



UNIVERSIDAD DE BUENOS AIRES  
Facultad de Ciencias Exactas y Naturales  
Departamento de Matemática

Tesis de Licenciatura

Tests de igualdad de matrices de dispersión basados en signos en el contexto de  
alta dimensión

Gonzalo Chebi

Directora: Dra. Graciela Boente Boente

Marzo de 2015

## Tests de igualdad de matrices de dispersión basados en signos en el contexto de alta dimensión

El desarrollo de métodos en contextos de alta dimensión surgió recientemente debido a grandes avances en el campo del almacenamiento y procesamiento de datos. Los test clásicos de máxima verosimilitud muchas veces asumen que el tamaño de la muestra  $n$  es mayor que la dimensión de cada vector de la muestra  $p$ . Además, necesitan que  $n$  sea mucho mayor que  $p$  para ser informativos. En este trabajo estudiaremos el problema de testear igualdad de matrices de covarianza. En los últimos años se propusieron nuevos métodos para este problema adaptados al contexto de alta dimensión. Schott (2007) propone un estadístico cuya distribución  $(n, p)$ -asintótica es normal estándar cuando las matrices de covarianza entre grupos son iguales y las observaciones provienen de una distribución normal multivariada. En este trabajo vamos a suponer que tenemos dos muestras con distribución elíptica  $\mathcal{E}_p(\mathbf{0}, \tilde{\Sigma}_i, \phi)$ ,  $i = 1, 2$  y presentaremos propuestas para testear  $H_0 : \tilde{\Sigma}_1 = \tilde{\Sigma}_2$ . En todas ellas nos basaremos en los signos de las observaciones originales. Esto permite considerar una familia de distribuciones para las observaciones más amplia que las consideradas en otros tests. En particular, no hace falta pedir que estas distribuciones tengan segundo momento. Otra de sus ventajas es que, al basarse en los signos, los estadísticos poseen buenas propiedades de robustez. Esta misma técnica fue usada para el testeo de otras hipótesis en contextos de alta dimensión. Por ejemplo, Pandaveine y Verdebout (2013) utilizan esta técnica para testear esfericidad e independencia entre componentes. Finalmente mostramos los resultados de una simulación: comprobaremos la validez empírica de los tests propuestos y compararemos sus potencias estimadas con las de otros tests para las mismas hipótesis.

*Palabras Clave: Test de igualdad de matrices de covarianza, Test en contexto de alta dimensión, Test basado en signos.*

# Agradecimientos

A Graciela, por la dedicación, la paciencia, las ayudas y los consejos. También por la seriedad para trabajar y por compartir su infinita capacidad matemática. Por mantener un excelente nivel en sus clases y nunca subestimar a sus alumnos.

A Ana y Daniela, por tomarse el trabajo de leer esta tesis.

A mi mamá, por ser tan inspiradora con su trabajo. Por preguntarme sobre qué estaba escribiendo. A mi papá, por acompañarme a mis primeras olimpiadas en Mar del Plata. A mi hermana, por estar para todo y por soportarme cuando sólo podía hablar sobre la tesis. A mis abuelos, por ofrecerme su apoyo económico y por interesarse en lo que estudio. A toda mi familia y familiastra, por bancarme siempre.

A toda la OMA, especialmente a Patricia y Flora, por mostrarme que la matemática no es lo que se enseña en el colegio. Sin su trabajo no hubiera estudiado esta carrera. A Carlos Di Fiore y Gabriel Carvajal, por ir a los entrenamientos a contar cosas que todavía me acuerdo.

A los profesores que tuve en la carrera, especialmente a Norberto Fava, Graciela Boente, Mariela Sued, Guillermo Duran, Claudia Lederman, Ana Bianco y Lucas Bali.

A mis amigos de la facultad, especialmente a los de aplicada, por hacer la carrera mucho más liviana.

A mis amigos de afuera, por haber sido capaces de escucharme en los momentos más tensos. A Pablo, por tener la posta siempre. A Lola, por hacerme caminar sobre las rocas.

A todos los de BrightSector, especialmente a Leandro Groisman, por la confianza, por enseñarme tanto y por las milanesas de los viernes.

A mis alumnos de ORT, por ponerle tanta onda a las clases. A Graciela Paso Viola, por la confianza.

Al que corresponda, por permitirme estudiar gratis lo que me gusta.

# Indice

<b>1</b>	<b>Introducción</b>	<b>1</b>
<b>2</b>	<b>Preliminares</b>	<b>4</b>
2.1	Función Característica, Cumulantes y Teorema Central del Límite . . . . .	4
2.2	Distribuciones esféricas y elípticas . . . . .	6
2.3	Distribución Wishart . . . . .	8
2.4	Desarrollos asintóticos . . . . .	10
2.4.1	Desarrollo para el caso de tamaño de muestra grande . . . . .	11
2.4.2	Desarrollo para el caso de alta dimensión . . . . .	13
2.5	Martingalas . . . . .	14
<b>3</b>	<b>Tests de igualdad de matrices de covarianza</b>	<b>16</b>
3.1	El caso $p < n$ . Test de cociente de máxima verosimilitud . . . . .	16
3.2	El caso $(n, p)$ –asintótico . . . . .	21
3.2.1	Distribución $(n, p)$ –asintótica del test de cociente de máxima verosimilitud . . . . .	21
3.2.2	El test de Schott (2007) . . . . .	23
3.2.3	Otros tests para el contexto de alta dimensión . . . . .	29
<b>4</b>	<b>Tests basados en signos</b>	<b>32</b>
4.1	El concepto de signo y matriz de covarianza de signos espaciales . . . . .	32
4.2	Tests basados en signos en el contexto de alta dimensión . . . . .	35
4.2.1	El test de Paindaveine y Verdebout (2013) para testear esfericidad . . . . .	36
4.2.2	Otros tests de esfericidad basados en signos para el contexto de alta dimensión . . . . .	39
<b>5</b>	<b>Tests de igualdad de matrices de covarianza basados en signos</b>	<b>40</b>
5.1	Un test basado en signos con $p$ fijo . . . . .	40

5.1.1	Algunos lemas previos y deducción del estadístico . . . . .	40
5.1.2	Test para el caso $\tilde{\Sigma} = \tilde{\Sigma}_1 = \tilde{\Sigma}_2$ diagonal . . . . .	43
5.1.3	Test para el caso $\tilde{\Sigma} = \tilde{\Sigma}_1 = \tilde{\Sigma}_2$ no diagonal . . . . .	46
5.2	Una heurística para el caso $(n, p)$ -asintótico . . . . .	48
5.2.1	Test para el caso $\tilde{\Sigma} = \tilde{\Sigma}_1 = \tilde{\Sigma}_2$ diagonal . . . . .	48
5.2.2	Test para el caso $\tilde{\Sigma} = \tilde{\Sigma}_1 = \tilde{\Sigma}_2$ no diagonal . . . . .	49
5.3	Test de permutaciones . . . . .	52
<b>6</b>	<b>Simulación</b> . . . . .	<b>54</b>
6.1	Condiciones de la simulación . . . . .	54
6.2	Simulación en el caso $\Sigma$ diagonal . . . . .	55
6.2.1	Observaciones sobre los niveles de significación estimados . . . . .	56
6.2.2	Observaciones sobre las potencias estimadas . . . . .	57
6.3	Simulación en el caso $\Sigma$ no diagonal . . . . .	65
6.3.1	Observaciones sobre los niveles de significación estimados . . . . .	65
6.3.2	Observaciones sobre las potencias estimadas . . . . .	66
<b>7</b>	<b>Código</b> . . . . .	<b>68</b>
7.1	auxiliares.R . . . . .	68
7.2	TCMV.R . . . . .	70
7.3	schott.R . . . . .	72
7.4	frobenius.R . . . . .	74
7.5	np.R . . . . .	79
7.6	np3.R . . . . .	81
7.7	permutaciones.R . . . . .	84

# Capítulo 1

## Introducción

En este capítulo vamos a hacer una breve descripción de los nuevos problemas que surgen al introducir el contexto de la alta dimensión y vamos a describir la organización de este trabajo.

El análisis multivariado tiene como objetivo obtener información de muestras aleatorias provenientes de una distribución multidimensional. Además de las generalizaciones que se deben hacer del caso univariado, aparecen nuevos problemas como por ejemplo el de reducción de la dimensión y análisis de las correlaciones entre las distintas variables.

En los últimos tiempos, el desarrollo tecnológico está permitiendo realizar nuevos estudios en el campo del análisis de datos. Principalmente, contamos con una capacidad en constante crecimiento en cuanto a las nuevas formas de obtener datos, almacenarlos y procesarlos. Esto hace que cada vez la información que se pretende obtener de ellos sea más ambiciosa. A estos problemas que manejan un gran volumen de datos se los suele encuadrar como problemas de **Big Data**. Por su masividad Internet suele ser la mejor fuente, originando nuevos métodos para obtener todo tipo de información. Por ejemplo, Aramaki *et al.* (2011) propusieron un método basado en publicaciones de Twitter para predecir epidemias de gripe en distintas partes del mundo.

Entre las disciplinas con problemas de alta dimensión podemos nombrar las siguientes:

- **Genética.** Nuevos desarrollos tecnológicos permitieron obtener medidas del genoma humano entero en forma barata y esto marcó una diferencia en el terreno de la biología. Por ejemplo, saber qué parte de la codificación genética es la que determina que alguien padezca una determinada enfermedad es un paso clave para próximos avances médicos. Los desafíos estadísticos consisten en crear nuevos métodos potentes en este contexto de alta dimensión.
- **Neurociencias.** Patologías y desórdenes como el Alzheimer o la esquizofrenia, entre muchas otras, tienen que ver con el funcionamiento de las redes de conectividad del cerebro. Entender las jerarquías, complejidades e interrelaciones entre sus distintas partes sirve para detectar los patrones que se repiten en los portadores de las enfermedades. Hubo un reciente desarrollo del diagnóstico por imágenes en tres dimensiones que muestra en tiempo real el funcionamiento de cada región del cerebro. Se puede pensar a cada imagen como el conjunto de sus voxels (el equivalente de píxel en el caso de tres dimensiones) y por lo tanto para poder inferir sobre ellas es necesario el desarrollo de métodos de alta dimensión. A este ejemplo también se lo

suele pensar en el marco de los datos funcionales, debido a que en este contexto se aprovechan las propiedades de suavidad en las imágenes.

- **Economía y finanzas.** Muchas grandes empresas están haciendo uso del análisis de grandes masas de datos con el fin de mejorar la eficiencia y reducir riesgos. Por ejemplo, conocer el historial del valor de ciertas acciones y valores de divisas, entre muchos otros indicadores, sirve para medir el riesgo que tienen sus posibles decisiones. Los métodos tradicionales de series de tiempo muchas veces no están diseñados para manejar dimensiones altas y pueden producir overfitting.

Tanto en el análisis multivariado como en el univariado, es de suma importancia el estudio asintótico de los estadísticos que se usan para el testeo de las diferentes hipótesis. Supongamos que observamos una muestra  $\mathbf{X}_1, \dots, \mathbf{X}_n$  y que para cada  $i$ ,  $\mathbf{X}_i$  es un vector aleatorio de dimensión  $p$ . Tanto el tamaño de la muestra  $n$  como la dimensión  $p$  de cada vector de la muestra, van a ser esenciales en el estudio asintótico que hagamos a lo largo de este trabajo.

Para muchos de los estadísticos propuestos para los diferentes problemas de inferencia se asume que  $n > p$ , es decir, que el tamaño de la muestra observada es más grande que la dimensión de cada uno de sus elementos. En este caso se hace un estudio  $n$ -asintótico, es decir, se fija un  $p \geq 2$  y se hace un análisis de la distribución de los estadísticos cuando  $n \rightarrow \infty$ .

Dentro del contexto de Big Data surgieron estudios que motivaron un nuevo problema. Puede ocurrir que la dimensión  $p$  sea muy grande y resulte prácticamente imposible encontrar una muestra con tamaño mayor que  $p$ . En estos casos los tests y métodos estadísticos que suponen  $n > p$  ya no son útiles y por lo tanto deben ser adaptados. El análisis asintótico, en este caso, debería suponer que  $p$  se va infinito junto con  $n$ . A este análisis se lo llama  $(n, p)$ -asintótico.

El objetivo de esta tesis es hacer un acercamiento a uno de los muchos temas a estudiar en la estadística para datos de alta dimensión. El test que nos va a interesar es el de igualdad de matrices de covarianza entre dos grupos (aunque a veces se puede generalizar a  $k > 2$  grupos). Más allá del interés intrínseco de este test, su importancia radica en que dicha igualdad es un supuesto para varios otros tests, como por ejemplo el de igualdad de medias entre varios grupos.

El procedimiento de inferencia que usaremos se basa en los vectores de norma 1 que define cada una de las observaciones. A estos vectores normalizados se los llama **signos**. Este enfoque ya fue utilizado para testear otras hipótesis sobre la matriz de covarianza, como por ejemplo esfericidad e independencia.

La organización de la tesis será de la siguiente forma:

En el **Capítulo 2** vamos a hacer un repaso de algunos temas de probabilidad y de análisis multivariado y profundizaremos algunos que vamos a necesitar para el resto del trabajo.

En el **Capítulo 3** vamos a enfocarnos en el problema de testear igualdad de matrices de covarianza entre dos o más grupos. Primero vamos a recordar el test clásico basado en el cociente de máxima verosimilitud y a mostrar algunos de sus posibles desarrollos asintóticos. Sin embargo, este test no es aplicable en el contexto de alta dimensión. Para dicho contexto presentaremos las propuestas de algunos autores, haciendo especial énfasis en el test de Schott (2007).

En el **Capítulo 4** introduciremos el concepto de signo y el de matriz de covarianza de signos espaciales. También vamos a ver algunas de sus propiedades y a mostrar de qué forma fue utilizada

la técnica de basarse en los signos de las observaciones para el problema de testear esfericidad.

El **Capítulo 5** contiene las ideas centrales de la tesis. Suponiendo que las muestras provienen de una distribución elíptica, se mostrarán algunas propuestas para testear igualdad de matrices de dispersión basándose en los signos de las observaciones originales.

En el **Capítulo 6** se presentan los resultados de una simulación realizada para comparar el nivel y potencia estimados de los tests propuestos en el **Capítulo 5** y algunos de los tests introducidos en el **Capítulo 3**.

Finalmente, en el **Capítulo 7** se muestra la implementación en R de los distintos tests utilizados en la simulación.

# Capítulo 2

## Preliminares

En este capítulo vamos a repasar algunos temas de probabilidad y de análisis multivariado (para  $p$  fijo) que serán de utilidad en el resto de esta tesis. Los temas de análisis multivariado pueden verse en Seber (1984) o Muirhead (1982) y los de probabilidad se pueden encontrar en Durrett (2010).

### 2.1 Función Característica, Cumulantes y Teorema Central del Límite

**Definición.** Dado un vector aleatorio  $\mathbf{X}$ , se define su **función característica** como

$$C_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{t}^T \mathbf{X}}).$$

Es sabido que hay una correspondencia biunívoca entre la función característica y la función de densidad, en el caso en que esta última exista. Para recuperar la densidad teniendo la función característica, se tiene el siguiente teorema:

**Teorema 2.1.1.** Si un vector aleatorio  $\mathbf{X} \in \mathbb{R}^p$  tiene una función de densidad  $f(\mathbf{x})$  y función característica  $C(\mathbf{t})$ , entonces vale la relación

$$f(\mathbf{x}) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-i\mathbf{t}^T \mathbf{x}} C(\mathbf{t}) d\mathbf{t}.$$

También vale el siguiente teorema, que da una forma alternativa de calcular esperanzas de productos entre componentes del vector aleatorio.

**Teorema 2.1.2.** Si  $E[|X_1|^{m_1} \dots |X_p|^{m_p}] < \infty$ , donde  $X_j$  representa la  $j$ -ésima componente del vector aleatorio  $\mathbf{X}$ , entonces  $C_{\mathbf{X}}(\mathbf{t})$  es  $(m_1, \dots, m_p)$  veces continuamente diferenciable y

$$\left. \frac{\partial^m}{\partial t_1^{m_1} \dots \partial t_p^{m_p}} C_{\mathbf{X}}(\mathbf{t}) \right|_{\mathbf{t}=0} = i^m E(X_1^{m_1} \dots X_p^{m_p}).$$

Para una demostración de los Teoremas 2.1.1 y 2.1.2 puede consultarse Lukacs (1970). Definamos ahora los cumulantes, que van a servir para obtener desarrollos asintóticos para ciertos estadísticos. Recordemos que en el caso univariado,

$$C_X(t) = \mathbb{E}(e^{itX}).$$

Por lo tanto, usando el desarrollo de Taylor

$$e^z = 1 + z + \frac{1}{2!}z^2 + \frac{1}{3!}z^3 + \dots$$

podemos escribir a la función característica como

$$C_X(t) = \mathbb{E} \left( \sum_{r=0}^{\infty} \frac{1}{r!} (itX)^r \right) = \sum_{r=0}^{\infty} \mu_r \frac{(it)^r}{r!}$$

donde  $\mu_r = \mathbb{E}(X^r)$  es el  $r$ -ésimo momento de  $X$ .

**Definición.** Se definen los **cumulantes** de  $X$  como los coeficientes  $\kappa_r$  en

$$\log C_X(t) = \sum_{r=1}^{\infty} \kappa_r \frac{(it)^r}{r!}.$$

Usando el desarrollo de Taylor para  $f(x) = \log(x+1)$  centrada en  $x=0$ , obtenemos las siguientes relaciones:

$$\begin{aligned} \kappa_1 &= \mu_1 = \mathbb{E}(X) \\ \kappa_2 &= \mu_2 - \mu_1^2 = \text{VAR}(X) \\ \kappa_3 &= \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \\ \kappa_4 &= \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4. \end{aligned}$$

Se definen la asimetría  $\kappa_{[3]}$  y la kurtosis  $\kappa_{[4]}$  como  $\kappa_{[3]} = \kappa_3/(\kappa_2)^{3/2}$  y  $\kappa_{[4]} = \kappa_4/(\kappa_2)^2$ . La primera mide la falta de simetría en la distribución y la segunda da información sobre los ‘picos o mesetas’ que tiene en comparación con la distribución normal.

