go=Search>.

Lema 2.2.2. $\binom{2m}{m} - \binom{2m}{m+1} = \frac{1}{m+1} \binom{2m}{m}.$

Demostración.

$$\begin{aligned}
\binom{2m}{m} - \binom{2m}{m+1} &= \frac{(2m)!}{m!m!} - \frac{(2m)!}{(m+1)!(2m-m-1)!} = \frac{(2m)!}{m!m!} - \frac{(2m)!}{(m+1)!(m-1)!} = \\
&= \frac{(2m)!(m+1)!(m-1)! - (2m)!m!m!}{(m+1)!(m-1)!m!m!} = \frac{(m-1)![(2m)!(m+1)! - (2m)!m!m]}{(m+1)!(m-1)!m!m!} = \\
&= \frac{[(2m)\dots(m+2)(m+1) - 2m(2m-1)\dots(m+2)m](m-1)!m!}{(m+1)!(m-1)!m!m!} = \\
&= \frac{(2m)\dots(m+2)(m+1) - (2m)(2m-1)\dots(m+2)m}{m!} = \frac{(2m)\dots(m+2)[(m+1) - m]}{m!} = \\
&= \frac{(2m)\dots(m+2)}{m!} = \frac{(2m)!}{m!(m+1)!} = \frac{(2m)!}{(m+1)m!m!} = \frac{1}{m+1} \binom{2m}{m}.
\end{aligned}$$

□

Lema 2.2.3. El número de Catalan C_m está acotado superiormente por 4^m .

Demostración.

$$\binom{2m}{m} \leq \binom{2m}{0} + \binom{2m}{1} + \dots + \binom{2m}{m} = \sum_{j=0}^m \binom{2m}{j} = 2^{2m} = 4^m.$$

Como $C_m = \frac{1}{m+1} \binom{2m}{m} \implies C_m \leq \frac{1}{m+1} 4^m \leq 4^m$.

□

Lema 2.2.4. Asintóticamente, el número de Catalan C_m es igual a $\frac{4^m}{(m+1)\sqrt{\pi m}}$.

Demostración. Vamos a utilizar la fórmula de Stirling $m! \approx \sqrt{2\pi m} \left(\frac{m}{e}\right)^m$.

$$C_m = \frac{1}{m+1} \binom{2m}{m} = \frac{1}{m+1} \cdot \frac{(2m)!}{m!m!} \approx \frac{1}{m+1} \frac{\sqrt{4\pi m} \left(\frac{2m}{e}\right)^{2m}}{2\pi m \left(\frac{m}{e}\right)^{2m}} = \frac{4^m}{(m+1)\sqrt{\pi m}}.$$

□

Adicionalmente, necesitaremos saber los momentos de la Ley Semicircular de Wigner, es decir, los momentos de la medida de probabilidad μ_{sc} .

Teorema 2.2.5 (Momentos de la Ley de Wigner). *Los momentos de la Ley Semicircular de Wigner son:*

$$\beta_k(\mu_{sc}) = \mathbf{E}_{\mu_{sc}}[x^k] = \int_{\mathbb{R}} x^k d\mu_{sc} = \begin{cases} \rho^{k/2} C_{k/2}, & \text{si } k \text{ es par} \\ 0, & \text{si } k \text{ es impar} \end{cases}$$

Demostración. Primero veremos que cuando k es par, estaremos integrando una función impar en un intervalo simétrico, por lo que la integral se anulará. Precizando, con un cambio de variables, veremos que $\int x^k \mu_{sc}(dx) = \rho^{k/2} \int x^k \sigma(dx)$, donde $\sigma(dx)$ es la Ley de Wigner con segundo momento unitario ($\mathbf{E}(\mathbf{Y}_{12}^2) = 1$). A continuación, veremos que $\beta_0(\sigma) = 1$, que $\beta_{2k+1}(\sigma) = 0$, y que $\beta_{2k}(\sigma) = \frac{2(2k-1)}{k+1} \beta_{2(k-1)}(\sigma)$. Finalmente, con un argumento de recursión, mostraremos que $\beta_{2k}(\sigma) = C_k = \frac{1}{k+1} \binom{2k}{k}$.

Como hemos dicho, empecemos desarrollando

$$\int x^k \mu_{sc}(dx) = \int_{-2\sqrt{\rho}}^{2\sqrt{\rho}} x^k \frac{1}{2\pi\rho} \sqrt{4\rho - x^2} dx.$$

Realizamos el cambio de variable $y = x/\sqrt{\rho} \implies x = \sqrt{\rho}y \implies dx = \sqrt{\rho}dy$. Los límites de integración cambian: $x = -2\sqrt{\rho} \implies y = -2$, $x = 2\sqrt{\rho} \implies y = 2$.

$$\int_{-2}^2 (\sqrt{\rho}y)^k \frac{1}{2\pi\rho} \sqrt{4\rho - (\sqrt{\rho}y)^2} \sqrt{\rho} dy = \rho^{k/2} \int_{-2}^2 y^k \frac{1}{2\pi} \sqrt{4 - y^2} dy = \rho^{k/2} \sigma(dx).$$

Ahora veamos que $\beta_0(\sigma) = 1$:

$$\beta_0(\sigma) = \int_{-2}^2 \frac{1}{2\pi} \sqrt{4 - x^2} dx = \frac{1}{2\pi} \int_{-2}^2 \sqrt{4 - x^2} dx.$$

Realizamos el cambio de variable $x = 2 \cos \theta \implies dx = -2 \sin \theta d\theta$. Los límites de integración cambian: $x = 2 \implies \cos \theta = 1 \implies \theta = 0$, $x = -2 \implies \cos \theta = -1 \implies \theta = \pi$. Esto, junto con que $\sqrt{4 - (2 \cos \theta)^2} = 2\sqrt{1 - \cos^2 \theta} = 2 \sin \theta$, nos lleva a

$$\frac{1}{2\pi} \int_{-2}^2 \sqrt{4 - x^2} dx = \frac{1}{2\pi} \int_{\pi}^0 2 \sin \theta (-2 \sin \theta) d\theta = \frac{2}{\pi} \int_0^{\pi} \sin^2 \theta d\theta = \frac{2}{\pi} \frac{\pi}{2} = 1.$$

Como habíamos predicho, podemos ver fácilmente que cuando k es impar $\beta_k(\sigma) = 0$ observando que el intervalo de integración es $[-2, 2]$ y que $f(x) = x^k \sqrt{4 - x^2}$ es impar, ya que $f(-x) = (-x)^k \sqrt{4 - (-x)^2} = (-1)^k x^k \sqrt{4 - x^2} = -f(x)$. Por lo tanto, por simetría, $\beta_k(\sigma) = 0$ para k impar.

Por otro lado, desarrollemos $\beta_k(\sigma)$ cuando k es par. Sin pérdida de generalidad y por comodidad, escribimos $\beta_{2k}(\sigma)$ para enfatizar que estamos trabajando con un índice par.

Realizamos el mismo cambio de variable $x = 2 \cos \theta$, obteniendo

$$\beta_{2k}(\sigma) = \frac{1}{2\pi} \int_{-2}^2 x^{2k} \sqrt{4 - x^2} dx = \frac{1}{2\pi} \int_0^{\pi} (2 \cos \theta)^{2k} 4 \sin^2 \theta d\theta = \frac{2}{\pi} \int_0^{\pi} (2 \cos \theta)^{2k} \sin^2 \theta d\theta.$$

Integramos por partes:

$$\begin{aligned} & \begin{cases} u = (2 \cos \theta)^{2k} \implies du = 2k(2 \cos \theta)^{2k-1}(-2 \sin \theta) d\theta \\ dv = \sin^2 \theta \implies v = \frac{1}{2}(\theta - \sin \theta \cos \theta) \end{cases} \\ \beta_{2k}(\sigma) &= \frac{2}{\pi} \left[\underbrace{\left[(2 \cos \theta)^{2k} \frac{1}{2}(\theta - \sin \theta \cos \theta) \right]_0^{\pi}}_{= \pi 2^{2k-1}} - \int_0^{\pi} \frac{1}{2}(\theta - \sin \theta \cos \theta)(-4k \sin \theta (2 \cos \theta)^{2k-1}) d\theta \right] = \\ &= \frac{2}{\pi} \left[\pi 2^{2k-1} - \left(\int_0^{\pi} \frac{-1}{2} 4k \theta \sin \theta (2 \cos \theta)^{2k-1} d\theta + \int_0^{\pi} \frac{1}{2} 4k \sin^2 \theta \cos \theta (2 \cos \theta)^{2k-1} d\theta \right) \right] = \\ &= 2^{2k} - \frac{2}{\pi} \left(\int_0^{\pi} -2k \theta \sin \theta (2 \cos \theta)^{2k-1} d\theta + \underbrace{\int_0^{\pi} k \sin^2 \theta (2 \cos \theta)^{2k} d\theta}_{= \frac{k\pi}{2} \beta_{2k}(\sigma)} \right) = \\ &= 2^{2k} + \frac{4k}{\pi} \underbrace{\int_0^{\pi} \theta \sin \theta (2 \cos \theta)^{2k-1} d\theta}_{:= \tilde{I}} - k \beta_{2k}(\sigma) \implies (k+1) \beta_{2k}(\sigma) = 2^{2k} + \frac{4k}{\pi} \tilde{I}. \end{aligned}$$

Nuestro siguiente objetivo será desentrañar \tilde{I} . Para ello volvemos a integrar por partes:

$$\begin{aligned} \begin{cases} u = (2 \cos \theta)^{2k-1} \implies du = (2k-1)(2 \cos \theta)^{2k-2}(-2 \sin \theta) d\theta \\ dv = \theta \sin \theta d\theta \implies v = \sin \theta - \theta \cos \theta \end{cases} \\ \tilde{I} = \underbrace{[(2 \cos \theta)^{2k-1}(\sin \theta - \theta \cos \theta)]_0^\pi}_{=-\pi 2^{2k-1}} - \int_0^\pi (\sin \theta - \theta \cos \theta)(-2(2k-1) \sin \theta (2 \cos \theta)^{2k-2}) d\theta = \\ = -\pi 2^{2k-1} - \left(\underbrace{\int_0^\pi -2(2k-1) \sin^2 \theta (2 \cos \theta)^{2k-2} d\theta}_{=-(2k-1)\pi \beta_{2k-2}(\sigma)} + \underbrace{\int_0^\pi (2k-1) \theta \sin \theta (2 \cos \theta)^{2k-1} d\theta}_{=(2k-1)\tilde{I}} \right) = \\ = -\pi 2^{2k-1} + (2k-1)\pi \beta_{2k-2}(\sigma) - (2k-1)\tilde{I} \implies \tilde{I} = \frac{-\pi}{k} 2^{2k-2} + \frac{(2k-1)\pi}{2k} \beta_{2k-2}(\sigma). \end{aligned}$$

