



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO
DIVISIÓN ACADÉMICA DE CIENCIAS BÁSICAS



**TEORÍA DE MATRICES ALEATORIAS EN EL
ESTUDIO DE LA MATRIZ DE COVARIANZA
POBLACIONAL DE DATOS DE DIMENSIÓN
ALTA**

TESIS

**PARA OBTENER EL TÍTULO DE
LICENCIADO EN MATEMÁTICAS**

**PRESENTA
DIDIER CORTEZ ELIZALDE**

**DIRECTOR DE TESIS
DRA. ADDY MARGARITA BOLÍVAR CIMÉ**

CUNDUACÁN, TAB, MEX.

JULIO 2017



**UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO**

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



División
Académica
de Ciencias
Básicas

DIRECCIÓN

Junio 14, de 2017.

C. DIDIER CORTEZ ELIZALDE
PASANTE DE LA LIC. EN MATEMATICAS
P R E S E N T E.

Por medio del presente y de la manera más cordial, me dirijo a usted para hacer de su conocimiento que proceda a la impresión del trabajo titulado **"TEORÍA DE MATRICES ALEATORIAS EN EL ESTUDIO DE LA MATRIZ DE COVARIANZA POBLACIONAL DE DATOS DE DIMENSIÓN ALTA"**, bajo la modalidad de titulación por TESIS, en virtud de que reúne los requisitos para el EXAMEN PROFESIONAL correspondiente.

Sin otro particular, reciba usted un cordial saludo.

A T E N T A M E N T E

DR. GERARDO DELGADILLO PIÑÓN
DIRECTOR



DIVISIÓN ACADÉMICA DE
CIENCIAS BÁSICAS

C.c.p. Archivo.

Miembro CUMEX desde 2008

Consortio de
Universidades
Mexicanas

UNA ALIANZA DE CALIDAD POR LA EDUCACIÓN SUPERIOR

Km. 1, Carretera Cunduacán - Jalpa de Méndez, A.P. 24, C.P. 86690, Cunduacán, Tabasco.
Tel/Fax (914)3360928, (993)3581500 Ext. 6702 E-mail: direccion.dacb@ujat.mx

CARTA DE AUTORIZACIÓN

El que suscribe, autoriza por medio del presente escrito a la Universidad Juárez Autónoma de Tabasco para que utilice tanto física como digitalmente la tesis de grado denominada "**Teoría de Matrices Aleatorias en el Estudio de la Matriz de Covarianza Poblacional de Datos de Dimensión Alta**", de la cual soy autor y titular de los Derechos de Autor.

La finalidad del uso por parte de la Universidad Juárez Autónoma de Tabasco de la tesis antes mencionada, será única y exclusivamente para difusión, educación y sin fines de lucro; autorización que se hace de manera enunciativa mas no limitativa para subirla a la Red Abierta de Bibliotecas Digitales (RABID) y a cualquier otra red académica con las que la Universidad tenga relación institucional.

Por lo antes manifestado, libero a la Universidad Juárez Autónoma de Tabasco de cualquier reclamación legal que pudiera ejercer respecto al uso y manipulación de la tesis mencionada y para los fines estipulados en este documento.

Se firma la presente autorización en la ciudad de Villahermosa, Tabasco a los 6 días del mes de Julio del año 2017

AUTORIZÓ



Didier Cortez Elizalde
102A11013

Dedicatoria

*"No puedes vivir tu vida para otras personas.
Tienes que hacer lo que es correcto para ti,
aunque le duela a tus seres queridos."*

Nicholas Sparks. The notebook.

Dedico este trabajo principalmente a Dios, por haberme dado vida y permitirme el haber llegado hasta este momento tan importante de mi formación profesional.

A mi madre por ser un pilar muy importante de mi familia y por demostrarme su cariño y apoyo incondicional sin importar la diferencia de opiniones.

A mis hermanos por el apoyo y el cariño que me han brindado durante todo este tiempo.

A mis amigos por el apoyo y por todos aquellos momentos agradables que pasamos juntos.

Agradecimientos

Primero quiero agradecer a Dios por darme la oportunidad de vivir y permitirme haber llegado a este momento tan importantes de mi vida.

Quiero agradecer a mi madre la Sra. Blanca Emma Elizalde por brindarme su amor y apoyo, sobre todo por darme la mejor herencia (el haber estudiado una licenciatura), también le agradezco la paciencia que me ha tenido, gracias. A mis hermanos Dania, Jorge, Lorena, Angel Eduardo gracias por estar presente en cada día de mi vida. A mis abuelos Martha Hurbieta (Q.E.P.D) y Eulogio Cortez gracias por querernos a mis hermanos y a mi.

A la División Académica de Ciencias Básicas de la UJAT gracias por el apoyo que recibí durante los congresos a los que asistí de los cuales me ayudaron a mi formación académica y tuve la oportunidad de conocer e intercambiar opiniones con muchas personas. A mis maestros, gracias por compartir sus enseñanzas.

Quiero agradecer a todas aquellas personas que con su ayuda y apoyo fueron partícipes de en este trabajo. A mi asesora de tesis la Dra. Addy Margarita Bolívar Cimé gracias por la orientación y supervisión, pero sobre todo por la motivación y el apoyo recibido durante todo este tiempo, también agradezco sus enseñanzas y la paciencia durante las sesiones de tesis. A mis amigos Veronica, Gabriela, Dorilian, Audy, Gustavo, Jareni, Williams gracias por acompañarme en los buenos momentos.

A todos ellos gracias.

Índice general

Introducción	1
1. Distribución Normal Multivariada	3
1.1. Preliminares	3
1.2. Función característica	6
1.3. Distribución marginal	10
1.4. Función de densidad	14
2. Distribución Wishart	17
2.1. Función de densidad	17
2.2. Función característica	19
2.3. Distribución marginal	23
2.4. Algunos resultados de la Teoría de Matrices Aleatorias	24
2.4.1. Teorema de Wigner	25
2.4.2. Teorema de Marchenko-Pastur (M-P)	26
2.4.3. Ley de Tracy-Widom (T-W)	29
3. Pruebas de hipótesis para la matriz de covarianza poblacional con datos de dimensión alta	32
3.1. Teorema de M-P en el estudio de la matriz de covarianza poblacional	32
3.2. Pruebas de hipótesis para la matriz de covarianza poblacional basadas en la ley de T-W	39
3.2.1. Prueba de hipótesis de esfericidad	39
3.2.2. Prueba de hipótesis para $H_0 : \Sigma = \Sigma_0$	40
3.3. Estudio de simulación para evaluar las pruebas de hipótesis basadas en la ley de T-W	42
3.4. Comparación de la prueba de hipótesis basada en la ley T-W con la basada en el estadístico de esfericidad	44
3.4.1. Prueba de hipótesis basada en el estadístico de esfericidad	44

3.4.2. Estudio de simulación para la comparación de las pruebas de hipótesis	45
Conclusiones	48
A. Algunas definiciones y propiedades de matrices	50
A.1. Traza de una matriz	52
A.2. Determinantes	53
A.3. Inversa de una matriz	53
A.4. Eigenvalores e eigenvectores	54
A.5. Matriz no negativa definida y positiva definida	55
A.6. Matriz Toeplitz	55
B. Algoritmos de simulaciones	58
B.1. Teorema de Marchenko-Pastur	58
B.2. Pruebas de hipótesis basadas en la ley de T-W	59
B.3. Pruebas de hipótesis basadas en la ley de T-W y el estadístico de esfericidad	60
Bibliografía	64

Introducción

La *Teoría de Matrices Aleatorias* (TMA) estudia matrices cuyas entradas son variables aleatorias (o equivalentemente, variables aleatorias que toman valores en un espacio de matrices). La primera distribución propuesta para matrices aleatorias fue la *distribución Wishart*, nombrada así en honor a John Wishart, quien en 1928 la introdujo como una generalización de la distribución chi-cuadrada. A inicios de los 50's los físicos usaron modelos de matrices aleatorias para estudiar fenómenos cuánticos, a partir de esto se desarrolló una buena parte de la Teoría de Matrices Aleatorias que se conoce hoy en día, algunos pioneros fueron Eugene Wigner, Vladimir Marchenko y Leonid Pastur. Posteriormente esta teoría ha sido ampliamente desarrollada por matemáticos y estadísticos.

Los datos multivariados de dimensión mayor al tamaño de la muestra (datos multivariados de dimensión alta) aparecen en diversos campos, algunos de ellos son genética, análisis funcional, finanzas, análisis de imágenes médicas, climatología, reconocimiento de texto, entre otros. Cabe mencionar que en el contexto de datos multivariados de dimensión alta la estimación de la matriz de covarianza poblacional es complicada, ya que se tienen que estimar muchos parámetros con pocos datos, por lo que la estimación de esta matriz y pruebas de hipótesis acerca de ella requieren técnicas estadísticas diferentes a las del caso clásico donde el tamaño de la muestra es mucho mayor que la dimensión de los datos.

La distribución Wishart es la distribución de la matriz de covarianza muestral de una muestra aleatoria de la distribución normal multivariada, debido a esto el estudio de la distribución Wishart ha sido de gran importancia en Análisis Multivariado. Algunos resultados clásicos de la Teoría de Matrices Aleatorias involucran convergencias asintóticas relacionadas con la distribución de Wishart, lo que facilita mucho el estudio de la matriz de covarianza poblacional de datos normales multivariados de dimensión alta. Uno de estos resultados es el Teorema de Marchenko-Pastur, el cual pro-

porciona la distribución asintótica de la *distribución empírica espectral* de matrices con distribución Wishart, la cual es conocida como la *distribución de Marchenko–Pastur*. Por otro lado también se encuentra un resultado de Johnstone en [7], que afirma que la distribución asintótica del eigenvalor más grande de una matriz aleatoria con distribución Wishart es la distribución de Tracy–Widom. Mediante estos resultados es posible estudiar la matriz de covarianza poblacional de datos normales multivariados de dimensión alta, más específicamente, estos resultados ayudan a determinar si la matriz de covarianza poblacional de un conjunto de datos normales multivariados de dimensión alta podría o no ser la matriz identidad.

En esta tesis se muestra cómo las herramientas de la Teoría de Matrices Aleatorias, en particular el Teorema de Marchenko–Pastur y la distribución de Tracy–Widom, pueden ser utilizadas para resolver el problema de Análisis Multivariado de determinar si la matriz de covarianza poblacional de datos normales multivariados de dimensión alta es la matriz identidad o una matriz específica. También se evalúa, mediante simulaciones, la conveniencia de utilizar estos procedimientos estadísticos que involucran la herramienta de la Teoría de Matrices Aleatorias.

En el Capítulo 1 se presentan las principales propiedades de la distribución normal multivariada. En el Capítulo 2 se presenta la definición y principales propiedades de la distribución Wishart, así como los resultados de la Teoría de Matrices Aleatorias que serán necesarios en la tesis. En el Capítulo 3 se presentan pruebas de hipótesis para la matriz de covarianza poblacional de datos de dimensión alta, basadas en resultados de la Teoría de Matrices Aleatorias, también se presentan estudios de simulación para evaluar el comportamiento de estas pruebas de hipótesis. Finalmente, se presenta un capítulo de conclusiones y un apéndice con algunas definiciones y propiedades de matrices utilizadas en la tesis.

Capítulo 1

Distribución Normal Multivariada

En este capítulo se dan algunas definiciones, teoremas y propiedades importantes de la distribución normal multivariada, las cuales fueron tomadas de [1] y [13].

1.1. Preliminares

Antes de definir la distribución normal multivariada, se presentan algunas definiciones básicas de Análisis Multivariado.

Definición 1.1. La **media o esperanza** de un vector aleatorio $X = (X_1, X_2, \dots, X_m)'$ de $m \times 1$ está definido como el siguiente vector

$$\mathbb{E}(X) = \begin{pmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \\ \vdots \\ \mathbb{E}(X_m) \end{pmatrix}.$$

Más general, si $Z = (Z_{ij})$ es una matriz aleatoria de $p \times q$, entonces $\mathbb{E}(Z)$ es la matriz cuyo ij -ésimo elemento es $\mathbb{E}(Z_{ij})$.

Es fácil mostrar que si B , C y D son matrices constantes de $m \times p$, $q \times n$ y $m \times n$ respectivamente, entonces

$$\mathbb{E}(BZC + D) = B\mathbb{E}(Z)C + D.$$

Definición 1.2. Si X tiene media μ , la **matriz de covarianza** de X se define como la siguiente matriz de $m \times m$

$$\Sigma = \text{Cov}(X) \equiv \mathbb{E}[(X - \mu)(X - \mu)'].$$

Notar que el elemento ij -ésimo de la matriz de covarianza Σ es

$$\sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)],$$

la covarianza entre X_i y X_j , y el ii -ésimo elemento es

$$\sigma_{ii} = \mathbb{E}[(X_i - \mu_i)^2],$$

por lo que los elementos de la diagonal de Σ no deben ser negativos. Por las propiedades de la covarianza de variables aleatorias, tenemos también que Σ es simétrica, es decir $Z = Z'$.

El siguiente lema relaciona a las matrices de covarianza con las *matrices no negativas definidas* (ver Definición A.1).

Lema 1.1. *La matriz Σ de $m \times m$ es una matriz de covarianza si y solo si es no negativa definida.*

Demostración. Supongamos que Σ es la matriz de covarianza de una variable aleatoria X , con media μ , entonces $\forall \alpha \in \mathbb{R}^m$

$$\begin{aligned} \text{Var}(\alpha'X) &= \mathbb{E}[(\alpha'X - \alpha'\mu)^2] \\ &= \mathbb{E}[(\alpha'(X - \mu))^2] \\ &= \mathbb{E}[\alpha'(X - \mu)(X - \mu)'\alpha] \\ &= \alpha'\mathbb{E}[(X - \mu)(X - \mu)']\alpha \\ &= \alpha'\Sigma\alpha \geq 0. \end{aligned} \tag{1.1}$$

Así $\alpha'\Sigma\alpha \geq 0$, $\forall \alpha \in \mathbb{R}^m$, ya que es igual a $\text{Var}(\alpha'X) \geq 0$ y por lo tanto Σ es no negativa definida.

Supongamos que Σ es una matriz no negativa definida de rango r , digamos $r \leq m$. Escribamos $\Sigma = CC'$, donde C es una matriz de $m \times r$ de rango r y sea Y un vector aleatorio de $r \times 1$ con media cero y $\text{Cov}(Y) = I$ y $X = CY$, entonces $\mathbb{E}(X) = C\mathbb{E}(Y) = C * \mathbf{0} = \mathbf{0}$ y

$$\begin{aligned} \text{Cov}(X) &= \mathbb{E}[XX'] = \mathbb{E}[CYY'C'] \\ &= C\mathbb{E}[YY']C' = CC\text{Cov}[Y]C' \\ &= CIC' = CC' = \Sigma. \end{aligned}$$

Así Σ es una matriz de covarianza. □

Como una consecuencia de la desigualdad (1.1), tenemos que si la matriz de covarianza Σ de un vector aleatorio X no es una *matriz positiva definida* (ver Definición A.2), entonces con probabilidad 1 las componentes de X

están relacionadas linealmente. Lo anterior es debido a que al no ser positiva definida existe un $\alpha \in \mathbb{R}^m$, con $\alpha \neq 0$, para el cual

$$\text{Var}(\alpha'X) = \alpha'\Sigma\alpha = 0.$$

Entonces con probabilidad 1 se tiene que $\alpha'X = k$, donde $k = \alpha'\mathbb{E}(X)$, lo que significa que X se encuentra en un hiperplano con probabilidad 1.

A menudo se hace una transformación lineal de vectores aleatorios y necesitamos saber como se transforma la matriz de covarianza. Supongamos que X es un vector aleatorio de $m \times 1$ con media μ_X y matriz de covarianza Σ_X y sea $Y = BX + b$, donde B y b son matrices constantes de dimensión $k \times m$ y $k \times 1$, respectivamente. La media de Y es $\mu_Y = B\mu_X + b$ y la matriz de covarianza de Y es

$$\begin{aligned}\Sigma_Y &= \mathbb{E}[(Y - \mu_Y)(Y - \mu_Y)'] \\ &= \mathbb{E}[(BX + b - (B\mu_X + b))(BX + b - (B\mu_X + b))'] \\ &= B\mathbb{E}[(X - \mu_X)(X - \mu_X)']B' \\ &= B\Sigma_X B'.\end{aligned}$$

El siguiente teorema será utilizado más adelante.

Teorema 1.1. *Si X es un vector aleatorio de $m \times 1$, entonces su distribución está determinada únicamente por la distribución de funciones lineales $\alpha'X$, $\forall \alpha \in \mathbb{R}^m$.*

Demostración. La función característica de $\alpha'X$ es

$$\Phi(t, \alpha) = \mathbb{E}(\exp(it\alpha'X)).$$

Así que

$$\Phi(1, \alpha) = \mathbb{E}(\exp(i\alpha'X))$$

es la función característica conjunta de las componentes de X . El resultado se sigue del hecho de que una distribución en \mathbb{R}^m es determinada únicamente por su función característica. \square

A continuación presentamos la definición de la distribución normal multivariada, la cual es la principal distribución multivariada considerada en este trabajo.

Definición 1.3. *El vector aleatorio X de $m \times 1$ se dice que tiene **distribución normal m -variada**, si para cada $\alpha \in \mathbb{R}^m$, la distribución de $\alpha'X$ es normal univariada.*

La distribución normal m -variada es denotada como $N_m(\mu, \Sigma)$ y escribiremos $X \sim N_m(\mu, \Sigma)$ para indicar que el vector aleatorio X tiene esta distribución. El siguiente resultado caracteriza una distribución normal multivariada.