En el caso de una distribución multivariada, es decir, cuando consideramos el vector aleatorio  $\mathbf{X} = (X_1, \dots, X_p)^T$ , podemos escribir a su función característica como

$$C_{\mathbf{X}}(\mathbf{t}) = \mathbb{E} \left( \sum_{r=0}^{\infty} \frac{1}{r!} (i\mathbf{t}^T \mathbf{X})^r \right) = \sum_{r=0}^{\infty} \sum_{r_1+\dots+r_p=r} \mu_{r_1\dots r_p}^{1\dots p} \frac{(it_1)_{r_1} \dots (it_p)_{r_p}}{r_1! \dots r_p!}$$

donde  $\mathbf{t} = (t_1, \dots, t_p)$  y  $\mu_{r_1\dots r_p}^{1\dots p} = \mathbb{E}(X_1^{r_1} \dots X_p^{r_p})$ . Se definen análogamente los cumulantes multivariados  $\kappa_{r_1\dots r_p}^{1\dots p}$  como los coeficientes del desarrollo

$$\log C_{\mathbf{X}}(\mathbf{t}) = \sum_{r=0}^{\infty} \sum_{r_1+\dots+r_p=r} \kappa_{r_1\dots r_p}^{1\dots p} \frac{(it_1)_{r_1} \dots (it_p)_{r_p}}{r_1! \dots r_p!}.$$

Un resultado esencial en el análisis asintótico multivariado es el Teorema Central del Límite:

**Teorema 2.1.3.** Sea  $\mathbf{X} \in \mathbb{R}^p$  un vector aleatorio con matriz de covarianza  $\Sigma \in \mathbb{R}^{p \times p}$ . Sean  $\mathbf{X}_1, \dots, \mathbf{X}_n$  vectores i.i.d. distribuidos como  $\mathbf{X}$ . Entonces, se tiene que

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{X}_i - \mathbb{E}(\mathbf{X}_i)) \xrightarrow{D} N(\mathbf{0}, \Sigma)$$

o equivalentemente

$$\sqrt{n}(\bar{\mathbf{X}} - \mathbb{E}(\mathbf{X})) \xrightarrow{D} N(\mathbf{0}, \Sigma).$$

## 2.2 Distribuciones esféricas y elípticas

**Definición.** Se dice que un vector aleatorio  $\mathbf{X} \in \mathbb{R}^p$  tiene una distribución **esférica** si para toda matriz  $\mathbf{H} \in \mathbb{R}^{p \times p}$  ortogonal vale que  $\mathbf{H}\mathbf{X}$  se distribuye igual que  $\mathbf{X}$ .

El siguiente teorema muestra otras formas equivalentes de definir las distribuciones esféricas. La demostración puede verse en Fujikoshi et al. (2010):

**Teorema 2.2.1.** Las siguientes afirmaciones son equivalentes:

- (i)  $\mathbf{X}$  tiene distribución esférica.
- (ii) Si  $\mathbf{X}$  tiene densidad, entonces ésta puede expresarse como  $g(\mathbf{x}^T \mathbf{x})$  para alguna función escalar  $g$ .
- (iii) La función característica de  $\mathbf{X}$  es de la forma  $\phi(\mathbf{t}^T \mathbf{t})$  para alguna función escalar  $\phi$ .

### Ejemplos.

- (i) Si  $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ , entonces es claro que su densidad  $f(\mathbf{x})$  puede expresarse en términos de  $\mathbf{x}^T \mathbf{x}$  y por lo tanto es esférica.
- (ii) Sea  $S$  es una variable aleatoria tal que  $mS^2 \sim \chi_m^2$  e  $\mathbf{Y} \sim N_p(0, \mathbf{I}_p)$  es independiente de  $S$ . Se dice que el vector aleatorio  $\mathbf{X} = \mathbf{Y}/S$  tiene una distribución  **$t$ -multivariada con  $m$  grados de libertad**. Se puede ver que su función de densidad  $f(\mathbf{x})$  es

$$f(\mathbf{x}) = \frac{\Gamma[(n+p)/2]}{\Gamma[n/2](n\pi)^{p/2}} \frac{1}{\left(1 + \frac{\mathbf{x}^T \mathbf{x}}{n}\right)^{(n+p)/2}},$$

por lo tanto, es esférica.

El siguiente teorema muestra una descomposición fundamental para las distribuciones esféricas. La demostración puede verse en Muirhead (1982).

**Teorema 2.2.2.** Sea  $\mathbf{X} \in \mathbb{R}^p$  esférica con  $\mathbb{P}(\mathbf{X} = 0) = 0$ . Entonces

$$\|\mathbf{X}\| \quad y \quad \mathbf{Z} = \frac{\mathbf{X}}{\|\mathbf{X}\|}$$

son variables aleatorias independientes y  $\mathbf{Z}$  tiene una distribución uniforme en la esfera unidad  $\mathcal{S}^{p-1} = \{\mathbf{u} \in \mathbb{R}^p : \|\mathbf{u}\| = 1\}$ .

El siguiente lema que involucra distribuciones esféricas va a ser de utilidad más adelante.

**Lema 2.2.3.** Sea  $\mathbf{X} \in \mathbb{R}^p$  una distribución esférica con  $\mathbb{P}(\mathbf{X} = 0) = 0$ . Si  $\mathbf{B}$  es una matriz simétrica idempotente de rango  $k$ , entonces

$$Z = \frac{\mathbf{X}^T \mathbf{B} \mathbf{X}}{\|\mathbf{X}\|^2} \sim \beta\left(\frac{k}{2}, \frac{p-k}{2}\right).$$

DEMOSTRACIÓN. Por la forma del estadístico  $Z$  y la esfericidad de  $\mathbf{X}$ , podemos asumir sin perder generalidad que  $\mathbf{X} \sim N_p(0, \mathbf{I}_p)$ . Sea  $\mathbf{H}$  la matriz ortogonal tal que

$$\mathbf{H} \mathbf{B} \mathbf{H}^T = \begin{bmatrix} \mathbf{I}_k & 0 \\ 0 & 0 \end{bmatrix}.$$

Consideramos  $\mathbf{U} = \mathbf{H} \mathbf{X}$ , entonces

$$Z = \frac{\mathbf{U}^T \mathbf{H} \mathbf{B} \mathbf{H}^T \mathbf{U}}{\mathbf{U}^T \mathbf{H} \mathbf{H}^T \mathbf{U}} = \frac{\sum_{i=1}^k \mathbf{U}_i^2}{\sum_{i=1}^p \mathbf{U}_i^2} = \frac{V_1}{V_1 + V_2}$$

donde  $V_1 = \sum_{i=1}^k \mathbf{U}_i^2$  y  $V_2 = \sum_{i=k+1}^p \mathbf{U}_i^2$ . Claramente  $V_1 \sim \chi_k^2$ ,  $V_2 \sim \chi_{p-k}^2$  y son independientes, por lo tanto  $Z \sim \beta(k/2, (p-k)/2)$ , como queríamos  $\square$

**Definición.** Se dice que vector aleatorio  $\mathbf{X} \in \mathbb{R}^p$  tiene una distribución **elíptica** si existe una matrix  $\mathbf{A} \in \mathbb{R}^{p \times p}$  y un vector  $\boldsymbol{\mu} \in \mathbb{R}^p$  tal que  $\mathbf{X} = \mathbf{A} \mathbf{Z} + \boldsymbol{\mu}$  donde  $\mathbf{Z}$  tiene una distribución esférica. En ese caso se indica  $\mathbf{X} \sim \mathcal{E}_p(\boldsymbol{\mu}, \mathbf{A} \mathbf{A}^T)$

Se puede probar el siguiente teorema que da definiciones alternativas para las distribuciones elípticas. La demostración puede verse en Fujikoshi *et al.* (2010).

**Teorema 2.2.4.** Las siguientes afirmaciones son equivalentes:

- (i)  $\mathbf{X} \sim \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ .
- (ii) Si  $\mathbf{X}$  tiene densidad, entonces ésta es de la forma  $|\boldsymbol{\Lambda}|^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\mathbf{x} - \boldsymbol{\mu}))$  para alguna función  $g$ .
- (iii) La función característica  $\mathbf{X}$  es de la forma  $\exp(it^T \boldsymbol{\mu}) \phi(\mathbf{t}^T \boldsymbol{\Lambda} \mathbf{t})$ .

A veces también vamos a indicar  $\mathbf{X} \sim \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \phi)$ , donde  $\phi$  es la función del Teorema 2.2.4.(iii) y de esta forma queda unívocamente determinada la distribución de  $\mathbf{X}$ . A la matriz  $\boldsymbol{\Lambda}$  se la denomina **matriz de dispersión** y es sabido que si  $\mathbb{E}(\|\mathbf{X}\|^2) < \infty$ , la matriz de covarianza de  $\mathbf{X}$  es un múltiplo de  $\boldsymbol{\Lambda}$ .

## 2.3 Distribución Wishart

Supongamos tener una muestra univariada  $X_1, \dots, X_n$  proveniente de una distribución  $N(\mu, \sigma^2)$ . Es sabido que  $s^2 = (1/(n-1)) \sum_{i=1}^n (X_i - \bar{X})^2$  es un estimador insesgado para  $\sigma^2$  y que  $(n-1)s^2/\sigma^2$  tiene una distribución  $\chi_{n-1}^2$ . Además, si supiéramos que  $\mu = 0$ , entonces  $\tilde{s}^2 = (1/n) \sum_{i=1}^n X_i^2$  también es insesgado para  $\sigma^2$  y  $n\tilde{s}^2/\sigma^2$  tiene distribución  $\chi_n^2$ . Este hecho tiene una generalización para el caso multivariado. Supongamos ahora que  $\mathbf{X}_i \in \mathbb{R}^p$  y  $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Podemos definir

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$$

y en este caso  $\mathbf{S}$  resulta también un estimador insesgado para  $\boldsymbol{\Sigma}$ . Análogamente, si se supiera que  $\boldsymbol{\mu} = \mathbf{0}$ , entonces el estadístico

$$\tilde{\mathbf{S}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \quad (2.1)$$

provee un estimador insesgado de  $\boldsymbol{\Sigma}$ . Las distribuciones de  $\mathbf{S}$  y de  $\tilde{\mathbf{S}}$  son una generalización multivariada de la distribución  $\chi^2$ .

**Definición.** Dadas  $\mathbf{Z}_i \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ ,  $\mathbf{Z}_i \in \mathbb{R}^p$ ,  $1 \leq i \leq n$  independientes, se dice que  $\mathbf{W} = \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T$  tiene una distribución **Wishart** no central con covarianza  $\boldsymbol{\Sigma}$  y parámetro de no centralidad  $\boldsymbol{\Delta} = \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T + \dots + \boldsymbol{\mu}_n \boldsymbol{\mu}_n^T$  y se indica  $\mathbf{W} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ . Si  $\boldsymbol{\Delta} = \mathbf{0}$ , entonces se indica  $\mathbf{W} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$

Se puede probar que la distribución del estimador de covarianza muestral  $\mathbf{S}$  es  $\mathcal{W}_p(n-1, \boldsymbol{\Sigma})$  mientras que la del estimador  $\tilde{\mathbf{S}}$  es  $\mathcal{W}_p(n, \boldsymbol{\Sigma})$ . Además, por la forma en la que se definen a las matrices Wishart centrales, es fácil encontrar la distribución asintótica de  $\mathbf{S}$ .

El siguiente lema permite calcular los dos primeros momentos de una matriz aleatoria con distribución Wishart.

**Lema 2.3.1.** Sean  $\mathbf{Y}_1, \dots, \mathbf{Y}_n \in \mathbb{R}^p$  vectores aleatorios i.i.d. con  $\mathbf{Y}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ ,  $1 \leq i \leq n$ . Entonces, si  $\mathbf{A} = \sum_{i=1}^n \mathbf{Y}_i \mathbf{Y}_i^T$  (i.e.  $\mathbf{A} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$ )

$$\mathbb{E}(\mathbf{A}) = n\boldsymbol{\Sigma}$$

$$\text{COV}(a_{ij}, a_{kl}) = n(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) \quad (2.2)$$

donde  $\mathbf{A} = (a_{ij})$ ,  $\boldsymbol{\Sigma} = (\sigma_{ij})$ .

DEMOSTRACIÓN. Calculemos el primer momento de  $\mathbf{A}$ . Consideramos el vector aleatorio  $\mathbf{Y} = (y_1, \dots, y_p)^T$  tal que  $\mathbf{Y} \sim \mathbf{Y}_i$ . Notemos que

$$\mathbb{E}(\mathbf{A}_{ij}) = \sum_{w=1}^n \mathbb{E}((\mathbf{Y}_w \mathbf{Y}_w^T)_{ij}) = n \mathbb{E}(\mathbf{Y} \mathbf{Y}^T)_{ij}.$$

Como  $\mathbb{E}(\mathbf{Y}) = \mathbf{0}$ ,

$$\mathbb{E}(\mathbf{Y} \mathbf{Y}^T)_{ij} = \mathbb{E}(y_i y_j) = \text{COV}(y_i, y_j) = \sigma_{ij},$$

de donde  $\mathbb{E}(\mathbf{A}) = n\boldsymbol{\Sigma}$ .

Calculemos ahora el segundo momento de  $\mathbf{A}$ . Es claro que  $\text{Cov}(a_{ij}, a_{k\ell}) = n \text{Cov}((\mathbf{Y}\mathbf{Y}^T)_{ij}, (\mathbf{Y}\mathbf{Y}^T)_{k\ell})$ . Luego, basta ver que  $\text{Cov}((\mathbf{Y}\mathbf{Y}^T)_{ij}, (\mathbf{Y}\mathbf{Y}^T)_{k\ell}) = \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}$ . Observemos que

$$\text{Cov}((\mathbf{Y}\mathbf{Y}^T)_{ij}, (\mathbf{Y}\mathbf{Y}^T)_{k\ell}) = \text{Cov}(y_i y_j, y_k y_\ell) = \mathbb{E}(y_i y_j y_k y_\ell) - \mathbb{E}(y_i y_j) \mathbb{E}(y_k y_\ell).$$

Supongamos primero que  $\mathbf{\Sigma} = \mathbf{I}_p$  y notemos que en este caso las coordenadas de  $\mathbf{Y}$  son  $N(0, 1)$  independientes. Analicemos los siguientes cuatro casos:

- (i) Supongamos que entre los cuatro numeros  $i, j, k, \ell$  hay uno que es distinto a los otros tres y, sin perder generalidad, supongamos que este índice es  $i$ . En este caso, tenemos que

$$\text{Cov}((\mathbf{Y}\mathbf{Y}^T)_{ij}, (\mathbf{Y}\mathbf{Y}^T)_{k\ell}) = \mathbb{E}(y_i) \mathbb{E}(y_j y_k y_\ell) - \mathbb{E}(y_i) \mathbb{E}(y_j) \mathbb{E}(y_k y_\ell) = 0.$$

- (ii) Supongamos ahora que  $i = j, k = \ell$  y  $i \neq k$ . En este caso,

$$\text{Cov}((\mathbf{Y}\mathbf{Y}^T)_{ij}, (\mathbf{Y}\mathbf{Y}^T)_{k\ell}) = \mathbb{E}(y_i y_j) \mathbb{E}(y_k y_\ell) - \mathbb{E}(y_i y_j) \mathbb{E}(y_k y_\ell) = 0.$$

- (iii) Por otro lado, si  $i = k, j = \ell$  y  $i \neq j$ ,

$$\text{Cov}((\mathbf{Y}\mathbf{Y}^T)_{ij}, (\mathbf{Y}\mathbf{Y}^T)_{k\ell}) = \mathbb{E}(y_i^2 y_j^2) - \mathbb{E}(y_i^2) \mathbb{E}(y_j^2) = 1.$$

Esto se debe a que  $y_i^2$  e  $y_j^2$  se distribuyen como una  $\chi_1^2$  y son independientes. Por simetría, pasa lo mismo si  $i = \ell, j = k$  y  $i \neq j$ .

- (iv) Por último, supongamos que  $i = j = k = \ell$ . Observemos que en este caso

$$\text{Cov}((\mathbf{Y}\mathbf{Y}^T)_{ij}, (\mathbf{Y}\mathbf{Y}^T)_{k\ell}) = \mathbb{E}(y_i^4) - \mathbb{E}(y_i^2)^2 = \text{VAR}(y_i^2) = 2.$$

Notemos que en todos los casos se tiene que  $\text{Cov}((\mathbf{Y}\mathbf{Y}^T)_{ij}, (\mathbf{Y}\mathbf{Y}^T)_{k\ell}) = \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}$ .

Supongamos ahora que  $\mathbf{\Sigma}$  es simétrica y definida positiva. Consideramos  $\mathbf{Q}$  tal que  $\mathbf{Q}\mathbf{Q}^T = \mathbf{\Sigma}$  (es decir, la matriz correspondiente a la descomposición de Cholesky de  $\mathbf{\Sigma}$ ). Tomamos  $\mathbf{Z} = (z_1, \dots, z_p)^T \sim N(\mathbf{0}, \mathbf{I}_p)$ . Luego,  $\mathbf{Q}\mathbf{Z} \sim N(\mathbf{0}, \mathbf{\Sigma}) \sim \mathbf{Y}$ . Por lo tanto, si indicamos con  $q_{ij}$  al elemento  $(i, j)$  de la matriz  $\mathbf{Q}$ , tenemos que

$$\begin{aligned} \text{Cov}((\mathbf{Y}\mathbf{Y}^T)_{ij}, (\mathbf{Y}\mathbf{Y}^T)_{k\ell}) &= \text{Cov}\left(\left((\mathbf{Q}\mathbf{Z})(\mathbf{Q}\mathbf{Z})^T\right)_{ij}, \left((\mathbf{Q}\mathbf{Z})(\mathbf{Q}\mathbf{Z})^T\right)_{k\ell}\right) \\ &= \text{Cov}\left(\left(\mathbf{Q}\mathbf{Z}\right)_i \left(\mathbf{Q}\mathbf{Z}\right)_j, \left(\mathbf{Q}\mathbf{Z}\right)_k \left(\mathbf{Q}\mathbf{Z}\right)_\ell\right) = \text{Cov}\left(\sum_{t,v=1}^p q_{it} q_{jv} z_t z_v, \sum_{s,r=1}^p q_{ks} q_{\ell r} z_s z_r\right) \\ &= \sum_{t,v,r,s=1}^p q_{it} q_{jv} q_{ks} q_{\ell r} \text{Cov}(z_i z_j, z_k z_\ell) = \sum_{t,v,r,s=1}^p q_{it} q_{jv} q_{ks} q_{\ell r} (\delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{kj}) \\ &= \sum_{\substack{t,v=1 \\ t \neq v}}^p q_{it} q_{jv} q_{kt} q_{\ell v} + \sum_{\substack{t,v=1 \\ t \neq v}}^p q_{it} q_{jv} q_{kv} q_{\ell t} + 2 \sum_{t=1}^p q_{it} q_{jt} q_{kt} q_{\ell t} \\ &= \sum_{t,v=1}^p q_{it} q_{kt} q_{jv} q_{\ell v} + \sum_{t,v=1}^p q_{it} q_{\ell t} q_{jv} q_{kv} = \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}, \end{aligned}$$

como queríamos ver.  $\square$

Si  $\mathbf{M} \in \mathbb{R}^{p \times p}$ , notemos con  $\text{vec}(\mathbf{M}) \in \mathbb{R}^{p^2}$  al vector formado por las columnas de la matriz  $\mathbf{M}$  (ubicadas en orden). En virtud del lema anterior, de la definición de la distribución Wishart central y del teorema central del límite multivariado, se tiene el siguiente teorema:

**Teorema 2.3.2.** *Si  $\mathbf{M}_n$  son matrices aleatorias tales que  $n\mathbf{M}_n \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$ , entonces  $\sqrt{n}(\text{vec}(\mathbf{M}_n) - \text{vec}(\boldsymbol{\Sigma})) \xrightarrow{D} N(0, \boldsymbol{\Phi})$ , donde  $\boldsymbol{\Phi} \in \mathbb{R}^{p^2 \times p^2}$ . Si denotamos a  $\phi_{ij,kl}$  al elemento de  $\boldsymbol{\Phi}$  que corresponde a la covarianza entre las entradas  $(i, j)$  y  $(k, l)$  de  $\sqrt{n}(\mathbf{M}_n - \boldsymbol{\Sigma})$ , entonces  $\phi_{ij,kl} = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$ , donde  $\sigma_{ij}$ ,  $i, j = 1, \dots, p$ , son los elementos de la matriz  $\boldsymbol{\Sigma}$ .*

Obtenemos por lo tanto el siguiente corolario:

**Corolario 2.3.3.** *Sea  $\tilde{\mathbf{S}}$  definida como en (2.1). Entonces,  $\sqrt{n}(\text{vec}(\tilde{\mathbf{S}}) - \text{vec}(\boldsymbol{\Sigma})) \xrightarrow{D} N(0, \boldsymbol{\Phi})$  con  $\boldsymbol{\Phi}$  definida como en el Teorema 2.3.2.*

Veamos ahora cuáles son la función de densidad y la característica de la distribución Wishart. Ambas demostraciones pueden encontrarse en Fujikoshi *et al.* (2010). Para una matriz  $\mathbf{A} \in \mathbb{R}^{p \times p}$  llamaremos  $\text{etr}(\mathbf{A}) = \exp(\text{tr}(\mathbf{A}))$ .