Sustituimos ahora \tilde{I} en la expresión de $\beta_{2k}(\sigma)$, obteniendo

$$\begin{aligned} (k+1)\beta_{2k}(\sigma) &= 2^{2k} + \frac{4k}{\pi} \left(\frac{-\pi}{k} 2^{2k-2} + \frac{(2k-1)\pi}{2k} \beta_{2k-2}(\sigma) \right) = \\ &= 2^{2k} - 2^{2k} + 2(2k-1)\beta_{2k-2}(\sigma) \implies \beta_{2k}(\sigma) = \frac{2(2k-1)}{k+1} \beta_{2k-2}(\sigma). \end{aligned}$$

Finalmente, podemos demostrar que $\beta_{2k}(\sigma) = C_k := \frac{1}{k+1} \binom{2k}{k}$ utilizando la recursión construida anteriormente.

$$\begin{aligned} \beta_{2k}(\sigma) &= \frac{2(2k-1)}{k+1} \beta_{2k-2}(\sigma) \implies \beta_{2k}(\sigma) = \frac{2(2k-1)}{k+1} \cdot \frac{2(2k-3)}{k} \beta_{2k-4}(\sigma) \implies \\ \implies \beta_{2k}(\sigma) &= \frac{2(2k-1)}{k+1} \cdot \frac{2(2k-3)}{k} \cdot \frac{2(2k-5)}{k-1} \cdot \dots \cdot \frac{2 \cdot 3}{3} \cdot \frac{2 \cdot 1}{2} \beta_0(\sigma). \end{aligned}$$

Es fácil ver que hay un total de k factores. En la expresión anterior realizamos las siguientes agrupaciones:

$$\underbrace{2 \cdot \dots \cdot 2}_{k \text{ factores}} \cdot \frac{1}{k+1} \cdot \frac{1}{k!} \left(\prod_{\substack{j=1 \\ j \equiv 1 \pmod{2}}}^{2k} j \right) \beta_0(\sigma) = 2^k \cdot \frac{1}{k+1} \cdot \frac{1}{k!} \frac{(2k)!}{\prod_{\substack{j=1 \\ j \equiv 0 \pmod{2}}}^{2k} j} \cdot \beta_0(\sigma).$$

Notemos que $\prod_{\substack{j=1 \\ j \equiv 0 \pmod{2}}}^{2k} j = 2^k \prod_{j=1}^k j = k!$, y entonces

$$\beta_{2k} = 2^k \cdot \frac{1}{k+1} \cdot \frac{1}{k!} \cdot \frac{(2k)!}{2^k k!} = \frac{1}{k+1} \frac{(2k)!}{k! k!} = \frac{1}{k+1} \binom{2k}{k} = C_k.$$

□

2.3. Grafos y caminos de Dyck

Empecemos exponiendo unos resultados básicos sobre grafos, que pueden ser encontrados en cualquier libro de Teoría de Grafos, como en [3].

Teorema 2.3.1. Sea $G = \{V, E\}$ un grafo formado por el conjunto de vértices V y el conjunto de enlaces E . Si G es un grafo conexo entonces $|E| \geq |V| - 1$.

Teorema 2.3.2. Sea $G = \{V, E\}$. Si $|E| = |V| - 1$, entonces G es un árbol.

Proposición 2.3.3. Si G es un árbol, entonces G no tiene ciclos.

Ahora vamos a introducir una nueva estructura combinatoria: los caminos de Dyck.

Definición 2.3.4 (Caminos de Dyck). Un camino de Dyck de longitud n es la lista (d_1, \dots, d_n) con $d_i = \pm 1 \ \forall i \in \{1, \dots, n\}$ que cumple las siguientes propiedades:

- Las sumas parciales son no negativas, es decir, $\sum_{j=1}^k d_j \geq 0$ para $1 \leq k \leq n$.
- El número de incrementos es igual al de decrementos, es decir, $\sum_{j=1}^n d_j = 0$.

Denotamos como \mathcal{D}_n al conjunto de caminos de Dyck de longitud n . Sea $d = (d_1, \dots, d_n) \in \mathcal{D}_n$ un camino de Dyck cualquiera, el cual podemos representar en \mathbb{Z}^2 de la siguiente forma:

- El camino siempre empieza en el punto $(1, 0)$.
- Si $d_j = 1$, nos desplazamos una unidad a la derecha en el movimiento j -ésimo con $1 \leq j \leq n$.
- Si $d_j = -1$, nos desplazamos una unidad arriba en el movimiento j -ésimo con $1 \leq j \leq n$.

Notemos que, como las sumas parciales son siempre no negativas, el primer movimiento va a ser a la derecha ($d_1 = 1$), y nunca vamos a tocar la diagonal $(q, q) \in \mathbb{Z}^2$. Adicionalmente, como d es de longitud n , tenemos $m = \frac{n}{2}$ desplazamientos a la derecha y $m = \frac{n}{2}$ movimientos hacia arriba, por lo que el punto final siempre va a ser el $(m+1, m)$.

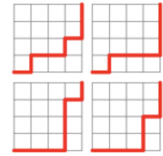


Figura 2.3:
Algunos
caminos
de Dyck
($n = 8$)

Teorema 2.3.5. $|\mathcal{D}_n| = C_{n/2}$, donde $C_{n/2}$ es el número de Catalan.

Demostración. Llamamos a W_m^s al conjunto de todos los caminos de longitud $n = 2m$ desde $(1, 0)$ a $(m-s+1, m+s)$. El cardinal de este conjunto es $|W_m^s| = \binom{2m}{m+s}$, ya que entre los $2m$ movimientos hay que escoger $m+s$ desplazamientos hacia arriba (y por consiguiente, $m-s$ desplazamientos a la derecha). Observar que W_m^s no tiene por qué tener más movimientos a la derecha que hacia arriba, por lo que puede tocar o cruzar la diagonal.

Los posibles caminos de Dyck de longitud $n = 2m$ es igual al conjunto W_m^0 , pero quitando aquellos caminos de W_m^0 que toquen o crucen la diagonal $(q, q) \in \mathbb{Z}^2$. Afirmamos ahora que el conjunto de caminos que tocan o cruzan la diagonal (llamémosle \widetilde{W}) es isomorfo a W_m^1 . Para ver que existe un isomorfismo, podemos encontrar una biyección entre \widetilde{W} y W_m^1 por un argumento de reflexión sobre la propia diagonal. Notar que el punto final de W_m^1 es $(m, m+1)$ y su reflejado con respecto a la diagonal es el punto $(m+1, m)$ que es el punto final de W_m^0 , \widetilde{W} y \mathcal{D}_n .

El cardinal del conjunto W_m^1 sí lo conocemos (cosa que no sabíamos de \widetilde{W}), y por lo tanto:

$$|\mathcal{D}_n| = |W_m^0| - |W_m^1| = \binom{2m}{m} - \binom{2m}{m+1} = C_m = C_{n/2}.$$

□

2.4. Demostración de la versión 3 del Teorema de Wigner

Nuestro primer objetivo será demostrar la 3ª versión de la Ley de Wigner (2.1.6). Para ello enunciamos el siguiente resultado de gran importancia.

Proposición 2.4.1. *Sea X_n una sucesión de variables aleatorias con $\mathbf{E}(X_n^2) < \infty$ para cada n . Supongamos que $\lim_{n \rightarrow \infty} \mathbf{E}(X_n) = m$, y que $\lim_{n \rightarrow \infty} \mathbf{V}(X_n) = 0$. Entonces $X_n \rightarrow m$ en probabilidad, es decir, dado $\epsilon > 0$, $\lim_{n \rightarrow \infty} \mathbf{P}(|X_n - m| > \epsilon) = 0$.*

Demostración. Aplicamos la desigualdad de Markov (3.2.1) vista en el Apéndice a $Y = |X_n - m|$ (con ϵ^2 en lugar de ϵ),

$$\mathbf{P}(|X_n - m| > \epsilon) = \mathbf{P}((X_n - m)^2 > \epsilon^2) \leq \frac{\mathbf{E}((X_n - m)^2)}{\epsilon^2}.$$

Ahora, usando la desigualdad triangular para la norma $L^2(\mathbf{P})$, obtenemos

$$\begin{aligned} \mathbf{E}((X_n - m)^2) &= \|X_n - m\|_{L^2(\mathbf{P})}^2 = \|X_n - \mathbf{E}(X_n) + \mathbf{E}(X_n) - m\|_{L^2(\mathbf{P})}^2 \leq \\ &\leq (\|X_n - \mathbf{E}(X_n)\|_{L^2(\mathbf{P})} + \|\mathbf{E}(X_n) - m\|_{L^2(\mathbf{P})})^2. \end{aligned}$$

Como $\mathbf{E}(X_n)$ y m son constantes, la segunda norma es simplemente el valor absoluto $|\mathbf{E}(X_n) - m|$. Por otro lado, la primera norma es $\|X_n - \mathbf{E}(X_n)\|_{L^2(\mathbf{P})} = \sqrt{\mathbf{E}[(X_n - \mathbf{E}(X_n))^2]} = \sqrt{\mathbf{V}(X_n)}$. Finalmente, por la desigualdad de Markov, llegamos a

$$\mathbf{P}(|X_n - m| > \epsilon) \leq \frac{(\sqrt{\mathbf{V}(X_n)} + |\mathbf{E}(X_n) - m|)^2}{\epsilon^2}.$$

La parte derecha tiende a cero por hipótesis, lo que prueba el resultado. \square

Ahora ya estamos listos para empezar con la demostración de (2.1.6). Empezamos desarrollando $\mathbf{E}[x^k]$:

$$\mathbf{E}[x^k] = \int_{\mathbb{R}} x^k d\mu_{\mathbf{X}_n} = \frac{1}{n} \sum_{i=1}^n \lambda_i^k(\mathbf{X}_n) = \frac{1}{n} \text{traza}(\mathbf{X}_n^k). \quad (2.3)$$

Aplicaremos la proposición (2.4.1) a los términos del teorema (2.1.6), donde uno de ellos viene de la expresión (2.3), para obtener un procedimiento basado en los dos siguientes pasos:

- Paso 1: para cada k impar mostrar que $\frac{1}{n} \mathbf{E}[\text{traza}(\mathbf{X}_n)] \rightarrow 0$ cuando $n \rightarrow \infty$, y para cada k par ver que $\frac{1}{n} \mathbf{E}[\text{traza}(\mathbf{X}_n)] \rightarrow \rho^{k/2} C_{k/2}$ cuando $n \rightarrow \infty$.
- Paso 2: para cada k , mostrar que $\mathbf{V}(\frac{1}{n} \text{traza}(\mathbf{X}_n)) \rightarrow 0$ cuando $n \rightarrow \infty$.