Teorema 1.2. *Si X tiene una distribución normal m -variada, entonces tanto $\mu \equiv \mathbb{E}[X]$ como $\Sigma \equiv \text{Cov}[X]$ existen y la distribución de X está determinada por μ y Σ .*

Demostración. Si $X = (X_1, X_2, \dots, X_m)'$, entonces para cada $i = 1, 2, \dots, m$, X_i es normal univariada (por la definición anterior), así que $\mathbb{E}(X_i)$ y $\text{Var}(X_i)$ existen y son finitas. Así la $\text{Cov}(X_i, X_j)$ existe, ya que si ponemos a $\mu = \mathbb{E}(X)$ y $\Sigma = \text{Cov}(X)$ tenemos que

$$\begin{aligned} \mathbb{E}(\alpha'X) &= \mathbb{E} \left[(\alpha_1, \alpha_2, \dots, \alpha_n) \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} \right] \\ &= \mathbb{E}[\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_m] \\ &= \alpha_1 \mathbb{E}[X_1] + \alpha_2 \mathbb{E}[X_2] + \dots + \alpha_n \mathbb{E}[X_m] \\ &= (\alpha_1, \alpha_2, \dots, \alpha_n) \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix} \\ &= \alpha' \mu. \end{aligned}$$

$$\begin{aligned} \text{Var}(\alpha'X) &= \mathbb{E}[(\alpha'X - \alpha'\mu)^2] \\ &= \mathbb{E}[(\alpha'(X - \mu))^2] \\ &= \mathbb{E}[\alpha'(X - \mu)(X - \mu)'\alpha] \\ &= \alpha \mathbb{E}[(X - \mu)(X - \mu)']\alpha \\ &= \alpha \Sigma \alpha'. \end{aligned}$$

Así la distribución de $\alpha'X \sim N(\alpha'\mu, \alpha \Sigma \alpha')$, para cada $\alpha \in \mathbb{R}^m$. Estas distribuciones univariadas están determinadas por μ y Σ , y también la distribución de X por el Teorema 1.1. \square

1.2. Función característica

En lo subsecuente denotaremos a la función característica de una variable o vector aleatorio X por Φ_X . El siguiente teorema proporciona la forma de

la función característica de la distribución normal multivariada.

Teorema 1.3. *Si $X \sim N_m(\mu, \Sigma)$, entonces la función característica de X es*

$$\Phi_X(t) = \exp\left(i\mu't - \frac{1}{2}t'\Sigma t\right).$$

Demostración. Sea

$$\Phi_X(t) = \mathbb{E}[\exp(it'X)] = \Phi_{t'X}(1),$$

en donde el lado derecho indica la función característica de la variable $t'X$ evaluada en 1. Puesto que $X \sim N_m(\mu, \Sigma)$, entonces $t'X \sim N(t'\mu, t'\Sigma t)$ así que

$$\Phi_{t'X}(1) = \exp\left(it'\mu - \frac{1}{2}t'\Sigma t\right).$$

□

Como aun no se ha establecido la existencia de la distribución normal multivariada, podría pasar que la Definición 1.3 no se cumpla para ningún vector aleatorio, por lo cual demostraremos que la función del teorema anterior, es de hecho la función característica de un vector aleatorio.

Sea Σ una matriz de covarianza de $m \times m$ (matriz no negativa definida) de rango r y sean U_1, U_2, \dots, U_r variables aleatorias normales estándar independientes. El vector $U = (U_1, U_2, \dots, U_r)'$ tiene una función característica de la siguiente forma

$$\begin{aligned} \Phi_U(t) &= \mathbb{E}(\exp(it'U)) \\ &= \mathbb{E}(\exp(it_1U_1 + \dots + it_rU_r)) \\ &= \prod_{j=1}^r \mathbb{E}[\exp(it_jU_j)] \\ &= \prod_{j=1}^r \exp\left(-\frac{t_j^2}{2}\right) \\ &= \exp\left(-\frac{1}{2}\sum_{j=1}^r t_j^2\right) \\ &= \exp\left(-\frac{1}{2}t't\right). \end{aligned} \tag{1.2}$$

Sea $X = CU + \mu$, donde C es una matriz de $m \times r$ de rango r tal que $\Sigma = CC'$, y $\mu \in \mathbb{R}^m$. Tenemos que

$$\begin{aligned}\mathbb{E}[\exp(it'X)] &= \mathbb{E}[\exp(it'(CU + \mu))] \\ &= \mathbb{E}(\exp(it'CU + it'\mu)) \\ &= \mathbb{E}[\exp(it'CU)]\exp[it'\mu].\end{aligned}$$

Como U tiene función característica (1.2), se tiene de la expresión anterior que

$$\begin{aligned}\mathbb{E}[\exp(it'X)] &= \Phi_U(C't)\exp[it'\mu] \\ &= \exp\left(-\frac{1}{2}t'CC't\right)\exp[it'\mu] \\ &= \exp\left(-\frac{1}{2}t'\Sigma t\right)\exp[it'\mu] \\ &= \exp\left(it'\mu - \frac{1}{2}t'\Sigma t\right),\end{aligned}$$

por lo tanto X tiene distribución $N_m(\mu, \Sigma)$, pues la expresión anterior corresponde a la función característica de esa distribución.

El siguiente resultado afirma que la distribución de una transformación lineal de un vector normal multivariado es también normal multivariado.

Teorema 1.4. *Supongamos que $X \sim N_m(\mu, \Sigma)$ y sean B, b de $k \times m, k \times 1$, respectivamente, entonces $Y = BX + b \sim N_k(B\mu + b, B\Sigma B')$.*

Demostración. Queremos mostrar que Y es normal k -variada. Sea $\alpha \in \mathbb{R}^k$, entonces

$$\begin{aligned}\alpha'Y &= (\alpha_1, \alpha_2, \dots, \alpha_k) \left[\begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1m} \\ B_{21} & B_{22} & \cdots & B_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ B_{k1} & B_{k2} & \cdots & B_{km} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix} \right] \\ &= (\alpha_1, \alpha_2, \dots, \alpha_k) \begin{pmatrix} B_{11}X_1 + B_{12}X_2 + \dots + B_{1m}X_m + b_1 \\ B_{21}X_1 + B_{22}X_2 + \dots + B_{2m}X_m + b_2 \\ \vdots \\ B_{k1}X_1 + B_{k2}X_2 + \dots + B_{km}X_m + b_k \end{pmatrix}.\end{aligned}$$

La expresión anterior es igual a

$$\begin{aligned}&\alpha_1(B_{11}X_1 + B_{12}X_2 + \dots + B_{1m}X_m + b_1) + \alpha_2(B_{21}X_1 + B_{22}X_2 + \dots \\ &\quad + B_{2m}X_m + b_2) + \dots + \alpha_k(B_{k1}X_1 + B_{k2}X_2 + \dots + B_{km}X_m + b_k) \\ &= X_1(\alpha_1B_{11} + \alpha_2B_{21} + \dots + \alpha_kB_{k1}) + X_2(\alpha_1B_{12} + \alpha_2B_{22} + \dots + \alpha_kB_{k2}) \\ &\quad + \dots + X_m(\alpha_1B_{1m} + \alpha_2B_{2m} + \dots + \alpha_kB_{km}) + (\alpha_1b_1 + \alpha_2b_2 + \dots + \alpha_kb_k).\end{aligned}$$

Así $\alpha'Y$ se puede escribir como una combinación lineal de las entradas del vector X más una constante. Como X es normal multivariada, por la Definición 1.3 las combinaciones lineales de sus entradas son normales univariadas, por lo tanto $\alpha'Y$ es normal univariada. De este modo por la Definición 1.3 se tiene que Y es normal k -variada.

Ahora mostraremos que $\mathbb{E}(Y) = B\mu + b$ y la $\text{Var}(Y) = B\Sigma B'$. Tenemos que

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}(BX + b) = B\mathbb{E}(X) + b \\ &= B\mu + b, \\ \text{Var}(Y) &= \mathbb{E}[(Y - \mu)(Y - \mu)'] \\ &= \mathbb{E}[(BX + b - B\mu - b)(BX + b - B\mu - b)'] \\ &= \mathbb{E}[B(X - \mu)(X - \mu)'B'] \\ &= B\mathbb{E}[(X - \mu)(X - \mu)']B' \\ &= B\Sigma B'.\end{aligned}$$

Por lo tanto $Y = BX + b \sim N_k(B\mu + b, B\Sigma B')$. □

Mediante el siguiente resultado, tenemos que la distribución de una combinación lineal de vectores normales multivariados independientes es también normal multivariada.

Teorema 1.5. *Si X_1, X_2, \dots, X_N son vectores independientes, con $X_i \sim N_m(\mu_i, \Sigma_i)$ para $i = 1, 2, \dots, N$, entonces para constantes fijas a_1, a_2, \dots, a_N ,*

$$\sum_{i=1}^N a_i X_i \sim N_m \left(\sum_{i=1}^N a_i \mu_i, \sum_{i=1}^N a_i^2 \Sigma_i \right).$$

Demostración. Por la independencia de los vectores X_i 's, tenemos que los vectores $a_i X_i$'s también son independientes. Por el Teorema 1.4 tenemos que $a_i X_i \sim N_m(a_i \mu_i, a_i^2 \Sigma_i)$ para $i = 1, 2, \dots, N$. Así, utilizando el Teorema 1.3

tenemos que la función característica de $U = \sum_{i=1}^N a_i X_i$ está dada por

$$\begin{aligned}\Phi_U(t) &= \prod_{j=1}^N \Phi_{a_j X_j}(t) = \prod_{j=1}^N \exp(ia_j \mu_j' t - \frac{1}{2} t' a_j^2 \Sigma_j t) \\ &= \exp(i \sum_{j=1}^N a_j \mu_j' t - \frac{1}{2} t' \sum_{j=1}^N a_j^2 \Sigma_j t).\end{aligned}$$

Debido a que la última expresión corresponde a la función característica de la distribución $N_m \left(\sum_{i=1}^N a_i \mu_i, \sum_{i=1}^N a_i^2 \Sigma_i \right)$, se tiene el resultado. \square

Antes de dar un corolario del teorema anterior presentamos las siguientes definiciones de media y matriz de covarianza muestral.

Definición 1.4. Sean X_1, X_2, \dots, X_N vectores aleatorios independientes e idénticamente distribuidos (una muestra aleatoria de tamaño N), la **media muestral** y **matriz de covarianza muestral** se definen, respectivamente, como

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad y \quad S_n = \frac{1}{n} A,$$

donde $n = N - 1$ y

$$A = \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})'.$$

Corolario 1.1. Si X_1, X_2, \dots, X_N son vectores independientes con distribución $N_m(\mu, \Sigma)$, entonces la distribución de la media muestral $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ es $N_m(\mu, \Sigma/N)$.

El corolario anterior afirma que la distribución de la media muestral de una muestra aleatoria normal multivariada, es también normal multivariada. En el Capítulo 2 se proporcionará la distribución de la matriz de covarianza muestral de datos normales multivariados.

1.3. Distribución marginal

A continuación presentamos resultados sobre las distribuciones marginales de un vector aleatorio con distribución normal multivariada.

Teorema 1.6. Si $X \sim N_m(\mu, \Sigma)$, entonces la distribución marginal de cualquier subconjunto de k ($< m$) componentes de X es normal k -variada.

Demostración. Esta demostración es un consecuencia del teorema anterior, ya que por ejemplo si particionamos X , μ y Σ como

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad y \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

donde X_1 , μ_1 son de $k \times 1$ y Σ_{11} es $k \times k$, definiendo

$$B = [I_k : \mathbf{0}] \text{ de } k \times m \quad \text{y} \quad b = \mathbf{0} \text{ de } k \times 1,$$

por el Teorema 1.4 tenemos que $Y = BX + b$ es normal k -variado. Notemos que

$$Y = [I_k : \mathbf{0}] \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \mathbf{0} = X_1,$$

$$E(Y) = [I_k : \mathbf{0}] \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \mathbf{0} = \mu_1,$$

$$\begin{aligned} \text{Cov}(Y) &= B\Sigma B' \\ &= [I_k : \mathbf{0}] \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} [I_k : \mathbf{0}]' \\ &= \Sigma_{11}. \end{aligned}$$

Por lo tanto $X_1 \sim N_k(\mu_1, \Sigma_{11})$. Así la distribución marginal de cualquier subvector de k componentes de X es normal, donde la media y la matriz de covarianza se obtienen tomando el subvector y la submatriz correspondiente de una manera similar a lo anterior. \square

Como consecuencia del teorema anterior se tiene que la distribución marginal de cada componente de X es normal univariada, lo contrario no es cierto en general, es decir, el hecho de que cada componente de un vector aleatorio sea normal univariada no implica que el vector tenga una distribución normal multivariada. Como contraejemplo, supongamos que U_1 , U_2 y U_3 son variables aleatorias independientes con distribución $N(0, 1)$ y Z es una variable aleatoria cualquiera independiente de U_1 , U_2 y U_3 . Definimos

$$X_1 = \frac{U_1 + ZU_3}{\sqrt{1 + Z^2}} \quad \text{y} \quad X_2 = \frac{U_2 + ZU_3}{\sqrt{1 + Z^2}}.$$

Suponiendo que Z es un valor fijo tenemos que

$$\begin{aligned}\mathbb{E}(X_1) &= \mathbb{E}\left(\frac{U_1 + ZU_3}{\sqrt{1 + Z^2}}\right) \\ &= \frac{\mathbb{E}(U_1) + Z\mathbb{E}(U_3)}{\sqrt{1 + Z^2}} \\ &= 0. \\ \text{Var}(X_1) &= \text{Var}\left(\frac{U_1 + ZU_3}{\sqrt{1 + Z^2}}\right) \\ &= \text{Var}\left(\frac{U_1}{\sqrt{1 + Z^2}}\right) + \text{Var}\left(\frac{ZU_3}{\sqrt{1 + Z^2}}\right) \\ &= \frac{1}{1 + Z^2} + \frac{Z^2}{1 + Z^2} \\ &= 1.\end{aligned}$$

Por lo tanto, condicionado a Z , $X_1 \sim N(0, 1)$, y ya que esta distribución no depende de Z , es la distribución incondicional de X_1 . Análogamente, condicionado a Z , $X_2 \sim N(0, 1)$. La distribución conjunta de X_1 y X_2 condicionada a Z es normal bivariada, ya que si consideramos

$$W = (X_1, X_2)' \quad \text{y} \quad \alpha = (\alpha_1, \alpha_2)',$$

entonces

$$\begin{aligned}\alpha'W &= (\alpha_1, \alpha_2) \begin{pmatrix} \frac{U_1 + ZU_3}{\sqrt{1 + Z^2}} \\ \frac{U_2 + ZU_3}{\sqrt{1 + Z^2}} \end{pmatrix} \\ &= \frac{\alpha_1 U_1}{\sqrt{1 + Z^2}} + \frac{\alpha_2 U_2}{\sqrt{1 + Z^2}} + \left(\frac{\alpha_1 Z + \alpha_2 Z}{\sqrt{1 + Z^2}}\right) U_3,\end{aligned}$$

la cual, si Z es un valor fijo, es una combinación lineal de variables aleatorias normales independientes y por lo tanto es normal, además

$$\begin{aligned}\mathbb{E}(\alpha'W) &= \alpha_1 \mathbb{E}(U_1) \left(\frac{1}{\sqrt{1 + Z^2}}\right) + \alpha_2 \mathbb{E}(U_2) \left(\frac{1}{\sqrt{1 + Z^2}}\right) \\ &\quad + \left(\frac{\alpha_1 Z + \alpha_2 Z}{\sqrt{1 + Z^2}}\right) \mathbb{E}(U_3) \\ &= 0,\end{aligned}$$

$$\begin{aligned}\text{Var}(\alpha'W) &= \text{Var}(U_1) \left(\frac{\alpha_1}{\sqrt{1+Z^2}} \right)^2 + \text{Var}(U_2) \left(\frac{\alpha_2}{\sqrt{1+Z^2}} \right)^2 \\ &\quad + \left(\frac{\alpha_1 Z + \alpha_2 Z}{\sqrt{1+Z^2}} \right)^2 \text{Var}(U_3) \\ &= \frac{\alpha_1^2 + \alpha_2^2 + (\alpha_1 Z + \alpha_2 Z)^2}{1+Z^2}.\end{aligned}$$

Por lo tanto, si Z es un valor fijo

$$\alpha'W \sim N \left(0, \frac{\alpha_1^2 + \alpha_2^2 + (\alpha_1 Z + \alpha_2 Z)^2}{1+Z^2} \right).$$

De este modo, la distribución de W condicionada a Z es normal bivariada, pero claramente la distribución incondicional de W no necesariamente es normal. La distribución de W sería normal multivariada si las componentes de X fueran todas normales e independientes.

Hay que recordar que la independencia de dos variables aleatorias implica que la covarianza entre éstas, si existe, es cero. Lo contrario no es cierto en general. En el caso de un vector con distribución normal multivariada, la independencia entre dos subvectores es equivalente a que todas las covarianzas entre las componentes de un subvector con las del otro sean iguales a cero, como se afirma en el siguiente teorema.

Teorema 1.7. *Si $X \sim N_m(\mu, \Sigma)$ y X , μ , Σ son particionadas como*

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad y \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

donde X_1 , μ_1 son de $k \times 1$ y Σ_{11} es $k \times k$, entonces los subvectores X_1 y X_2 son independientes si y solo si $\Sigma_{12} = \mathbf{0}$.

Demostración. Sabemos que si X_1 y X_2 son independientes, entonces la independencia implica de que la covarianza entre las entradas de X_1 y las entradas de X_2 es igual a cero, es decir que $\Sigma_{12} = \mathbf{0}$.