**Lema 2.3.4.** *Si  $\mathbf{W} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ , entonces la función característica de  $\mathbf{W}$  está dada por*

$$C_{\mathbf{W}}(\mathbf{T}) = |\mathbf{I}_p - 2i\boldsymbol{\Sigma}\mathbf{T}|^{-\frac{n}{2}} \text{etr}\{i\boldsymbol{\Delta}\mathbf{T}(\mathbf{I}_p - 2i\boldsymbol{\Sigma}\mathbf{T})\}.$$

En particular, si  $\mathbf{W} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$ , entonces  $C_{\mathbf{W}}(\mathbf{T}) = |\mathbf{I}_p - 2i\boldsymbol{\Sigma}\mathbf{T}|^{-n/2}$ .

**Lema 2.3.5.** *Si  $\mathbf{W} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$  entonces, si  $n \geq p$ , la función de densidad de  $\mathbf{W}$  viene dada por*

$$f_{\mathbf{W}}(\mathbf{T}) = \frac{1}{2^{\frac{pn}{2}} \Gamma_p(\frac{n}{2}) |\boldsymbol{\Sigma}|^{\frac{n}{2}}} \text{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{T}\right) |\mathbf{T}|^{\frac{n-p-1}{2}} \mathbb{I}_{(\mathbf{T} > 0)}.$$

En particular, si  $\mathbf{W} \sim \mathcal{W}_p(n, \mathbf{I}_p)$ , se tiene  $f_{\mathbf{W}}(\mathbf{T}) = \frac{1}{2^{\frac{pn}{2}} \Gamma_p(\frac{n}{2})} \text{etr}\left(-\frac{1}{2}\mathbf{T}\right) |\mathbf{T}|^{\frac{n-p-1}{2}} \mathbb{I}_{(\mathbf{T} > 0)}$ .

Por último, enunciaremos el Teorema de Descomposición de Bartlett que va a ser de utilidad más adelante. Su demostración puede verse en Muirhead (1982)

**Teorema 2.3.6.** *Sea  $\mathbf{W} \sim \mathcal{W}_p(n, \mathbf{I}_p)$ . Entonces,  $\mathbf{W} = \mathbf{C}\mathbf{C}^T$  donde  $\mathbf{C}$  es una matriz aleatoria triangular inferior, los elementos  $C_{i,j}$  con  $i > j$  son  $N(0, 1)$  y los elementos diagonales al cuadrado  $C_{i,i}^2$  se distribuyen como una  $\chi_{n+1-i}^2$  y además todos estos  $p(p+1)/2$  elementos son independientes.*

## 2.4 Desarrollos asintóticos

En esta sección nos vamos a concentrar sólomente en estadísticos que tienen un tipo específico de momentos. Esto va a servir para hacer desarrollos asintóticos para el test de cociente de máxima verosimilitud para testear igualdad de matrices de covarianza entre distintos grupos. Vamos a mostrar dos desarrollos: el primera es de utilidad cuando el tamaño de la muestra es grande pero deja de aproximar bien cuando la dimensión  $p$  también es grande. El segundo contempla también el caso en que ambos parámetros sean grandes.

### 2.4.1 Desarrollo para el caso de tamaño de muestra grande

Consideremos un estadístico  $W$  con momentos de la forma

$$E[W^h] = K \left( \frac{\prod_{k=1}^b y_k^{y_k}}{\prod_{j=1}^a x_j^{x_j}} \right)^h \frac{\prod_{j=1}^a \Gamma(x_j(1+h) + \xi_j)}{\prod_{k=1}^b \Gamma(y_k(1+h) + \eta_k)}, \quad h = 0, 1, \dots \quad (2.3)$$

donde  $\Gamma(\cdot)$  es la función Gamma,  $\sum_{k=1}^b y_k = \sum_{j=1}^a x_j$  y  $K$  se define para que  $\mathbb{E}(W^0) = 1$ , es decir,

$$K = \frac{\prod_{k=1}^b \Gamma(y_k + \eta_k)}{\prod_{j=1}^a \Gamma(x_j + \xi_j)}.$$

Este estadístico, aunque parezca artificial, aparece en varios problemas de interés que involucran a  $k$  poblaciones con distribuciones  $N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ . Por ejemplo, el estadístico del test de cociente de máxima verosimilitud para testear igualdad de matrices de covarianza (que vamos a estudiar en el siguiente capítulo) o el del test  $H_0 : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_k$  cuando  $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma} \forall j = 1, \dots, k$ .

Consideremos  $V = -2\rho \log W$  con  $0 < \rho < 1$  una constante que varía con  $N$  y definamos  $\alpha_j = (1 - \rho)x_j$ ,  $\beta_k = (1 - \rho)y_k$ .

En lo que sigue, vamos a asumir que  $\alpha_j$  y  $\beta_k$  tienen límites cuando  $x_j$  e  $y_k$  tienden a infinito. Notemos que esto implica que  $\rho = \rho_N \rightarrow 1$ . Supongamos también que  $x_j = c_j N$  para  $j = 1, \dots, a$  y  $y_k = d_k N$  para  $k = 1, \dots, b$ .

Consideremos la función característica  $C(t)$  de  $V$ , es decir,  $C(t) = \mathbb{E}(e^{itV}) = \mathbb{E}(e^{-2it\rho \log W}) = \mathbb{E}(W^{-2it\rho})$ .

Tomando logaritmo de ambos lados y usando (2.3) se puede ver que

$$\begin{aligned} \log C(t) &= (2it\rho) \left( \sum_{j=1}^a x_j \log x_j - \sum_{k=1}^b y_k \log y_k \right) + \sum_{j=1}^a \log \Gamma(x_j(1 - 2it\rho) + \xi_j) \\ &\quad - \sum_{k=1}^b \log \Gamma(y_k(1 - 2it\rho) + \eta_k) + \sum_{k=1}^b \log \Gamma(y_k + \eta_k) - \sum_{j=1}^a \log \Gamma(x_j + \xi_j). \end{aligned}$$

Vamos a necesitar la siguiente fórmula de Barnes (1899):

$$\log \Gamma(z + h) = \log \sqrt{2\pi} + (z + h - 1/2) \log z - z - \sum_{r=1}^m (-1)^r \frac{B_{r+1}(h)}{r(r+1)z^r} + R_{m+1}(z),$$

donde  $R_{m+1}(z) \in O(z^{-(m+1)})$  y  $B_r(h)$  son los polinomios de Bernoulli de grado  $r$  definidos como

$$\frac{\tau e^{h\tau}}{e^\tau} = \sum_{r=0}^{\infty} \frac{\tau^r}{r!} B_r(h).$$

Algunos de dichos polinomios son  $B_0(h) = 1$ ,  $B_1(h) = h - 1/2$ ,  $B_2(h) = h^2 - h + 1/6$ ,  $B_3(h) = h^3 - 3h^2/2 + h/2$ ,  $B_4(h) = h^4 - 2h^3 + h^2 - 1/30$ ,  $B_5(h) = h^5 - 5h^4/2 + 5h^3/3 - h/6$ .

Usando esta fórmula, podemos desarrollar la expresión que obtuvimos para  $\log C(t)$ . Agrupando convenientemente, obtenemos

$$\log C(t) = -\frac{f}{2} \log(1 - 2it) + \sum_{l=1}^r \omega_r \{(1 - 2it)^{-l} - 1\} + R_{r+1}, \quad (2.4)$$

donde

$$f = -2 \left\{ \sum_{j=1}^a \xi_j - \sum_{k=1}^b \eta_k - \frac{1}{2}(a - b) \right\} \quad (2.5)$$

$$\omega_r = \frac{(-1)^r}{r(r+1)} \left\{ \sum_{j=1}^a \frac{B_{r+1}(\alpha_j + \xi_j)}{(\rho x_j)^r} - \sum_{k=1}^b \frac{B_{r+1}(\beta_k + \eta_k)}{(\rho y_k)^r} \right\} \quad (2.6)$$

$$R_{r+1} \in \sum_{j=1}^a O(x_j^{-(r+1)}) + \sum_{k=1}^b O(y_k^{-(r+1)}).$$

Usando el desarrollo (2.4) e invirtiendo la función característica para obtener la densidad (y luego la acumulada), se tiene el siguiente teorema de Box (1949) cuya demostración puede verse en Fujikoshi *et al.* (2010).

**Teorema 2.4.1.** *Supongamos que los momentos en una variable aleatoria  $0 \leq W \leq 1$  están dados por (2.3) con  $x_j = c_j N$ ,  $y_k = d_k N$  con  $\sum_{j=1}^a c_j = \sum_{k=1}^b d_k$ . Entonces, tenemos el siguiente desarrollo para la variable aleatoria  $V = -2\rho \log W$ ,*

$$\begin{aligned} P(V \leq x) &= G_f(x) + \frac{\gamma_2}{m^2} \{G_{f+4}(x) - G_f(x)\} + \frac{\gamma_3}{m^3} \{G_{f+6}(x) - G_f(x)\} \\ &\quad + \frac{\gamma_4}{m^4} \{G_{f+8}(x) - G_f(x)\} + \frac{\gamma_2^2}{2m^4} \{G_{f+8}(x) - 2G_{f+4}(x) + G_f(x)\} + O(m^{-5}) \end{aligned}$$

donde  $G_f(x)$  es la función de distribución acumulada de una distribución  $\chi_f^2$  con  $f$  definida en (2.5),  $\rho$  y  $m$  definidos como  $\rho = 1 - \gamma/N$ ,  $m = N\rho = N - \gamma$  y  $\gamma, \gamma_i, i = 2, 3, 4$  están dados por

$$\begin{aligned} \gamma &= \frac{1}{f} \left\{ \sum_{j=1}^a c_j^{-1} B_2(\xi_j) - \sum_{k=1}^b d_k^{-1} B_2(\eta_k) \right\} \\ \gamma_2 &= -\frac{1}{6} \left\{ \frac{3}{2} \gamma^2 f + \sum_{j=1}^a c_j^{-2} B_3(\xi_j) - \sum_{k=1}^b d_k^{-2} B_3(\eta_k) \right\} \\ \gamma_3 &= \frac{1}{12} \left\{ 2\gamma^3 f - 24\gamma\gamma_2 + \sum_{j=1}^a c_j^{-3} B_4(\xi_j) - \sum_{k=1}^b d_k^{-3} B_4(\eta_k) \right\} \\ \gamma_4 &= -\frac{1}{20} \left\{ \frac{5}{2} \gamma^4 f + 60\gamma^2\gamma_2 + 60\gamma\gamma_3 + \sum_{j=1}^a c_j^{-4} B_5(\xi_j) - \sum_{k=1}^b d_k^{-4} B_5(\eta_k) \right\}. \end{aligned}$$

Notemos que se eligió  $\rho$  de forma que  $\omega_1 = 0$  en (2.6).

Como mencionamos al principio de la sección, esta expresión deja de ser útil a medida que  $p$  crece. Veremos ahora otra desarrollo asintótico que sirve para cuando el tamaño de la muestra y la dimensión de cada elemento son ambos grandes.

### 2.4.2 Desarrollo para el caso de alta dimensión

Consideremos un estadístico  $\widetilde{W}$  con momentos de la forma

$$\mathbb{E}(\widetilde{W}^h) = \frac{\prod_{j=1}^a \Gamma(x_j + h\tau_j + \xi_j) \prod_{k=1}^b \Gamma(y_k + \eta_k)}{\prod_{k=1}^b \Gamma(y_k + h + \eta_k) \prod_{j=1}^a \Gamma(x_j + \xi_j)}$$

y llamemos  $\widetilde{V} = -\log(\widetilde{W})$ . Indicamos por  $\widetilde{C}$  a la función característica de  $\widetilde{V}$ . Este tipo de estadísticos, bajo algunas hipótesis, provienen de transformar estadísticos como los que consideramos al principio de la sección.

Usando nuevamente que  $\widetilde{C}(t) = \mathbb{E}(e^{-it \log \widetilde{W}}) = \mathbb{E}(\widetilde{W}^{-it})$  obtenemos que la función generadora de cumulantes se puede escribir como

$$\log \widetilde{C}(t) = \sum_{j=1}^a \log \Gamma(x_j + \xi_j - it\tau_j) - \log \Gamma(x_j + \xi_j) - \sum_{k=1}^b \log \Gamma(y_k + \eta_k - it) - \log \Gamma(y_k + \eta_k). \quad (2.7)$$

Para esta desarrollo en vez de la fórmula de Barnes para el logaritmo de la función  $\Gamma$ , vamos a usar el desarrollo de Taylor. De esta forma, tenemos que

$$\log \Gamma(a + b) = \log \Gamma(a) + \sum_{k=1}^{\infty} \frac{1}{k!} \psi^{(k-1)}(a) b^k,$$

donde  $\psi^{(k-1)}(z) = \frac{d^k}{dz^k} \log \Gamma(z)$  es la función polygamma. Esta función puede expresarse como

$$\begin{aligned} \psi^{(0)}(a) &= -C + \sum_{k=0}^{\infty} \left( \frac{1}{1+k} - \frac{1}{k+a} \right), \\ \psi^{(s)}(a) &= \sum_{k=0}^{\infty} \frac{(-1)^{s+1} s!}{(k+a)^{s+1}} \quad s = 1, 2, \dots \end{aligned}$$

Como los cumulantes  $\kappa^{(s)}$  cumplen  $\log \widetilde{C}(t) = \sum_{s=1}^{\infty} \kappa^{(s)} (it)^s / s!$ , usando la expresión (2.7) obtenemos la siguiente expresión para los cumulantes,

$$\kappa^{(s)} = (-1)^s \left\{ \sum_{j=1}^a \psi^{(s-1)}(x_j + \xi_j) \tau_j^s - \sum_{k=1}^b \psi^{(s-1)}(y_k + \eta_k) \right\}, \quad s = 1, 2, \dots$$

Consideremos al estadístico estandarizado

$$Z = \frac{\widetilde{V} - \kappa^{(1)}}{(\kappa^{(2)})^{1/2}}$$

y sean  $\widetilde{\kappa}^{(s)} = \kappa^{(s)} / (\kappa^{(2)})^{s/2}$  los cumulantes estandarizados. Es fácil ver que la función característica de  $Z$  puede escribirse como

$$C_Z(t) = \exp \left\{ -\frac{t^2}{2} + \sum_{s=3}^{\infty} \frac{\widetilde{\kappa}^{(s)}}{s!} (it)^s \right\}.$$

Por lo tanto, haciendo un desarrollo de  $\exp(z)$  para  $z \in \mathbb{C}$  obtenemos

$$\begin{aligned}
C_Z(t) &= \exp\left\{-\frac{t^2}{2}\right\} \exp\left\{(it)^3 \sum_{s=0}^{\infty} \frac{\tilde{\kappa}^{(s+3)}}{(s+3)!} (it)^s\right\} \\
&= \exp\left\{-\frac{t^2}{2}\right\} \left\{1 + \sum_{k=1}^{\infty} \frac{1}{k!} (it)^{3k} \left(\sum_{s=0}^{\infty} \frac{\tilde{\kappa}^{(s+3)}}{(s+3)!} (it)^s\right)^k\right\} \\
&= \exp\left\{-\frac{t^2}{2}\right\} \left\{1 + \sum_{k=1}^{\infty} \frac{1}{k!} (it)^{3k} \sum_{j=0}^{\infty} \gamma_{k,j} (it)^j\right\}
\end{aligned} \tag{2.8}$$

donde

$$\gamma_{k,j} = \sum_{s_1 + \dots + s_k = j} \frac{\tilde{\kappa}^{(s_1+3)} \dots \tilde{\kappa}^{(s_k+3)}}{(s_1+3)! \dots (s_k+3)!}.$$

Para obtener un desarrollo asintótico para  $Z$  usando (2.8) necesitamos que  $\gamma_{k,j} \in O(m^{-(j+k)})$  para algún parámetro  $m$ . En ese caso tendríamos

$$C_Z = \exp\left\{-\frac{t^2}{2}\right\} \left\{1 + \sum_{k=1}^s \frac{1}{k!} (it)^{3k} \left(\sum_{j=0}^{s-k} \gamma_{k,j} (it)^j\right)^k\right\} + O(m^{-(s+1)}).$$

Usando formalmente el Teorema 2.1.1 se obtiene que la función de distribución para  $Z$  hasta el orden  $m^{-s}$  se escribe como

$$Q_s(x) = \Phi(x) - \phi(x) \left\{ \sum_{k=1}^s \frac{1}{k!} \sum_{j=0}^{s-k} \gamma_{k,j} H_{3k+j-1}(x) \right\}, \tag{2.9}$$

donde  $\Phi$  y  $\phi$  son las funciones de distribución y la densidad de una normal estándar, respectivamente,  $H_r$  es el polinomio de Hermite de orden  $r$ , definido como

$$\frac{d^r}{dx^r} \exp\left(-\frac{x^2}{2}\right) = (-1)^r H_r(x) \exp\left(-\frac{x^2}{2}\right).$$

Es importante tener en claro que estas expresiones se obtuvieron formalmente, es decir, hay que probar que efectivamente estos desarrollos asintóticos tienen el orden deseado.