Utilizamos este procedimiento porque resulta que los momentos de la DEE están concentrados alrededor de la DEE de esperanza de \mathbf{X}_n . Por lo que calcularemos los momentos $\beta_k(\mu_{\mathbf{X}_n})$, lo cual es más sencillo que calcular $\beta_k(\mu_{\mathbf{X}_n})$, y demostraremos que asintóticamente estos coinciden con el segundo paso, viendo que la varianza tiende a cero cuando $n \rightarrow \infty$.

Paso 1: Empezamos desarrollando (2.3)

$$\frac{1}{n} \mathbf{E}[\text{traza}(\mathbf{X}_n^k)] = \frac{1}{n} \left[\text{traza} \left(\left(\frac{1}{\sqrt{n}} \mathbf{Y}_n \right)^k \right) \right] = n^{-\frac{k}{2}-1} \mathbf{E}[\text{traza}(\mathbf{Y}_n^k)]. \quad (2.4)$$

Usando la definición del producto de matrices tenemos

$$(\mathbf{Y}_n^k)_{ij} = \sum_{1 \leq i_2, \dots, i_k \leq n} Y_{ii_2} Y_{i_2 i_3} \dots Y_{i_{k-1} i_k} Y_{i_k j}.$$

Centrándonos en la diagonal:

$$(\mathbf{Y}_n^k)_{ii} = \sum_{1 \leq i_2, \dots, i_k \leq n} Y_{ii_2} Y_{i_2 i_3} \dots Y_{i_{k-1} i_k} Y_{i_k i}.$$

Y por tanto

$$\mathbf{E} [\text{traza} (\mathbf{Y}_n^k)] = \sum_{i_1=1}^n \mathbf{E} [(\mathbf{Y}_n^k)_{i_1 i_1}] = \sum_{1 \leq i_1, i_2, \dots, i_k \leq n} Y_{i_1 i_2} Y_{i_2 i_3} \dots Y_{i_{k-1} i_k} Y_{i_k i_1} := \sum_{\mathbf{i} \in [n]^k} \mathbf{E}(\mathbf{Y}_{\mathbf{i}}),$$

donde $[n] = \{1, \dots, n\}$, $\mathbf{i} = (i_1, i_2, \dots, i_k)$, $\mathbf{Y}_{\mathbf{i}} = Y_{i_1 i_2} Y_{i_2 i_3} \dots Y_{i_{k-1} i_k} Y_{i_k i_1}$.

Para ser capaces de lidiar con esta suma, vamos a realizar una codificación de \mathbf{i} basada en grafos. Para cada \mathbf{i} tenemos su grafo asociado $G_{\mathbf{i}}$, donde \mathbf{i} define un camino cerrado.

$$G_{\mathbf{i}} = \{V_{\mathbf{i}}, E_{\mathbf{i}}\},$$

$$V_{\mathbf{i}} = \{i_1, i_2, \dots, i_k\} \text{ (conjunto de vértices),}$$

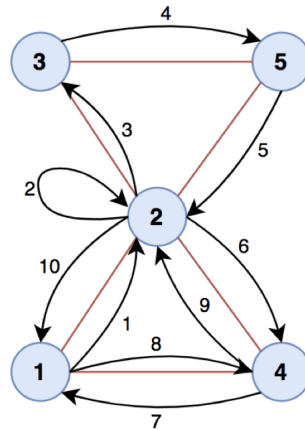
$$E_{\mathbf{i}} = \{\{i_1, i_2\}, \{i_2, i_3\}, \dots, \{i_{k-1}, i_k\}, \{i_k, i_1\}\} \text{ (conjunto de enlaces).}$$

Por ejemplo, para $\mathbf{i} = (1, 2, 2, 3, 5, 2, 4, 1, 4, 2)$ tenemos su grafo asociado $G_{\mathbf{i}} = \{V_{\mathbf{i}}, E_{\mathbf{i}}\}$ con $V_{\mathbf{i}} = \{1, 2, 3, 4, 5\}$ y $E_{\mathbf{i}} = \{\{1, 2\}, \{1, 4\}, \{2, 2\}, \{2, 3\}, \{2, 4\}, \{2, 5\}, \{3, 5\}\}$.

Entonces $w_{\mathbf{i}}$ es un camino cerrado en $G_{\mathbf{i}}$ formado por la sucesión:

$$w_{\mathbf{i}} = (\{i_1, i_2\}, \{i_2, i_3\}, \dots, \{i_{k-1}, i_k\}, \{i_k, i_1\}).$$

Para el ejemplo anterior, el conjunto de índices \mathbf{i} también define el camino cerrado $w_{\mathbf{i}}$, que codifica $Y_{\mathbf{i}} = Y_{12} Y_{22} Y_{23} Y_{35} Y_{52} Y_{24} Y_{41} Y_{14} Y_{42} Y_{21}$, el cual puede ser interpretado acorde la siguiente representación.



Si ahora tenemos en cuenta que $Y_{ij} = Y_{ji} \forall i, j$ (simetría), entonces podemos considerar el enlace $\{i, j\} = \{j, i\} \forall i, j = 1, \dots, n$. Adicionalmente, escribiremos $w_{\mathbf{i}}(\{i, j\})$ para referirnos

al número de veces que se atraviesa el enlace $\{i, j\}$, por lo que podemos expresar $\mathbf{Y}_i = \prod_{1 \leq i, j \leq n} Y_{ij}^{w_i(\{i, j\})}$.

Como Y_{ij} son independientes $\forall i, j$, y como la diagonal puede tener distinta distribución de las entradas fuera de la diagonal, vamos a distinguir los enlaces propios $E_i^s = \{\{i, i\} \in E_i\}$ de los enlaces de conexión $E_i^c = \{\{i, j\} \in E_i : i \neq j\}$. Con esto en mente podemos reescribir $\mathbf{E}(\mathbf{Y}_i)$:

$$\mathbf{E}(\mathbf{Y}_i) = \prod_{1 \leq i, j \leq n} \mathbf{E}\left(Y_{ij}^{w_i(\{i, j\})}\right) = \prod_{e_s \in E_i^s} \mathbf{E}\left(Y_{11}^{w_i(e_s)}\right) \cdot \prod_{e_c \in E_i^c} \mathbf{E}\left(Y_{12}^{w_i(e_c)}\right).$$

Por tanto, el valor de $\mathbf{E}(\mathbf{Y}_i)$ está determinado por (E_i, w_i) . Denotemos como $\Pi(G_i, w_i)$ al valor

$$\Pi(G_i, w_i) := \mathbf{E}(\mathbf{Y}_i) = \prod_{e_s \in E_i^s} \mathbf{E}\left(Y_{11}^{w_i(e_s)}\right) \cdot \prod_{e_c \in E_i^c} \mathbf{E}\left(Y_{12}^{w_i(e_c)}\right).$$

Notemos que para un valor k fijado, el grafo orientado conexo G_i tiene k vértices como máximo. Como w_i recoge todos los valores no nulos de \mathbf{Y}_i tenemos que $|w_i| = \sum_{e \in E_i} w_i(e) = k$.

Denotemos como \mathcal{G}_k al conjunto de pares (G, w) con $G = \{V, E\}$ un grafo conexo con un máximo de k vértices, y w es un camino cerrado por G de longitud k , es decir, $|w| = k$.

$$\mathbf{E}[\text{traza}(\mathbf{Y}_n^k)] = \sum_{(G, w) \in \mathcal{G}_k} \sum_{\substack{\mathbf{i} \in [n]^k \\ (G_i, w_i) = (G, w)}} \mathbf{E}(\mathbf{Y}_i) = \sum_{(G, w) \in \mathcal{G}_k} \Pi(G, w) \cdot \#\{\mathbf{i} \in [n]^k : (G_i, w_i) = (G, w)\}.$$

Por lo tanto, sustituyendo en (2.4):

$$\frac{1}{n} \mathbf{E}[\text{traza}(\mathbf{X}_n^k)] = \sum_{(G, w) \in \mathcal{G}_k} \Pi(G, w) \cdot \frac{\#\{\mathbf{i} \in [n]^k : (G_i, w_i) = (G, w)\}}{n^{k/2+1}}. \quad (2.5)$$

Nos centramos ahora en calcular el valor $\#\{\mathbf{i} \in [n]^k : (G_i, w_i) = (G, w)\}$. Para cualquier $(G, w) \in \mathcal{G}_k$, un \mathbf{i} que se corresponde con el grafo G y el camino w está completamente determinado por los distintos valores de $[n]$ que aparecen en los vértices de G . El número de valores distintos que pueden aparecer en \mathbf{i} es $|V|$, es decir, el número total de vértices de G . Para el primero de los valores tenemos n posibilidades, para el segundo $n - 1$, y así hasta escoger $|V|$ valores diferentes entre $[n] = \{1, \dots, n\}$ para etiquetar los vértices de G . Así que:

$$\#\{\mathbf{i} \in [n]^k : (G_i, w_i) = (G, w)\} = \underbrace{n(n-1) \dots (n - |V| + 1)}_{|V| \text{ factores}}.$$