Ahora supongamos que $\Sigma_{12} = \mathbf{0}$. Sean Y_1 y Y_2 vectores aleatorios independientes donde $Y_1 \sim N_k(\mu_1, \Sigma_{11})$ y $Y_2 \sim N_{m-k}(\mu_2, \Sigma_{22})$ y ponemos a $Y = (Y_1', Y_2')$, entonces tanto X como Y son $N_m(\mu, \Sigma)$, donde Σ es de la forma

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix}.$$

Por lo tanto, al tener X la misma distribución de Y , se tiene que X_1 y X_2 son independientes. \square

Una demostración alternativa del teorema anterior sería usando las funciones características de X_1 y X_2 , es decir, mostrando que la función característica conjunta de X_1 y X_2 es el producto de sus funciones características, cuando $\Sigma_{12} = \mathbf{0}$. Sea $t' = (t'_1, t'_2)$, donde t_1 es de $k \times 1$ y t_2 es de $(m - k) \times 1$, entonces por el Teorema 1.3 tenemos que la función característica de X es

$$\begin{aligned}
 \Phi_X(t) &= \mathbb{E}[\exp(it'X)] = \exp\left(i\mu't - \frac{1}{2}t'\Sigma t\right) \\
 &= \exp\left[i\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}' \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} t'_1 & t'_2 \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}\right] \\
 &= \exp\left[i(\mu'_1 t_1 + \mu'_2 t_2) - \frac{1}{2}(t'_1 \Sigma_{11} t_1 + t'_2 \Sigma_{22} t_2)\right] \\
 &= \exp\left[i\mu'_1 t_1 - \frac{1}{2}t'_1 \Sigma_{11} t_1 + i\mu'_2 t_2 - \frac{1}{2}t'_2 \Sigma_{22} t_2\right] \\
 &= \exp\left[i\mu'_1 t_1 - \frac{1}{2}t'_1 \Sigma_{11} t_1\right] \exp\left[i\mu'_2 t_2 - \frac{1}{2}t'_2 \Sigma_{22} t_2\right] \\
 &= \mathbb{E}[\exp(it'X_1)]\mathbb{E}[\exp(it'X_2)] \\
 &= \Phi_{X_1}(t)\Phi_{X_2}(t).
 \end{aligned}$$

1.4. Función de densidad

Como se mencionó en la Sección 1.1, en el caso en que la matriz de covarianza de un vector aleatorio X con distribución normal multivariada no es positiva definida, el vector X se encuentra en un hiperplano con probabilidad 1, por lo tanto no existe una función de densidad para X . Por el contrario, si la matriz de covarianza de X es positiva definida, la función de densidad de X existe y es proporcionada en el siguiente teorema.

Teorema 1.8. *Si $X \sim N_m(\mu, \Sigma)$ y Σ es positiva definida entonces la función de densidad de X es*

$$f_X(X) = (2\pi)^{-\frac{m}{2}} (\det\Sigma)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(X - \mu)'\Sigma^{-1}(X - \mu)\right].$$

Demostración. Sea $\Sigma = CC'$ donde C es una matriz no singular de $m \times m$ y sea $X = CU + \mu$, donde U es un vector de $m \times 1$ con variables aleatorias independientes $N(0, 1)$, es decir, $U \sim N_m(\mathbf{0}, I_m)$. Para obtener la función de densidad de X damos uso al *método de la transformación* para el caso multivariado (ver [12], Sección 6.2). La función de densidad conjunta de U

es

$$\begin{aligned} f_U(U) &= \prod_{i=1}^m (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}U_i^2\right) \\ &= (2\pi)^{-\frac{m}{2}} \exp\left(-\frac{1}{2}U'U\right). \end{aligned}$$

Consideremos la siguiente transformación con su transformación inversa

$$\begin{aligned} X &= CU + \mu, \\ U &= B(X - \mu), \text{ con } B = C^{-1}. \end{aligned}$$

El Jacobiano de la transformación inversa es

$$\begin{aligned} |J| &= \det \begin{pmatrix} \frac{\partial U_1}{\partial X_1} & \frac{\partial U_1}{\partial X_2} & \cdots & \frac{\partial U_1}{\partial X_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial U_m}{\partial X_1} & \frac{\partial U_m}{\partial X_2} & \cdots & \frac{\partial U_m}{\partial X_m} \end{pmatrix} \\ &= \det \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1m} \\ \vdots & \vdots & \cdots & \vdots \\ B_{m1} & B_{m2} & \cdots & B_{mm} \end{pmatrix} \\ &= \det B = \det C^{-1} = (\det C)^{-1} \\ &= [\det(CC')]^{-1/2} = (\det \Sigma)^{-1/2}. \end{aligned}$$

Para las últimas igualdades se utilizaron las propiedades de los determinantes e inversas de matrices dadas en las secciones A.2 y A.3. Por otro lado, usando las propiedades de inversa de una matriz de la Sección A.3 tenemos que

$$\begin{aligned} \frac{1}{2}U'U &= \frac{1}{2}(C^{-1}(x - \mu))'(C^{-1}(x - \mu)) \\ &= \frac{1}{2}[(x - \mu)'(C')^{-1}C^{-1}(x - \mu)] \\ &= \frac{1}{2}[(x - \mu)'\Sigma^{-1}(x - \mu)]. \end{aligned}$$

Así la función de densidad de X es

$$f_X(X) = f_U(U)|J| = (2\pi)^{-\frac{m}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(X - \mu)'\Sigma^{-1}(X - \mu)\right].$$

□

Vale la pena mostrar, de forma explícita, la distribución normal bivariada ($m = 2$). Sean

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

y

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & -\rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

donde $\text{Var}(X_1) = \sigma_1^2$, $\text{Var}(X_2) = \sigma_2^2$ y la correlación entre X_1 y X_2 es $\rho = \text{Cov}(X_1, X_2)/(\sigma_1\sigma_2)$. Para que Σ sea positiva definida y X tenga función de densidad, se necesita que $\sigma_1^2, \sigma_2^2 > 0$ y

$$\det\Sigma = \sigma_1^2\sigma_2^2(1 - \rho^2) > 0,$$

por lo que $-1 < \rho < 1$. En ese caso

$$\Sigma^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}.$$

Ahora sustituimos lo que tenemos en la función de densidad normal multivariada para el caso de $m = 2$ y tenemos que

$$f_X(X_1, X_2) = \frac{1}{2\pi\sigma_1^2\sigma_2^2(1 - \rho^2)^{1/2}} * \exp \left[-\frac{1}{2(1 - \rho^2)} (X - \mu)' \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} (X - \mu) \right],$$

donde $(X - \mu) = \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix}$. Por lo tanto la densidad conjunta de X_1 y X_2 es

$$f_X(X_1, X_2) = \frac{1}{2\pi\sigma_1^2\sigma_2^2(1 - \rho^2)^{1/2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left[\left(\frac{X_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{X_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \frac{(X_1 - \mu_1)(X_2 - \mu_2)}{\sigma_1\sigma_2} \right] \right\}.$$

La función de densidad normal estándar se obtiene considerando a las variables aleatorias estandarizadas $Z_i = (X_i - \mu_i)/\sigma_i$ para $i = 1, 2$, es decir, la función de densidad conjunta de Z_1 y Z_2 es

$$f_Z(Z_1, Z_2) = \frac{1}{2\pi(1 - \rho^2)^{1/2}} \exp \left[-\frac{1}{2(1 - \rho^2)} (Z_1^2 + Z_2^2 - 2\rho Z_1 Z_2) \right].$$

Capítulo 2

Distribución Wishart

En este capítulo se exponen algunos resultados y propiedades importantes de la distribución Wishart, los cuales son tomados de [1] y [13]. También se presentan algunos resultados y conceptos de la Teoría de Matrices Aleatorias, como el Teorema de Marchenko-Pastur y la ley de Tracy-Widom, los cuales serán utilizados en el siguiente capítulo.

2.1. Función de densidad

La distribución de Wishart, propuesta por John Wishart en 1928, fue la primera distribución propuesta para matrices aleatorias. A continuación se presenta la definición de esta distribución.

Definición 2.1. Sea $A = Z'Z$, donde las filas de la matriz Z de $n \times p$ son *i.i.d.* $N_p(\mathbf{0}, \Sigma)$, entonces se dice que A tiene **distribución Wishart** con n grados de libertad y matriz de covarianza Σ . Notación: $A \sim W_p(n, \Sigma)$.

Comentario 2.1. Puede verse que la matriz A de la definición anterior se puede escribir de la forma

$$A = \sum_{\alpha=1}^n Z_{\alpha} Z'_{\alpha},$$

donde los vectores Z_{α} son *i.i.d.* $N_p(\mathbf{0}, \Sigma)$. De hecho los vectores Z_{α} son las filas de la matriz Z .

En el caso en que $n \geq p$ la función de densidad de la distribución $W_p(n, \Sigma)$ existe y está dada por el siguiente teorema, cuya demostración puede consultarse en [1].

Teorema 2.1. Si $A \sim W_p(n, \Sigma)$ con $n \geq p$, entonces la función de densidad de A es

$$f(A) = \frac{1}{2^{pn/2} \Gamma_p\left(\frac{1}{2}n\right) (\det \Sigma)^{n/2}} \text{etr} \left(-\frac{1}{2} \Sigma^{-1} A \right) (\det A)^{(n-p-1)/2}, \quad (A > 0)$$

donde $\Gamma_p(\cdot)$ es la función gamma multivariada y $\text{etr}(\cdot) = \exp(\text{tr}(\cdot))$ es la composición de la traza y la exponencial.

Como se vió al final de la Sección 1.2, la distribución de la media muestral de una muestra aleatoria de una distribución normal multivariada es también normal multivariada. El siguiente corolario nos proporciona la función de densidad de la matriz de covarianza muestral cuando los datos provienen de la distribución $N_p(\mu, \Sigma)$, y nos dice también que la matriz de covarianza muestral tiene distribución $W_p(n, \Sigma/n)$.

Corolario 2.1. Si X_1, \dots, X_N son vectores aleatorios independiente $N_p(\mu, \Sigma)$ y $N > p$, la función de densidad de la matriz de covarianza muestral

$$S_n = \frac{1}{n} \sum_{\alpha=1}^N (X_\alpha - \bar{X})(X_\alpha - \bar{X})', \quad (n = N - 1)$$

es

$$f(S) = \frac{1}{\Gamma_p\left(\frac{1}{2}n\right) (\det \Sigma)^{n/2} \left(\frac{1}{2}n\right)^{pn/2}} \text{etr} \left(-\frac{1}{2} n \Sigma^{-1} S \right) (\det S)^{(n-p-1)/2},$$

donde $S > 0$ (S es positiva definida).

Demostración. Notemos que podemos escribir S_n como

$$S_n = \sum_{\alpha=1}^n [(1/\sqrt{n})Z_\alpha][(1/\sqrt{n})Z_\alpha]',$$

donde $(1/\sqrt{n})Z_1, \dots, (1/\sqrt{n})Z_n$ son i.i.d. $N_p(\mathbf{0}, \Sigma/n)$. Así por el Comentario 2.1 tenemos que $S_n \sim W_p(n, \Sigma/n)$. □

Como caso particular del corolario anterior, para $p = 1$ la distribución Wishart es igual a una distribución chi-cuadrada con n grados de libertad (χ_n^2), lo cual se muestra a continuación. Sean

$$p = 1, \quad \Sigma = \sigma^2, \quad S_n = s_n^2 = \frac{1}{n} \sum_{i=1}^N (X_i - \bar{X})^2,$$

luego del corolario anterior se sigue que

$$f_{s_n^2}(s^2) = \frac{1}{\Gamma\left(\frac{1}{2}n\right) (\det\sigma^2)^{n/2}} \left(\frac{1}{2}n\right)^{n/2} \exp\left(-\frac{ns^2}{2\sigma^2}\right) (\det s^2)^{n/2-1}.$$

Sea $v = \frac{ns^2}{\sigma^2}$, usando el *método de la transformación* (ver [12], Sección 5.1) tenemos que

$$f_v(v) = f_{s_n^2}\left(\frac{v\sigma^2}{n}\right) \left| \frac{d\left(\frac{v\sigma^2}{n}\right)}{dv} \right|,$$

así la función de densidad de v es

$$\begin{aligned} f_v(v) &= \frac{1}{\Gamma\left(\frac{1}{2}n\right) 2^{n/2}} \left(\frac{n}{\sigma^2}\right)^{n/2} \exp\left(-\frac{nv\sigma^2}{2n\sigma^2}\right) \left(\frac{v\sigma^2}{n}\right)^{n/2-1} \left(\frac{\sigma^2}{n}\right) \\ &= \frac{1}{\Gamma\left(\frac{1}{2}n\right) 2^{n/2}} e^{-v/2} \left(\frac{n}{\sigma^2}\right)^{n/2-1} v^{n/2-1} \left(\frac{\sigma^2}{n}\right)^{n/2-1} \\ &= \frac{1}{\Gamma\left(\frac{1}{2}n\right) 2^{n/2}} e^{-v/2} v^{n/2-1}, \quad (v > 0). \end{aligned}$$

Por lo tanto v tiene distribución χ_n^2 . Esto nos muestra que si $A \sim W_1(n, \sigma^2/n)$ entonces $nA/\sigma^2 \sim \chi_n^2$.

2.2. Función característica

El siguiente teorema proporciona la función característica de la distribución Wishart.

Teorema 2.2. *Si $A \sim W_p(n, \Sigma)$, entonces la función característica de A (la función característica de las $\frac{1}{2}p(p+1)$ variables a_{ij} , $1 \leq i \leq j \leq p$) es*

$$\begin{aligned} \Phi(\Theta) &= \mathbb{E} \left[\exp \left(i \sum_{j \leq k} \Theta_{jk} a_{jk} \right) \right] = \mathbb{E} \left[\text{etr} \left(\frac{i}{2} A \Gamma \right) \right] \\ &= \det(I_p - i\Gamma\Sigma)^{-n/2}, \end{aligned}$$

donde $\Theta = (\Theta_{ij})$ es una matriz simétrica de $p \times p$, $\Gamma = (\gamma_{ij})$ es una matriz de $p \times p$, con $\gamma_{ij} = (1 + \delta_{ij})\Theta_{ij}$, $\Theta_{ij} = \Theta_{ji}$, y δ_{ij} es la delta Kronecker

$$\delta_{ij} = \begin{cases} 1, & \text{si } i = j; \\ 0, & \text{si } i \neq j. \end{cases}$$

Demostración. La función característica puede ser escrita como

$$\Phi(\Theta) = \mathbb{E} \left\{ \exp \left[\frac{i}{2} \sum_{j,k=1}^p (1 + \delta_{jk}) \Theta_{jk} a_{jk} \right] \right\},$$

ya que si consideramos a

$$\frac{i}{2} \sum_{j,k=1}^p (1 + \delta_{jk}) \Theta_{jk} a_{jk}$$

entonces se tienen los siguientes casos:

i) Si $j = k$,

$$\frac{i}{2} \sum_{j=1}^p (1 + \delta_{jj}) \Theta_{jj} a_{jj} = \frac{i}{2} \sum_{j=1}^p 2 \Theta_{jj} a_{jj}.$$

ii) Si $j < k$,

$$\frac{i}{2} \sum_{j < k}^p (1 + \delta_{jk}) \Theta_{jk} a_{jk} = \frac{i}{2} \sum_{j < k}^p \Theta_{jk} a_{jk}.$$

iii) Si $j > k$

$$\frac{i}{2} \sum_{j > k}^p (1 + \delta_{jk}) \Theta_{jk} a_{jk} = \frac{i}{2} \sum_{j > k}^p \Theta_{jk} a_{jk}.$$

Así tenemos que

$$\begin{aligned} \frac{i}{2} \sum_{j,k=1}^p (1 + \delta_{jk}) \Theta_{jk} a_{jk} &= \frac{i}{2} \sum_{j=1}^p 2 \Theta_{jj} a_{jj} + \frac{i}{2} \sum_{j < k}^p \Theta_{jk} a_{jk} + \frac{i}{2} \sum_{j > k}^p \Theta_{jk} a_{jk} \\ &= \frac{i}{2} \left(\sum_{j=1}^p 2 \Theta_{jj} a_{jj} + \sum_{j < k}^p 2 \Theta_{jk} a_{jk} \right) \\ &= i \sum_{j < k}^p \Theta_{jk} a_{jk}. \end{aligned}$$

Debido a que $\Gamma = (\gamma_{ij})$, $i, j = 1, \dots, p$, con $\gamma_{ij} = (1 + \delta_{ij})\Theta_{ij}$, $\Theta_{ij} = \Theta_{ji}$ tenemos que

$$\frac{i}{2} \sum_{j,k=1}^p (1 + \delta_{jk})\Theta_{jk}a_{jk} = \frac{i}{2} \sum_{j,k=1}^p \gamma_{jk}a_{jk},$$

y sabemos que γ_{jk} y a_{jk} son matrices simétricas de $p \times p$, así que

$$\frac{i}{2} \sum_{j,k=1}^p \gamma_{jk}a_{jk} = \frac{i}{2} \text{tr}(A\Gamma).$$

Supongamos que $A = \sum_{j=1}^n Z_j Z_j'$, donde Z_1, \dots, Z_n son vectores aleatorios independientes $N_p(\mathbf{0}, \Sigma)$. Entonces

$$\begin{aligned} \Phi(\Theta) &= \mathbb{E} \left[\text{etr} \left(\frac{i}{2} A\Gamma \right) \right] \\ &= \mathbb{E} \left[\text{etr} \left(\frac{i}{2} \sum_{j=1}^n Z_j Z_j' \Gamma \right) \right] = \mathbb{E} \left[\text{etr} \left(\frac{i}{2} \sum_{j=1}^n Z_j' \Gamma Z_j \right) \right] \\ &= \prod_{j=1}^n \mathbb{E} \left[\exp \frac{i}{2} Z_j' \Gamma Z_j \right] = \left[\mathbb{E} \left(\exp \frac{i}{2} Z_1' \Gamma Z_1 \right) \right]^n. \end{aligned}$$