## 2.5 Martingalas

En esta sección vamos a definir el concepto de martingala y luego vamos a enunciar un Teorema Central del Límite. Su importancia en nuestro trabajo radica en que este teorema es uno de los principales métodos en las demostraciones de resultados  $(n, p)$ -asintóticos.

**Definiciones.**

- Se dice que  $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$  es una **filtración** si es una sucesión creciente de  $\sigma$ -álgebras.

- Una sucesión de variables aleatorias  $X_n$  se dice **adaptada a**  $\{\mathcal{F}_n\}$  si para cada  $n \in \mathbb{N}$  vale que  $X_n$  es medible respecto de  $\mathcal{F}_n$ .
- $X_n$  se dice una **martingala respecto a**  $\{\mathcal{F}_n\}$  si
  - $\mathbb{E}(|X_n|) < \infty$
  - $X_n$  es adaptada a  $\{\mathcal{F}_n\}$
  - $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$
- Si en la condición (iii) se reemplaza el  $=$  por  $\leq$  o  $\geq$ ,  $X_n$  se dice respectivamente una **supermartingala** o una **submartingala** respecto a  $\{\mathcal{F}_n\}$ . Además, si la condición (iii) se reemplaza por  $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = 0$ ,  $X_n$  se dice una **martingala diferencia** respecto a  $\{\mathcal{F}_n\}$ .
- Se llama **arreglo de martingalas** a las sucesiones dobles  $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$  donde las variables  $S_{ni}$  tienen media cero, varianza finita y son una martingala para cada  $n$  fijo. Se define de forma análoga el **arreglo de martingalas diferencia**.

**Ejemplo.** Consideremos las tiradas sucesivas e independientes de una moneda equilibrada. Llamemos  $\xi_n$  a la variable aleatoria que vale 1 si en la  $n$ -ésima tirada la moneda sale cara y -1 si en la  $n$ -ésima tirada sale ceca. Sea  $X_n = \xi_1 + \dots + \xi_n$  y  $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$  la  $\sigma$ -álgebra generada por las variables  $\xi_1, \dots, \xi_n$ . Entonces,  $X_n$  es una martingala respecto a  $\{\mathcal{F}_n\}$ . Efectivamente, las dos primeras condiciones son triviales de verificar. Para la tercera, notemos que  $X_{n+1} = X_n + \xi_{n+1}$ . Por la linealidad de la esperanza condicional y la independencia de  $\xi_{n+1}$  con  $\{\mathcal{F}_n\}$ , tenemos que

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}(X_n|\mathcal{F}_n) + \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = X_n + \mathbb{E}(\xi_{n+1}) = X_n$$

Notemos que si tenemos una martingala  $S_n$  respecto a  $\{\mathcal{F}_n\}$  entonces la sucesión  $X_n = S_n - S_{n-1}$  es una martingala diferencia, donde  $S_0 = \mathbb{E}(X_1)$ .

Enunciemos ahora el teorema central del límite para martingalas

**Teorema 2.5.1.** Sea  $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$  un arreglo de martingalas y consideramos  $X_{ni} = S_{ni} - S_{n,i-1}$ . Sea  $\eta^2$  una variable aleatoria finita en casi todo punto. Supongamos que

$$\forall \epsilon > 0, \quad \sum_i \mathbb{E}(X_{ni}^2 \mathbb{I}(|X_{ni}| > \epsilon) | \mathcal{F}_{n,i-1}) \xrightarrow{p} 0 \quad (2.10)$$

$$V_{n,k_n}^2 = \sum_i \mathbb{E}(X_{ni}^2 | \mathcal{F}_{n,i-1}) \xrightarrow{p} \eta^2 \quad (2.11)$$

y que las  $\sigma$ -álgebras están anidadas en el siguiente sentido,

$$\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i} \quad \forall \quad 1 \leq i \leq k_n, n \geq 1.$$

Entonces,  $S_{n,k_n} = \sum_i X_{ni} \xrightarrow{D} Z$  donde la variable aleatoria  $Z$  tiene función característica  $\mathbb{E}(e^{-t^2 \eta^2 / 2})$ .

Notemos que las variables aleatorias  $Z$  que cumplen que su función característica es  $\mathbb{E}(e^{-t^2 \eta^2 / 2})$  corresponden a una mezcla de normales, es decir, son de la forma  $\eta^2 N$  donde  $N \sim N(0, 1)$  y  $N$  independiente de  $\eta^2$ . Si  $\eta^2$  es una constante entonces la distribución límite del teorema es normal. A la condición (2.10) se la llama **condición de Lindeberg**.

## Capítulo 3

# Tests de igualdad de matrices de covarianza

En este Capítulo vamos a introducir algunos test de igualdad de matrices de covarianza. En primer lugar, vamos a recordar el test de cociente de máxima verosimilitud para el caso en que  $p < n$  y  $p$  fijo, donde  $p$  es la dimensión de cada elemento de la muestra y  $n$  es el tamaño de la misma. Luego veremos cómo se ha hecho para adaptar este test en el caso  $(n, p)$ -asintótico.

### 3.1 El caso $p < n$ . Test de cociente de máxima verosimilitud

El test cociente de máxima verosimilitud es un procedimiento para obtener test de hipótesis que ha sido ampliamente estudiado. Supongamos que la hipótesis a testear es  $H_0$  y la alternativa es  $H_1$ . Llamemos  $\mathbf{x}_1, \dots, \mathbf{x}_n$  a los valores observados de la muestra,  $\boldsymbol{\theta} \in \Theta$  al vector de parámetros desconocidos e indiquemos  $\Theta_0$  y  $\Theta_1$  a los subconjuntos de  $\Theta$  que determinan  $H_0$  y  $H_1$  respectivamente. Sea además  $L(\mathbf{x}_1, \dots, \mathbf{x}_n, \boldsymbol{\theta})$  la función de verosimilitud. Si el número

$$\frac{\sup_{\Theta_0} L(\mathbf{x}_1, \dots, \mathbf{x}_n, \boldsymbol{\theta})}{\sup_{\Theta_1} L(\mathbf{x}_1, \dots, \mathbf{x}_n, \boldsymbol{\theta})}$$

fuese pequeño, entonces sería un motivo razonable para sospechar que  $H_0$  es falsa. Muchas veces, restringirse a  $H_0$  determina un subespacio de dimensión menor respecto a la del espacio paramétrico original  $\Theta$ , por lo que el cociente de máxima verosimilitud se puede reescribir como

$$\frac{\sup_{\Theta_0} L(\mathbf{x}_1, \dots, \mathbf{x}_n, \boldsymbol{\theta})}{\sup_{\Theta} L(\mathbf{x}_1, \dots, \mathbf{x}_n, \boldsymbol{\theta})}. \quad (3.1)$$

En la situación que nos interesa, supongamos que tenemos  $k$  grupos y para cada grupo  $i$ , una muestra aleatoria  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$  proveniente de una distribución normal  $N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ . Queremos estudiar el problema

$$H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \dots = \boldsymbol{\Sigma}_k \quad \text{vs.} \quad H_1 : \exists i \neq j \text{ tal que } \boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j. \quad (3.2)$$

Por lo tanto, para hallar el test de cociente de máxima verosimilitud podemos utilizar el estadístico dado en (3.1). Sea  $n = \sum_{i=1}^k n_i$ ,  $\mathbf{S}_i = \mathbf{V}_i/n_i$  la covarianza muestral para el grupo  $i$ -ésimo, donde  $\mathbf{V}_i = \sum_{j=1}^{n_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^T$  y  $\mathbf{V} = \sum_{i=1}^k \mathbf{V}_i$ .

Primero enunciaremos el siguiente lema auxiliar.

**Lema 3.1.1.** *Si  $\mathbf{X}_1, \dots, \mathbf{X}_n$  es una muestra aleatoria con  $\mathbf{X}_i \in \mathbb{R}^p$  donde  $\mathbf{X}_i \sim N(\mathbf{0}, \mathbf{\Sigma})$ ,  $\mathbf{\Sigma} > 0$  y  $n \geq p$  entonces el estadístico de máxima verosimilitud para  $\mathbf{\Sigma}$  es  $\widehat{\mathbf{\Sigma}} = \mathbf{V}/n$ .*

DEMOSTRACIÓN. La función de verosimilitud para esta muestra la podemos escribir como

$$L(\mathbf{\Sigma}) = L(\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{\Sigma}) = \left(\frac{1}{2\pi}\right)^{\frac{pn}{2}} \left(\frac{1}{|\mathbf{\Sigma}|}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{V})\right\}.$$

Tenemos que ver que  $L(\widehat{\mathbf{\Sigma}}) \geq L(\mathbf{\Sigma})$  para toda  $\mathbf{\Sigma}$  simétrica y definida positiva. Observemos que

$$\begin{aligned} L(\widehat{\mathbf{\Sigma}}) \geq L(\mathbf{\Sigma}) &\iff \log L(\widehat{\mathbf{\Sigma}}) \geq \log L(\mathbf{\Sigma}) \\ &\iff -\frac{n}{2} \log(|\widehat{\mathbf{\Sigma}}|) - \frac{1}{2}np \geq -\frac{n}{2} \log(|\mathbf{\Sigma}|) - \frac{n}{2} \text{tr}\left(\mathbf{\Sigma}^{-1} \frac{\mathbf{V}}{n}\right) \\ &\iff \log(|\widehat{\mathbf{\Sigma}}|) + p \leq \log(|\mathbf{\Sigma}|) + \text{tr}(\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}}). \end{aligned}$$

Como  $\widehat{\mathbf{\Sigma}}$  es simétrica y definida positiva con probabilidad 1, existe una matriz  $\widehat{\mathbf{C}}$  tal que  $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{C}}\widehat{\mathbf{C}}^T$ . Luego,  $\text{tr}(\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}}) = \text{tr}(\widehat{\mathbf{C}}^T\mathbf{\Sigma}^{-1}\widehat{\mathbf{C}})$ . Llamemos  $\theta_1, \dots, \theta_p$  a los autovalores de  $\widehat{\mathbf{C}}^T\mathbf{\Sigma}^{-1}\widehat{\mathbf{C}}$  y notemos que

$$\log(|\mathbf{\Sigma}|) = \log\left(\frac{|\mathbf{\Sigma}|}{|\widehat{\mathbf{\Sigma}}|} |\widehat{\mathbf{\Sigma}}|\right) = \log\left(\frac{|\mathbf{\Sigma}|}{|\widehat{\mathbf{\Sigma}}|}\right) + \log(|\widehat{\mathbf{\Sigma}}|)$$

y que

$$\prod_{i=1}^p \theta_i = |\widehat{\mathbf{C}}^T\mathbf{\Sigma}^{-1}\widehat{\mathbf{C}}| = |\widehat{\mathbf{C}}| |\widehat{\mathbf{C}}^T| |\mathbf{\Sigma}^{-1}| = \frac{|\widehat{\mathbf{\Sigma}}|}{|\mathbf{\Sigma}|}.$$

Entonces,

$$\begin{aligned} L(\widehat{\mathbf{\Sigma}}) \geq L(\mathbf{\Sigma}) &\iff \log(|\widehat{\mathbf{\Sigma}}|) + p \leq \log(|\mathbf{\Sigma}|) + \text{tr}(\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}}) \\ &\iff p \leq \log\left(\frac{|\mathbf{\Sigma}|}{|\widehat{\mathbf{\Sigma}}|}\right) + \text{tr}(\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}}) \\ &\iff 0 \leq -\sum_{i=1}^p \log(\theta_i) + \sum_{i=1}^p \theta_i - p \\ &\iff 0 \leq \sum_{i=1}^p \theta_i - (\log(\theta_i) + 1). \end{aligned}$$

Recordemos que si  $x \geq 0$  entonces  $x \geq \log(x) + 1$ , por lo tanto  $\sum_{i=1}^p \theta_i - (\log(\theta_i) + 1) \geq 0$ , como queríamos ver.  $\square$

**Teorema 3.1.2.** *Consideramos las  $k$  muestras aleatorias  $\mathbf{X}_{ij} \in \mathbb{R}^p$ ,  $i = 1, \dots, k$ ,  $j = 1, \dots, n_i$  tales que  $\mathbf{X}_{ij} \sim N(0, \mathbf{\Sigma}_i)$ ,  $\mathbf{\Sigma}_i > 0$  y  $n_i \geq p$ . Entonces, el test de cociente de máxima verosimilitud para*

(3.2) se basa en el estadístico

$$\lambda = \frac{n^{pn/2} \prod_{i=1}^k |\mathbf{V}_i|^{\frac{n_i}{2}}}{\prod_{i=1}^k n_i^{pn_i/2} |\mathbf{V}|^{\frac{n}{2}}} \quad (3.3)$$

y la distribución asintótica de  $-2 \log \lambda$  es una  $\chi_f^2$  donde  $f = \frac{1}{2}p(p+1)(k-1)$ .

DEMOSTRACIÓN. Llamamos  $L_i(\boldsymbol{\Sigma}_i)$  a la función de verosimilitud para el grupo  $i$ -ésimo y  $L(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_p)$  a la función de verosimilitud global. Sea  $\boldsymbol{\Theta}$  el espacio paramétrico y denotamos  $\boldsymbol{\Theta}_0$  y  $\boldsymbol{\Theta}_1$  a los subespacios de  $\boldsymbol{\Theta}$  que determinan  $H_0$  y  $H_1$  respectivamente. Es claro que  $\dim(\boldsymbol{\Theta}) > \dim(\boldsymbol{\Theta}_0)$ , por lo tanto podemos calcular el estadístico del test usando la fórmula (3.1). Por otro lado, como el denominador  $\sup L(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_p) = \prod_{i=1}^p L_i(\boldsymbol{\Sigma}_i)$ , entonces podemos usar el Lema 3.1.1 para calcularlo. Tenemos entonces que

$$\begin{aligned} \sup_{\boldsymbol{\Theta}} L(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_p) &= \sup \prod_{i=1}^p L_i(\boldsymbol{\Sigma}_i) = \prod_{i=1}^p L_i\left(\frac{\mathbf{V}_i}{n_i}\right) \\ &= \prod_{i=1}^p \left[ \left(\frac{1}{2\pi}\right)^{\frac{pn_i}{2}} \left(\frac{1}{|\mathbf{V}_i|}\right)^{\frac{n_i}{2}} n_i^{\frac{pn_i}{2}} e^{-\frac{1}{2}n_i p} \right] \\ &= \left(\frac{1}{2\pi}\right)^{\frac{pn}{2}} e^{-\frac{1}{2}np} \prod_{i=1}^p \left[ \left(\frac{1}{|\mathbf{V}_i|}\right)^{\frac{n_i}{2}} n_i^{\frac{pn_i}{2}} \right]. \end{aligned}$$

Por otro lado, para el numerador, tenemos que calcular

$$\begin{aligned} \sup_{\boldsymbol{\Theta}_0} L(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_p) &= \sup_{\boldsymbol{\Sigma}} L(\boldsymbol{\Sigma}, \dots, \boldsymbol{\Sigma}) = \sup_{\boldsymbol{\Sigma}} \left(\frac{1}{2\pi}\right)^{\frac{pn}{2}} \left(\frac{1}{|\boldsymbol{\Sigma}|}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2} \sum_{i,j} \mathbf{x}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}_{ij}\right\} \\ &= \sup_{\boldsymbol{\Sigma}} \left(\frac{1}{2\pi}\right)^{\frac{pn}{2}} \left(\frac{1}{|\boldsymbol{\Sigma}|}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{V})\right\}. \end{aligned}$$

Entonces, observando la demostración del Lema 3.1.1, tenemos que

$$\sup_{\boldsymbol{\Theta}_0} L(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_p) = L\left(\frac{\mathbf{V}}{n}, \dots, \frac{\mathbf{V}}{n}\right) = \left(\frac{1}{2\pi}\right)^{\frac{pn}{2}} e^{-\frac{1}{2}np} \left(\frac{1}{|\mathbf{V}|}\right)^{\frac{n}{2}} n^{\frac{pn}{2}}.$$

Dividiendo las expresiones obtenidas para el numerador y el denominador se obtiene el estadístico  $\lambda$ , como queríamos. Por otra parte, la diferencia entre las dimensiones de  $\boldsymbol{\Theta}$  y el espacio  $\boldsymbol{\Theta}_0$  definido por  $H_0$  es

$$f = \dim(\boldsymbol{\Theta}) - \dim(\boldsymbol{\Theta}_0) = k \frac{p(p+1)}{2} - \frac{p(p+1)}{2} = (k-1) \frac{p(p+1)}{2},$$

de donde  $-2 \log(\lambda) \xrightarrow{D} \chi_f^2$ , como queríamos ver.  $\square$

Para el caso en que la media  $\boldsymbol{\mu}$  es desconocida, el estadístico del test es muy parecido. En primer lugar, las matrices  $\mathbf{V}_i$  se reemplazan por  $\mathbf{A}_i = \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}})(\mathbf{x}_{ij} - \bar{\mathbf{x}})^T$  y se puede probar que de esta forma se pierde un grado de libertad en la matriz Wishart. Es decir,  $\mathbf{A}_i \sim \mathcal{W}_p(\boldsymbol{\Sigma}, n_i - 1)$ . Es

fácil demostrar que, si llamamos  $\mathbf{A} = \mathbf{A}_1 + \dots + \mathbf{A}_k$ , entonces el estadístico del test de máxima verosimilitud es

$$\widetilde{\lambda}^* = \frac{n^{pn/2} \prod_{i=1}^k |\mathbf{A}_i|^{\frac{n_i}{2}}}{\prod_{i=1}^k n_i^{pn_i/2} |\mathbf{A}|^{\frac{n}{2}}}.$$

Se puede ver que este test es sesgado y que cambiando  $n_i$  por  $n_i - 1$  y  $n$  por  $n - k$  (es decir, por los correspondientes grados de libertad de  $\mathbf{A}$  y  $\mathbf{A}_i$ ), el nuevo test resulta insesgado. Por este motivo, suele considerarse el estadístico

$$\lambda^* = \frac{(n - k)^{p(n-k)/2} \prod_{i=1}^k |\mathbf{A}_i|^{\frac{n_i-1}{2}}}{\prod_{i=1}^k (n_i - 1)^{p(n_i-1)/2} |\mathbf{A}|^{\frac{n-k}{2}}}. \quad (3.4)$$