Por lo tanto

$$\frac{1}{n} \mathbf{E}[\text{traza}(\mathbf{X}_n^k)] = \sum_{(G, w) \in \mathcal{G}_k} \Pi(G, w) \cdot \frac{n(n-1) \dots (n - |V| + 1)}{n^{k/2+1}}.$$

Adicionalmente, supongamos que existe un enlace $e = \{i, j\}$ con $w(e) = 1$. Como $\mathbf{E}(Y_{ij}) = 0 \forall i, j$, tenemos que $\mathbf{E}(\mathbf{Y}_i) = \prod_{1 \leq i, j \leq n} \mathbf{E}(Y_{ij}^{w_i(\{i, j\})}) = 0$, por lo que reducimos la suma anterior teniendo en cuenta ahora que $w \geq 2$:

$$\frac{1}{n} \mathbf{E}[\text{traza}(\mathbf{X}_n^k)] = \sum_{\substack{(G, w) \in \mathcal{G}_k \\ w \geq 2}} \Pi(G, w) \cdot \frac{n(n-1) \dots (n - |V| + 1)}{n^{k/2+1}} \quad (2.6)$$

Ahora nos preocupamos de desentrañar el valor $\Pi(G, w)$. La condición $w \geq 2$ restringe el número de grafos que puede aparecer. Como $|w_i| = k$, cada enlace de G_i es atravesado como mínimo dos veces, es decir, el número de enlaces en G_i es $\leq k/2$. Como además el grafo es conexo, el número de enlaces mínimo es $|V| - 1$:

$$|V| - 1 \leq |E| \leq k/2 \implies |V| \leq k/2 + 1.$$

Notemos que el producto $n(n-1)\dots(n-|V|+1)$ es como mucho $n^{|V|}$. Si k es impar, $|V| \leq k/2 + 1 \implies |V| \leq \frac{k-1}{2} + \frac{1}{2}$, por lo que obtenemos

$$\begin{aligned} (2.6) &= \sum_{\substack{(G,w) \in \mathcal{G}_k \\ w \geq 2}} \Pi(G, w) \cdot \frac{n(n-1)\dots(n-|V|+1)}{n^{k/2+1}} \leq \\ &\leq \sum_{\substack{(G,w) \in \mathcal{G}_k \\ w \geq 2}} \Pi(G, w) \cdot \frac{n^{|V|}}{n^{k/2+1}} \leq \sum_{\substack{(G,w) \in \mathcal{G}_k \\ w \geq 2}} \Pi(G, w) \cdot \frac{n^{k/2+1/2}}{n^{k/2+1}} \sim O(n^{-1/2}). \end{aligned}$$

Por tanto, (2.6) tiende a cero cuando $n \rightarrow \infty$ si k es impar, lo que implica que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{E} [\text{traza}(\mathbf{X}_n^k)] = 0 \text{ si } k \text{ es impar.}$$

A partir de ahora supondremos que k es par.

El último cálculo nos dice que los únicos caminos cerrados a considerar son los que pasan por $k/2 + 1$ vértices $\implies |V| = k/2 + 1$, y por consiguiente las desigualdades $|V| - 1 \leq |E| \leq k/2$ se convierten en igualdades: $|E| = |V| - 1 = k/2$. Ahora es cuando recordamos que los grafos cuyo número de enlaces es igual a su número de vértices menos uno son los árboles.

Argumentando como antes obtenemos que $n(n-1)\dots(n-|V|+1) \leq n^{k/2+1}$, y entonces $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{E} [\text{traza}(\mathbf{X} : n^k)] = \sum_{(G,w) \in \mathcal{G}_k^{k/2+1}} \Pi(G, w)s$, donde $\mathcal{G}_k^{k/2+1}$ es el conjunto de pares (G, w) con G grafo conexo con $k/2 + 1$ vértices, sin enlaces propios (al ser un árbol), y el camino cerrado w recorre cada enlace de G exactamente 2 veces.

Sea $(G, w) \in \mathcal{G}_k^{k/2+1}$, gracias a toda la información anterior obtenemos

$$\Pi(G, w) = \prod_{e \in E^c} \mathbf{E}(Y_{12}^{w(e_c)}) = \prod_{e_c \in E^c} \mathbf{E}(Y_{12}^2) = \rho^{|E|} = \rho^{k/2}.$$

Por consiguiente

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{E} [\text{traza}(\mathbf{X}_n^k)] = \rho^{k/2} |\mathcal{G}_k^{k/2+1}|.$$

Finalmente debemos calcular el valor $|\mathcal{G}_k^{k/2+1}|$, es decir, contar el número de posibles grafos conexos con $k/2 + 1$ vértices y recorridos por un cierto camino w de longitud k .

Para esto usaremos una estructura combinatoria llamada *Dyck path* o camino de Dyck, introducida en la sección (2.3).

Dado $(G, w) \in \mathcal{G}_k^{k/2+1}$, definimos $d = d(G, w) \in \{1, -1\}^k$ como sigue:

$$\begin{aligned} d_1 &= 1 \\ d_j &= \begin{cases} 1, & \text{si } w_j \notin \{w_1, \dots, w_{j-1}\} \\ -1, & \text{en otro caso} \end{cases} \quad \forall j \in \{2, \dots, k\} \end{aligned}$$

$$\implies d = d(G, w) = (d_1, d_2, \dots, d_k) \in \mathcal{D}_k.$$

Además, d cumple las siguientes propiedades:

- a) El número de incrementos $+1$ es igual al número de incrementos -1 : $d_1 + \dots + d_k = 0$.
- b) Sumas parciales no negativas: $d_1 + \dots + d_j \geq 0 \ \forall j \leq k$.

Por otro lado, dado un árbol G con $k/2 + 1$ vértices y un camino de Dyck $d \in \mathcal{D}_k$, podemos definir a partir de d un camino w de longitud k tal que $(G, w) \in \mathcal{G}_k^{k/2+1}$:

$$w_j = \begin{cases} w_1 = 1, \\ w_j = w_s + 1 \text{ con } w_s = \max\{w_1, \dots, w_{j-1}\}, & \text{si } d_{j-1} = 1 \\ w_j = \text{Padre}(w_{j-1}), & \text{si } d_{j-1} = -1 \end{cases} \quad \forall j \in \{2, \dots, k\}$$

Por consiguiente, existe una biyección entre $\mathcal{G}_k^{k/2+1}$ y \mathcal{D}_k . Esto nos permite calcular el valor $|\mathcal{G}_k^{k/2+1}|$ debido a que $|\mathcal{D}_k|$ es conocido gracias al teorema (2.3.5) $\implies |\mathcal{G}_k^{k/2+1}| = |\mathcal{D}_k| = C_{k/2}$.

Tras todo este razonamiento llegamos a que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{E} [\text{traza}(\mathbf{X}_n^k)] = \rho^{k/2} C_{k/2} \text{ si } k \text{ es par.}$$

Resumiendo, hemos probado el primer paso, que afirma que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{E} [\text{traza}(\mathbf{X}_n^k)] = \begin{cases} \rho^{k/2} C_{k/2}, & \text{si } k \text{ es par} \\ 0, & \text{si } k \text{ es impar} \end{cases} \quad (2.7)$$

Paso 2: Empecemos desarrollando $\mathbf{V}(\frac{1}{n} \text{traza}(\mathbf{X}_n^k))$:

$$\begin{aligned} \mathbf{V} \left(\frac{1}{n} \text{traza}(\mathbf{X}_n^k) \right) &= \mathbf{E} \left[\left(\frac{1}{n} \frac{1}{n^{k/2}} \text{traza}(\mathbf{Y}_n) \right)^2 \right] - \left(\mathbf{E} \left[\frac{1}{n} \frac{1}{n^{k/2}} \text{traza}(\mathbf{Y}_n) \right] \right)^2 = \\ &= \frac{1}{n^{k+2}} \left\{ \mathbf{E} [\text{traza}(\mathbf{Y}_n^k)]^2 - (\mathbf{E} [\text{traza}(\mathbf{Y}_n^k)])^2 \right\}. \end{aligned}$$

Por un lado, $\mathbf{E} [\text{traza}(\mathbf{Y}_n^k)]^2 = \sum_{\mathbf{i}, \mathbf{j} \in [n]^k} \mathbf{E}(\mathbf{Y}_{\mathbf{i}} \mathbf{Y}_{\mathbf{j}})$.

Por otro lado, $(\mathbf{E} [\text{traza}(\mathbf{Y}_n^k)])^2 = \sum_{\mathbf{i}, \mathbf{j} \in [n]^k} \mathbf{E}(\mathbf{Y}_{\mathbf{i}}) \mathbf{E}(\mathbf{Y}_{\mathbf{j}})$.

Entonces la varianza es

$$\mathbf{V} \left(\frac{1}{n} \text{traza}(\mathbf{X}_n^k) \right) = \frac{1}{n^{k+2}} \sum_{\mathbf{i}, \mathbf{j} \in [n]^k} [\mathbf{E}(\mathbf{Y}_{\mathbf{i}} \mathbf{Y}_{\mathbf{j}}) - \mathbf{E}(\mathbf{Y}_{\mathbf{i}}) \mathbf{E}(\mathbf{Y}_{\mathbf{j}})].$$

Los razonamientos que se exponen a continuación son análogos a algunos del paso 1.

Primero, los valores $\mathbf{E}(\mathbf{Y}_{\mathbf{i}} \mathbf{Y}_{\mathbf{j}})$ y $\mathbf{E}(\mathbf{Y}_{\mathbf{i}}) \mathbf{E}(\mathbf{Y}_{\mathbf{j}})$ solo dependen de \mathbf{i} y \mathbf{j} a través de una cierta estructura de un grafo que representa la $2k$ -tupla \mathbf{i}, \mathbf{j} .

Sea $G_{\mathbf{i} \# \mathbf{j}} \equiv G_{\mathbf{i}} \cup G_{\mathbf{j}}$, donde la unión de dos grafos es el grafo cuyos vértices son la unión de los vértices de $G_{\mathbf{i}}$ y $G_{\mathbf{j}}$, y sus enlaces son la unión de los enlaces de $G_{\mathbf{i}}$ y $G_{\mathbf{j}}$.