Sea $Y = \Sigma^{-1/2} Z_1$, por el Teorema 1.4 tenemos que $Y \sim N_p(\mathbf{0}, I_p)$, entonces

$$\Phi(\Theta) = \left(\mathbb{E} \left[\exp \frac{i}{2} Y' \Sigma^{1/2} \Gamma \Sigma^{1/2} Y \right] \right)^n.$$

Debido a que $\Sigma^{1/2} \Gamma \Sigma^{1/2}$ es una matriz simétrica real, por el Teorema A.4 existe una matriz ortogonal H de $p \times p$ tal que

$$H \Sigma^{1/2} \Gamma \Sigma^{1/2} H' = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n),$$

donde $\lambda_1, \dots, \lambda_n$ son los valores propios de $\Sigma^{1/2} \Gamma \Sigma^{1/2}$. Sea $U = HY$, por el Teorema 1.4 se tiene que $U \sim N_p(\mathbf{0}, I_p)$. Así, si $U = (U_1, U_2, \dots, U_p)'$ tenemos

que

$$\begin{aligned}
 \Phi(\Theta) &= \left(\mathbb{E} \left[\exp \frac{i}{2} U' \Lambda U \right] \right)^n \\
 &= \left(\mathbb{E} \left[\exp \frac{i}{2} \sum_{j=1}^m \lambda_j U_j^2 \right] \right)^n \\
 &= \prod_{j=1}^m \mathbb{E} \left(\exp \frac{i}{2} \lambda_j U_j^2 \right)^n \\
 &= \prod_{j=1}^m (1 - i \lambda_j)^{-n/2},
 \end{aligned}$$

donde se usa el hecho que U_j^2 , $j = 1, \dots, p$, son variables aleatorias independientes con distribución χ_1^2 , cuya función característica es de la forma $\Phi(t) = (1 - 2it)^{-1/2}$. Por propiedades del determinante dadas en la Sección A.2, tenemos que

$$\begin{aligned}
 \prod_{j=1}^p (1 - i \lambda_j) &= \det(I_p - i \Lambda) \\
 &= \det(I_p - i H \Sigma^{1/2} \Gamma \Sigma^{1/2} H') = \det(I_p - i \Sigma^{1/2} \Gamma \Sigma^{1/2} H' H) \\
 &= \det(I_p - i \Sigma^{1/2} \Gamma \Sigma^{1/2}) = \det(I_p - i \Gamma \Sigma^{1/2} \Sigma^{1/2}) \\
 &= \det(I_p - i \Gamma \Sigma).
 \end{aligned}$$

Por lo tanto

$$\Phi(\Theta) = \det(I_p - i \Gamma \Sigma)^{-n/2}.$$

□

Como consecuencia del teorema anterior, se tiene el siguiente resultado que nos dice que la suma de matrices de Wishart independientes con la misma matriz de covarianza seguirá siendo Wishart.

Teorema 2.3. Sean A_1, \dots, A_r matrices aleatorias de $p \times p$ independientes donde $A_i \sim W_p(n_i, \Sigma)$, $i = 1, \dots, r$, entonces $\sum_{i=1}^r A_i \sim W_p(n, \Sigma)$, donde $n = \sum_{i=1}^r n_i$.

Demostración. Como las A_i , con $i = 1, 2, \dots, r$, son independientes, entonces la función característica de $U = \sum_{i=1}^r A_i$ es el producto de las funciones características de las A_i , $i = 1, \dots, r$. Así, por el Teorema 2.2 tenemos que

$$\begin{aligned}\Phi_U(\Theta) &= \prod_{i=1}^r \Phi_{A_i}(\Theta) \\ &= \prod_{i=1}^r \det(I_p - i\Gamma\Sigma)^{-n_i/2} \\ &= \det(I_p - i\Gamma\Sigma)^{-n/2},\end{aligned}$$

la cual es la función característica de la distribución Wishart con n grados de libertad y matriz de covarianza Σ . \square

2.3. Distribución marginal

El siguiente teorema nos servirá para demostrar resultados sobre la distribución marginal de submatrices de una matriz Wishart.

Teorema 2.4. *Si $A \sim W_p(n, \Sigma)$ y M es una matriz de $k \times p$ y de rango k , entonces $MAM' \sim W_k(n, M\Sigma M')$.*

Demostración. La función característica de MAM' es

$$\begin{aligned}\Phi_{MAM'}(\Theta) &= \mathbb{E} \left[\text{etr} \left(\frac{i}{2} MAM'\Gamma \right) \right] \\ &= \mathbb{E} \left[\text{etr} \left(\frac{i}{2} AM'\Gamma M \right) \right] \quad (\text{propiedades de la traza, ver Sección A.1}) \\ &= \det(I_p - iM'\Gamma M\Sigma)^{-n/2} \quad (\text{por el Teorema 2.2}) \\ &= \det(I_k - i\Gamma M\Sigma M')^{-n/2} \quad (\text{propiedades del determinante, ver Sección A.2}).\end{aligned}$$

Por lo tanto $MAM' \sim W_k(n, M\Sigma M')$. \square

Como consecuencia del teorema anterior tenemos el siguiente corolario, que nos dice que la distribución marginal de una submatriz localizada en la diagonal de una matriz Wishart es también Wishart.

Corolario 2.2. *Si $A \sim W_p(n, \Sigma)$ y A y Σ están particionadas como*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

donde A_{11} y Σ_{11} son matrices de $k \times k$, entonces $A_{11} \sim W_k(n, \Sigma_{11})$.

Demostración. Sea $M = [I_k : \mathbf{0}]$ de dimensión $k \times p$. Por el Teorema 2.4 tenemos que

$$\begin{aligned} MAM' &= [I_k : \mathbf{0}] \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I_k \\ \mathbf{0} \end{bmatrix} \\ &= A_{11} \end{aligned}$$

$$\begin{aligned} M\Sigma M' &= [I_k : \mathbf{0}] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_k \\ \mathbf{0} \end{bmatrix} \\ &= \Sigma_{11}, \end{aligned}$$

por lo tanto $A_{11} \sim W_k(n, \Sigma_{11})$. \square

El resultado del corolario anterior también es válido para la submatriz A_{22} , en ese caso la distribución marginal es $W_{p-k}(n, \Sigma)$. A continuación tenemos el siguiente teorema que proporciona una condición suficiente para la independencia de A_{11} y A_{22} .

Teorema 2.5. *Si $A \sim W_p(n, \Sigma)$, donde A y Σ están particionadas como en el corolario anterior y $\Sigma_{12} = \mathbf{0}$, entonces A_{11} y A_{22} son independientes y sus respectivas distribuciones son $W_k(n, \Sigma_{11})$ y $W_{p-k}(n, \Sigma_{22})$.*

Demostración. Sea $A = \sum_{\alpha=1}^n Z_\alpha Z_\alpha'$, donde Z_1, \dots, Z_n son i.i.d. $N_p(\mathbf{0}, \Sigma)$. Sea Z_α particionada como

$$Z_\alpha = (Z_\alpha^{(1)'}, Z_\alpha^{(2)'})',$$

donde $Z_\alpha^{(1)'}$, $Z_\alpha^{(2)'}$ son de dimensión k y $p - k$, respectivamente. Si $\Sigma_{12} = \mathbf{0}$, entonces $Z_\alpha^{(1)'}$ y $Z_\alpha^{(2)'}$ son independientes y como

$$A_{11} = \sum_{\alpha=1}^n Z_\alpha^{(1)} Z_\alpha^{(1)'} \text{ y } A_{22} = \sum_{\alpha=1}^n Z_\alpha^{(2)} Z_\alpha^{(2)'},$$

se tiene que A_{11} y A_{22} son independientes con distribución $W_k(n, \Sigma_{11})$ y $W_{p-k}(n, \Sigma_{22})$, respectivamente. \square

2.4. Algunos resultados de la Teoría de Matrices Aleatorias

En esta sección se presenta el comportamiento asintótico de los eigenvalores de la matriz de covarianza muestral para datos normales multivariados, en

particular se presenta la distribución asintótica de la *Distribución Empírica Espectral* de la matriz de covarianza muestral dada por el *Teorema de Marchenko-Pastur*, y la distribución asintótica del eigenvalor más grande, la cual es la *ley de Tracy-Widom*.

2.4.1. Teorema de Wigner

Las siguientes definiciones y resultados, así como sus demostraciones, pueden ser consultados en [5].

Definición 2.2. *Supongamos que A es una matriz de $n \times n$ con eigenvalores λ_j , $j = 1, 2, \dots, n$. Si todos los eigenvalores son reales (es decir, A es una matriz hermitiana), la función de distribución empírica de estos eigenvalores está dada por*

$$F_A(x) = \frac{1}{n} \#\{i \leq n : \lambda_i \leq x\} = \frac{\sum_{i=1}^n \mathbf{I}_{[\lambda_i, \infty)}(x)}{n} = \frac{\#\{1 \leq i \leq n : \lambda_i \leq x\}}{n},$$

y es llamada **Distribución Empírica Espectral (DEE)** de la matriz A .

Definición 2.3. *Una **matriz de Wigner** es una matriz aleatoria A^n de $n \times n$ con elementos reales o complejos tales que sus entradas A_{ij}^n , con $1 \leq i \leq j \leq n$, son independientes y A^n es hermitiana, es decir $A_{ij}^n = \overline{A_{ji}^n}$.*

Consideraremos matrices Wigner que cumplen las siguientes tres condiciones:

$$E(A_{ij}^n) = 0, \tag{2.1}$$

$$E(|A_{ij}^n|^2) = 1/n, \tag{2.2}$$

$$M_k \equiv \sup_{n \in \mathbb{N}} \max_{1 \leq i \leq j \leq n} E(|\sqrt{n}A_{ij}^n|^k) < \infty. \tag{2.3}$$

El siguiente teorema, conocido como el Teorema de Wigner, proporciona la convergencia casi segura de la distribución empírica espectral de una sucesión de matrices de Wigner.

Teorema 2.6 (Teorema de Wigner). *Sea A^n , $n \geq 1$, una sucesión de matrices de Wigner que cumplen las condiciones (2.1)–(2.3), entonces cuando $n \rightarrow \infty$*

$$F_{A^n}(x) := \frac{1}{n} \#\{j \leq n : \lambda_j \leq x\} \xrightarrow{c.s.} \int_{-\infty}^x \frac{1}{2\pi} \sqrt{4 - t^2} dt, \quad \forall x \in [-2, 2].$$

La distribución límite del Teorema de Wigner es llamada la **distribución del semicírculo**. La siguiente proposición nos proporciona los momentos de esta distribución.

Proposición 2.1. *Los momentos pares de la distribución del semicírculo están dados por los **números de Catalan***

$$\int_{-2}^2 y^{2k} \frac{1}{2\pi} \sqrt{4 - y^2} dy = \frac{1}{k+1} \binom{2k}{k}.$$

Los momentos impares son cero por la simetría de la distribución.

2.4.2. Teorema de Marchenko-Pastur (M-P)

Un resultado análogo al Teorema de Wigner pero considerando matrices Wishart es el Teorema de Marchenko-Pastur, el cual puede ser consultado en [2] y [3], y es presentado a continuación.

Teorema 2.7 (Teorema de Marchenko-Pastur (M-P)). *Sea $S_n = A_n/n$ una matriz de covarianza muestral, donde $A_n \sim W_p(n, \sigma^2 I_p)$, $p = p(n)$ y $p/n \rightarrow \gamma > 0$, cuando $n \rightarrow \infty$, entonces para toda $x \in \mathbb{R}$*

$$F_{S_n}(x) \rightarrow F(x) \quad \text{cuando } n \rightarrow \infty$$

*casi seguramente, donde F es la **ley de Marchenko-Pastur (M-P)**, cuya densidad está dada por*

$$f(x) = \frac{1}{2\pi\gamma\sigma^2} \sqrt{(b-x)(x-a)} \mathbf{I}_{(a,b)}(x) + \mathbf{I}_{(1,\infty)}(\gamma) \left(1 - \frac{1}{\gamma}\right) \delta_0(x),$$

donde $a = \sigma^2(1 - \sqrt{\gamma})^2$, $b = \sigma^2(1 + \sqrt{\gamma})^2$, y δ_0 es la función delta de Dirac en cero.

Notemos que si $\gamma > 1$ (caso de dimensión mayor al tamaño de muestra) la ley de Marchenko-Pastur tiene un átomo en $x = 0$ con masa $1 - 1/\gamma$. En el teorema anterior σ^2 es el parámetro de escala. Si $\sigma^2 = 1$ decimos que la ley de M-P es la ley de M-P estándar. En lo sucesivo consideraremos a la ley de M-P estándar.

La siguiente proposición debida a Gemman [6] y Silverstein [14], afirma que los eigenvalores positivos de la matriz de covarianza muestral tienden a estar dentro del soporte de la ley de M-P, ya que el eigenvalor positivo más pequeño converge casi seguramente al extremo inferior del soporte, mientras que el eigenvalor más grande converge casi seguramente al extremo superior del soporte.

Proposición 2.2 (Gemman (1980) y Silverstein (1985)). *Bajo las mismas hipótesis del teorema anterior, si l_1 y l_r son el eigenvalor más grande y el eigenvalor más pequeño no cero de S_n respectivamente, con $r = \min(n, p)$, entonces*

$$l_r \rightarrow (1 - \sqrt{\gamma})^2 \quad y \quad l_1 \rightarrow (1 + \sqrt{\gamma})^2$$

casi seguramente, cuando $p/n \rightarrow \gamma > 0$ y $n \rightarrow \infty$.

En el siguiente resultado se presenta la forma que tienen los momentos de la ley de M-P.

Proposición 2.3. *Si $\gamma \leq 1$, los momentos de la distribución de Marchenko-Pastur están dados por*

$$\int_a^b x^r f(x) dx = \sum_{k=0}^{r-1} \frac{\gamma^k}{k+1} \binom{r}{k} \binom{r-1}{k}, \quad \forall r \geq 1.$$

Demostración. Tenemos que $a + b = 2(1 + \gamma)$ y $ab = (1 - \gamma)^2$, entonces

$$\begin{aligned} \sqrt{(b-x)(x-a)} &= \sqrt{bx - ab - x^2 + ax} \\ &= \sqrt{-\left(x - \frac{a+b}{2}\right)^2 + \frac{(a+b)^2}{4} - ab} \\ &= \sqrt{-\left(x - \frac{2(1+\gamma)}{2}\right)^2 + \frac{4(1+\gamma)^2}{4} - (1-\gamma)^2} \\ &= \sqrt{4\gamma - [x - (1+\gamma)]^2}. \end{aligned}$$

Así tenemos que

$$\begin{aligned} \int_a^b x^r f(x) dx &= \int_a^b \frac{x^r}{2\pi x \gamma} \sqrt{(b-x)(x-a)} dx \\ &= \int_a^b \frac{x^{r-1}}{2\pi \sqrt{\gamma}} \sqrt{4 - \frac{(x - (1+\gamma))^2}{\gamma}} dx. \end{aligned}$$

Haciendo $x = (y\sqrt{\gamma} + 1 + \gamma)$, $dx = \sqrt{\gamma} dy$ y usando el *teorema del binomio* se tiene

$$\begin{aligned} \int_a^b x^r f(x) dx &= \int_{-2}^2 \frac{1}{2\pi} (y\sqrt{\gamma} + 1 + \gamma)^{r-1} \sqrt{4 - y^2} dy \\ &= \int_{-2}^2 \sum_{k=0}^{r-1} \binom{r-1}{k} (1+\gamma)^{r-1-k} (y\sqrt{\gamma})^k \frac{1}{2\pi} \sqrt{4 - y^2} dy \\ &= \sum_{k=0}^{(r-1)/2} \binom{r-1}{2k} (1+\gamma)^{r-1-2k} \gamma^k \int_{-2}^2 y^{2k} \frac{1}{2\pi} \sqrt{4 - y^2} dy. \end{aligned}$$

La última igualdad es porque los momentos impares de la ley del semicírculo son cero. La Proposición 2.1 implica que la expresión anterior es igual a

$$\begin{aligned}
 & \sum_{k=0}^{(r-1)/2} (1+\gamma)^{r-1-2k} \gamma^k \binom{r-1}{2k} \binom{2k}{k} \frac{1}{k+1} \\
 &= \sum_{k=0}^{(r-1)/2} (1+\gamma)^{r-1-2k} \gamma^k \frac{(r-1)!}{k!(r-1-2k)!(k+1)!} \\
 &= \sum_{k=0}^{(r-1)/2} \sum_{s=0}^{r-1-2k} \gamma^s \gamma^k \binom{r-1-2k}{s} \frac{(r-1)!}{k!(r-1-2k)!(k+1)!} \\
 &= \sum_{k=0}^{(r-1)/2} \sum_{s=0}^{r-1-2k} \gamma^{s+k} \frac{(r-1)!}{s!(r-1-2k-s)!k!(k+1)!}.
 \end{aligned}$$

Tomando $t = k + s$, la expresión anterior es igual a

$$\begin{aligned}
 & \sum_{k=0}^{(r-1)/2} \sum_{t=k}^{r-1-k} \gamma^t \frac{(r-1)!}{(t-k)!(r-1-k-t)!k!(k+1)!} \\
 &= \sum_{t=0}^{r-1} \gamma^t \sum_{k=0}^{\min(t, r-1-t)} \frac{1}{r} \binom{r}{t} \binom{t}{k} \binom{r-t}{k+1} \\
 &= \frac{1}{r} \sum_{t=0}^{r-1} \gamma^t \binom{r}{t} \sum_{k=0}^{\min(t, r-1-t)} \binom{t}{k} \binom{r-t}{r-t-k-1}.
 \end{aligned}$$

Usando la *identidad de Vandermonde* $\binom{m+n}{r} = \sum_{k=0}^r \binom{m}{k} \binom{n}{r-k}$, la última expresión es igual a

$$\begin{aligned}
 \frac{1}{r} \sum_{t=0}^{r-1} \gamma^t \binom{r}{t} \binom{r}{r-t-1} &= \frac{1}{r} \sum_{t=0}^{r-1} \gamma^t \binom{r}{t} \binom{r}{t+1} \\
 &= \sum_{t=0}^{r-1} \gamma^t \frac{1}{t+1} \binom{r}{t} \binom{r-1}{t}.
 \end{aligned}$$

□

La siguiente proposición muestra la relación entre la ley del semicírculo y la ley de M-P.