Veamos ahora que podemos aplicar el teorema de Box para obtener desarrollos asintóticos de estos estadísticos. Necesitamos primero el siguiente lema,

**Lema 3.1.3.** *Bajo  $H_0$ , el  $h$ -ésimo momento de  $W = |\mathbf{V}|^{n/2} \prod_{i=1}^k |\mathbf{V}_i|^{n_i/2}$  es*

$$\mathbb{E}(W^h) = \prod_{j=1}^p \frac{\Gamma(\frac{1}{2}(n - j + 1))}{\Gamma(\frac{1}{2}(n + nh - j + 1))} \left( \prod_{i=1}^k \frac{\Gamma(\frac{1}{2}(n_i + n_i h - j + 1))}{\Gamma(\frac{1}{2}(n_i - j + 1))} \right).$$

DEMOSTRACIÓN. Notemos que bajo  $H_0$ ,  $\mathbf{V}_i \sim \mathcal{W}_p(n_i, \Sigma)$  y  $\mathbf{V} \sim \mathcal{W}_p(n, \Sigma)$ . Como  $W$  es invariante bajo transformaciones no singulares del tipo  $\mathbf{V}_i \mapsto \mathbf{A}\mathbf{V}_i\mathbf{A}^T$  y  $\mathbf{V} \mapsto \mathbf{A}\mathbf{V}\mathbf{A}^T$ , se puede suponer sin perder generalidad que  $\Sigma = \mathbf{I}_p$  (pues  $\Sigma^{-\frac{1}{2}}\mathbf{V}_i\Sigma^{-\frac{1}{2}} \sim \mathcal{W}_p(n_i, \mathbf{I}_p)$  y  $\Sigma^{-\frac{1}{2}}\mathbf{V}\Sigma^{-\frac{1}{2}} \sim \mathcal{W}_p(n, \mathbf{I}_p)$ ). Llamemos  $\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma(a - (i-1)/2)$  y  $K(p, m) = (2^{pm/2} \Gamma_p(m/2))^{-1}$ . Teniendo en cuenta la función de densidad de una matriz Wishart, se tiene que

$$\begin{aligned} \mathbb{E}(W^h) &= \mathbb{E} \left( \left| \sum_{i=1}^k \mathbf{V}_i \right|^{-\frac{nh}{2}} \prod_{i=1}^k |\mathbf{V}_i|^{\frac{n_i h}{2}} \right) \\ &= \int \cdots \int_{\mathbf{V}_i > 0} \left| \sum_{i=1}^k \mathbf{V}_i \right|^{-\frac{nh}{2}} \prod_{i=1}^k |\mathbf{V}_i|^{\frac{n_i h}{2}} \frac{1}{2^{\frac{pn_i}{2}} \Gamma_p(\frac{n_i}{2})} \text{etr}(-\frac{1}{2} \mathbf{V}_i) |\mathbf{V}_i|^{\frac{n_i - p - 1}{2}} d\mathbf{V}_1 \dots d\mathbf{V}_k \\ &= \prod_{i=1}^k \frac{K(p, n_i)}{K(p, n_i h + n_i)} \int \cdots \int_{\mathbf{V}_i > 0} \left| \sum_{i=1}^k \mathbf{V}_i \right|^{-\frac{nh}{2}} \prod_{i=1}^k |\mathbf{V}_i|^{\frac{n_i + n_i h - p - 1}{2}} \text{etr}(-\frac{1}{2} \mathbf{V}_i) K(p, n_i h + n_i) d\mathbf{V}_1 \dots d\mathbf{V}_k. \end{aligned} \quad (3.5)$$

Ahora, notemos que (3.5) es la esperanza de  $|\mathbf{Q}|^{-\frac{nh}{2}}$  con  $\mathbf{Q} = \sum_{i=1}^k \mathbf{Q}_i$  donde  $\mathbf{Q}_i \sim \mathcal{W}_p(n_i h + n_i, \mathbf{I}_p)$ , es decir,  $\mathbf{Q} \sim \mathcal{W}_p(nh + n, \mathbf{I}_p)$ . Por el teorema de descomposición de Bartlett, tenemos que  $\mathbf{Q} = \mathbf{C}\mathbf{C}^T$  donde  $\mathbf{C}$  es una matriz triangular inferior cuyo  $j$ -ésimo elemento diagonal  $c_j$  cumple que  $c_j^2$  se distribuye como una  $\chi_{nh+n+1-j}^2$  y todas las  $c_j^2$  son independientes. Recordemos que el  $q$ -ésimo momento de una variable  $A$  distribuida como  $\chi_n^2$  se obtiene como

$$\mathbb{E}(A^q) = 2^q \frac{\Gamma(n/2 + q)}{\Gamma(n/2)} \quad \text{para } n/2 + q > 0.$$

Entonces, como  $(nh + n + 1 - j)/2 - nh/2 = (n + 1 - j)/2 > 0$  para  $j = 1, \dots, p$  (pues  $n \geq p$ ), se tiene que

$$\begin{aligned} \mathbb{E} \left( |\mathbf{Q}|^{-\frac{nh}{2}} \right) &= \mathbb{E} \left( \prod_{j=1}^p (c_j^2)^{-nh/2} \right) \\ &= 2^{-\frac{pnh}{2}} \prod_{j=1}^p \frac{\Gamma(\frac{1}{2}(n+1-j))}{\Gamma(\frac{1}{2}(n+nh+1-j))}. \end{aligned}$$

La fórmula del enunciado sale de que

$$\prod_{i=1}^k \frac{K(p, n_i)}{K(p, n_i h + n_i)} = 2^{\frac{pnh}{2}} \prod_{i=1}^k \prod_{j=1}^p \frac{\Gamma(\frac{1}{2}(n_i + n_i h + 1 - j))}{\Gamma(\frac{1}{2}(n_i + 1 - j))} \quad \square$$

Usando este lema, sale inmediatamente la siguiente proposición:

**Proposición 3.1.4.** *Bajo  $H_0$ , los momentos del estadístico  $\lambda$  son*

$$\mathbb{E} \left( \lambda^h \right) = K \frac{\prod_{j=1}^p (n/2)^{n/2}}{\prod_{i=1}^k \prod_{l=1}^p (n_i/2)^{n_i/2}} \frac{\prod_{i=1}^k \prod_{l=1}^p \Gamma(\frac{1}{2}(n_i + n_i h - l + 1))}{\prod_{j=1}^p \Gamma(\frac{1}{2}(n + nh - j + 1))}.$$

En este caso se cumplen las hipótesis del Teorema de Box 2.4 tomando  $x_{il} = n_i/2$ ,  $y_j = n/2$ ,  $\xi_{il} = (1 - l)/2$  y  $\eta_j = (1 - j)/2$ , luego tenemos el siguiente desarrollo asintótico,

**Proposición 3.1.5.** *Bajo  $H_0$  y asumiendo que  $n_i/n \rightarrow r_i \in (0, 1) \forall i = 1, \dots, k$ , se tiene que*

$$\begin{aligned} \mathbb{P}(-2\rho \log \lambda < x) &= G_f(x) + \frac{\gamma}{m^2} (G_{f+4}(x) - G_f(x)) + O(m^{-3}) \\ &= G_f(x) - \frac{2\gamma}{m^2} g_f(x) \frac{x}{f} \left( 1 + \frac{x}{f+2} \right) + O(m^{-3}), \end{aligned}$$

donde  $G_f$ ,  $g_f$  son respectivamente las funciones de probabilidad acumulada y densidad de una  $\chi_f^2$ . Las constantes  $m$ ,  $\rho$ ,  $f$  y  $\gamma$  se definen como

$$\begin{aligned} m = \rho n &= n - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \left( \sum_{i=1}^k \frac{n}{n_i} - 1 \right), \quad f = (k-1) \frac{p(p+1)}{2} \\ \gamma &= \frac{p(p+1)}{48} \left\{ (p-1)(p+2) \left( \sum_{i=1}^k \frac{n^2}{n_i^2} - 1 \right) - 6(k-1)n^2(1-\rho)^2 \right\}. \end{aligned}$$

Este desarrollo pierde potencia de aproximación cuando  $p$  se hace grande. Vamos a mostrar otro desarrollo asintótico que sirve para el caso en que tanto  $n$  como  $p$  son grandes. Para esto consideremos al estadístico transformado  $\widetilde{W} = W^{2/n}$ , donde  $W$  está definido como en el Lema 3.1.3. Se ve fácilmente que los momentos de  $\widetilde{W}$  son

$$\mathbb{E} \left( \widetilde{W}^h \right) = \prod_{j=1}^p \prod_{i=1}^k \frac{\Gamma[\frac{1}{2}(n-j+1)] \Gamma[\frac{1}{2}(n_i-j+1) + h \frac{n_i}{n}]}{\Gamma[\frac{1}{2}(n-j+1) + h] \Gamma[\frac{1}{2}(n_i-j+1)]}.$$

Si suponemos que  $\frac{n_i}{n}$  es constantemente  $\tau_i$ , entonces estos momentos tienen la forma descrita en la Sección 2.4.2. De esta forma, los cumulantes del estadístico  $V = -\log \widetilde{W}$  son

$$\kappa_V^{(s)} = (-1)^s \sum_{j=1}^p \left\{ -\psi^{(s-1)}\left(\frac{1}{2}(n-j+1)\right) + \sum_{i=1}^k \psi^{(s-1)}\left(\frac{1}{2}(n_i-j+1)\right) \tau_i^s \right\}.$$

Siguiendo argumentos análogos a los de la Sección 2.4.2, definimos el estadístico y cumulantes estandarizados como

$$\widetilde{V} = \frac{V - \kappa_V^{(1)}}{(\kappa_V^{(2)})^{1/2}}, \quad \widetilde{\kappa}^{(s)} = \frac{\kappa_V^{(s)}}{(\kappa_V^{(2)})^{s/2}}. \quad (3.6)$$

Indiquemos con  $O_j^*$  a un término de orden  $j$  respecto a  $(n^{-1}, p^{-1})$ . En Wakaki (2007) se prueba que el desarrollo formal dado en (2.9) vale en este caso y obtenemos entonces el siguiente teorema:

**Teorema 3.1.6.** *Sea  $\widetilde{V}$  dado como en (3.6). Entonces la función de distribución de  $\widetilde{V}$  se puede expandir como  $\mathbb{P}(\widetilde{V} \leq x) = Q_s(s) + O_{s+1}^*$  donde  $Q_s(x) = \Phi(x) - \phi(x) \left\{ \sum_{r=1}^s (1/r!) \sum_{j=0}^{s-r} \gamma_{r,j} H_{3r+j-1}(x) \right\}$ , con*

$$\gamma_{r,j} = \sum_{s_1 + \dots + s_r = j} \frac{\widetilde{\kappa}^{(s_1+3)} \dots \widetilde{\kappa}^{(s_r+3)}}{(s_1+3)! \dots (s_r+3)!}.$$

## 3.2 El caso $(n, p)$ –asintótico

En esta sección vamos a mostrar algunos test de igualdad de matrices de covarianza entre grupos para el caso de alta dimensión. En los casos en que no hagamos las demostraciones de los resultados  $(n, p)$ –asintóticos, mencionaremos cuáles fueron las ideas principales para establecer este tipo de resultados.

### 3.2.1 Distribución $(n, p)$ –asintótica del test de cociente de máxima verosimilitud

Recientemente se obtuvo un resultado  $(n, p)$ –asintótico para el test de cociente de máxima verosimilitud que acabamos de estudiar. Notemos que el estadístico considera los determinantes de las matrices  $\mathbf{V}$  y  $\mathbf{V}_i$ . En particular, este test no tendría sentido si alguna de las matrices  $\mathbf{V}_i$  fuese singular con probabilidad 1. Teniendo esto en cuenta y recordando la definición de las  $\mathbf{V}_i$ , es necesario pedir que  $n_i \geq p \quad \forall \quad i = 1, \dots, k$ .

El siguiente es un teorema  $(n, p)$ –asintótico sobre el test de cociente de máxima verosimilitud con  $\boldsymbol{\mu}$  desconocido, cuya demostración puede verse en Yang y Jiang (2013).

**Teorema 3.2.1.** *Sean  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$  muestras aleatorias provenientes de distribuciones  $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ . Supongamos que  $p$  tiende a infinito y los tamaños de muestra  $n_i = n_i(p)$  cumplen que  $\min n_i > p + 1$  y  $\lim_{p \rightarrow \infty} p/n_i = y_i \in (0, 1]$ . Consideremos el estadístico  $\lambda^* = \lambda_n^*$  del test de máxima verosimilitud con media desconocida definido en (3.4). Entonces,*

$$\frac{\log(\lambda_n^*) - \mu_n}{(n-k)\sigma_n} \xrightarrow{D} N(0, 1), \quad (3.7)$$

donde

$$\begin{aligned}\mu_n &= \frac{1}{4} \left[ (n-k)(2n-2p-2k-1) \log \left( 1 - \frac{p}{n-k} \right) \right. \\ &\quad \left. - \sum_{i=1}^k (n_i-1)(2n_i-2p-3) \log \left( 1 - \frac{p}{n_i-1} \right) \right] \\ \sigma_n^2 &= \frac{1}{2} \left[ \log \left( 1 - \frac{p}{n-k} \right) - \sum_{i=1}^k \left( \frac{n_i-1}{n-k} \right)^2 \log \left( 1 - \frac{p}{n_i-1} \right) \right].\end{aligned}$$

La idea de la demostración es mostrar que la sucesión de funciones características de

$$\frac{\log(\lambda_n^*) - \mu_n}{(n-k)\sigma_n}$$

tiende puntualmente a la función característica de una  $N(0, 1)$ .

Es importante tener en cuenta que los estadísticos  $\lambda$  y  $\lambda^*$  definidos en (3.3) y (3.4) tienen esencialmente la misma estructura y los únicos cambios son los grados de libertad de las matrices Wishart a las que se les toma el determinante dentro del estadístico. Por eso tenemos como corolario el siguiente resultado para el estadístico  $\lambda$ .

**Corolario 3.2.2.** Sean  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}, i = 1, \dots, k$  muestras aleatorias provenientes de distribuciones  $N_p(0, \Sigma_i)$ . Supongamos que  $p$  tiende a infinito y los tamaños de muestra  $n_i = n_i(p)$  cumplen que  $\min n_i > p$  y  $\lim_{p \rightarrow \infty} p/n_i = y_i \in (0, 1]$ . Consideremos el estadístico  $\lambda = \lambda_n$  del test de máxima verosimilitud con media cero definido en (3.3). Entonces,

$$C_n = \frac{\log(\lambda_n) - \mu_n}{n\sigma_n} \xrightarrow{D} N(0, 1), \quad (3.8)$$

donde

$$\begin{aligned}\mu_n &= \frac{1}{4} \left[ n(2n-2p-1) \log \left( 1 - \frac{p}{n} \right) - \sum_{i=1}^k n_i(2n_i-2p-1) \log \left( 1 - \frac{p}{n_i} \right) \right] \\ \sigma_n^2 &= \frac{1}{2} \left[ \log \left( 1 - \frac{p}{n} \right) - \sum_{i=1}^k \left( \frac{n_i}{n} \right)^2 \log \left( 1 - \frac{p}{n_i} \right) \right].\end{aligned}$$

Teniendo en cuenta este Corolario, se propone un test basado en el estadístico  $C_n$  definido en (3.8). Recordemos que bajo la hipótesis alternativa se espera que  $\lambda_n$  sea chico. Luego, dado un nivel  $0 < \alpha < 1$ , en este test rechazaremos si  $C_n$  es menor que el percentil  $(1 - \alpha)100\%$  de una distribución normal estándar.

Este test tiene dos supuestos fuertes. En primer lugar, se supone la normalidad de los datos y por otro lado se asume que el tamaño de muestra de cada uno de los grupos es mayor a la dimensión  $p$  de cada elemento de las muestras.

En la simulación del Capítulo 6 incluiremos un estudio de nivel y potencia para el Test de Cociente de Máxima Verosimilitud de la Sección 3.1 y para el test presentado en esta subsección.

Vamos a ver algunas otras ideas que surgieron para realizar este tipo de tests y que debilitan las hipótesis del Teorema 3.2.1.

### 3.2.2 El test de Schott (2007)

Supongamos que tenemos  $k \geq 2$  grupos de variables aleatorias  $\{\mathbf{X}_{ij}, i = 1, \dots, k, j = 1, \dots, N_i\}$  tal que la variable aleatoria  $\mathbf{X}_{ij}$  tiene distribución normal y matriz de covarianza  $\Sigma_i$  y media  $\mu_i$ . Las hipótesis a testear son

$$H_0 : \Sigma_1 = \dots = \Sigma_k \quad \text{vs.} \quad H_1 : \exists i, j \text{ tal que } \Sigma_i \neq \Sigma_j.$$

Supongamos tener estimadores independientes  $\mathbf{S}_1, \dots, \mathbf{S}_k$  para las matrices  $\Sigma_1, \dots, \Sigma_k$  donde  $n_i \mathbf{S}_i \sim \mathcal{W}_p(\Sigma_i, n_i)$ . Notemos que si las medias  $\mu_i$  fueran conocidas,  $n_i$  coincide con  $N_i$  tomando  $\mathbf{S}_i = \sum_{j=1}^{N_i} (\mathbf{X}_{ij} - \mu_i)(\mathbf{X}_{ij} - \mu_i)^\top$  y si las medias fueran desconocidas, entonces se puede tomar  $n_i = N_i - 1$  tomando  $\mathbf{S}_i = \sum_{j=1}^{N_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)^\top$ .

Dada una matriz  $\mathbf{A} \in \mathbb{R}^{p \times p}$ ,  $\mathbf{A} = (a_{ij})$ , se define su norma de Frobenius como  $\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}\mathbf{A}^\top)}$ . Es fácil ver que esta norma también se puede definir como  $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$ . El estadístico propuesto se basa en la norma al cuadrado de Frobenius de la diferencias de las matrices  $\mathbf{S}_i$ . La ventaja de trabajar con esta norma es que los estadísticos suelen resultar invariantes por rotaciones y esto permite hacer algunos supuestos que luego facilitan las cuentas.