$\mathbf{Y}_{\mathbf{i}} \mathbf{Y}_{\mathbf{j}}$ está, de nuevo, representado/codificado por los caminos $w_{\mathbf{i}}$ y $w_{\mathbf{j}}$. Observar que $w_{\mathbf{i}}$ y $w_{\mathbf{j}}$ no constituyen un camino de $2k$ pasos ya que, a priori, no tiene por qué acabar el primer camino donde empieza el segundo.

Como podemos recuperar (G_i, w_i) y (G_j, w_j) de $(G_{i\#j}, w_i, w_j)$, el mismo razonamiento que en el paso anterior puede ser aplicado para ver que la cantidad $\mathbf{E}(\mathbf{Y}_i \mathbf{Y}_j) - \mathbf{E}(\mathbf{Y}_i) \mathbf{E}(\mathbf{Y}_j)$ está determinada por $(G_{i\#j}, w_i, w_j)$. Denotemos a este valor como

$$\Pi(G_{i\#j}, w_i, w_j) = \mathbf{E}(\mathbf{Y}_i \mathbf{Y}_j) - \mathbf{E}(\mathbf{Y}_i) \mathbf{E}(\mathbf{Y}_j).$$

Sea ahora $\mathcal{G}_{k,k}$ el conjunto de grafos conexos G con $|V| \leq 2k$, junto con dos caminos de longitud k cada uno cuya unión cubre G . Esto significa que podemos escribir:

$$\mathbf{V} \left(\frac{1}{n} \text{traza}(\mathbf{X}_n^k) \right) = \frac{1}{n^{k+2}} \sum_{(G, w, w') \in \mathcal{G}_{k,k}} \Pi(G, w, w') \cdot \#\{(\mathbf{i}, \mathbf{j}) \in [n]^{2k} : (G_{i\#j}, w_i, w_j) = (G, w, w')\}.$$

Como hicimos anteriormente, expandimos la expresión anterior separando los enlaces propios de los enlaces de conexión. Denotemos como $E_{i\#j}^c$ a los enlaces de conexión en $G_{i\#j}$, como $E_{i\#j}^s$ los enlaces propios, y $w_{i\#j}(e)$ al número de veces que el enlace e es atravesado por al menos uno de los caminos w_i y w_j .

Entonces podemos reescribir $\mathbf{E}(\mathbf{Y}_i \mathbf{Y}_j) - \mathbf{E}(\mathbf{Y}_i) \mathbf{E}(\mathbf{Y}_j)$ como

$$\begin{aligned} \mathbf{E}(\mathbf{Y}_i \mathbf{Y}_j) - \mathbf{E}(\mathbf{Y}_i) \mathbf{E}(\mathbf{Y}_j) &= \prod_{e_s \in E_{i\#j}^s} \mathbf{E}(Y_{11}^{w_{i\#j}(e_s)}) \cdot \prod_{e_c \in E_{i\#j}^c} \mathbf{E}(Y_{12}^{w_{i\#j}(e_c)}) - \\ &- \prod_{e_s \in E_i^s} \mathbf{E}(Y_{11}^{w_i(e_s)}) \cdot \prod_{e_c \in E_i^c} \mathbf{E}(Y_{12}^{w_i(e_c)}) \cdot \prod_{e_s \in E_j^s} \mathbf{E}(Y_{11}^{w_j(e_s)}) \cdot \prod_{e_c \in E_j^c} \mathbf{E}(Y_{12}^{w_j(e_c)}). \end{aligned}$$

Es importante notar que $\sum_{e \in G_{i\#j}} w_{i\#j}(e) = 2k = \sum_{e \in G_i} w_i(e) + \sum_{e \in G_j} w_j(e)$, es decir, la suma de todos los exponentes en $\mathbf{E}(\mathbf{Y}_i \mathbf{Y}_j)$ o en $\mathbf{E}(\mathbf{Y}_i) \mathbf{E}(\mathbf{Y}_j)$ es la longitud total de w_i y w_j , que es $2k$.

Cada uno de estos términos está acotado en módulo por algo de la forma r_{m_1}, \dots, r_{m_l} para ciertos enteros positivos m_1, \dots, m_l tal que $m_1 + \dots + m_l = 2k$. Recordar que $r_m = \max\{\mathbf{E}(Y_{11}^m), \mathbf{E}(Y_{12}^m)\}$. Hay un número finito de particiones de longitud $2k$, y el máximo M_{2k} sobre todas ellas es finito. Por lo tanto tenemos la (potencialmente inmensa) acotación

$$|\Pi(G_{i\#j}, w_i, w_j)| = |\mathbf{E}(\mathbf{Y}_i \mathbf{Y}_j) - \mathbf{E}(\mathbf{Y}_i) \mathbf{E}(\mathbf{Y}_j)| \leq 2M_{2k} \forall \mathbf{i}, \mathbf{j} \in [n]^k.$$

Dicho esto, podemos mostrar que muchos términos en $\Pi(G_{i\#j}, w_i, w_j)$ son idénticamente cero. Por construcción, cada enlace de $G_{i\#j}$ es atravesado al menos una vez por la unión de los dos caminos w_i y w_j .

Supongamos que e es un enlace atravesado una sola vez. Esto significa que $w_{i\#j}(e) = 1 \implies w_i, w_j \in \{0, 1\}$. Por consiguiente, la expresión anterior y el hecho de que $\mathbf{E}(Y_{11}) = \mathbf{E}(Y_{12}) = 0$ implican que $\Pi(G_{i\#j}, w_i, w_j) = 0$. Por lo tanto, nos podemos centrar en los grafos de $\mathcal{G}_{k,k}$ que cada enlace es recorrido al menos dos veces:

$$\mathbf{V} \left(\frac{1}{n} \text{traza}(\mathbf{X}_n^k) \right) = \sum_{\substack{(G, w, w') \in \mathcal{G}_{k,k} \\ w + w' \geq 2}} \Pi(G, w, w') \cdot \frac{\#\{(\mathbf{i}, \mathbf{j}) \in [n]^{2k} : (G_{i\#j}, w_i, w_j) = (G, w, w')\}}{n^{k+2}}.$$

La enumeración de las $2k$ -tuplas que describe cierto grafo con dos caminos es igual que en la sección anterior: la estructura (G, w, w') está determinada una vez escogemos $|V|$ índices distintos para etiquetar los vértices. Por tanto, como antes, tenemos $n(n-1) \dots (n-|V|+1)$ posibles elecciones. Recordar además que $n(n-1) \dots (n-|V|+1) \leq n^{|V|}$.

Retomando la condición $w + w' \geq 2$ (cada enlace es atravesado al menos dos veces), podemos sostener que hay como máximo k vértices: $|V| - 1 \stackrel{(*)}{\leq} |E| \stackrel{(**)}{\leq} k \implies |V| \leq k + 1$. Donde en $(*)$ se ha usado que G es un grafo conexo, y en $(**)$ que $w + w' \geq 2$.

Con todo esto en mente podemos proceder de la siguiente manera

$$\begin{aligned} n(n-1) \dots (n-|V|+1) &\leq n^{|V|} \leq n^{k+1} \implies \\ \implies \mathbf{V} \left(\frac{1}{n} \text{traza}(\mathbf{X}_n^k) \right) &\leq \sum_{\substack{(G,w,w') \in \mathcal{G}_{k,k} \\ w+w' \geq 2}} \Pi(G,w,w') \cdot \frac{n^{k+1}}{n^{k+2}} \leq \frac{1}{n} \underbrace{2M_{2k} \cdot |\mathcal{G}_{k,k}|}_{\text{independiente de } n} \implies \\ &\implies \mathbf{V} \left(\frac{1}{n} \text{traza}(\mathbf{X}_n^k) \right) \sim O(1/n) \xrightarrow{n \rightarrow \infty} 0. \end{aligned} \quad (2.8)$$

Esto es suficiente para probar el teorema (2.1.6). \square

2.5. Demostración de la versión 2 del Teorema de Wigner

Resulta que si las densidades $\mu_{\mathbf{X}_n}$ estuviesen todas ellas concentradas en un único intervalo acotado para n finito, entonces la demostración de la versión 2 (2.1.5) a partir de la versión 3 (2.1.6) sería inmediata simplemente usando el Teorema de aproximación de Weierstrass. Sin embargo, $\mu_{\mathbf{X}_n}$ no está acotada para un n finito, por lo que tendremos que usar adicionalmente un argumento de truncación.

Antes de comenzar la demostración del teorema (2.1.5) enunciamos el siguiente resultado, que nos será de utilidad para la misma.

Lema 2.5.1. Sea $k \in \mathbb{N}$ y $\epsilon > 0$. Entonces para cualquier $B > 4$,

$$\limsup_{n \rightarrow \infty} \mathbf{P} \left(\int_{|x| > B} |x|^k \mu_{\mathbf{X}_n}(dx) > \epsilon \right) = 0.$$

Demostración. Primero, por la desigualdad de Markov (3.2.1), tenemos

$$\mathbf{P} \left(\int_{|x| > B} |x|^k \mu_{\mathbf{X}_n}(dx) > \epsilon \right) \leq \frac{1}{\epsilon} \mathbf{E} \left(\int_{|x| > B} |x|^k \mu_{\mathbf{X}_n}(dx) \right).$$

Ahora, sea $\nu(dx) = |x|^k \mu_{\mathbf{X}_n}(dx)$. La desigualdad de Markov (que se aplica para todas las medidas positivas, no solo para las medidas de probabilidad) muestra que

$$\int_{|x| > B} |x|^k \mu_{\mathbf{X}_n}(dx) = \nu\{x : |x| > B\} = \nu\{x : |x|^k > B^k\} \leq \frac{1}{B^k} \int |x|^k \nu(dx) = \frac{1}{B^k} \int x^{2k} \mu_{\mathbf{X}_n}(dx).$$

Tomando esperanzas obtenemos

$$\mathbf{P} \left(\int_{|x| > B} |x|^k \mu_{\mathbf{X}_n}(dx) > \epsilon \right) \leq \frac{1}{\epsilon B^k} \mathbf{E} \left(\int x^{2k} \mu_{\mathbf{X}_n}(dx) \right) = \frac{1}{\epsilon B^k} \cdot \frac{1}{n} \mathbf{E}[\text{traza}(\mathbf{X}_n^k)].$$

Por la expresión (2.7) sabemos que la parte derecha converge a $C_k/\epsilon B^k$, donde C_k es el número de Catalan, el cual está acotado por 4^k . Por tanto, se sigue que

$$\limsup_{n \rightarrow \infty} \mathbf{P} \left(\int_{|x| > B} |x|^k \mu_{\mathbf{X}_n}(dx) > \epsilon \right) \leq \frac{1}{\epsilon} \left(\frac{4}{B} \right)^k.$$

Por otra parte, cuando $|x| > B > 4 > 1$, la función $k \mapsto |x|^k$ es estrictamente creciente, lo que quiere decir que la sucesión de límites superiores es creciente. Pero esta sucesión decrece exponencialmente porque $4/B < 1$. La única forma de que esto sea posible es que la sucesión de límites superiores es constatemente 0. \square

Ya estamos listos para enfrentarnos a la demostración del teorema (2.1.5) a partir de lo que ya sabemos del teorema (2.1.6).