Proposición 2.4. *Si X es una variable aleatoria con la distribución del semicírculo, entonces X^2 tiene la distribución de M-P.*

Demostración. Sea $Y = X^2$, por el método de la transformación para transformaciones uno a uno por pedazos (ver [12], Sección 5.1), tenemos que

$$\begin{aligned} f_Y(y) &= \frac{1}{2\pi 2\sqrt{y}} \left[\sqrt{4 - (\sqrt{y})^2} + \sqrt{4 - (\sqrt{-y})^2} \right] \\ &= \frac{1}{2\pi y} \sqrt{y} \sqrt{4 - y} \\ &= \frac{1}{2\pi y} \sqrt{(4 - y)y}, \end{aligned}$$

la cual es la densidad de M-P con $\gamma = 1$. Si $\gamma = 1$, la función de densidad de M-P tiene soporte $a = 0$ y $b = 4$. \square

El Teorema de Marchenko-Pastur se cumple para matrices más generales de la forma $B_n = YY^T$, donde las entradas de la matriz Y de $p \times n$ son variables aleatorias i.i.d. con media cero y varianza finita (ver [2]).

2.4.3. Ley de Tracy-Widom (T-W)

El siguiente teorema, debido a Johnstone [7], muestra cuál es la distribución asintótica del eigenvalor más grande de una matriz Wishart. Este resultado será utilizado en el siguiente capítulo para realizar pruebas de hipótesis acerca de la matriz de covarianza poblacional de datos normales multivariados de dimensión alta.

Teorema 2.8 (Johnstone (2001)). *Sea $A \sim W_p(n, I_p)$ y sea l_1 el eigenvalor más grande de A . Si $p/n \rightarrow \gamma > 0$, entonces*

$$\frac{l_1 - \mu_{np}}{\sigma_{np}} \xrightarrow{\text{dist}} F_1,$$

donde las constantes de centralización y escala son

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

y F_1 es la función de distribución de la **ley de Tracy-Widom** definida como

$$F_1(s) = \exp \left(-\frac{1}{2} \int_s^\infty q(x) + (x-s)q^2(x) dx \right), \quad s \in \mathbb{R},$$

donde q son las soluciones de la ecuación diferenciable de Painlevé II

$$q''(x) = xq(x) + 2q^3(x), \quad q(x) \sim Ai(x), \quad \text{cuando } x \rightarrow +\infty,$$

y $Ai(x)$ es la función Airy.

La función de distribución F_1 proviene de una familia de distribuciones F_β , donde $\beta = 1, 2, 4$. La función F_β aparece como la distribución del eigenvalor más grande de los conjuntos **Gaussian Othogonal Ensemble**, **Gaussian Unitary Ensemble** y **Gaussian Symplectic Ensemble** (GOE, GUE y GSE), respectivamente (ver [11]). A continuación describiremos estos conjuntos.

- **Gaussian Othogonal Ensemble** (GOE). Este conjunto está compuesta por matrices aleatorias reales y simétricas. Sus elementos H_{jk} tales que $j \leq k$ son estadísticamente independientes y su distribución de probabilidad es invariante bajo todas las transformaciones reales y ortogonales de H , es decir, bajo la transformación

$$H \rightarrow W'HW,$$

donde W es cualquier matriz real ortogonal.

- **Gaussian Unitary Ensemble** (GUE). Este conjunto está compuesto por matrices aleatorias hermitianas. Sus elementos diagonales H_{jj} y las partes real e imaginaria de sus elementos no diagonales H_{jk} con $j \leq k$ son estadísticamente independientes y su distribución de probabilidad es invariante bajo todas las transformaciones unitarias de H , es decir, bajo la transformación

$$H \rightarrow U^{-1}HU,$$

donde U es cualquier matriz unitaria.

- **Gaussian Symplectic Ensemble** (GSE). Este conjunto está compuesto por matrices hermitianas y autoduales. Sus elementos diagonales y sus cuatro elementos cuaterniónicos de los elementos no diagonales H_{jk} con $j > k$ son estadísticamente independientes y su distribución de probabilidad es invariante bajo todas las transformaciones simplécticas de H , es decir, bajo la transformación

$$H \rightarrow W^RHW,$$

donde W es cualquier matriz simpléctica y W^R es el dual de W .

En [8] se sugiere usar las **constantes de centralización y escala de segundo orden** para una convergencia más rápida, dadas por

$$\mu'_{np} = \left(\sqrt{n - \frac{1}{2}} + \sqrt{p - \frac{1}{2}} \right)^2,$$

$$\sigma'_{np} = \left(\sqrt{n - \frac{1}{2}} + \sqrt{p - \frac{1}{2}} \right) \left(\frac{1}{\sqrt{n - \frac{1}{2}}} + \frac{1}{\sqrt{p - \frac{1}{2}}} \right)^{1/3}.$$

El Teorema 2.8 es válido usando las constantes μ'_{np} y σ'_{np} , es decir, se tiene el siguiente resultado.

Teorema 2.9. *Bajo las mismas condiciones del teorema anterior, si $p/n \rightarrow \gamma > 0$, entonces*

$$\frac{l_1 - \mu'_{np}}{\sigma'_{np}} \xrightarrow{\text{dist}} F_1,$$

donde μ'_{np} y σ'_{np} son las constantes de centralización y escala definidas anteriormente.

Capítulo 3

Pruebas de hipótesis para la matriz de covarianza poblacional con datos de dimensión alta

En este capítulo se presentan pruebas de hipótesis para la matriz de covarianza poblacional al considerar datos de dimensión alta. Se utilizan resultados de la Teoría de Matrices Aleatorias presentados en el capítulo anterior, específicamente el Teorema de Marchenko-Pastur y la distribución de Tracy-Widom, para determinar si la matriz de covarianza poblacional de datos normales multivariados de dimensión alta tienen una cierta matriz de covarianza poblacional.

3.1. Teorema de M-P en el estudio de la matriz de covarianza poblacional

El siguiente ejemplo, tomado de Johnstone [8], ilustra la forma en que el Teorema de Marchenko-Pastur y la Proposición 2.2 explican la dispersión de los eigenvalores muestrales cuando los datos tienen distribución normal estándar multivariada.

Ejemplo 3.1 (Johnstone (2007)). *Se obtuvieron observaciones de $n = 10$ vectores independientes X_1, X_2, \dots, X_n con distribución $N_p(\mathbf{0}, I_p)$, con $p = 10$. En este caso los eigenvalores de la matriz de covarianza poblacional son todos iguales a uno, pero los eigenvalores observados de la matriz de covarianza*

muestral $S_n = \sum_{i=1}^n X_i X_i' / n$ fueron

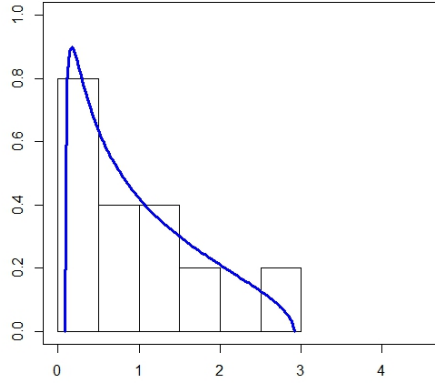
0.003, 0.036, 0.095, 0.16, 0.30, 0.51, 0.78, 1.12, 1.40, 3.07.

Notar que en el cálculo de S_n no se resta la media muestral, esto es común hacerlo cuando la media poblacional es cero. Se puede observar una dispersión extrema en los eigenvalores muestrales y no todos ellos están cerca de uno. Este fenómeno puede explicarse por el Teorema 2.7 y la Proposición 2.2, porque el límite de la distribución empírica espectral de la matriz de covarianza muestral cuando $p = n$, es decir cuando $\gamma = 1$, es la ley de M-P con soporte

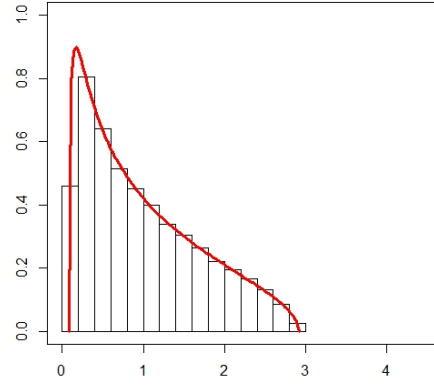
$$a = (1 - \sqrt{1})^2 = 0 \quad y \quad b = (1 + \sqrt{1})^2 = 4,$$

que corresponde al rango de los eigenvalores observados.

Para estudiar con más detalle lo visto en el ejemplo anterior, se hicieron simulaciones de muestras aleatorias de una distribución normal $N_p(\mathbf{0}, I_p)$. Se consideraron varios valores de n y p , que contemplan el caso clásico en que $\gamma < 1$ y el de dimensión alta en que $\gamma \geq 1$ en el Teorema de Marchenko-Pastur, y se calcularon los eigenvalores muestrales para ver su dispersión. Los resultados se muestran en las figuras 3.1–3.4, donde se grafica el histograma de los eigenvalores muestrales (escalado de tal forma que tiene área total igual a uno) y la función de densidad M-P para comparar el ajuste de la distribución empírica espectral con la distribución de M-P. En todas las gráficas se observa un buen ajuste, en particular cuando n y p son grandes. Se observa también el átomo en $x = 0$ de la distribución empírica espectral para los casos en que $\gamma > 1$. La probabilidad del átomo $x = 0$ es $1 - 1/\gamma$, sin embargo en los histogramas no se observa esta altura debido a que se está haciendo un acercamiento para que se pueda apreciar bien la densidad de M-P.

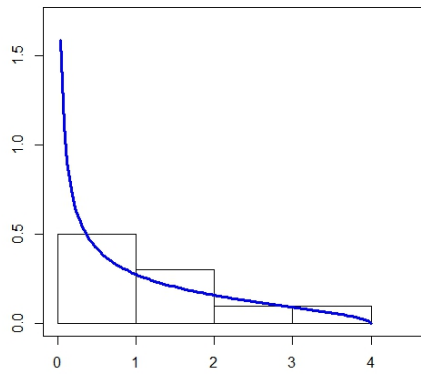


(a) $p = 10, n = 20, \gamma = 1/2$

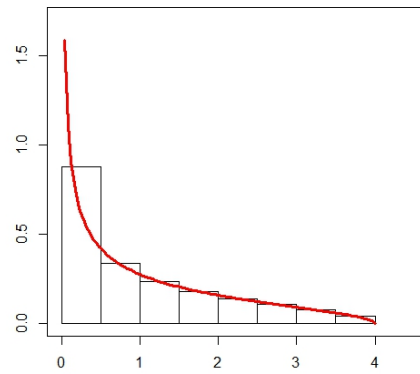


(b) $p = 1000, n = 2000, \gamma = 1/2$

Figura 3.1: Histograma de los eigenvalores muestrales y la función de densidad de M-P, cuando la matriz de covarianza poblacional es la identidad.

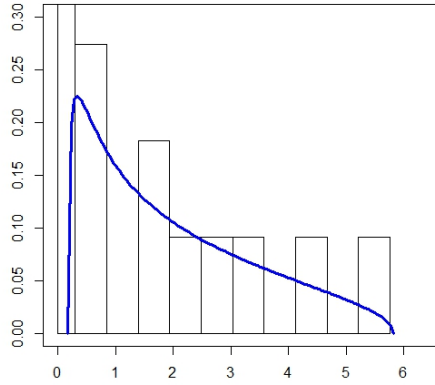


(a) $p = n = 10, \gamma = 1$

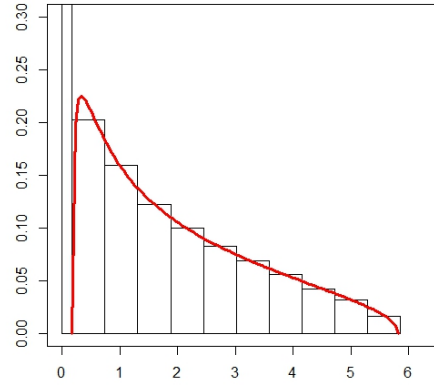


(b) $p = n = 1000, \gamma = 1$

Figura 3.2: Histograma de los eigenvalores muestrales y la función de densidad de M-P, cuando la matriz de covarianza poblacional es la identidad.

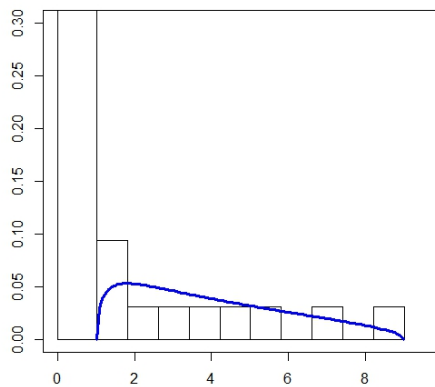


(a) $p = 20, n = 10, \gamma = 2$

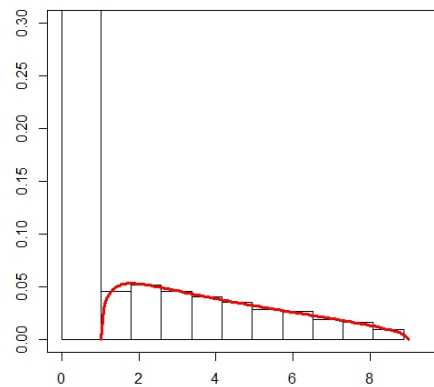


(b) $p = 2000, n = 1000, \gamma = 2$

Figura 3.3: Histograma de los eigenvalores muestrales y la función de densidad de M-P, cuando la matriz de covarianza poblacional es la identidad.



(a) $p = 40, n = 10, \gamma = 4$



(b) $p = 2000, n = 500, \gamma = 4$

Figura 3.4: Histograma de los eigenvalores muestrales y la función de densidad de M-P, cuando la matriz de covarianza poblacional es la identidad.

Con el análisis anterior podemos concluir que si se observan eigenvalores muestrales muy alejados del soporte de la ley de M-P se tendría una evidencia de que la matriz de covarianza poblacional no es la identidad. Un modelo de matriz de covarianza que puede producir eigenvalores muestrales muy alejados del soporte de la ley de M-P se presenta continuación.

Ejemplo 3.2. *Considérese el modelo de covarianza Spiked dado por*

$$\Sigma = \text{diag}(\lambda_1, 1, 1, \dots, 1),$$

con $\lambda_1 > 1$. Si $A_n \sim W_p(n, \Sigma)$, el Teorema de Marchenko-Pastur sigue siendo cierto (ver [10]). Sin embargo si l_1 es el eigenvalor muestral más grande, entonces

$$l_1 \rightarrow \begin{cases} (1 + \sqrt{\gamma})^2, & \lambda_1 \leq 1 + \sqrt{\gamma}; \\ \lambda_1 \left(1 + \frac{\gamma}{\lambda_1 - 1}\right), & \lambda_1 > 1 + \sqrt{\gamma}; \end{cases}$$

casi seguramente cuando $n, p \rightarrow \infty$ y $p/n \rightarrow \gamma$ (ver [3]). Cuando $\lambda_1 > 1 + \sqrt{\gamma}$ se tiene que

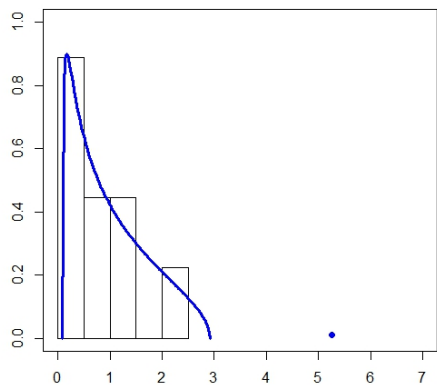
$$(1 + \sqrt{\gamma})^2 < \lambda_1 \left(1 + \frac{\gamma}{\lambda_1 - 1}\right),$$

es decir, ya no se cumple que los eigenvalores muestrales tienden a permanecer en el soporte de la ley M-P.