Más específicamente, consideramos el estadístico  $\sum_{i < j} \text{tr}((\mathbf{S}_i - \mathbf{S}_j)^2)$ . Notemos que

$$\mathbb{E}(\text{tr}((\mathbf{S}_i - \mathbf{S}_j)^2)) = \mathbb{E}(\text{tr}(\mathbf{S}_i^2)) + \mathbb{E}(\text{tr}(\mathbf{S}_j^2)) - 2\mathbb{E}(\text{tr}(\mathbf{S}_i \mathbf{S}_j)).$$

Llamemos  $s_{i,tq}$  al elemento  $(t, q)$  de matriz aleatoria  $\mathbf{S}_i$  y  $\sigma_{i,tq}$  al elemento  $(t, q)$  de la matriz  $\Sigma_i$ . Usando el Teorema 2.3.1 y que  $n_i \mathbf{S}_i \sim \mathcal{W}_p(n_i, \Sigma_i)$ , tenemos que

$$\begin{aligned} \mathbb{E}(\text{tr}(\mathbf{S}_i^2)) &= \mathbb{E}\left(\sum_{t,q=1}^p s_{i,tq}^2\right) = \sum_{t,q=1}^p \mathbb{E}^2(s_{i,tq}) + \text{VAR}(s_{i,tq}) \\ &= \sum_{t,q=1}^p \sigma_{i,tq}^2 + \frac{1}{n_i}(\sigma_{i,tt}\sigma_{i,qq} + \sigma_{i,tq}^2) = \text{tr}(\Sigma_i^2) + \frac{1}{n_i}(\text{tr}(\Sigma_i^2) + \text{tr}^2(\Sigma_i)). \end{aligned}$$

Por otro lado,

$$\mathbb{E}(\text{tr}(\mathbf{S}_i \mathbf{S}_j)) = \mathbb{E}\left(\sum_{t,q=1}^p s_{i,tq} s_{j,tq}\right) = \sum_{t,q=1}^p \mathbb{E}(s_{i,tq})\mathbb{E}(s_{j,tq}) = \sum_{t,q=1}^p \sigma_{i,tq} \sigma_{j,tq} = \text{tr}(\Sigma_i \Sigma_j).$$

Combinando estas ecuaciones, tenemos que

$$\mathbb{E}\left(\sum_{i < j} \text{tr}((\mathbf{S}_i - \mathbf{S}_j)^2)\right) = \sum_{i < j} \text{tr}((\Sigma_i - \Sigma_j)^2) + \frac{1}{n_i} \{\text{tr}(\Sigma_i^2) + \text{tr}^2(\Sigma_i)\} + \frac{1}{n_j} \{\text{tr}(\Sigma_j^2) + \text{tr}^2(\Sigma_j)\}.$$

Notemos que

$$\begin{aligned} \mathbb{E}(\text{tr}^2(\mathbf{S}_i)) &= \sum_{t,q=1}^p \mathbb{E}(s_{i,tt} s_{i,qq}) = \sum_{t,q=1}^p \text{COV}(s_{i,tt}, s_{i,qq}) + \mathbb{E}(s_{i,tt})\mathbb{E}(s_{i,qq}) \\ &= \sum_{t,q=1}^p \frac{2}{n_i} \sigma_{i,tq}^2 + \sigma_{i,tt}\sigma_{i,qq} = \frac{2}{n_i} \text{tr}(\Sigma_i^2) + \text{tr}^2(\Sigma_i). \end{aligned}$$

Además, si  $\eta_i = (n_i + 2)(n_i - 1)$ , tenemos que

$$\begin{aligned} \mathbb{E} \left( \frac{1}{\eta_i} [n_i(n_i - 2)\text{tr}(\mathbf{S}_i^2) + n_i^2\text{tr}^2(\mathbf{S}_i)] \right) &= \mathbb{E} \left( \frac{n_i(n_i - 2)}{(n_i + 2)(n_i - 1)} \left\{ \text{tr}(\boldsymbol{\Sigma}_i^2) + \frac{1}{n_i}(\text{tr}(\boldsymbol{\Sigma}_i^2) + \text{tr}^2(\boldsymbol{\Sigma}_i)) \right\} \right) \\ &= \text{tr}(\boldsymbol{\Sigma}_i^2) + \text{tr}^2(\boldsymbol{\Sigma}_i). \end{aligned}$$

Combinando estos resultados, podemos fácilmente llegar al Lema 3.2.3.

**Lema 3.2.3.** Sean  $\mathbf{S}_1, \dots, \mathbf{S}_k$  estimadores independientes para las matrices  $\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_k$  donde  $n_i \mathbf{S}_i \sim \mathcal{W}_p(\boldsymbol{\Sigma}_i, n_i)$ , entonces se tiene que

$$\begin{aligned} t_{np} &= \sum_{i < j} \text{tr}((\mathbf{S}_i - \mathbf{S}_j)^2) - \frac{1}{n_i \eta_i} [n_i(n_i - 2)\text{tr}(\mathbf{S}_i^2) + n_i^2\text{tr}^2(\mathbf{S}_i)] - \frac{1}{n_j \eta_j} [n_j(n_j - 2)\text{tr}(\mathbf{S}_j^2) + n_j^2\text{tr}^2(\mathbf{S}_j)] \\ &= \sum_{i < j} \left\{ 1 - \frac{n_i - 2}{\eta_i} \right\} \text{tr}(\mathbf{S}_i^2) + \left\{ 1 - \frac{n_j - 2}{\eta_j} \right\} \text{tr}(\mathbf{S}_j^2) - 2\text{tr}(\mathbf{S}_i \mathbf{S}_j) - \frac{n_i}{\eta_i} \text{tr}^2(\mathbf{S}_i) - \frac{n_j}{\eta_j} \text{tr}^2(\mathbf{S}_j) \end{aligned} \quad (3.9)$$

es un estimador insesgado de  $\sum_{i < j} \text{tr}((\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_j)^2)$ . En particular,  $t_{np}$  tiene esperanza cero si y sólo si  $H_0$  definida en (3.2) es cierta.

El siguiente resultado permite obtener la varianza del estadístico  $t_{np}$  definido en (3.9). Su demostración puede verse en Schott (2007).

**Lema 3.2.4.** Bajo  $H_0 : \boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$ , la varianza de  $t_{np}$  se puede expresar como

$$\sigma_{np}^2 = c_1 \text{tr}(\boldsymbol{\Sigma}^4) + c_2 \{\text{tr}(\boldsymbol{\Sigma}^2)\}^2, \quad (3.10)$$

donde

$$\begin{aligned} c_1 &= 4 \sum_{i < j} \frac{n_i n_j (n_i + n_j)^2 + (2n_i - n_j)(2n_j - n_i)(n_i + n_j) - 2(3n_i^2 + n_i n_j + 3n_j^2) + 8}{n_i n_j \eta_i \eta_j} \\ &\quad + 4(k-1)(k-2) \sum_{i=1}^k \frac{n_i - 2}{n_i \eta_i}, \\ c_2 &= 4 \sum_{i < j} \frac{n_i n_j (n_i + n_j)^2 + 3n_i n_j (n_i + n_j) - 2(2n_i^2 + n_i n_j + 2n_j^2) - 4(n_i + n_j) + 8}{n_i n_j \eta_i \eta_j} \\ &\quad + 4(k-1)(k-2) \sum_{i=1}^k \frac{1}{\eta_i}. \end{aligned}$$

El resultado principal del trabajo de Schott (2007) es la normalidad  $(n, p)$ -asintótica bajo  $H_0$  del estadístico  $t_{np}$ . En cuanto a los parámetros de la normal, la única candidata a la media puede ser cero. Para hallar un candidato a varianza, tenemos que calcular el límite de las varianzas dadas por el Lema 3.2.4. Consideremos  $h$  un índice para coordinar el movimiento de los  $n_i$  y  $p$ . Para calcular este límite asumiremos lo siguiente:

**A1.** Cuando  $h \rightarrow \infty$ ,

$$\frac{p_h}{n_{ih}} \rightarrow b_i \in [0, +\infty).$$

**A2.** Bajo  $H_0$ , si llamamos  $\Sigma_h = \Sigma_{1h} = \dots = \Sigma_{kh}$  entonces

$$\lim_{h \rightarrow \infty} \frac{\text{tr}(\Sigma_h^r)}{p_h} = \gamma_r \in (0, +\infty) \quad r = 1, \dots, 8.$$

**Lema 3.2.5.** Bajo **A1** y **A2**,  $\lim_{h \rightarrow \infty} \sigma_{np}^2 = \sum_{i < j} 4(b_i + b_j)^2 \gamma_2^2 + (k-1)(k-2) \sum_{i=1}^k 4b_i^2 \gamma_2^2$ .

DEMOSTRACIÓN. Tenemos que  $\lim_{h \rightarrow \infty} \sigma_{np}^2 = \lim_{h \rightarrow \infty} c_{1h} \text{tr}(\Sigma_h^4) + \lim_{h \rightarrow \infty} c_{2h} \text{tr}^2(\Sigma_h^2)$ . Para el primer término del límite notemos que

$$\lim_{h \rightarrow \infty} c_{1h} \text{tr}(\Sigma_h^4) = \lim_{h \rightarrow \infty} c_{1h} p_h \frac{\text{tr}(\Sigma_h^4)}{p_h} = \gamma_4 \lim_{h \rightarrow \infty} c_{1h} p_h.$$

Haciendo un análisis sencillo de los órdenes de cada  $n_{ih}$  en la expresión de  $c_1$  y observando que la condición **A1** implica que  $p_h \in O(n_{ih})$  para todo  $i$  se puede ver que este límite vale cero.

Para el segundo término, observemos que

$$\lim_{h \rightarrow \infty} c_{2h} \text{tr}^2(\Sigma_h^2) = \lim_{h \rightarrow \infty} c_{2h} p_h^2 \left\{ \frac{\text{tr}(\Sigma_h^2)}{p_h} \right\}^2 = \gamma_2^2 \lim_{h \rightarrow \infty} c_{2h} p_h^2.$$

Haciendo el mismo tipo de análisis que el hecho para el primer término, se ve que los únicos términos del límite que no valen cero son los

$$4\gamma_2^2 \lim_{h \rightarrow \infty} \sum_{i < j} \frac{p^2 n_i n_j (n_i + n_j)^2}{n_i n_j \eta_i \eta_j} = 4\gamma_2^2 \sum_{i < j} \lim_{h \rightarrow \infty} \left( \frac{p(n_i + n_j)}{\sqrt{\eta_i \eta_j}} \right)^2 = 4\gamma_2^2 \sum_{i < j} (b_i + b_j)^2$$

y

$$4(k-1)(k-2)\gamma_2^2 \lim_{h \rightarrow \infty} \sum_i \frac{p^2}{\eta_i} = 4(k-1)(k-2)\gamma_2^2 \sum_{i=1}^k b_i^2.$$

Juntando estos dos términos se obtiene el resultado.  $\square$

El siguiente teorema muestra que efectivamente la distribución  $(n, p)$ -asintótica del estadístico  $t_{np}$  es, bajo  $H_0$ , una normal con la media cero y varianza dada en el Lema 3.2.5.

**Teorema 3.2.6.** Sean  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_{ih}}, i = 1, \dots, k$  muestras aleatorias provenientes de distribuciones  $N_{p_h}(\mathbf{0}, \Sigma_{ih})$ . Sean  $\mathbf{S}_{1h}, \dots, \mathbf{S}_{kh}$  estimadores de las matrices  $\Sigma_{1h}, \dots, \Sigma_{kh}$  respectivamente, tales que  $n_{ih} \mathbf{S}_{ih} \sim \mathcal{W}_{p_h}(n_{ih}, \Sigma_{ih})$ . Bajo **A1** y **A2**, se tiene que  $t_{np} \xrightarrow{D} N(0, \theta^2)$  cuando  $h \rightarrow \infty$ , donde

$$\theta^2 = \sum_{i < j} 4(b_i + b_j)^2 \gamma_2^2 + (k-1)(k-2) \sum_{i=1}^k 4b_i^2 \gamma_2^2.$$

DEMOSTRACIÓN. Vamos a omitir el índice  $h$  en esta demostración para simplificar un poco la notación. Observemos que el estadístico  $t_{np}$  es invariante bajo transformaciones de la forma  $\mathbf{S}_i \mapsto \mathbf{P}\mathbf{S}_i\mathbf{P}$  donde  $\mathbf{P}$  es una matriz ortogonal, por lo tanto, podemos asumir que bajo  $H_0$ ,  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_k$  es una matriz diagonal. Si fijamos un grupo  $i$ , entonces  $n_i\mathbf{S}_i \sim \mathcal{W}_p(n_i, \boldsymbol{\Sigma})$  con  $\boldsymbol{\Sigma}$  diagonal. Recordemos que, si  $\mathbf{w}_1, \dots, \mathbf{w}_n$  son i.i.d. con distribución  $N_p(\mathbf{0}, \boldsymbol{\Sigma})$  con  $\boldsymbol{\Sigma}$  diagonal, entonces los  $p$  elementos de cada  $\mathbf{w}_j$  son independientes. Entonces podemos escribir

$$\begin{aligned} (n\mathbf{S})_{jk} &= \sum_{t=1}^n (\mathbf{w}_t \mathbf{w}_t^\top)_{jk} = \sum_{t=1}^n w_{tj} w_{tk} && \text{donde } w_{tq} \sim N(0, \sigma_{qq}) \text{ independientes, } q = 1, \dots, p \\ &= \sum_{t=1}^n \sigma_{jj}^{1/2} \sigma_{kk}^{1/2} z_{tj} z_{tk} && \text{donde } z_{tq} \sim N(0, 1) \text{ independientes, } q = 1, \dots, p \\ &= \sigma_{jj}^{1/2} \sigma_{kk}^{1/2} \mathbf{Z}_j^\top \mathbf{Z}_k && \text{donde } \mathbf{Z}_s \sim N_n(0, \mathbf{I}_n) \text{ independientes, } s = 1, \dots, p \end{aligned}$$

y por lo tanto, obtenemos la siguiente expresión para  $s_{i,jk}$ ,

$$s_{i,jk} = n_i^{-1} \sigma_{jj}^{1/2} \sigma_{kk}^{1/2} \mathbf{Z}_{i,j}^\top \mathbf{Z}_{i,k}, \quad (3.11)$$

donde  $\mathbf{Z}_{i,1}, \dots, \mathbf{Z}_{i,p}$  son independientes con distribución  $N_{n_i}(\mathbf{0}, \mathbf{I}_{n_i})$ .

Para  $\ell = 1, \dots, p$  definamos  $X_{n,\ell} = t_{n,\ell} - t_{n,\ell-1}$ . Nuestro objetivo ahora es ver que esta sucesión doble es un arreglo de martingalas diferencia. Vamos a encontrar una expresión para  $X_{n,\ell}$ . Para eso primero notemos que si llamamos  $\mathbf{S}_i^{(\ell)}$  al bloque de  $\ell \times \ell$  superior izquierdo de la matriz  $\mathbf{S}_i$ , entonces

$$\begin{aligned} \text{tr} \left( \left( \mathbf{S}_i^{(\ell)} \right)^2 \right) - \text{tr} \left( \left( \mathbf{S}_i^{(\ell-1)} \right)^2 \right) &= 2 \sum_{h=1}^{\ell-1} s_{i,h\ell}^2 + s_{i,\ell\ell}^2 \\ \text{tr} \left( \mathbf{S}_i^{(\ell)} \mathbf{S}_j^{(\ell)} \right) - \text{tr} \left( \mathbf{S}_i^{(\ell-1)} \mathbf{S}_j^{(\ell-1)} \right) &= 2 \sum_{h=1}^{\ell-1} s_{i,h\ell} s_{j,h\ell} + s_{i,\ell\ell} s_{j,\ell\ell} \\ \text{tr}^2 \left( \mathbf{S}_i^{(\ell)} \right) - \text{tr}^2 \left( \mathbf{S}_i^{(\ell-1)} \right) &= 2 \sum_{h=1}^{\ell-1} s_{i,hh} s_{i,\ell\ell} + s_{i,\ell\ell}^2. \end{aligned}$$

Usando (3.9), tenemos que

$$\begin{aligned} t_{n,\ell} - t_{n,\ell-1} &= \sum_{i < j} \left\{ 1 - \frac{n_i - 2}{\eta_i} \right\} \left\{ \text{tr} \left( \left( \mathbf{S}_i^{(\ell)} \right)^2 \right) - \text{tr} \left( \left( \mathbf{S}_i^{(\ell-1)} \right)^2 \right) \right\} \\ &\quad + \left\{ 1 - \frac{n_j - 2}{\eta_j} \right\} \left\{ \text{tr} \left( \left( \mathbf{S}_j^{(\ell)} \right)^2 \right) - \text{tr} \left( \left( \mathbf{S}_j^{(\ell-1)} \right)^2 \right) \right\} - 2 \left\{ \text{tr} \left( \mathbf{S}_i^{(\ell)} \mathbf{S}_j^{(\ell)} \right) - \text{tr} \left( \mathbf{S}_i^{(\ell-1)} \mathbf{S}_j^{(\ell-1)} \right) \right\} \\ &\quad - \frac{n_i}{\eta_i} \left\{ \text{tr}^2 \left( \mathbf{S}_i^{(\ell)} \right) - \text{tr}^2 \left( \mathbf{S}_i^{(\ell-1)} \right) \right\} - \frac{n_j}{\eta_j} \left\{ \text{tr}^2 \left( \mathbf{S}_j^{(\ell)} \right) - \text{tr}^2 \left( \mathbf{S}_j^{(\ell-1)} \right) \right\}. \end{aligned}$$

Luego, podemos expresar  $X_{n,\ell} = \sum_{i < j} \tau_{ij,\ell}$  donde

$$\begin{aligned}\tau_{ij,\ell} &= \rho_{1i,\ell} + \rho_{1j,\ell} - \rho_{2ij,\ell} - \rho_{3i,\ell} - \rho_{3j,\ell} \\ \rho_{1i,\ell} &= \left\{ 1 - \frac{n_i - 2}{\eta_i} \right\} \left( 2 \sum_{h=1}^{\ell-1} s_{i,h\ell}^2 + s_{i,\ell\ell}^2 \right) \\ \rho_{2ij,\ell} &= \left( 2 \left[ \sum_{h=1}^{\ell-1} s_{i,h\ell} s_{j,h\ell} \right] + s_{i,\ell\ell} s_{j,\ell\ell} \right) \\ \rho_{3i,\ell} &= \frac{n_i}{\eta_i} \left( 2 \left[ \sum_{h=1}^{\ell-1} s_{i,hh} s_{i,\ell\ell} \right] + s_{i,\ell\ell}^2 \right).\end{aligned}$$