Demostración. Fijamos una función continua acotada $f \in C_a(\mathbb{R})$, un $\epsilon > 0$, y $B > 4$. Por el Teorema de aproximación de Weierstrass sabemos que existe un polinomio P_ϵ tal que

$$\sup_{|x| \leq B} |f(x) - P_\epsilon(x)| \leq \frac{\epsilon}{6}.$$

Ahora, por la desigualdad triangular tenemos

$$\left| \int f d\mu_{\mathbf{X}_n} - \int f d\sigma \right| \leq \left| \int f d\mu_{\mathbf{X}_n} - \int P_\epsilon d\mu_{\mathbf{X}_n} \right| + \left| \int P_\epsilon d\mu_{\mathbf{X}_n} - \int P_\epsilon d\sigma \right| + \left| \int P_\epsilon d\sigma - \int f d\sigma \right|.$$

Por lo tanto, el evento $\{|\int f d\mu_{\mathbf{X}_n} - \int f d\sigma| > \epsilon\}$ está contenido en la unión de los tres eventos anteriores, los cuales están acotados por $\epsilon/3$. esto significa que

$$\begin{aligned} \mathbf{P} \left(\left| \int f d\mu_{\mathbf{X}_n} - \int f d\sigma \right| > \epsilon \right) &\leq \mathbf{P} \left(\left| \int f d\mu_{\mathbf{X}_n} - \int P_\epsilon d\mu_{\mathbf{X}_n} \right| > \epsilon/3 \right) + \\ &+ \mathbf{P} \left(\left| \int P_\epsilon d\mu_{\mathbf{X}_n} - \int P_\epsilon d\sigma \right| > \epsilon/3 \right) + \mathbf{P} \left(\left| \int P_\epsilon d\sigma - \int f d\sigma \right| > \epsilon/3 \right). \end{aligned}$$

Por construcción, $|P_\epsilon - f| < \epsilon/6$ en $[-B, B]$, lo que incluye el intervalo $[-2, 2]$ de σ , por lo que el último término es idénticamente 0. Adicionalmente, para el primer término, separamos la integral sobre $[-B, B]$ y su complementario:

$$\left| \int (f - P_\epsilon) d\mu_{\mathbf{X}_n} \right| \leq \int_{|x| \leq B} |f(x) - P_\epsilon(x)| d\mu_{\mathbf{X}_n}(dx) + \int_{|x| > B} |f(x) - P_\epsilon(x)| d\mu_{\mathbf{X}_n}(dx).$$

Con un razonamiento análogo al realizado anteriormente, podemos estimar:

$$\begin{aligned} \mathbf{P} \left(\left| \int f d\mu_{\mathbf{X}_n} - \int P_\epsilon d\mu_{\mathbf{X}_n} \right| > \epsilon/3 \right) &\leq \\ &\leq \mathbf{P} \left(\int |f - P_\epsilon| \mathbf{1}_{|x| \leq B} d\mu_{\mathbf{X}_n} > \epsilon/6 \right) + \mathbf{P} \left(\int |f - P_\epsilon| \mathbf{1}_{|x| > B} d\mu_{\mathbf{X}_n} > \epsilon/6 \right). \end{aligned} \quad (2.9)$$

El último término en la la expresión anterior tiende a 0 cuando $n \rightarrow \infty$ gracias a la tercera versión de la Ley de Wigner (2.1.6) (debido a que la convergencia en probabilidad se conserva bajo sumas y multiplicaciones por escalares). Nos faltaría controlar el primer término de la parte derecha de (2.9). Para ello, hacemos lo siguiente: sea $k = \deg(P_\epsilon) = \text{grado}(P_\epsilon)$. Como f es acotada, tenemos que $|f(x) - P_\epsilon(x)| \leq \|f\|_\infty + |P_\epsilon(x)|$, y en el conjunto $\{x : |x| > B\}$, $\|f\|_\infty + |P_\epsilon(x)| \leq c|x|^k$ para una cierta constante c (debido a que $B > 0$). Esto significa que

$$\mathbf{P} \left(\int |f - P_\epsilon| \mathbf{1}_{|x| \geq B} d\mu_{\mathbf{X}_n}(dx) > \epsilon/6 \right) \leq \mathbf{P} \left(\int c|x|^k \mathbf{1}_{|x| \geq B} d\mu_{\mathbf{X}_n}(dx) > \epsilon/6 \right).$$

Por el lema (2.5.1), esta sucesión tiene $\limsup_{n \rightarrow \infty} = 0$, que concluye la demostración. \square

2.6. Demostración de la versión 1 del Teorema de Wigner

Para terminar queremos demostrar la versión 1 (2.1.4) de la Ley Semicircular de Wigner a partir de las versiones 2 y 3 ya demostradas. Sabemos que la variable aleatoria $E_n(I)$, con $[a, b] = I \subset \mathbb{R}$, se puede escribir como:

$$E_n(I) = \frac{1}{n} \cdot \#\{j \in [n] : \lambda_j(\mathbf{X}_n) \in I\} = \mu_{\mathbf{X}_n}(I).$$

Demostración. La conclusión deseada es que $\mu_{\mathbf{X}_n}(I) \rightarrow \mu_{sc}(I)$ en probabilidad $\forall I \subset \mathbb{R}$. Por un sencillo argumento de escala podemos suponer $\rho = 1$, por lo que en el resto de este desarrollo, μ_{sc} será la Ley semicircular de Wigner con parámetro $\rho = 1$.

Razonando como en la expresión (2.1) podemos desarrollar $E_n(I)$:

$$E_n(I) = \mu_{\mathbf{X}_n} = \int \mathbf{1}_I d\mu_{\mathbf{X}_n} = \frac{1}{n} \text{traza}(\mathbf{1}_I(\mathbf{X}_n)) .$$

Ahora definimos la función $f_\epsilon \in C_b(\mathbb{R})$ de la siguiente manera

$$f_\epsilon(x) = \begin{cases} 1, & \text{si } x \in I = [a, b] \\ \frac{x-a+\epsilon}{\epsilon}, & \text{si } x \in [a-\epsilon, a] \\ \frac{b-x+\epsilon}{\epsilon}, & \text{si } x \in (b, b+\epsilon] \\ 0, & \text{en otro caso} \end{cases}$$

Por construcción, $\mathbf{1}_I(x) \leq f_\epsilon(x) \forall x \in \mathbb{R}$, y por tanto, usando la 2ª versión del Teorema de Wigner, obtenemos

$$\frac{1}{n} \text{traza}(\mathbf{1}_I(\mathbf{X}_n)) \leq \frac{1}{n} \text{traza}(f_\epsilon(\mathbf{X}_n)) \implies \mu_{\mathbf{X}_n}(I) \leq \int f_\epsilon d\mu_{\mathbf{X}_n} \xrightarrow{n \rightarrow \infty} \int f_\epsilon d\mu_{sc}.$$

Por consiguiente

$$\limsup \mu_{\mathbf{X}_n}(I) \leq \int f_\epsilon d\mu_{sc} \xrightarrow[\epsilon \rightarrow 0]{(*)} \int \mathbf{1}_I d\mu_{sc} = \mu_{sc}(I), \quad (2.10)$$

donde en $(*)$ se ha usado el teorema de convergencia dominada.

Considerando ahora la función \tilde{f}_ϵ dada por

$$\tilde{f}_\epsilon(x) = \begin{cases} 1, & \text{si } x \in I = [a+\epsilon, b-\epsilon] \\ \frac{x-a}{\epsilon}, & \text{si } x \in [a, a+\epsilon) \\ \frac{b-x}{\epsilon}, & \text{si } x \in (b-\epsilon, b] \\ 0, & \text{en otro caso} \end{cases}$$

Por construcción, $\tilde{f}_\epsilon \leq \mathbf{1}_I$, y por un argumento análogo al anterior llegamos a

$$\liminf \mu_{\mathbf{X}_n}(I) \geq \int \tilde{f}_\epsilon d\mu_{sc} \xrightarrow[\epsilon \rightarrow 0]{} \int \mathbf{1}_I d\mu_{sc} = \mu_{sc}(I). \quad (2.11)$$

Finalmente, juntando (2.10) y (2.11), llegamos a que $\mu_{\mathbf{X}_n}(I) \xrightarrow{n \rightarrow \infty} \mu_{sc}(I)$ en probabilidad. \square

2.7. Ley de Tracy-Widom

En la Teoría de Matrices Aleatorias es común preguntarse no sólo por la distribución límite de los autovalores de las matrices aleatorias, sino también por el comportamiento límite del mayor autovalor de las mismas.

Como hemos visto, el soporte de μ_{sc} con parámetro $\rho = 1$ es $[-2, 2]$, lo que sugiere que el mayor autovalor de una matriz de Wigner $\mathbf{X}_n = 1/\sqrt{n}\mathbf{Y}_n$ converge a 2 cuando $n \rightarrow \infty$.