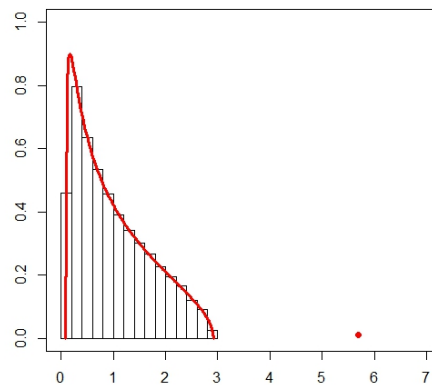
Para este modelo se hicieron simulaciones de muestras aleatorias de una distribución normal multivariada $N_p(\mathbf{0}, \Sigma)$, donde Σ es el modelo de covarianza Spiked con $\lambda_1 = 5$, considerando los mismos valores de n y p que usamos anteriormente, y se calcularon los eigenvalores muestrales para ver su dispersión. Los resultados se muestran en las figuras 3.5–3.8. En las figuras se observa que el eigenvalor muestral más grande se encuentra alejado del soporte de la ley de M-P, el histograma del resto de los eigenvalores muestrales se ajusta bien a la ley de M-P.

Por lo tanto, el Teorema de M-P nos proporciona un método para determinar si hay evidencia para suponer que la matriz de covarianza poblacional de datos multivariados normales de dimensión alta es la matriz identidad, es decir, en el caso en que existan eigenvalores muestrales muy alejados del soporte de la ley de M-P concluimos que hay evidencia en contra de que la matriz de covarianza poblacional es la identidad.

El problema con el método anterior es que es subjetiva la decisión de si un eigenvalor muestral está muy alejado del soporte de la ley de M-P, por lo que se requieren llevar a cabo pruebas de hipótesis formales para la matriz de covarianza poblacional, las cuales se presentan en la siguiente sección.

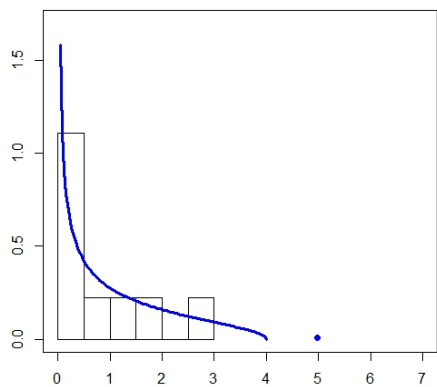


(a) $p = 10, n = 20, \gamma = 1/2$

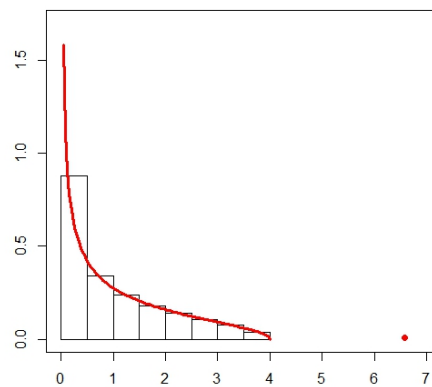


(b) $p = 1000, n = 2000, \gamma = 1/2$

Figura 3.5: Histograma de los eigenvalores muestrales y la función de densidad de M-P, cuando la matriz de covarianza poblacional es el modelo de covarianza Spiked con $\lambda_1 = 5$. Se marca con “•” el eigenvalor muestral más grande.

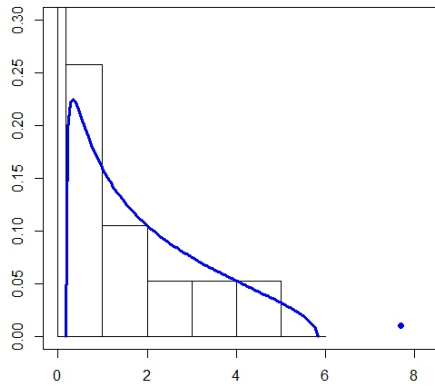


(a) $p = n = 10, \gamma = 1$

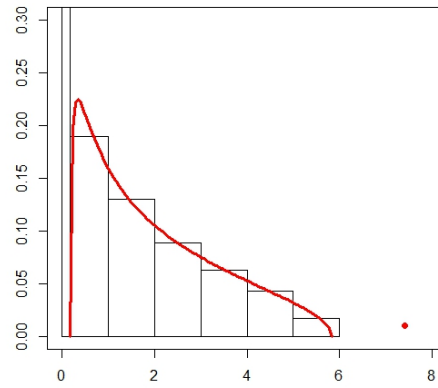


(b) $p = n = 1000, \gamma = 1$

Figura 3.6: Histograma de los eigenvalores muestrales y la función de densidad de M-P, cuando la matriz de covarianza poblacional es el modelo de covarianza Spiked con $\lambda_1 = 5$. Se marca con “•” el eigenvalor muestral más grande.

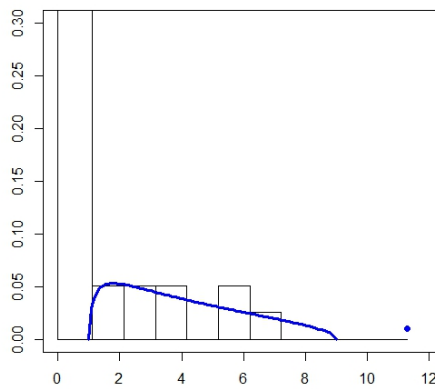


(a) $p = 20, n = 10, \gamma = 2$

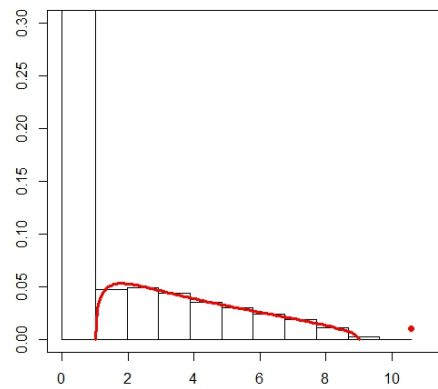


(b) $p = 2000, n = 1000, \gamma = 2$

Figura 3.7: Histograma de los eigenvalores muestrales y la función de densidad de M-P, cuando la matriz de covarianza poblacional es el modelo de covarianza Spiked con $\lambda_1 = 5$. Se marca con “•” el eigenvalor muestral más grande.



(a) $p = 40, n = 10, \gamma = 4$



(b) $p = 2000, n = 500, \gamma = 4$

Figura 3.8: Histograma de los eigenvalores muestrales y la función de densidad de M-P, cuando la matriz de covarianza poblacional es el modelo de covarianza Spiked con $\lambda_1 = 5$. Se marca con “•” el eigenvalor muestral más grande.

3.2. Pruebas de hipótesis para la matriz de covarianza poblacional basadas en la ley de T-W

En esta sección se verá la aplicación de la distribución Tracy-Widom de la Teoría de Matrices Aleatorias, presentada en el capítulo anterior, para llevar a cabo pruebas de hipótesis sobre la matriz de covarianza poblacional de datos normales multivariados.

3.2.1. Prueba de hipótesis de esfericidad

A continuación presentamos una prueba de hipótesis de esfericidad, es decir, una prueba para la hipótesis de que la matriz de covarianza poblacional es proporcional a la matriz identidad, considerando datos normales multivariados de dimensión alta.

El resultado del Teorema 2.8 proporciona una herramienta para llevar a cabo pruebas de hipótesis de la matriz de covarianza a partir del eigenvalor más grande de la matriz de covarianza muestral (eigenvalor muestral más grande), en particular para probar

$$H_0 : \Sigma = I_p \quad \text{vs} \quad H_1 : \Sigma \neq I_p. \quad (3.1)$$

Para ilustrar esto consideremos el siguiente ejemplo de Johnstone [8], donde se lleva a cabo el contraste de las hipótesis anteriores utilizando la distribución de Tracy-Widom.

Ejemplo 3.3 (Johnstone (2007)). *Dada una muestra aleatoria de tamaño n de la distribución $N_p(\mathbf{0}, I_p)$, con $n = p = 10$, se observó que el eigenvalor muestral más grande era 4.25, el cual se encuentra fuera del soporte $[0, 4]$ de la ley de M-P. Utilizando la aproximación de la distribución del eigenvalor muestral más grande a ley T-W, con las constantes de centralización y escala de segundo orden del Teorema 2.9, se obtuvo que la probabilidad de observar un valor más grande que 4.25 es del 6%, por lo que considerando un nivel de significancia del 5% no se rechaza la hipótesis nula de que la matriz de covarianza poblacional sea la identidad.*

En el ejemplo anterior el eigenvalor muestral más grande se salió del soporte de la ley de M-P, pero esto no indica que la matriz de covarianza poblacional no sea la identidad y es necesario realizar pruebas de hipótesis para determinar si la matriz de covarianza poblacional podría o no ser la matriz identidad.

En resumen, debido al Teorema 2.9, si tenemos las observaciones de una muestra aleatoria de tamaño n de una normal multivariada $N_p(\mathbf{0}, \Sigma)$ y l_1 es el eigenvalor muestral más grande, una prueba de hipótesis de nivel α para contrastar las hipótesis (3.1) es rechazar H_0 si

$$\frac{nl_1 - \mu'_{np}}{\sigma'_{np}}$$

es mayor que el punto porcentual superior α de la distribución Tracy-Widom F_1 , denotado por $F_1(\alpha)$, donde μ'_{np} y σ'_{np} son las constantes de centralización y escala del Teorema 2.9.

Sean X_1, X_2, \dots, X_N vectores aleatorios i.i.d. con distribución $N_p(\mu, \Sigma)$. Supongamos ahora que estamos interesados en el juego de hipótesis

$$H_0 : \Sigma = \sigma^2 I_p \quad \text{vs} \quad H_1 : \Sigma \neq \sigma^2 I_p. \quad (3.2)$$

Sea S_n la matriz de covarianza muestral dada por la Definición 1.4, donde $n = N - 1$. Si σ es conocido, entonces bajo H_0 se tiene que $A = nS_n/\sigma \sim W_p(n, I_p)$ (ver Sección 2.1). Por lo que en este caso, para contrastar las hipótesis (3.2) se puede aplicar la prueba de hipótesis basada en la ley de T-W utilizando al eigenvalor más grande de A . Como se menciona en [9], en el caso en que σ es desconocido puede ser estimado con $\hat{\sigma} = \text{tr}S_n/p$, y al considerar que $\hat{A} = nS_n/\hat{\sigma}$ tiene distribución aproximada $W_p(n, I_p)$, se puede aplicar la prueba hipótesis basada en la ley de T-W utilizando el eigenvalor más grande de \hat{A} para contrastar el juego de hipótesis (3.2).

3.2.2. Prueba de hipótesis para $H_0 : \Sigma = \Sigma_0$

Sean X_1, X_2, \dots, X_N vectores aleatorios i.i.d. con distribución $N_p(\mu, \Sigma)$. Consideremos el juego de hipótesis

$$H_0 : \Sigma = \Sigma_0 \quad \text{vs} \quad H_1 : \Sigma \neq \Sigma_0, \quad (3.3)$$

donde Σ_0 es una matriz positiva definida especificada. Definamos $Y_i = \Sigma_0^{-1/2} X_i$, $i = 1, 2, \dots, N$, los cuales son vectores aleatorios i.i.d. con distribución $N_p(\Sigma_0^{-1/2} \mu, \Sigma_0^{-1/2} \Sigma \Sigma_0^{-1/2})$. Observemos que bajo la hipótesis nula las Y_i son i.i.d. con distribución $N_p(\Sigma_0^{-1/2} \mu, I_p)$. Debido a lo anterior, para contrastar las hipótesis (3.3) se puede aplicar la prueba de hipótesis basada en la ley de T-W utilizando al eigenvalor más grande de la matriz de covarianza muestral de las Y_i , que como se mostrará en el siguiente lema, es igual al eigenvalor más grande de $\Sigma_0^{-1} S_n$, donde S_n es la matriz de covarianza muestral de las X_i .

Lema 3.1. Sean X_i y Y_i , con $i = 1, 2, \dots, N$, como anteriormente. Denotamos por $l_1(A)$ al eigenvalor más grande de una matriz A . Sea S_n la matriz de covarianza muestral de las X_i dada por la Definición 1.4. Entonces la matriz de covarianza muestral de las Y_i está dada por $\Sigma_0^{-1/2} S_n \Sigma_0^{-1/2}$ y

$$l_1(\Sigma_0^{-1} S_n) = l_1(\Sigma_0^{-1/2} S_n \Sigma_0^{-1/2}).$$

Demostración. Veamos que la matriz de covarianza muestral de las Y_i está dada por $\Sigma_0^{-1/2} S_n \Sigma_0^{-1/2}$. Tenemos que

$$\begin{aligned} \Sigma_0^{-1/2} S_n \Sigma_0^{-1/2} &= \Sigma_0^{-1/2} \frac{1}{n} \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})' \Sigma_0^{-1/2} \\ &= \frac{1}{n} \sum_{i=1}^N \Sigma_0^{-1/2} (X_i - \bar{X})(X_i - \bar{X})' \Sigma_0^{-1/2} \\ &= \frac{1}{n} \sum_{i=1}^N (\Sigma_0^{-1/2} X_i - \Sigma_0^{-1/2} \bar{X})(\Sigma_0^{-1/2} X_i - \Sigma_0^{-1/2} \bar{X})'. \end{aligned}$$

Notar que $\Sigma_0^{-1/2} \bar{X} = \frac{1}{N} \sum_{i=1}^N \Sigma_0^{-1/2} X_i = \bar{Y}$, por lo que la última expresión de de arriba es igual a

$$\hat{S} = \frac{1}{n} \sum_{i=1}^N (Y_i - \bar{Y})(Y_i - \bar{Y})',$$

la cual es la matriz de covarianza muestral de las Y_i . Ahora mostraremos que

$$l_1(\Sigma_0^{-1} S_n) = l_1(\Sigma_0^{-1/2} S_n \Sigma_0^{-1/2}).$$

Observemos que si λ es un eigenvalor de $\Sigma_0^{-1/2} S_n$, entonces para algún $v \neq 0$ se tiene que

$$\begin{aligned} \Sigma_0^{-1} S_n v &= \lambda v \\ \iff \Sigma_0^{-1/2} \Sigma_0^{-1/2} S_n v &= \lambda v \\ \iff \Sigma_0^{-1/2} S_n v &= \lambda \Sigma_0^{1/2} v \\ \iff \Sigma_0^{-1/2} S_n \Sigma_0^{-1/2} (\Sigma_0^{1/2} v) &= \lambda (\Sigma_0^{1/2} v). \end{aligned}$$

Tomando $\hat{v} = \Sigma_0^{1/2} v$ en la expresión anterior, tenemos que

$$\Sigma_0^{-1/2} S_n \Sigma_0^{-1/2} \hat{v} = \lambda \hat{v},$$

Ahora veamos que $\widehat{v} \neq 0$. Debido a que $\Sigma_0^{1/2}$ es una matriz invertible sus columnas son linealmente independientes, por lo que si $\widehat{v} = \Sigma_0^{1/2}v = 0$, entonces $v = 0$, lo cual es una contradicción ya que habíamos supuesto que $v \neq 0$. Así

$$\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}\widehat{v} = \lambda\widehat{v}, \quad \text{con } \widehat{v} \neq 0,$$

por lo tanto λ es un eigenvalor de $\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}$. Análogamente se puede mostrar que si λ es un eigenvalor de $\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}$ entonces también es un eigenvalor de $\Sigma_0^{-1/2}S_n$. De este modo los eigenvalores de $\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}$ y $\Sigma_0^{-1/2}S_n$ son equivalentes y se tiene que $l_1(\Sigma_0^{-1}S_n) = l_1(\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2})$. \square

Como se mencionó anteriormente, el problema de probar la hipótesis nula en (3.3) se traduce en probar la hipótesis nula de que la matriz de covarianza de las Y_i es la identidad. Para realizar la prueba de hipótesis con la ley T-W se debe calcular entonces el eigenvalor más grande de la matriz de covarianza muestral de las Y_i , que por el lema anterior es precisamente $l_1(\Sigma_0^{-1}S_n)$. A partir de aquí el procedimiento para aplicar la prueba de hipótesis usando la ley T-W es como se describió en la Sección 3.2.1.

En resumen, para contrastar las hipótesis (3.3) lo que se hace es lo siguiente:

- Se calcula $l_1(\Sigma_0^{-1}S_n)$, que es igual al eigenvalor más grande de la matriz de covarianza muestral de las Y_i .
- Debido a que

$$\frac{nl_1(\Sigma_0^{-1}S_n) - \mu'_{np}}{\sigma'_{np}} \xrightarrow{dist} F_1, \quad (3.4)$$

donde μ'_{np} y σ'_{np} son las constantes de centralización y escala de segundo orden del Teorema 2.9, entonces rechazamos H_0 con un nivel de significancia α si el lado izquierdo de (3.4) es mayor que el punto porcentual superior α de la distribución Tracy-Widom F_1 , denotado por $F_1(\alpha)$.

3.3. Estudio de simulación para evaluar las pruebas de hipótesis basadas en la ley de T-W

Por medio de simulaciones se evaluó que tan buena es la prueba basada en la ley Tracy-Widom para determinar si muestras aleatorias normales multivariadas tienen una cierta matriz de covarianza.

Se consideró el juego de hipótesis (3.1) debido a que es el más simple. Para distintos valores de n y p se simularon $M = 1000$ muestras aleatorias de tamaño n con distribución $N_p(\mathbf{0}, I_p)$ y se calculó la proporción de veces que se rechazó la hipótesis nula para varios niveles de significancia, si la prueba es buena estas proporciones deben parecerse a los niveles de significancia. Se consideraron casos en que $\gamma = p/n \geq 1$ (caso de dimensión alta) y en que $\gamma = p/n < 1$ (caso clásico). Los niveles de significancia considerados fueron $\alpha = 0.1, 0.05, 0.01$. Los resultados se exponen en los cuadros 3.1–3.4.

α	$n=20, p=10$	$n=200, p=100$	$n=1000, p=500$	$n=2000, p=1000$
0.1	0.1	0.091	0.102	0.097
0.05	0.052	0.044	0.044	0.04
0.01	0.009	0.013	0.013	0.006

Cuadro 3.1: Resultados de una muestra aleatoria con $\gamma = 1/2$.