Consideremos la filtración  $\mathcal{F}_{n,\ell-1} = \{\mathbf{Z}_{i,1}, \dots, \mathbf{Z}_{i,\ell-1} \mid i = 1, \dots, k\}$ . Notemos que es la misma para cada  $n$ , por lo que están anidadas en la forma del Teorema Central del Límite para Martingalas. Recordando la expresión (3.11) que teníamos para  $s_{i,jk}$ , demostraremos las siguientes identidades:

$$\mathbb{E}(\mathbf{S}_{i,h\ell}^2 | \mathcal{F}_{n,\ell-1}) = \frac{\sigma_{hh}\sigma_{\ell\ell}}{n_i^2} \mathbf{Z}_{i,h}^\top \mathbf{Z}_{i,h} \quad (3.12)$$

$$\mathbb{E}(\mathbf{S}_{i,\ell\ell}^2 | \mathcal{F}_{n,\ell-1}) = \frac{\sigma_{\ell\ell}^2}{n_i} (n_i + 2) \quad (3.13)$$

$$\mathbb{E}(\mathbf{S}_{i,h\ell} \mathbf{S}_{j,h\ell} | \mathcal{F}_{n,\ell-1}) = 0 \quad (3.14)$$

$$\mathbb{E}(\mathbf{S}_{i,\ell\ell} \mathbf{S}_{j,\ell\ell} | \mathcal{F}_{n,\ell-1}) = \sigma_{\ell\ell}^2 \quad (3.15)$$

$$\mathbb{E}(\mathbf{S}_{i,hh} \mathbf{S}_{i,\ell\ell} | \mathcal{F}_{n,\ell-1}) = \frac{\sigma_{hh}\sigma_{\ell\ell}}{n_i} \mathbf{Z}_{i,h}^\top \mathbf{Z}_{i,h}. \quad (3.16)$$

Indiquemos  $\mathbf{Z}_{i,h} = (z_{ih,1}, \dots, z_{ih,n_i})^\top$ . Para probar (3.12) observemos que

$$\begin{aligned}\mathbb{E}(s_{i,h\ell}^2 | \mathcal{F}_{n,\ell-1}) &= n_i^{-2} \sigma_{hh}\sigma_{\ell\ell} \mathbb{E}((\mathbf{Z}_{i,h}^\top \mathbf{Z}_{i,h})^2 | \mathcal{F}_{n,\ell-1}) = n_i^{-2} \sigma_{hh}\sigma_{\ell\ell} \mathbb{E}\left(\sum_{r,s=1}^{n_i} z_{ih,r} z_{il,r} z_{ih,s} z_{il,s} \mid \mathcal{F}_{n,\ell-1}\right) \\ &= n_i^{-2} \sigma_{hh}\sigma_{\ell\ell} \sum_{r,s=1}^{n_i} z_{ih,r} z_{ih,s} \mathbb{E}(z_{il,r} z_{il,s}) = n_i^{-2} \sigma_{hh}\sigma_{\ell\ell} \sum_{r=1}^{n_i} z_{ih,r}^2 \mathbb{E}(z_{il,r}^2) = n_i^{-2} \sigma_{hh}\sigma_{\ell\ell} \mathbf{Z}_{i,h}^\top \mathbf{Z}_{i,h}.\end{aligned}$$

Veamos ahora (3.13).

$$\begin{aligned}\mathbb{E}(s_{i,\ell\ell}^2 | \mathcal{F}_{n,\ell-1}) &= \mathbb{E}(s_{i,\ell\ell}^2) = n_i^{-2} \sigma_{\ell\ell}^2 \mathbb{E}((\mathbf{Z}_{i,\ell}^\top \mathbf{Z}_{i,\ell})^2) = n_i^{-2} \sigma_{\ell\ell}^2 \mathbb{E}\left(\sum_{r,s=1}^{n_i} z_{i\ell,r}^2 z_{i\ell,s}^2\right) \\ &= n_i^{-2} \sigma_{\ell\ell}^2 \left\{ \sum_{\substack{r,s=1 \\ r \neq s}}^{n_i} \mathbb{E}(z_{i\ell,r}^2) \mathbb{E}(z_{i\ell,s}^2) + \sum_{r=1}^{n_i} \mathbb{E}(z_{i\ell,r}^4) \right\} = n_i^{-2} \sigma_{\ell\ell}^2 \{n_i(n_i - 1) + 3n_i\} \\ &= n_i^{-1} \sigma_{\ell\ell}^2 (n_i + 2).\end{aligned}$$

La identidad (3.14) se deduce fácilmente de la simetría de  $\mathbf{Z}_{i,l}$  ya que

$$\begin{aligned}\mathbb{E}(s_{i,\ell}^2|\mathcal{F}_{n,\ell-1}) &= n_i^{-1}n_j^{-1}\sigma_{hh}\sigma_{\ell\ell}\mathbb{E}\left(\sum_{r=1}^{n_i}\sum_{s=1}^{n_j}z_{ih,r}z_{il,r}z_{jh,s}z_{j\ell,s}|\mathcal{F}_{n,\ell-1}\right) \\ &= n_i^{-1}n_j^{-1}\sigma_{hh}\sigma_{\ell\ell}\sum_{r=1}^{n_i}\sum_{s=1}^{n_j}z_{ih,r}z_{jh,s}\mathbb{E}(z_{il,r})\mathbb{E}(z_{j\ell,s}) = 0.\end{aligned}$$

Por otra parte, como  $\mathbf{Z}_{i,l} \sim N(\mathbf{0}, \mathbf{I}_{n_i})$  tenemos que

$$\mathbb{E}(s_{i,\ell}s_{j,\ell}|\mathcal{F}_{n,\ell-1}) = n_i^{-1}n_j^{-1}\sigma_{\ell\ell}^2\sum_{r=1}^{n_i}\sum_{s=1}^{n_j}\mathbb{E}(z_{il,r}^2)\mathbb{E}(z_{j\ell,s}^2) = n_i^{-1}n_j^{-1}\sigma_{\ell\ell}^2n_in_j = \sigma_{\ell\ell}^2$$

lo que prueba (3.15). Finalmente, (3.16) se deduce de

$$\mathbb{E}(s_{i,hh}s_{i,\ell}|\mathcal{F}_{n,\ell-1}) = n_i^{-1}\sigma_{hh}\sigma_{\ell\ell}\sum_{r,s=1}^{n_i}z_{ih,r}^2\mathbb{E}(z_{il,s}^2) = n_i^{-1}\sigma_{hh}\sigma_{\ell\ell}\mathbf{Z}_{i,h}^T\mathbf{Z}_{i,h}.$$

Estas fórmulas permiten calcular con facilidad las esperanzas condicionales de  $\rho_{1i,\ell}$ ,  $\rho_{2ij,\ell}$  y  $\rho_{3i,\ell}$ . Cálculos inmediatos permiten obtener

$$\begin{aligned}\mathbb{E}(\rho_{1i,\ell}|\mathcal{F}_{n,\ell-1}) &= \left\{1 - \frac{n_i - 2}{\eta_i}\right\} \left\{2\sum_{h=1}^{\ell-1}\frac{\sigma_{hh}\sigma_{\ell\ell}}{n_i^2}\mathbf{Z}_{i,h}^T\mathbf{Z}_{i,h} + \frac{\sigma_{\ell\ell}^2}{n_i}(n_i + 2)\right\} \\ \mathbb{E}(\rho_{2ij,\ell}|\mathcal{F}_{n,\ell-1}) &= 2\sigma_{\ell\ell}^2 \\ \mathbb{E}(\rho_{3i,\ell}|\mathcal{F}_{n,\ell-1}) &= \frac{n_i}{\eta_i} \left\{2\sum_{h=1}^{\ell-1}\frac{\sigma_{\ell\ell}\sigma_{hh}}{n_i}\mathbf{Z}_{i,h}^T\mathbf{Z}_{i,h} + \frac{\sigma_{\ell\ell}^2}{n_i}(n_i + 2)\right\}.\end{aligned}$$

Usando estas igualdades es fácil probar que  $X_{n,\ell}$  es una martingala diferencia respecto a  $\mathcal{F}_{n,\ell}$  porque

$$\mathbb{E}(X_{n,\ell}|\mathcal{F}_{n,\ell-1}) = \sum_{\substack{i,j=1 \\ i < j}}^k \mathbb{E}(\rho_{1i,\ell}) + \mathbb{E}(\rho_{1j,\ell}) - \mathbb{E}(\rho_{2ij,\ell}) - \mathbb{E}(\rho_{3i,\ell}) - \mathbb{E}(\rho_{3j,\ell}) = 0.$$

Para poder usar el Teorema Central del Límite para Martingalas, es necesario ver que vale

$$\sum_{\ell=1}^p \mathbb{E}(X_{n,\ell}^2|\mathcal{F}_{n,\ell-1}) \xrightarrow{p} \theta^2 \quad (3.17)$$

y que vale la condición de Lindeberg

$$\sum_{\ell=1}^p \mathbb{E}(X_{n,\ell}^2 \mathbb{I}(|X_{n,\ell}| > \epsilon)|\mathcal{F}_{n,\ell-1}) \xrightarrow{p} 0. \quad (3.18)$$

Las demostraciones de (3.17) y (3.18) pueden verse en Schott (2007). Por lo tanto, estamos en condiciones de usar el Teorema Central del Límite para Martingalas. Luego,  $t_{np} \xrightarrow{D} N(0, \theta^2)$ , como queríamos probar.  $\square$

Notemos que el valor de  $\theta^2$  es desconocido y por lo tanto necesitamos estimarlo. Llamamos  $n = \sum_{i=1}^k n_i$  y  $\mathbf{S} = (\sum_{i=1}^k n_i \mathbf{S}_i)/n$ . Schott (2007) prueba que

$$\hat{\theta}^2 = 4 \left\{ \sum_{\substack{i,j=1 \\ i < j}}^k \left( \frac{n_i + n_j}{n_i n_j} \right)^2 + (k-1)(k-2) \sum_{i=1}^k \frac{1}{n_i^2} \right\} a^2$$

donde

$$a = \frac{n^2}{(n+2)(n-1)} \left( \text{tr}(\mathbf{S}^2) - \frac{\text{tr}^2(\mathbf{S})}{n} \right)$$

es un estimador consistente de  $\theta^2$ . Por lo tanto, usando el Teorema de Slutsky y el Teorema 3.2.6 obtenemos que el estadístico  $t_{np}^* = t_{np}/\hat{\theta}$  converge en distribución a una  $N(0, 1)$ . Luego, fijado un  $0 < \alpha < 1$ , el test rechaza  $H_0$  cuando el valor de  $t_{np}^*$  excede al percentil  $100(1-\alpha)$  % de una normal estándar.

### 3.2.3 Otros tests para el contexto de alta dimensión

Como mencionamos en la Introducción, el problema de la alta dimensión cobró importancia en los últimos años debido a avances informáticos. Vamos a describir algunas ideas para nuevos tests y sus resultados  $(n, p)$ -asintóticos. Consideremos dos grupos con respectivas matrices de covarianza  $\Sigma_1$  y  $\Sigma_2$ . Al igual que antes, tenemos dos muestras  $\mathbf{X}_1, \dots, \mathbf{X}_{n_1}$  y  $\mathbf{Y}_1, \dots, \mathbf{Y}_{n_2}$  con medias respectivas  $\boldsymbol{\mu}_1 = (\mu_{11}, \dots, \mu_{1p})^T$ ,  $\boldsymbol{\mu}_2 = (\mu_{21}, \dots, \mu_{2p})^T$  (en principio desconocidas) y matrices de covarianza  $\Sigma_1$  y  $\Sigma_2$ . Nos interesan las hipótesis

$$H_0 : \Sigma_1 = \Sigma_2 \quad \text{vs.} \quad \Sigma_1 \neq \Sigma_2.$$

Cuando  $H_0$  es rechazada a veces es importante saber en qué coordenadas las matrices difieren. Por ejemplo, en análisis biológico de microarrays, un cambio en la interrelación de un gen con algún otro sirve para detectar posibles mutaciones. Cai *et al.* (2013) proponen un test que permite inferir sobre esto bajo  $H_1$ . Es claro que  $H_0$  es equivalente a que  $\max_{1 \leq i \leq j \leq p} |\sigma_{1,ij} - \sigma_{2,ij}| = 0$ , donde llamamos  $\sigma_{t,ij}$  al elemento  $(i, j)$  de  $\Sigma_t$ ,  $t = 1, 2$ . Teniendo esto en mente tendría sentido basar el test en el máximo de los valores estimados para  $|\sigma_{1,ij} - \sigma_{2,ij}|$ . Si usamos la matriz de covarianza muestral para estimar las coordenadas de  $\Sigma_1$  y  $\Sigma_2$ , entonces resulta que los estimadores  $|\hat{\sigma}_{1,ij} - \hat{\sigma}_{2,ij}|$  son heterocedásticos y por lo tanto hay que estandarizarlos.

Indicaremos por  $\mathbf{X} = (X_1, \dots, X_p)$  a la distribución de la muestra  $\mathbf{X}_1, \dots, \mathbf{X}_{n_1}$  y por  $\mathbf{Y} = (Y_1, \dots, Y_p)$  a la distribución de la muestra  $\mathbf{Y}_1, \dots, \mathbf{Y}_{n_2}$ . La varianza de  $\hat{\sigma}_{1,ij} - \hat{\sigma}_{2,ij}$  puede estimarse como  $\hat{\theta}_{1,ij}/n_1 + \hat{\theta}_{2,ij}/n_2$  donde  $\hat{\theta}_{1,ij}$  y  $\hat{\theta}_{2,ij}$  son estimadores de  $\text{VAR}((X_i - \mu_{1i})(X_j - \mu_{1j}))$  y de  $\text{VAR}((Y_i - \mu_{2i})(Y_j - \mu_{2j}))$  respectivamente. Indicamos con  $X_{ki}$  al elemento  $i$ -ésimo de  $\mathbf{X}_k$ ,  $Y_{ki}$  al elemento  $i$ -ésimo de  $\mathbf{Y}_k$ ,  $\bar{\mathbf{X}} = (\bar{X}_1, \dots, \bar{X}_p)^T$  al promedio de las observaciones  $\mathbf{X}_1, \dots, \mathbf{X}_{n_1}$  y

$\bar{\mathbf{Y}} = (\bar{Y}_1, \dots, \bar{Y}_p)^T$  al promedio de las observaciones  $\mathbf{Y}_1, \dots, \mathbf{Y}_{n_2}$ . Entonces podemos tomar

$$\begin{aligned}\hat{\theta}_{1,ij} &= \frac{1}{n_1} \sum_{k=1}^{n_1} [(X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) - \hat{\sigma}_{1,ij}]^2 \\ \hat{\theta}_{2,ij} &= \frac{1}{n_2} \sum_{k=1}^{n_2} [(Y_{ki} - \bar{Y}_i)(Y_{kj} - \bar{Y}_j) - \hat{\sigma}_{2,ij}]^2.\end{aligned}$$

Así, se define

$$W_{ij} = \frac{\hat{\sigma}_{1,ij} - \hat{\sigma}_{2,ij}}{\sqrt{\frac{\hat{\theta}_{1,ij}}{n_1} + \frac{\hat{\theta}_{2,ij}}{n_2}}} \quad \text{y} \quad M_{ij} = W_{ij}^2.$$

El estadístico del test es  $M_n = \max_{1 \leq i \leq j \leq p} M_{ij}$  y se rechaza  $H_0$  cuando  $M_n$  es grande. Bajo ciertas hipótesis para las observaciones que generalizan al caso normal, Cai *et al.* (2013) prueban que  $M_n - 4 \log p + \log \log p$  converge en distribución a una variable aleatoria  $V$  con función de distribución acumulada

$$F_V(x) = \exp \left\{ -\frac{1}{\sqrt{8\pi}} \exp \left( -\frac{x}{2} \right) \right\}$$

y de esta forma se pueden constriuir tests asintóticos de nivel  $\alpha$  para  $0 < \alpha < 1$ .

Otro test que se restringe al caso de muestras normales multivariada fue propuesto por Srivastava y Yanagihara (2010). Dadas dos matrices  $\Sigma_1$  y  $\Sigma_2$ , una posible idea para detectar su diferencia es considerar  $\text{tr}(\Sigma_1) - \text{tr}(\Sigma_2)$  y  $\text{tr}(\Sigma_1^2) - \text{tr}(\Sigma_2^2)$ . Una distancia que toma en cuenta ambas es  $d = \text{tr}(\Sigma_1^2)/\text{tr}(\Sigma_1)^2 - \text{tr}(\Sigma_2^2)/\text{tr}(\Sigma_2)^2$ . Se consideran para  $i = 1, 2$  estimadores consistentes  $\hat{\gamma}_i$  de  $\gamma_i = \text{tr}(\Sigma_i^2)/\text{tr}(\Sigma_i)^2$ . Estandarizando el estadístico  $\hat{\gamma}_1 - \hat{\gamma}_2$  se propone basar el test en

$$Q_n = \frac{\hat{\gamma}_1 - \hat{\gamma}_2}{\sqrt{\hat{\xi}_1^2 + \hat{\xi}_2^2}},$$

donde  $\hat{\xi}_i^2$  son estimadores consistentes de la varianza asintótica de  $\hat{\gamma}_i$ . En Srivastava (2005) se prueba que

$$\frac{\hat{\gamma}_i - \gamma_i}{\xi_i} \xrightarrow{D} N(0, 1). \quad (3.19)$$

Por lo tanto, bajo  $H_0$ ,  $Q_n \xrightarrow{D} N(0, 1)$  y de esta forma se obtienen tests asintóticos. En la demostración de (3.19), para llegar a la normalidad asintótica de  $\hat{\gamma}_i - \gamma_i$  se usa la siguiente versión del Teorema Central del Límite dada en Rao (1973, p.147 Problema 4.7).

**Teorema 3.2.7** Sean  $\mathbf{X}_1, \mathbf{X}_2, \dots$  vectores aleatorios independientes con  $\mathbb{E}(\mathbf{X}_i) = 0$  y tales que  $(1/n) \sum_{i=1}^n \text{VAR}(\mathbf{X}_i) \rightarrow \Sigma$  para alguna matriz  $\Sigma$  cuando  $n \rightarrow \infty$ . Supongamos que además vale que para cada  $\epsilon > 0$

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} (\|\mathbf{X}_i\|^2 \mathbb{I}(\|\mathbf{X}_i\| > \epsilon \sqrt{n})) \rightarrow 0.$$

Entonces,

$$\frac{1}{\sqrt{n}} (\mathbf{X}_1 + \dots + \mathbf{X}_n) \xrightarrow{D} N(0, \Sigma).$$

Otro test fue propuesto recientemente por Li y Chen (2012) en donde se propone, al igual que en el test de Schott, basarse en  $\text{tr}((\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2)$ . Sin embargo, a diferencia de este último test, obtienen estimadores insesgados y consistentes para cada uno de los términos en  $\text{tr}((\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2) = \text{tr}(\mathbf{\Sigma}_1^2) + \text{tr}(\mathbf{\Sigma}_2^2) - 2\text{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$ . De esta forma proponen un estadístico cuya esperanza es cero si y sólo  $H_0$  es cierta. Se prueba la normalidad  $(n, p)$ -asintótica usando el Teorema Central del Límite para Martingalas.