Teorema 2.7.1. *Sea \mathbf{X}_n una matriz de Wigner. Entonces $\lambda_n(\mathbf{X}_n)$ converge a 2 en probabilidad, es decir, para cualquier $\epsilon, \delta > 0$ tenemos*

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(n^{1/6-\epsilon} (\lambda_n(\mathbf{X}_n) - 2) > \delta \right) = 0.$$

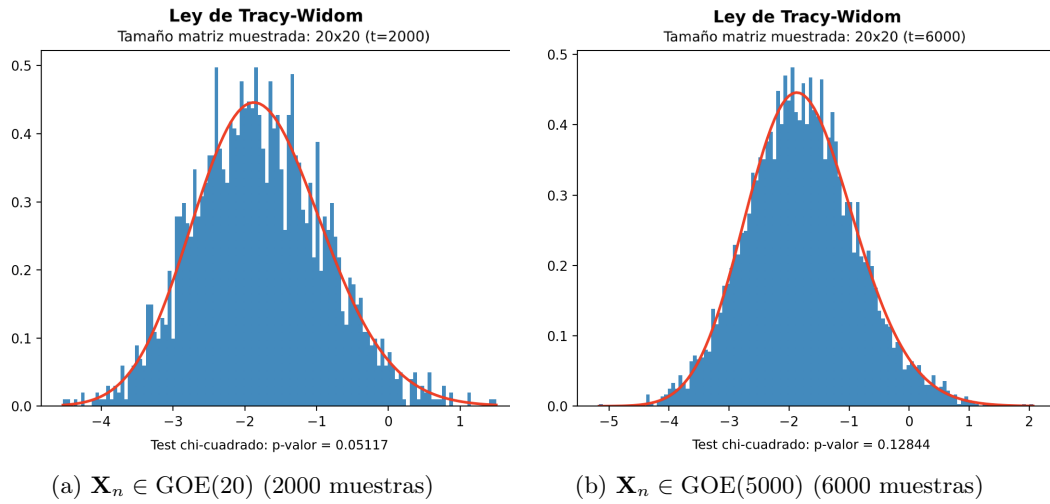


Figura 2.4: Simulación de la FDP de la Ley de Tracy-Widom para $\mathbf{X}_n \in \text{GOE}(n)$

Por otro lado, es intuitivo pensar que el menor autovalor de una matriz de Wigner \mathbf{X}_n converge a -2 cuando $n \rightarrow \infty$. Podemos argumentarlo recordando que si \mathbf{X}_n es una matriz de Wigner, entonces $-\mathbf{X}_n$ también lo es, por lo que si el mayor autovalor de \mathbf{X}_n converge a 2 cuando $n \rightarrow \infty$, el menor autovalor de $-\mathbf{X}_n$ convergerá a -2 cuando $n \rightarrow \infty$.

Teniendo en cuenta la propiedad de simetría anterior más el resultado (2.7.1), podemos afinar más y sostener que si \mathbf{X}_n es una matriz de Wigner, entonces $\lambda_1(\mathbf{X}_n)$ converge a -2 en probabilidad, es decir, dados $\epsilon, \delta > 0$, tenemos

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(n^{1/6-\epsilon} (\lambda_n(\mathbf{X}_n) + 2) > \delta \right) = 0.$$

Bibliografía

- [1] ANDERSON G. W., GUIONNET A., ZEITOUNI O.: *An Introduction to Random Matrices* (2009). Cambridge University Press.
- [2] BAI Z. D., SILVERSTEIN J. W.: *Spectral Analysis of Large Dimensional Random Matrices* (2010). 2nd ed., Springer.
- [3] BONDY A., MURTY U.S.: *Graph Theory* (2008). Springer-Verlag London.
- [4] BORDENAVE C.: *Lecture Notes on Random Matrix Theory* (2019).
- [5] COUILLET R., DEBBAH M.: *Random Matrix Methods for Wireless Communications* (2011). Cambridge University Press.
- [6] GHOSH M., SINHA B. K.: A Simple Derivation of the Wishart Distribution (2002). *The American Statistician*, 56:2, 100-101.
- [7] HOTELLING H.: The Generalization of Student's Ratio (1931). *Annals of Mathematical Statistics*, Vol. 2, No. 3.
- [8] JOHNSTONE I. M.: One the Distribution of the Largest Eigenvalue in Principal Component Analysis (2001). *The Annals of Statistics*, Vol. 29, No. 2, 295–327.
- [9] KEMP T.: Math 247A: Introduction to Random Matrix Theory (2016).
- [10] MARDIA K. V., KENT J. T., BIBBY J. M.: *Multivariate Analysis* (1979). Academic Press.
- [11] MUIRHEAD R. J.: *Aspects of Multivariate Statistical Theory* (1982). Wiley Series in Probability and Statistics.
- [12] *Lecture notes of Sungkyu Jung: The Wishart distribution* (2013). <https://www.stat.pitt.edu/sungkyu/course/2221Fall13/lec2.pdf>
- [13] WIGNER E.: Characteristic Vectors of Bordered Matrices With Infinite Dimensions (1955). *Annals of Mathematics*, Vol. 62, No. 3.
- [14] WIGNER E.: On the Distribution of the Roots of Certain Symmetric Matrices (1958). *Annals of Mathematics*, Vol. 67, No. 2.
- [15] ZAIONTZ C.: One Sample Hotelling's T-square. Real Statistics. Accedido el 27 de abril de 2021. <http://www.real-statistics.com/multivariate-statistics/hotellings-t-square-statistic/one-sample-hotellings-t-square/>

Apéndice

3.1. Matrices simétricas, definidas positivas

En este apéndice recogeremos las propiedades de las matrices simétricas y/o definidas positivas que ya conocemos, pero que no podemos olvidar para entender por completo el resto del trabajo.

3.1.1. Resultados sobre diagonalización

Teorema 3.1.1 (Teorema espectral). *Toda matriz simétrica A (hermítica en el caso complejo) es ortodiagonalizable, es decir, podemos escribir $A = O^{-1}\Lambda O$ con O matriz ortogonal formada por los autovectores (todos reales) de A y Λ matriz diagonal formada por los autovalores de A . Como O es ortogonal $O^{-1} = O^T$ y podemos escribir $A = O^{-1}\Lambda O = O^T\Lambda O$.*

Recordemos un concepto que extiende la noción de raíz cuadrada de los números a las matrices.

Definición 3.1.2. *Una matriz A se dice que admite una raíz cuadrada cuando existe matriz B tal que $A = B^2 = BB^T$.*

Proposición 3.1.3. *Toda matriz A simétrica semidefinida positiva admite una raíz cuadrada.*

Proposición 3.1.4. *Toda matriz A simétrica definida positiva admite una raíz cuadrada B no singular, es decir, invertible. Además, B es también simétrica definida positiva.*

Ahora exponemos un resultado que une el teorema (3.1.1) con la noción de matriz definida positiva.

Teorema 3.1.5. *Toda matriz A simétrica definida positiva es ortodiagonalizable y todos sus autovalores son reales y positivos, es decir, $A = O^{-1}\Lambda O$ con O matriz ortogonal formada por los autovectores de A y Λ matriz diagonal formada por los autovalores reales positivos de A .*

Lema 3.1.6. *La raíz cuadrada de una matriz A simétrica definida positiva se puede expresar como $B = O^{-1}\Lambda^{1/2}O$.*

El lema anterior se deriva del teorema (3.1.5), de la proposición (3.1.4) y del hecho de que las potencias de una matriz diagonal es la matriz con las potencias de los elementos de su diagonal.

$$A = O^{-1}\Lambda O = O^{-1}\Lambda^{1/2}\Lambda^{1/2}O = \underbrace{(O^{-1}\Lambda^{1/2}O)}_B \underbrace{(O^{-1}\Lambda^{1/2}O)}_B.$$

Adicionalmente, es fácil ver que $A^{-1} = (O^{-1}\Lambda O)^{-1} = O^{-1}\Lambda^{-1}(O^{-1})^{-1} = O^{-1}\Lambda^{-1}O$, y por lo tanto podemos obtener una expresión para la raíz cuadrada inversa de una matriz

$$A^{-\frac{1}{2}} = (A^{1/2})^{-1} = (O^{-1}\Lambda^{1/2}O)^{-1} = O^{-1}\Lambda^{-\frac{1}{2}}O.$$

Definición 3.1.7. *El complemento de Schur de un bloque de matriz se define de la manera siguiente. Supongamos que A, B, C y D son respectivamente matrices $p \times p$, $p \times q$, $q \times p$ y $q \times q$, y que D sea invertible. Sea*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

de modo que M es una matriz $(p+q) \times (p+q)$. Entonces el complemento de Schur del bloque D de la matriz M es la matriz $p \times p$

$$M/D = A - BD^{-1}C.$$

Gracias a la definición anterior podemos sacar una expresión peculiar para el determinante de una matriz.

Lema 3.1.8. *Sea la matriz M como en la definición anterior, entonces:*

$$\det(M) = \det(A)\det(D - CA^{-1}B).$$

3.1.2. Funciones características de variables aleatorias con valores en matrices simétricas

La definición de las funciones características se puede extender a las matrices aleatorias. Sea $A \in \mathcal{M}_p$, la función de característica de A evaluada en una matriz $T \in \mathcal{S}_p$ se define como

$$\varphi_A(T) = \mathbf{E}[\exp(i\text{traza}(TA))]. \quad (3.12)$$

Como ejemplo, en el caso particular $\mathbf{M} \sim W_p(n, \Sigma)$ tenemos el siguiente resultado.

Teorema 3.1.9 (Función característica de la distribución de Wishart). *Sea $\mathbf{M} \sim W_p(n, \Sigma)$, entonces*

$$\varphi_{\mathbf{M}}(T) = \det(I_p - 2i\Sigma T)^{-n/2}. \quad (3.13)$$

Demostración. Primero, usamos que por definición Σ es definida positiva y calculamos su descomposición de Cholesky, es decir, $\Sigma = LL^T$ con L matriz triangular inferior.

Fijémonos en que podemos escribir $\mathbf{M} = L \left(\sum_{j=1}^n \mathbb{Z}_j \mathbb{Z}_j^T \right) L^T$, donde $\mathbb{Z}_j \sim N_p(0, I_p)$, gracias a la propiedad (1.11).