α	$n=10, p=10$	$n=100, p=100$	$n=500, p=500$	$n=1000, p=1000$
0.1	0.097	0.102	0.11	0.102
0.05	0.047	0.045	0.061	0.051
0.01	0.016	0.009	0.013	0.009

Cuadro 3.2: Resultados de una muestra aleatoria con $\gamma = 1$.

α	$n=10, p=20$	$n=100, p=200$	$n=500, p=1000$	$n=1000, p=2000$
0.1	0.1	0.103	0.102	0.103
0.05	0.049	0.054	0.052	0.054
0.01	0.012	0.01	0.009	0.012

Cuadro 3.3: Resultados de una muestra aleatoria con $\gamma = 2$.

α	$n=10, p=40$	$n=50, p=200$	$n=100, p=400$	$n=500, p=2000$
0.1	0.093	0.086	0.108	0.102
0.05	0.044	0.043	0.048	0.039
0.01	0.012	0.01	0.014	0.007

Cuadro 3.4: Resultados de una muestra aleatoria con $\gamma = 4$.

En general se observa que las proporciones en que se rechaza la hipótesis nula es cercana al nivel de significancia y que la aproximación no varía significativamente al aumentar los valores de n y p , manteniendo fija la razón

$\gamma = p/n$. Se concluye que la prueba de hipótesis basada en la ley Tracy-Widom es una prueba aceptable, tanto en el caso de dimensión alta, que es el de mayor interés en este trabajo, como en el caso clásico.

3.4. Comparación de la prueba de hipótesis basada en la ley T-W con la basada en el estadístico de esfericidad

3.4.1. Prueba de hipótesis basada en el estadístico de esfericidad

Una prueba muy utilizada en Análisis Multivariado clásico para contrastar las hipótesis (3.2) es la basada en el estadístico de esfericidad, el cual se define a continuación.

Definición 3.1. Sean X_1, X_2, \dots, X_N vectores aleatorios i.i.d. con distribución $N_p(\mu, \Sigma)$, y sea

$$S_n = \frac{1}{n} \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})', \quad \text{con } n = N - 1,$$

la matriz de covarianza muestral. El estadístico

$$V = \frac{\det S_n}{(\text{tr} S_n / p)^p},$$

es llamado *estadístico de esfericidad*.

El estadístico de esfericidad surge de la prueba de razón de verosimilitud para $H_0 : \Sigma = \sigma^2 I_p$, la cual consiste en rechazar la hipótesis nula con un nivel de significancia α si $V \leq k_\alpha$, donde k_α es el punto porcentual inferior α de la distribución de V (ver [1], [13]). Si la hipótesis nula es verdadera es claro que V debe ser cercano a uno. El siguiente resultado, que puede ser consultado en [13], proporciona una aproximación a la distribución de $-n\rho \log V$ a través de la distribución chi-cuadrada.

Teorema 3.1. Cuando la hipótesis $H_0 : \Sigma = \sigma^2 I$ es verdadera, la distribución de $-n\rho \log V$, donde $\rho = 1 - (2p^2 + p + 2)/6pn$, sigue aproximadamente una distribución chi-cuadrada con $f = (p+2)(p-1)/2$ grados de libertad, cuando n es grande, es decir,

$$\mathbb{P}(-n\rho \log V \leq x) \approx \mathbb{P}(\chi_f^2 \leq x), \quad \forall x \in \mathbb{R}.$$

Utilizando esta aproximación, una prueba de hipótesis de nivel α para contrastar las hipótesis (3.2) es rechazar H_0 si $-n\rho \log(V) > \chi_f^2(\alpha)$, donde $\chi_f^2(\alpha)$ es el punto porcentual superior α de la distribución chi-cuadrada con f grados libertad.

La prueba de esfericidad con el estadístico V es usado únicamente cuando $n \geq p$. Nótese que en el caso en que $n < p$ el determinante de S_n es cero y por lo tanto $V = 0$.

Si nos interesa probar $H_0 : \Sigma = \Sigma_0$, donde Σ_0 es una matriz especificada, como se vió anteriormente esto es equivalente a mostrar que la matriz de covarianza de los datos transformados $Y_i = \Sigma_0^{-1/2} X_i$, $i = 1, 2, \dots, N$, es la identidad. Ya que la matriz de covarianza muestral de las Y_i es $\Sigma_0^{-1/2} S_n \Sigma_0^{-1/2}$, donde S_n es la matriz de covarianza de las X_i , se tiene por propiedades de la traza y del determinante (secciones A.1 y A.2) que el estadístico de esfericidad para las Y_i es

$$V = \frac{\det(\Sigma_0^{-1} S_n)}{(\text{tr}(\Sigma_0^{-1} S_n)/p)^p}.$$

De este modo, una prueba de hipótesis de nivel α para $H_0 : \Sigma = \Sigma_0$ contra $H_1 : \Sigma \neq \Sigma_0$ es rechazar H_0 si $-n\rho \log(V) > \chi_f^2(\alpha)$, donde $\chi_f^2(\alpha)$ es el punto porcentual superior α de la distribución chi-cuadrada con $f = (p+2)(p-1)/2$ grados libertad.

3.4.2. Estudio de simulación para la comparación de las pruebas de hipótesis

A continuación se presenta un estudio de simulación para comparar el desempeño de la prueba de hipótesis que utiliza el estadístico de esfericidad con la prueba de hipótesis basada en la ley T-W. Recordemos que esta última prueba puede ser empleada cuando $n < p$ o cuando $n \geq p$.

Para estas simulaciones se consideró el juego de hipótesis (3.3), en lugar de considerar el juego de hipótesis más simple (3.1), esto con el fin de ilustrar la utilidad de la ley de T-W para llevar a cabo pruebas de hipótesis de la matriz de covarianza poblacional, no solo en el caso de pruebas de esfericidad sino también en pruebas más generales. Para esto se consideraron varios valores de p y n , y se simularon $M = 1000$ muestras aleatorias de tamaño $N = n + 1$ con distribución $N_p(\mathbf{0}, \Sigma)$, donde Σ es una matriz Toeplitz (ver Sección A.6) de la forma $\Sigma = \text{Toeplitz}(1, c, c^2, \dots, c^{p-1})$ con $c = 0.5$, la cual por el Lema A.1 es positiva definida. Los escenarios en las simulaciones contemplan un valor pequeño de p (10) y un valor grande (100), y para cada valor de p se consideraron cinco valores de n tales que las mismas razones p/n son tomadas

en cuenta con los dos valores de p , con el fin de ver el comportamiento de las pruebas hipótesis al hacer crecer tanto p como n , pero manteniendo la misma razón $\gamma = p/n$. Se calculó la proporción de veces que se rechazaba la hipótesis nula con ambas pruebas para los niveles de significancia $\alpha = 0.1, 0.05, 0.01$.

La mejor prueba es aquella que tiene la proporción de veces que se rechaza la hipótesis nula más cercana al nivel de significancia. Los resultados de las simulaciones se exponen en los cuadros 3.5 y 3.6. En estos cuadros la prueba χ_f^2 se refiere a la prueba de hipótesis que se basa en el estadístico de esfericidad y la prueba T-W se refiere a la prueba de hipótesis basada en la ley T-W.

$p=10$	α			
	Pruebas	0.1	0.05	0.01
$n=11$	χ_f^2	0.352	0.250	0.130
	T-W	0.092	0.047	0.009
$n=15$	χ_f^2	0.172	0.097	0.025
	T-W	0.089	0.048	0.012
$n=20$	χ_f^2	0.133	0.073	0.017
	T-W	0.092	0.043	0.010
$n=50$	χ_f^2	0.089	0.046	0.010
	T-W	0.097	0.053	0.011
$n=100$	χ_f^2	0.112	0.054	0.009
	T-W	0.087	0.044	0.012

Cuadro 3.5: Comparación de las pruebas considerando p y n pequeñas

$p=100$	α			
	Pruebas	0.1	0.05	0.01
$n=110$	χ_f^2	1	1	1
	T-W	0.098	0.050	0.009
$n=150$	χ_f^2	0.896	0.820	0.611
	T-W	0.109	0.048	0.007
$n=200$	χ_f^2	0.452	0.320	0.120
	T-W	0.116	0.064	0.012
$n=500$	χ_f^2	0.110	0.061	0.011
	T-W	0.112	0.060	0.009
$n=1000$	χ_f^2	0.107	0.050	0.012
	T-W	0.084	0.037	0.008

Cuadro 3.6: Comparación de las pruebas considerando p y n grandes

Notemos que en las simulaciones solo se consideró el caso en que $\gamma =$

$p/n < 1$ y no el caso de interés, es decir, el caso de dimensión alta cuando $\gamma \geq 1$, debido a que cuando $p > n$ algunos de los eigenvalores de la matriz $\Sigma_0^{-1}S_n$ son cero y el estadístico de esfericidad es cero y por tanto $-n\rho\log V$ no existe.

Las proporciones de rechazo de la prueba basada en el estadístico de esfericidad tienden a ser buenas cuando n es lo suficientemente grande con respecto a p , pero son muy malas en caso contrario, ya que estas proporciones no se aproximan a los niveles de significancia deseados. Los peores resultados de esta prueba se observan cuando $p = 100$ y $n = 110, 150, 200$ (p y n grandes, con n no tan alejado de p), en este caso las proporciones de rechazo son mucho más grandes que el nivel de significancia correspondiente. En el caso en que $p = 10$ y $n = 11, 15, 20$ (p y n pequeños, con n no tan alejado de p), las proporciones de rechazo son también más grandes que el nivel de significancia correspondiente pero son un poco más cercanas a él.

Con respecto a la prueba basada en la ley de T-W las proporciones de rechazo se aproximan a los niveles de significancia aún cuando n no es tan grande con respecto a p . Los resultados considerando p y n pequeños son semejantes a cuando son grandes, es decir, al hacer crecer p y n manteniendo la misma razón $\gamma = p/n$ los resultados no varían significativamente.

Los resultados de los cuadros 3.5 y 3.6 nos indican que la prueba basada en la ley de T-W resultó ser mucho mejor que la prueba basada en el estadístico de esfericidad para este tipo de datos en que $n > p$. Se observa por tanto que la aproximación a la distribución T-W dada por el Teorema 2.8 es mucho mejor que la aproximación a la distribución chi-cuadrada dada por el Teorema 3.1. Cabe mencionar que en los libros de texto clásicos de Análisis Multivariado las pruebas que se presentan para la matriz de covarianza poblacional son únicamente las basadas en el estadístico de esfericidad, y lo que se vió con los resultados de las simulaciones es que la prueba basada en la ley de T-W es una mejor alternativa.

Conclusiones

La Teoría de Matrices Aleatoria (TMA) nos brinda una herramienta para comprender el comportamiento asintótico de los eigenvalores de una matriz de covarianza muestral de ciertos datos de dimensión alta, en particular para el caso de datos normales multivariados. En este caso la distribución de la matriz de covarianza muestral es la distribución Wishart, la cual es una generalización de la distribución chi-cuadrada. La TMA proporciona resultados asintóticos de los eigenvalores de una matriz Wishart, en especial la distribución asintótica del eigenvalor más grande. Estos resultados pueden ser usados para estudiar la forma que tiene la matriz de covarianza poblacional de datos normales multivariados de dimensión alta, en particular para llevar a cabo pruebas de hipótesis de la matriz de covarianza poblacional.

El Teorema de Marchenko-Pastur (M-P) proporciona el límite de la distribución empírica espectral de la matriz de covarianza muestral cuando la razón entre la dimensión p de los datos y el tamaño de la muestra n tiende a una constante positiva. Este resultado es de utilidad para comprender el comportamiento disperso de los eigenvalores de la matriz de covarianza muestral (eigenvalores muestrales). Si los datos provienen de una distribución normal multivariada con matriz de covarianza identidad, se esperaría ver a los eigenvalores muestrales dispersos alrededor del soporte de la distribución de Marchenko-Pastur $[(1 - \sqrt{\gamma})^2, (1 + \sqrt{\gamma})^2]$, con $\gamma = p/n$. Si por el contrario, la matriz de covarianza poblacional no es la identidad, no hay garantía de que los eigenvalores muestrales pertenezcan a este intervalo o estén cerca de él, por lo que si en la práctica observamos que los eigenvalores muestrales están muy lejos de este intervalo, tendríamos evidencias para suponer que los datos no tienen matriz de covarianza poblacional igual a la matriz identidad. Un modelo de matriz de covarianza poblacional que produce eigenvalores muestrales muy alejados del soporte de la distribución de M-P es el modelo de covarianza Spiked. En este trabajo se llevaron a cabo simulaciones para estudiar la dispersión de los eigenvalores muestrales considerando datos normales multivariados, principalmente de dimensión alta. Con estas simulaciones se

comprueba la utilidad del Teorema de Marchenko-Pastur para determinar si hay evidencia para suponer que la matriz de covarianza poblacional de datos multivariados normales de dimensión alta es la matriz identidad. Sin embargo, como la decisión de si los eigenvalores muestrales están muy alejados del soporte de la ley de M-P es subjetiva, es necesario llevar a cabo pruebas de hipótesis formales de la matriz de covarianza poblacional.

La distribución asintótica del eigenvalor más grande de la matriz de covarianza muestral de una muestra aleatoria de la distribución normal multivariada es la ley de Tracy-Widom (T-W). Este resultado permite realizar pruebas de hipótesis de esfericidad y de que la matriz de covarianza poblacional es una matriz de covarianza específica, considerando datos normales multivariados. Las simulaciones llevadas a cabo en esta tesis para verificar el comportamiento de las pruebas de hipótesis basadas en la ley de T-W indican que esta prueba es aceptable, tanto en el caso de dimensión alta como en el caso clásico, ya que las proporciones de veces en que se rechazó la hipótesis nula fueron cercanas al nivel de significancia considerado. También se realizaron simulaciones, considerando únicamente el caso clásico, para comparar la prueba de hipótesis basada en la ley de T-W con la basada en el estadístico de esfericidad. En estas simulaciones se observó que la prueba basada en la ley de T-W resultó tener un mejor comportamiento que la basada en el estadístico de esfericidad, debido a que las proporciones de rechazo generalmente fueron cercanas al nivel de significancia considerado, mientras que con la prueba basada en el estadístico de esfericidad muchas veces las proporciones de rechazo resultaron muy alejadas del nivel de significancia considerado, en particular en casos en que la dimensión es cercana al tamaño de la muestra. Esto muestra que la TMA es una herramienta muy útil en el caso de dimensión alta y en el caso clásico para analizar la matriz de covarianza poblacional de datos normales multivariados.

Apéndice A

Algunas definiciones y propiedades de matrices

En este apéndice se presentan las principales definiciones y propiedades de matrices que se utilizan a lo largo de esta tesis. Iniciamos con el siguiente conjunto de definiciones.

- Si A es una matriz de $p \times q$, una **matriz rectangular** de números reales o complejos se escribe como

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix},$$

el término a_{ij} representa el elemento del i -ésimo renglón y la j -ésima columna, con esto la matriz A de $p \times q$ se escribe de la forma $A = (a_{ij})$.

- Si $p = q$, entonces A es llamada una **matriz cuadrada**.
- Si $q = 1$, entonces A es un **vector columna** y si $p = 1$, entonces A es un **vector fila**.
- Si $a_{ij} = 0$ para $i = 1, 2, \dots, p$ y $j = 1, 2, \dots, q$, entonces A es llamada la **matriz cero** y se escribe como $A = \mathbf{0}$.
- Si $p = q$, $a_{ii} = 1$ para $i = 1, 2, \dots, p$ y $a_{ij} = 0$ para $i \neq j$, entonces A es llamada la **matriz identidad** de orden p y se escribe como $A = I$ o $A = I_p$.
- Los **elementos de la diagonal** de una matriz A de $p \times p$ son a_{11}, \dots, a_{pp} .

- La **transpuesta** de una matriz A de $p \times q$, denotada como A' , es la matriz de $q \times p$ obtenida al intercambiar las filas y las columnas de A , es decir, si $A = (a_{ij})$, entonces $A' = (b_{ij})$ donde $b_{ij} = a_{ji}$.
- Si A es una matriz cuadrada de orden p esta es llamada **simétrica** si $A = A'$ y **anti-simétrica** si $A = -A'$. Si A es anti-simétrica, entonces los elementos de la diagonal son ceros.
- La matriz A es llamada **triangular superior** si los elementos que están debajo de la diagonal son ceros, y es de la forma

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ 0 & a_{22} & \cdots & a_{2p} \\ \vdots & & & \\ 0 & 0 & \cdots & a_{pp} \end{pmatrix}.$$

Si todos los elementos por encima de la diagonal son ceros, entonces diremos que A es una matriz **triangular inferior**.

- Si A es una matriz de $p \times p$ y todos los elementos que están fuera de la diagonal son ceros, entonces ésta es llamada una matriz **diagonal**, es de la forma

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & a_{pp} \end{pmatrix},$$

y se escribe como $A = \text{diag}(a_{11}, \dots, a_{pp})$.

- La **suma** de dos matrices A y B de $p \times q$ es

$$A + B = (a_{ij} + b_{ij}).$$

- Si A y B son matrices de $p \times q$ y $q \times r$ respectivamente (el número de columnas de la matriz A es igual al número de filas de la matriz B), entonces el **producto** de A y B es la matriz de $p \times r$ definida por

$$AB = \left(\sum_{k=1}^q a_{ik} b_{kj} \right).$$

El producto de una matriz por un escalar α está dado por

$$\alpha A = \alpha(a_{ij}) = (\alpha a_{ij}).$$

- Si A es una matriz de $p \times p$, es llamada ortogonal si $AA' = I_p = A'A$.
- Si A es una matriz simétrica de $p \times p$ se cumple que $AA' = A^2$.

Las siguientes propiedades de matrices son elementales:

- $A + (-1)A = \mathbf{0}$.
- $(AB)' = B'A'$.
- $(A')' = A$.
- $(A + B)' = A' + B'$.
- $A(BC) = (AB)C$.
- $A(B + C) = AB + AC$.
- $(A + B)C = AC + BC$.
- $AI = A$.

A.1. Traza de una matriz

Si A es una matriz de $p \times p$, la **traza** de A se define como la suma de los elementos de la diagonal, es decir

$$\text{tr}(A) = \sum_{i=1}^p a_{ii} = a_{11} + \dots + a_{pp}.$$

En particular $\text{tr}(I_p) = p$. Si A y B son matrices de $p \times p$ las siguientes propiedades de la traza son elementales:

- $\text{tr}(kA) = k\text{tr}(A)$, $\forall k \in \mathbb{R}$.
- $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$.
- $\text{tr}(A') = \text{tr}(A)$.
- $\text{tr}(AB) = \text{tr}(BA)$.
- $\text{tr}(AB)' = \text{tr}(B'A')$.

A.2. Determinantes

El **determinante** de una matriz cuadrada de $p \times p$, denotada por $|A|$ o $\det A$, está definido como

$$\det A = \sum_{\pi} \varepsilon_{\pi} a_{1j_1} a_{2j_2} \cdots a_{pj_p},$$

donde \sum_{π} denota la suma sobre todas las $p!$ permutaciones $\pi = (j_1, \dots, j_p)$ de $(1, \dots, p)$ y $\varepsilon_{\pi} = 1$ ó -1 de acuerdo a si permutación π es par o impar. Las siguientes propiedades del determinante son elementales:

- Si todos los elementos de una fila o columna de A son cero, entonces el $\det A = 0$.
- $\det A = \det A'$.
- Si todos los elementos de una fila o columna de A se multiplican por un escalar α , entonces el determinante se multiplica por α .
- Si A es una matriz de $p \times p$ y α es un escalar, entonces $\det(\alpha A) = \alpha^p \det A$.
- Si B es una matriz obtenida de A al intercambiar dos filas o columnas, entonces $\det B = -\det A$.
- Si dos filas o columnas de A son iguales, entonces el $\det A = 0$.
- Si A y B son matrices de $p \times p$, entonces $\det(AB) = (\det A)(\det B)$.
- Si A es de $p \times q$ y B es de $q \times p$, entonces $\det(I_p + AB) = \det(I_q + BA)$.

A.3. Inversa de una matriz

Si $A = (a_{ij})$ es de dimensión $p \times p$, con $\det A \neq 0$, A es llamada **matriz no singular**. En este caso hay una única matriz B tal que $AB = I_p = BA$. La matriz B es llamada la **inversa** de A y es denotada por A^{-1} . Las siguientes propiedades son elementales:

- $(A^{-1})' = (A')^{-1}$.
- Si A y C son matrices no singulares de $p \times p$, entonces $(AC)^{-1} = C^{-1}A^{-1}$.
- $\det(A)^{-1} = (\det A)^{-1}$.
- Si A es una matriz ortogonal, entonces $A^{-1} = A'$.

A.4. Eigenvalores e eigenvectores

Si A es una matriz de $p \times p$, la **ecuación característica** de la matriz A está dada por

$$\det(A - \lambda I_p) = 0, \quad \lambda \in \mathbb{R}.$$

El lado izquierdo es un polinomio de grado p en λ , por lo cual este polinomio tiene exactamente p raíces las cuales son los **eigenvalores** (o **valores propios**) de A . Estas raíces no son necesariamente distintas y pueden ser reales o complejas. Si λ_1 es un valor propio de A , entonces

$$\det(A - \lambda_1 I) = 0,$$

por lo tanto $A - \lambda_1 I$ es singular. Si X_1 es un vector no nulo tal que $(A - \lambda_1 I)X_1 = 0$, este es un **eigenvector** (o **vector propio**) de A correspondiente a λ_1 . A continuación se presentan algunos resultados importantes sobre los eigenvalores y eigenvectores de una matriz, así como también de descomposición de matrices, los cuales pueden ser consultados en [13].

Teorema A.1. *Si A es una matriz simétrica real, entonces sus eigenvalores son todos reales.*

Teorema A.2. *Sea A una matriz simétrica. Si λ_i y λ_j son eigenvalores distintos de A , entonces los correspondientes eigenvectores X_i y X_j son ortogonales.*

Teorema A.3. *Si A es una matriz real de $p \times p$ con eigenvalores reales, entonces existe una matriz ortogonal H tal que $H'AH$ es una matriz triangular superior cuyos elementos de la diagonal son los eigenvalores de A .*

Teorema A.4. *Si A es una matriz real simétrica de $p \times p$ con eigenvalores $\lambda_1, \dots, \lambda_p$, entonces existe una matriz ortogonal H de $p \times p$ tal que*

$$H'AH = D \equiv \text{diag}(\lambda_1, \dots, \lambda_p).$$

A.5. Matriz no negativa definida y positiva definida

Definición A.1. *Sea A una matriz simétrica de $m \times m$, llamaremos a la matriz A **no negativa definida** si*

$$\alpha' A \alpha \geq 0, \quad \forall \alpha \in \mathbb{R}^m.$$

Definición A.2. Sea A una matriz simétrica de $m \times m$, llamaremos a la matriz A **positiva definida** si

$$\alpha' A \alpha > 0, \quad \forall \alpha \in \mathbb{R}^m, \quad \alpha \neq 0.$$

Criterio A.1. Sea $A = (a_{ij})$, $i, j = 1, 2, \dots, m$, una matriz simétrica. El l -ésimo menor principal de A es $|B_l| = \det(B_l)$ donde $B_l = (a_{ij})$, $i, j = 1, 2, \dots, l$. La matriz A es positiva definida si y sólo si todos sus menores principales son positivos.

Criterio A.2. Una matriz simétrica A es positiva definida (no negativa definida) si y solo si todos sus valores propios son positivos (no negativos).

A.6. Matriz Toeplitz

La siguiente definición puede ser consultada en [4].

Definición A.3. Una **matriz Toeplitz** $A = (a_{ij})$ es una matriz $d \times d$, tal que $a_{ij} = a_{i+1, j+1}$, $i, j = 1, 2, \dots, d-1$.

Por ejemplo, para $d = 5$, la matriz Toeplitz presenta la forma

$$A = \begin{pmatrix} a & b & c & d & e \\ f & a & b & c & d \\ g & f & a & b & c \\ h & g & f & a & b \\ i & h & g & f & a \end{pmatrix}.$$

Notemos que si la matriz A es simétrica, entonces la matriz Toeplitz está determinada por la primera fila y por ejemplo, para $d = 5$, se usa la notación $A = \text{toeplitz}(a, b, c, d, e)$ para indicar que la matriz A tiene la forma

$$A = \begin{pmatrix} a & b & c & d & e \\ b & a & b & c & d \\ c & b & a & b & c \\ d & c & b & a & b \\ e & d & c & b & a \end{pmatrix}.$$

Así la matriz Toeplitz $\Sigma_d = \text{toeplitz}(1, s, s^2, \dots, s^{d-1})$, con $0 < s < 1$, es

$$\Sigma_d = \begin{pmatrix} 1 & s & s^2 & s^3 & \dots & s^{d-1} \\ s & 1 & s & s^2 & \dots & s^{d-2} \\ s^2 & s & 1 & s & \dots & s^{d-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s^{d-1} & s^{d-2} & s^{d-3} & s^{d-4} & \dots & 1 \end{pmatrix}.$$

El siguiente resultado es útil en la Sección 3.4.2.

Lema A.1. *La matriz Toeplitz $\Sigma_d = \text{toeplitz}(1, s, s^2, \dots, s^{d-1})$, con $0 < s < 1$, es positiva definida.*

Demostración. Demostraremos por inducción que los menores principales $|B_l|$, satisfacen $|B_l| = (1 - s^2)^{l-1}$, $l = 1, 2, \dots, d$. Así, por el Criterio A.1, Σ_d es positiva definida.

Para $i = 1$, tenemos que $|B_l| = 1 = (1 - s^2)^0$. Supongamos que $|B_l| = (1 - s^2)^{l-1}$. Veamos ahora que $|B_{l+1}| = (1 - s^2)^l$. Por propiedades de determinantes tenemos que

$$\begin{aligned}
 |B_{l+1}| &= \begin{vmatrix} 1 & s & s^2 & s^3 & \dots & s^l \\ s & 1 & s & s^2 & \dots & s^{l-1} \\ s^2 & s & 1 & s & \dots & s^{l-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s^{l-1} & s^{l-2} & s^{l-3} & s^{l-4} & \dots & s \\ s^l & s^{l-1} & s^{l-2} & s^{l-3} & \dots & 1 \end{vmatrix} \\
 &= 1|B_l| - s \begin{vmatrix} s & s^2 & s^3 & \dots & s^{l-1} & s^l \\ s & 1 & s & \dots & s^{l-3} & s^{l-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s^{l-1} & s^{l-2} & s^{l-3} & \dots & s & 1 \end{vmatrix} \\
 &\quad + s^2 \begin{vmatrix} s & s^2 & \dots & s^{l-2} & s^{l-1} & s^l \\ 1 & s & \dots & s^{l-3} & s^{l-2} & s^{l-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s^{l-1} & s^{l-2} & \dots & s^2 & s & 1 \end{vmatrix} \\
 &\quad + \dots + (-1)^{l+1} s^l \begin{vmatrix} s & s^2 & \dots & s^{l-2} & s^{l-1} & s^l \\ 1 & s & \dots & s^{l-3} & s^{l-2} & s^{l-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s^{l-2} & s^{l-3} & \dots & s^3 & s^2 & s \end{vmatrix}.
 \end{aligned}$$

Notemos que a partir del segundo término de la última expresión, podemos factorizar s del primer renglón de la matriz en el determinante. A partir del tercer término, cada determinante es igual a cero debido a que la primera y la segunda fila de la matriz son iguales. Así

$$|B_{l+1}| = |B_l| - s^2|B_l| = |B_l|(1 - s^2) = (1 - s^2)^l,$$

por la hipótesis de inducción. De esta forma $|B_l| = (1 - s^2)^{l-1} > 0$, para todo $l \geq 1$.

□

Apéndice B

Algoritmos de simulaciones

B.1. Teorema de Marchenko-Pastur

Las gráficas mostradas en la sección 3.1 fueron obtenidas mediante los algoritmos que presentamos a continuación.

(a) Para $n \geq p$

```
library(MASS)
n=2000
p=1000
g=p/n
a=(1-sqrt(g))^2
b=(1+sqrt(g))^2
mu=rep(0,p)
Sigma=diag(rep(1,p))
X=mvrnorm(n,mu,Sigma)
Sn=t(X)%*%X/n
eival=sort(eigen(Sn)$values)
eival
hist(eival,freq=FALSE, xlab="", ylab="",
xlim=c(0,4.5), ylim=c(0,1), axes=FALSE,main="")
par(new=TRUE)
MP<-function(x)
{sqrt((b-x)*(x-a))/(2*pi*x*g)}
plot(MP1,a,b,xlab="", ylab="",col="red",
lwd= "3", xlim=c(0,4.5), ylim=c(0,1)).
```

(b) Para $n < p$


```

n1=500
p1=2000
g=p1/n1
a=(1-sqrt(g))^2
b=(1+sqrt(g))^2
mu1=rep(0,p1)
Sigma1=diag(rep(1,p1))
X1=mvrnorm(n1,mu1,Sigma1)
Sn1=t(X1)%*%X1/n1
eival1=sort(eigen(Sn1)$values)
eival1 #[(n1+1):p1]
a1=eival1[p1-n1+1]
ap=eival1[p1]
k=10
v=seq(a1-0.0001,ap, (ap-a1+0.0001)/k)
h=hist(eival1,freq=FALSE, breaks=c(0, v), xlab="", ylab="",
xlim=c(0,b+0.5), ylim=c(0,0.3), axes=FALSE,main="")
par(new=TRUE)
MP1<-function(x)
{sqrt((b-x)*(x-a))/(2*pi*x*g)}
plot(MP1,a,b,xlab="", ylab="",col="red",
lwd= "3", xlim=c(0,b+0.5),ylim=c(0,0.3)).

```

B.2. Pruebas de hipótesis basadas en la ley de T-W

Los resultados expuestos en la Sección 3.3 fueron obtenidos con el siguiente algoritmo. Se contempla el caso clásico $\gamma = p/n < 1$ y de dimensión alta $\gamma = p/n \geq 1$.

```

library(MASS)
library(RMTstat)
n=100
p=400
beta=1
pr1=0.9
pr05=0.95
pr01=0.99
Q1= qtw(pr1,beta, lower.tail= T)
Q05= qtw(pr05,beta, lower.tail= T)

```

```

Q01= qtw(pr01,beta, lower.tail= T)
cont1=0
cont05=0
cont01=0
mu=rep(0,p)
Sigma=diag(rep(1,p))
mnp=(sqrt(n-1/2)+ sqrt(p-1/2))^2
signp=sqrt(mnp)*(1/sqrt(n-1/2)+1/sqrt(p-1/2))^(1/3)
M=1000
for (i in 1:M){
X=mvrnorm(n,mu,Sigma)
Sn=t(X)%*%X/n
A=Sn*n
eival=eigen(A)$values
L1=max(eival)
F=(L1-mnp)/signp
pval=1-ptw(F,beta, lower.tail= T)
if (Q1<F){
cont1=cont1+1
}

if (Q05<F){
cont05=cont05+1
}

if (Q01<F){
cont01=cont01+1
}
}

cont1/M
cont05/M
cont01/M

```

B.3. Pruebas de hipótesis basadas en la ley de T-W y el estadístico de esfericidad

Los resultados expuestos en la Sección 3.4.2 fueron obtenidos con el siguiente algoritmo. Se contempla solo el caso clásico $\gamma = p/n < 1$.

```

library(MASS)
library(RMTstat)
n=150
p=100
N=n+1
mu=rep(0,p)
beta=1
f=p*(p+1)/2-1
ro=1-(2*p^2+p+2)/(6*p*n)

pr1=0.9
pr05=0.95
pr01=0.99

P1=qchisq(pr1,f, lower.tail = TRUE, log.p = FALSE)
P05=qchisq(pr05,f, lower.tail = TRUE, log.p = FALSE)
P01=qchisq(pr01,f, lower.tail = TRUE, log.p = FALSE)
contch1=0
contch05=0
contch01=0

Q1= qtw(pr1,beta, lower.tail= T)
Q05= qtw(pr05,beta, lower.tail= T)
Q01= qtw(pr01,beta, lower.tail= T)
cont1=0
cont05=0
cont01=0

g=0.5
y=g^(0:(p-1))
Sigma=toeplitz(y)

mnp=(sqrt(n-1/2)+ sqrt(p-1/2))^2
signp=sqrt(mnp)*(1/sqrt(n-1/2)+1/sqrt(p-1/2))^(1/3)

M=1000
for (i in 1:M){
X=mvrnorm(N,mu,Sigma)
S=var(X)
L=solve(Sigma)
C=L%*%S
}

```

```
eival=eigen(C)$values
trC=sum(eival)
detC=prod(eival)
V=(detC)/((trC/p)^p)
E=-n*ro*log(V)
pval0=1-pchisq(E,f)
pval0

L1=max(eival)
F=(n*L1-mnp)/signp
pval=1-ptw(F,beta, lower.tail= T)
pval

if(P1<E){
contch1=contch1+1
}
if(P05<E){
contch05=contch05+1
}
if(P01<E){
contch01=contch01+1
}
if (Q1<F){
cont1=cont1+1
}

if (Q05<F){
cont05=cont05+1
}

if (Q01<F){
cont01=cont01+1
}

}
contch1/M
contch05/M
contch01/M
cont1/M
cont05/M
cont01/M
```

Bibliografía

- [1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, second edition, 1984.
- [2] Z. D. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.
- [3] J. Baik and J. W. Silverstein. Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.
- [4] A. Böttcher and S. M. Grudsky. *Toeplitz Matrices, Asymptotic Linear Algebra, and Functional Analysis*. Birkhäuser, 2012.
- [5] J. A. Domínguez Molina and A. Rocha Arteaga. El Teorema de Wigner para Matrices Aleatorias. *Miscelánea Matemática*, 52:31–51, 2011.
- [6] S. Geman. A Limit Theorem for the Norm of Random Matrices. *Annals of Probability*, 8(2):252–261, 1980.
- [7] I. M. Johnstone. On the Distribution of the Largest Eigenvalue in Principal Components Analysis. *The Annals of Statistics*, 29(2):295–327, 2001.
- [8] I. M. Johnstone. High Dimensional Statistical Inference and Random Matrices. In *Proceedings of the International Congress of Mathematicians, Vol. I*, pages 307–333. European Mathematical Society, 2007.
- [9] Z. Ma. Accuracy of the Tracy–Widom Limits for the Extreme Eigenvalues in White Wishart Matrices. *Bernoulli*, 18(1):322–359, 2012.

- [10] V. A. Marčenko and L. A. Pastur. Distribution of Eigenvalues for Some Sets of Random Matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483, 1967.
- [11] M. L. Mehta. *Random Matrices*. Academic press, 2004.
- [12] A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics*. McGraw-hill, 1974.
- [13] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, 2009.
- [14] J. W. Silverstein. The Smallest Eigenvalue of a Large Dimensional Wishart Matrix. *Annals of Probability*, 13(4):1364–1368, 1985.