Por otro lado, fijamos una matriz $T \in \mathcal{S}_p$. La matriz $L^T T L$ es también simétrica, y por lo tanto podemos computar su descomposición espectral gracias al teorema (3.1.1):

$$L^T T L = U \Lambda U^T,$$

donde $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ es diagonal y $U U^T = I_p$. Ahora podemos escribir:

$$\begin{aligned} \text{traza}(T\mathbf{M}) &= \text{traza} \left(T L \left(\sum_{j=1}^n \mathbb{Z}_j \mathbb{Z}_j^T \right) L^T \right) \stackrel{(*)}{=} \text{traza} \left(U \Lambda U^T \left(\sum_{j=1}^n \mathbb{Z}_j \mathbb{Z}_j^T \right) \right) \stackrel{(*)}{=} \\ &\stackrel{(*)}{=} \text{traza} \left(\Lambda U^T \left(\sum_{j=1}^n \mathbb{Z}_j \mathbb{Z}_j^T \right) U \right) = \text{traza} \left(\Lambda \left(\sum_{j=1}^n (U^T \mathbb{Z}_j)(U^T \mathbb{Z}_j^T)^T \right) \right), \end{aligned} \quad (3.14)$$

donde en (\star) se ha usado que la traza del producto de matrices cuadradas es cíclica, es decir, si $A_1, \dots, A_k \in \mathcal{M}_p$, entonces $\text{traza}(A_1 A_2 \dots A_k) = \text{traza}(A_2 \dots A_k A_1) = \dots = \text{traza}(A_k A_1 \dots A_{k-1})$.

Notemos dos observaciones clave:

- $U^T \mathbf{Z}_j \sim N_p(0, I_p)$, ya que $\mathbf{Z}_j \sim N_p(0, I_p)$ y U^T es una matriz ortogonal.
- $\text{traza}(\Lambda \mathbb{Z}_j \mathbb{Z}_j^T) = \sum_{k=1}^p \lambda_k Z_{jk}^2$, ya que recordemos que $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$.

Juntando todo lo anterior, obtenemos:

$$\mathbf{E}[\exp(i \text{traza}(T A))] = \mathbf{E} \left[\exp \left(i \sum_{j=1}^n \sum_{k=1}^p \lambda_k Z_{jk}^2 \right) \right] = \prod_{j=1}^n \prod_{k=1}^p \mathbf{E}[\exp(i \lambda_k Z_{jk}^2)].$$

Pero $Z_{jk}^2 \sim \chi_1^2$, y entonces tenemos que

$$\varphi_{\mathbf{M}}(T) = \prod_{j=1}^n \prod_{k=1}^p \varphi_{\chi_1^2}(\lambda_k) = \prod_{j=1}^n \prod_{k=1}^p (1 - 2i \lambda_k)^{-1/2}.$$

Como $\prod_{k=1}^p (1 - 2i \lambda_k)^{-1/2} = \det(I_p - 2i \Lambda)^{-1/2}$, entonces

$$\begin{aligned} \varphi_{\mathbf{M}}(T) &= \prod_{j=1}^n \det(I_p - 2i \Lambda)^{-1/2} = \det(I_p - 2i \Lambda)^{-n/2} = \det(I_p - 2i U \Lambda U^T)^{-n/2} \\ &= \det(I_p - 2i L^T T L)^{-n/2} = \det(I_p - 2i \Sigma T)^{-n/2}. \end{aligned} \quad (3.15)$$

□

3.1.3. Normas de matrices simétricas

Definición 3.1.10 (Norma de Frobenius). *La norma de Frobenius de una matriz simétrica \mathbf{X} es la suma de sus elementos al cuadrado, es decir*

$$\frac{1}{n} \|\mathbf{X}\|_2^2 = \frac{1}{n} \sum_{i=1}^n \lambda_i(\mathbf{X})^2 = \frac{1}{n} \sum_{1 \leq i, j \leq n} X_{ij}^2.$$

La última igualdad la podemos ratificar viendo que ambas expresiones son iguales a $\frac{1}{n} \text{traza}(\mathbf{X}^2)$. Para ello utilizaremos que \mathbf{X} es simétrica, y por tanto ortogonalizable (teorema espectral): $\mathbf{X} = O^{-1} \Lambda O$, donde Λ es la matriz diagonal formada por los autovalores de \mathbf{X} , y O es una matriz ortogonal formada por los autovectores de \mathbf{X} .

$$\frac{1}{n} \text{traza}(\mathbf{X}^2) = \frac{1}{n} \text{traza}(O^{-1} \Lambda O O^{-1} \Lambda O) = \frac{1}{n} \text{traza}(O^{-1} \Lambda^2 O) = \frac{1}{n} \text{traza}(\Lambda^2) = \frac{1}{n} \sum_{i=1}^n \lambda_i(\mathbf{X})^2.$$

Por otro lado, cada entrada (i, i) de la diagonal de la matriz \mathbf{X} es de la forma $\sum_{j=1}^n X_{ij} X_{ji}$. Como \mathbf{X} es simétrica, cada entrada (i, i) de la diagonal es igual a $\sum_{j=1}^n X_{ij}^2$. Calculando la traza obtenemos

$$\frac{1}{n} \text{traza}(\mathbf{X}^2) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n X_{ij}^2 = \frac{1}{n} \sum_{1 \leq i, j \leq n} X_{ij}^2.$$

3.2. Desigualdad de Markov

Lema 3.2.1 (Desigualdad de Markov). Sea $Y \geq 0$ una variable aleatoria con $\mathbf{E}(Y) < \infty$. Entonces para cualquier $\epsilon > 0$,

$$\mathbf{P}(Y > \epsilon) \leq \frac{\mathbf{E}(Y)}{\epsilon}.$$

Demostración. Es fácil ver que $\epsilon \mathbf{1}_{\{Y \geq \epsilon\}} \leq Y$, ya que $\epsilon \mathbf{1}_{\{Y \geq \epsilon\}}$ toma el valor ϵ cuando $Y \geq \epsilon$, y $\epsilon \mathbf{1}_{\{Y \geq \epsilon\}} = 0$ cuando $Y < \epsilon$. Tomando esperanzas a ambos lados obtenemos que $\mathbf{E}(\epsilon \mathbf{1}_{\{Y \geq \epsilon\}}) \leq \mathbf{E}(Y) \implies \epsilon \mathbf{E}(\mathbf{1}_{\{Y \geq \epsilon\}}) \leq \mathbf{E}(Y)$. Como $\mathbf{E}(\mathbf{1}_{\{Y \geq \epsilon\}}) = \mathbf{P}(Y \geq \epsilon)$ y $\epsilon > 0$, llegamos a

$$\epsilon \mathbf{P}(Y \geq \epsilon) \leq \mathbf{E}(Y) \implies \mathbf{P}(Y \geq \epsilon) \leq \frac{\mathbf{E}(Y)}{\epsilon}.$$

□

3.3. Sobre el ensemble ortogonal gaussiano

El *ensemble* ortogonal gaussiano recibe esta denominación debido a que las matrices que pertenecen a este *ensemble* son invariantes por conjugación ortogonal, es decir, si $\mathbf{M} \in \text{GOE}(n)$, entonces $\mathbf{O}\mathbf{M}\mathbf{O}^T \in \text{GOE}(n)$ para cualquier matriz \mathbf{O} ortogonal.

Demostración. Empecemos viendo que si $\mathbf{X} \in \mathcal{S}_n$ tal que sus entradas son normales estándar independientes, entonces las entradas de $\mathbf{Y} = \mathbf{O}\mathbf{X}\mathbf{O}^T$ también son normales estándar independientes. Observar primero que

$$\mathbf{E}(X_{ij}X_{pq}) = \begin{cases} \mathbf{E}(X_{ij}^2) = \mathbf{V}(X_{ij}) = 1, & \text{si } i = p \text{ y } j = q \\ 0, & \text{en otro caso} \end{cases} \implies \mathbf{E}(X_{ij}X_{pq}) = \delta_{ip}\delta_{jq}.$$

Sea \mathbf{O} es una matriz ortogonal tal que $\mathbf{O} = (a_{ij})_{1 \leq i, j \leq n}$, entonces

$$Y_{ij} = \sum_{k,l=1}^n a_{ik} X_{kl} a_{jl},$$

por lo que cada entrada de \mathbf{Y} es una combinación lineal de las entradas de \mathbf{X} , y como éstas son normales independientes obtenemos que las entradas de \mathbf{Y} siguen una distribución normal con media cero. Nos preguntamos ahora por la varianza de las entradas de \mathbf{Y} .

$$\begin{aligned} \mathbf{E}(Y_{ij}Y_{pq}) &= \mathbf{E} \left[\sum_{k,l=1}^n a_{ik} X_{kl} a_{jl} \sum_{r,s=1}^n a_{pr} X_{rs} a_{qs} \right] = \sum_{k,l=1}^n a_{ik} a_{jl} \sum_{r,s=1}^n a_{pr} a_{qs} \mathbf{E}[X_{kl}X_{rs}] = \\ &= \sum_{k,l=1}^n a_{ik} a_{jl} \sum_{r,s=1}^n a_{pr} a_{qs} \delta_{kr}\delta_{ls} = \sum_{k,l=1}^n a_{ik} a_{jl} a_{pk} a_{ql} = \sum_{k=1}^n a_{ik} a_{pk} \sum_{l=1}^n a_{jl} a_{ql} = \delta_{ip}\delta_{jq}. \end{aligned}$$

Donde la última igualdad es debido a que \mathbf{O} es una matriz ortogonal. Con el cálculo anterior podemos calcular la varianza de las entradas \mathbf{Y} :

$$\mathbf{V}(Y_{ij}) = \mathbf{E}(Y_{ij}^2) = \delta_{ii}\delta_{jj} = 1.$$

Por consiguiente, hemos probado que si las entradas de \mathbf{X} son normales estándar independientes, entonces las entradas de $\mathbf{O}\mathbf{Y}\mathbf{O}^T$ son también normales estándar independientes.

Ahora, si $\mathbf{M} \in \text{GOE}(n)$ entonces \mathbf{M} se puede escribir como $\mathbf{M} = \frac{1}{\sqrt{n}}(\mathbf{X} + \mathbf{X}^T)$ con \mathbf{X} matriz cuyas entradas son normales estándar independientes. Si O es una matriz ortogonal entonces

$$O\mathbf{M}O^T = \frac{1}{\sqrt{n}}O(\mathbf{X} + \mathbf{X}^T)O^T = \frac{1}{\sqrt{n}}(O\mathbf{X}O^T + O\mathbf{X}^TO^T) \stackrel{d}{=} \frac{1}{\sqrt{n}}(\mathbf{X} + \mathbf{X}^T) = \mathbf{M} \in \text{GOE}(n).$$

□