En estos trabajos, después de definir los tests y mostrar los resultados  $(n, p)$ -asintóticos, se hacen estudios de nivel y de potencia fijando ciertos modelos. Aunque haya resultados asintóticos sobre el nivel, es importante hacerlos para ver cuan grandes deberían ser los tamaños de las muestras o la dimensión de cada elemento para que el nivel del test se acerque al nivel nominal fijado inicialmente.

Comparar potencias para diferentes tests es un tema delicado, porque depende mucho de qué modelo se esté tomando. Por ejemplo, en el trabajo de Srivastava (2010) se eligen unos modelos para los cuales la potencia estimada de su test resulta mayor que la del test de Schott (2007). Sin embargo, Schott (2007) fija otro modelo para estimar la potencia de su test y ésta resulta mayor que la estimada para el test de Srivastava (2010).

Cai *et al.* (2013) proponen un modelo esparso para estudiar potencia. Eligen matrices  $\mathbf{\Sigma}_1$  y  $\mathbf{\Sigma}_2$  iguales en todas las coordenadas salvo en algunas pocas elegidas al azar. Por cómo se definió este último test, tendría sentido que una diferencia de esta forma pueda ser detectada por el test definido en Cai *et al.* (2013) pero no por los tests basados en la norma de Frobenius. Esto es porque modificar pocas coordenadas de una matriz de dimensión grande no cambia significativamente sus autovalores.

Por estos motivos no es sencillo afirmar que cierto test es mejor que otro. Más adelante, al hacer el estudio de nivel y potencia para los tests que proponemos en este trabajo, vamos a comparar los resultados con los obtenidos al aplicar el test de Schott (2007).

# Capítulo 4

## Tests basados en signos

En este capítulo vamos a introducir el concepto de signo de un vector aleatorio y vamos a analizar algunos test que se pueden realizar basándose en ellos, estudiando la distribución  $n$  y  $(n, p)$ -asintótica de los estadísticos usados. Además de la importancia intrínseca de estos test, es importante analizar las técnicas usadas en las demostraciones para una posible adaptación para los tests que se presentan en el próximo capítulo.

### 4.1 El concepto de signo y matriz de covarianza de signos espaciales

**Definición.** Dado un vector aleatorio  $\mathbf{X} \in \mathbb{R}^p$ , definimos el **signo de  $\mathbf{X}$**  como

$$s(\mathbf{X}) = \frac{\mathbf{X}}{\|\mathbf{X}\|},$$

es decir, el vector de norma 1 definido por  $\mathbf{X}$ .

El nombre viene del caso  $p = 1$ , donde la función signo coincide con la habitual. La definición es muy sencilla y eso es en parte lo que la hace atractiva. Intuitivamente, al considerar los signos de una muestra aleatoria se está perdiendo información sobre las magnitudes de las observaciones pero no se pierde información sobre sus orientaciones.

Siendo un poco más formales, consideremos una muestra aleatoria  $\mathbf{X}_1, \dots, \mathbf{X}_n$  tal que  $\mathbf{X}_i \sim \mathbf{X}$ ,  $\mathbb{E}(\mathbf{X}) = \mathbf{0}$  y matriz de covarianza  $\Sigma = \mathbb{E}(\mathbf{X}\mathbf{X}^T)$ .

Podemos considerar una descomposición de la matriz  $\Sigma = \sigma(\Sigma)\Lambda$  donde  $\sigma$  es una función escalar tal que  $\sigma(\mathbf{I}_p) = 1$  y  $\sigma(a\Sigma) = a\sigma(\Sigma)$ . En general suele considerarse a  $\sigma(\Sigma)$  como la media aritmética o geométrica de los autovalores de  $\Sigma$ . Es decir, se toma

$$\sigma(\Sigma) = \frac{\text{tr}(\Sigma)}{p} \quad \text{o} \quad \sigma(\Sigma) = \det(\Sigma)^{1/p}.$$

La función  $\sigma(\Sigma)$  se llama la **escala** de la matriz  $\Sigma$  y a  $\Lambda = \Sigma/\sigma(\Sigma)$  se la llama la **forma** de  $\Sigma$ . Cuando nos restringimos a los signos de una muestra aleatoria, estamos perdiendo información sobre la escala pero no sobre la forma de la matriz de covarianza poblacional.

**Definición.** Dada  $\mathbf{X}_1, \dots, \mathbf{X}_n$  una muestra aleatoria tal que  $\mathbb{E}(\mathbf{X}_i) = \mathbf{0}$  y  $\mathbb{P}(\mathbf{X}_i = \mathbf{0}) = 0$ , se llama **matriz de covarianza de signos espaciales** (poblacional) a

$$\tilde{\Sigma} = \mathbb{E} \left( s(\mathbf{X}_1) s(\mathbf{X}_1)^T \right) = \mathbb{E} \left( \frac{\mathbf{X}_1 \mathbf{X}_1^T}{\|\mathbf{X}_1\|^2} \right).$$

Definimos además la matriz de covarianza de signos espaciales muestral como

$$\mathbf{V} = \frac{1}{n} \sum_{i=1}^n s(\mathbf{X}_i) s(\mathbf{X}_i)^T = \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{X}_i \mathbf{X}_i^T}{\|\mathbf{X}_i\|^2}.$$

Observemos que  $\tilde{\Sigma}$  está siempre definida aún cuando  $\mathbf{X}$  no tiene segundo momento finito.

En esta tesis consideraremos dos muestras aleatorias  $\mathbf{X}_{ij}$ ,  $i = 1, 2$ ,  $j = 1, \dots, n_i$  provenientes de poblaciones con matrices de covarianza de signos espaciales  $\tilde{\Sigma}_i$ ,  $i = 1, 2$ , respectivamente. Nos interesa testear

$$H_0 : \tilde{\Sigma}_1 = \tilde{\Sigma}_2 \quad vs. \quad H_1 : \tilde{\Sigma}_1 \neq \tilde{\Sigma}_2. \quad (4.1)$$

Una cuestión importante es ver como interpretar las hipótesis (4.1) en términos de las matrices de covarianza usuales  $\Sigma_1$  y  $\Sigma_2$ , en el caso en que estas existan. Veremos cómo se pueden relacionar los autovectores y autovalores de estas dos matrices. Necesitamos primero el siguiente lema auxiliar.

**Lema 4.1.1.** Sea  $\mathbf{X} \sim \mathcal{E}_p(\mathbf{0}, \Sigma, \phi)$ . Entonces, existe un vector aleatorio  $\mathbf{X}^* \sim N(\mathbf{0}, \Sigma)$  tal que  $s(\mathbf{X}) = s(\mathbf{X}^*)$ .

DEMOSTRACIÓN. Como  $\mathbf{X} \sim \mathcal{E}_p(\mathbf{0}, \Sigma, \phi)$  entonces  $\mathbf{X} = \beta \Lambda^{1/2} \mathbf{Y}$  donde  $\Sigma = \beta \Lambda \beta^T$ ,  $\beta$  es ortogonal,  $\Lambda$  es diagonal e  $\mathbf{Y}$  es esférica. Por lo tanto,  $\mathbf{U} = s(\mathbf{Y})$  es uniforme en la esfera  $\mathcal{S}^{p-1}$ . De esta forma,

$$s(\mathbf{X}) = \beta \Lambda^{1/2} \frac{\mathbf{Y}}{\|\Lambda^{1/2} \mathbf{Y}\|} = \beta \Lambda^{1/2} \frac{\mathbf{U}}{\|\Lambda^{1/2} \mathbf{U}\|}.$$

Sea  $Z^2 \sim \chi_1^2$  independiente de  $\mathbf{U}$ . Definamos  $\mathbf{Y}^* = Z\mathbf{U}$  y  $\mathbf{X}^* = \beta \Lambda^{1/2} \mathbf{Y}^*$ . Entonces,  $\mathbf{Y}^* \sim N(\mathbf{0}, \mathbf{I}_p)$  de donde  $\mathbf{X}^* \sim N(\mathbf{0}, \Sigma)$  y  $s(\mathbf{X}) = s(\mathbf{X}^*)$ .  $\square$

Teniendo en cuenta este lema podemos probar el siguiente resultado.

**Lema 4.1.2.** Sea  $\mathbf{X}$  un vector aleatorio y  $\tilde{\Sigma} = \mathbb{E} (s(\mathbf{X}) s(\mathbf{X})^T)$ . Supongamos que vale (a) o (b) donde

(a)  $\mathbf{X} \sim \mathcal{E}_p(\mathbf{0}, \Sigma, \phi)$ .

(b)  $\mathbb{E} (\|\mathbf{X}\|^2) < \infty$  y si  $\Sigma = \mathbb{E} (\mathbf{X} \mathbf{X}^T) = \beta^T \Lambda \beta$  con  $\beta$  ortogonal y  $\Lambda$  diagonal, entonces el vector

$$\boldsymbol{\xi} = (\xi_1, \dots, \xi_p)^T = \Lambda^{-1/2} \beta \mathbf{X} \quad (4.2)$$

cumple que es invariante por transformaciones ortogonales del tipo  $(\mathbf{e}_1, \dots, -\mathbf{e}_i, \dots, \mathbf{e}_p)$  para  $i = 1, \dots, p$ , donde  $\mathbf{e}_j$  es el  $j$ -ésimo vector canónico en  $\mathbb{R}^p$ .

Sean  $\lambda_1, \dots, \lambda_p$  y  $\psi_1, \dots, \psi_p$  los autovalores de  $\mathbf{\Sigma}$  y  $\tilde{\mathbf{\Sigma}}$  respectivamente. Entonces, las matrices  $\mathbf{\Sigma}$  y  $\tilde{\mathbf{\Sigma}}$  tienen los mismos autovectores y los autovalores se relacionan de la forma

$$\psi_j = \mathbb{E} \left( \frac{\lambda_j \xi_j^2}{\sum_{\ell=1}^p \lambda_\ell \xi_\ell^2} \right), \quad (4.3)$$

donde, si vale (a), los  $\xi_i$  son independientes tales que  $\xi_i^2 \sim \chi_1^2$  y, si vale (b), los  $\xi_i$  se definen como en (4.2).

DEMOSTRACIÓN. Sea  $\alpha_j$  el autovector correspondiente al autovalor  $\lambda_j$  de la matriz  $\mathbf{\Sigma}$ . Para  $j = 1, \dots, p$  llamamos

$$\xi_j = \frac{\alpha_j^T \mathbf{X}}{\lambda_j^{1/2}}.$$

Observemos que esta definición coincide con la dada en (4.2) para el caso (b). Además,  $\mathbf{X} = \sum_{i=1}^p \xi_i \lambda_i^{1/2} \alpha_i$  y por lo tanto  $\|\mathbf{X}\|^2 = \sum_{i=1}^p \xi_i^2 \lambda_i$ . Fijando un  $1 \leq i \leq p$  y usando estas expresiones, tenemos que

$$\begin{aligned} \tilde{\mathbf{\Sigma}} \alpha_i &= \mathbb{E} \left( \frac{\mathbf{X} \mathbf{X}^T \alpha_i}{\|\mathbf{X}\|^2} \right) = \mathbb{E} \left( \frac{\mathbf{X}}{\|\mathbf{X}\|^2} \xi_i \lambda_i^{1/2} \right) = \mathbb{E} \left( \frac{\left( \sum_{j=1}^p \xi_j \lambda_j^{1/2} \alpha_j \right) \xi_i \lambda_i^{1/2}}{\|\mathbf{X}\|^2} \right) \\ &= \sum_{j \neq i}^p \lambda_i^{1/2} \lambda_j^{1/2} \alpha_j \mathbb{E} \left( \frac{\xi_i \xi_j}{\|\mathbf{X}\|^2} \right) + \lambda_i \alpha_i \mathbb{E} \left( \frac{\xi_i^2}{\|\mathbf{X}\|^2} \right) \\ &= \sum_{j \neq i}^p \lambda_i^{1/2} \lambda_j^{1/2} \alpha_j \mathbb{E} \left( \frac{\xi_i \xi_j}{\sum_{i=1}^p \xi_i^2 \lambda_i} \right) + \lambda_i \alpha_i \mathbb{E} \left( \frac{\xi_i^2}{\sum_{i=1}^p \xi_i^2 \lambda_i} \right). \end{aligned}$$

Notemos que si vale (a) entonces por el Lema 4.1.1 podemos asumir que  $\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma})$ , lo que implica que el vector  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_p)$  tiene distribución  $N(\mathbf{0}, \mathbf{I}_p)$  y por lo tanto

$$(\xi_1, \dots, \xi_j, \dots, \xi_p) \sim (\xi_1, \dots, -\xi_j, \dots, \xi_p) \quad \text{para } j = 1, \dots, p, \quad (4.4)$$

pues las variables aleatorias  $\xi_1, \dots, \xi_p$  son independientes y simétricas. Por otra parte, si vale (b) entonces por hipótesis también se obtiene la simetría obtenida en la fórmula (4.4). Entonces, para ambos casos obtenemos

$$\frac{\xi_i \xi_j}{\sum_{k=1}^p \lambda_k \xi_k^2} \sim -\frac{\xi_i \xi_j}{\sum_{k=1}^p \lambda_k \xi_k^2} \quad \text{si } i \neq j.$$

Esto implica que para  $i \neq j$ ,

$$\mathbb{E} \left( \frac{\xi_i \xi_j}{\sum_{i=1}^p \xi_i^2 \lambda_i} \right) = 0.$$

Por lo tanto,

$$\tilde{\mathbf{\Sigma}} \alpha_i = \lambda_i \alpha_i \mathbb{E} \left( \frac{\xi_i^2}{\sum_{i=1}^p \xi_i^2 \lambda_i} \right),$$

es decir,  $\alpha_i$  es también autovector de  $\tilde{\mathbf{\Sigma}}$  y su autovalor correspondiente viene dado por (4.3). Por último, si vale (a) entonces podemos asumir que  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_p)^T \sim N(\mathbf{0}, \mathbf{I}_p)$  y por lo tanto  $\xi_1^2, \dots, \xi_p^2$  son independientes y distribuidas como una  $\chi_1^2$ , lo que concluye la demostración.  $\square$

Por una comunicación personal con David Tyler, tenemos el siguiente lema.

**Lema 4.2.3.** Sean  $X_1, \dots, X_p$  variables aleatorias no negativas. Entonces, dados dos vectores  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$  y  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_p)$  con entradas no negativas y distintos del vector cero, se tiene que

$$\mathbb{E} \left( \frac{\lambda_i X_i}{\sum_{j=1}^p \lambda_j X_j} \right) = \mathbb{E} \left( \frac{\nu_i X_i}{\sum_{j=1}^p \nu_j X_j} \right) \quad \forall i = 1, \dots, p \quad (4.5)$$

si y sólo si  $\boldsymbol{\nu} \propto \boldsymbol{\lambda}$ .

DEMOSTRACIÓN Si  $\boldsymbol{\nu} \propto \boldsymbol{\lambda}$ , entonces es trivial ver que vale (4.5). Por otro lado, supongamos que vale (4.5) y supongamos que los dos vectores no son proporcionales. Notemos que en este caso para cualquier  $i = 1, \dots, p$  vale que  $\lambda_i = 0$  si y solo si  $\nu_i = 0$ . Para cada  $i = 1, \dots, p$  definimos  $\rho_i = 0$  si  $\lambda_i = 0$  y  $\rho_i = \nu_i/\lambda_i$  si  $\lambda_i \neq 0$ . Sea  $Y_i = \lambda_i X_i$ . Luego,  $\rho_i Y_i = \nu_i X_i$ , de donde

$$\mathbb{E} \left( \frac{Y_i}{\sum_{j=1}^p Y_j} \right) = \mathbb{E} \left( \frac{\rho_i Y_i}{\sum_{j=1}^p \rho_j Y_j} \right) \quad \forall i = 1, \dots, p. \quad (4.6)$$

Supongamos sin perder generalidad que  $\rho_1 \geq \rho_2 \geq \dots \geq \rho_p$ . Como estamos suponiendo que  $\boldsymbol{\nu}$  y  $\boldsymbol{\lambda}$  no son proporcionales, entonces  $\rho_1 > \rho_p$ . De esta manera, tenemos que

$$\mathbb{E} \left( \frac{Y_1}{\sum_{j=1}^p Y_j} \right) = \mathbb{E} \left( \frac{\rho_1 Y_1}{\rho_1 \sum_{j=1}^p Y_j} \right) < \mathbb{E} \left( \frac{\rho_1 Y_1}{\sum_{j=1}^p \rho_j Y_j} \right),$$

que se contradice con (4.6). Se sigue que entonces  $\boldsymbol{\nu}$  y  $\boldsymbol{\lambda}$  son proporcionales.  $\square$

A partir de los Lemas 4.2.2 y 4.2.3 se deduce el siguiente corolario:

**Colorario 4.2.4.** Sean  $\mathbf{X}_1$  y  $\mathbf{X}_2$  dos vectores aleatorios con  $\mathbb{P}(\mathbf{X}_i = \mathbf{0}) = 0$ ,  $i = 1, 2$ . Sea  $s(\mathbf{X}) = \mathbf{X}/\|\mathbf{X}\|$ ,  $\tilde{\boldsymbol{\Sigma}}_i = \mathbb{E}(s(\mathbf{X}_i)s(\mathbf{X}_i)^T)$ ,  $i = 1, 2$ . Entonces,

(a) Supongamos que  $\mathbf{X}_1$  y  $\mathbf{X}_2$  tienen segundo momento finito. Para  $i = 1, 2$ , sea  $\boldsymbol{\Sigma}_i = \mathbb{E}(\mathbf{X}_i \mathbf{X}_i^T)$  con  $\boldsymbol{\Sigma}_i = \boldsymbol{\beta}_i^T \boldsymbol{\Lambda}_i \boldsymbol{\beta}_i$  donde  $\boldsymbol{\beta}_i$  es ortogonal y  $\boldsymbol{\Lambda}_i$  es diagonal. Supongamos que el vector aleatorio  $\boldsymbol{\xi}_i = \boldsymbol{\Lambda}_i^{-1/2} \boldsymbol{\beta}_i \mathbf{X}_i$  cumple que es invariante por transformaciones ortogonales del tipo  $(\mathbf{e}_1, \dots, -\mathbf{e}_i, \dots, \mathbf{e}_p)$  para  $i = 1, \dots, p$ , donde  $\mathbf{e}_j$  es el  $j$ -ésimo vector canónico en  $\mathbb{R}^p$ . Entonces,  $\tilde{\boldsymbol{\Sigma}}_1 = \tilde{\boldsymbol{\Sigma}}_2 \iff \boldsymbol{\Sigma}_1 \propto \boldsymbol{\Sigma}_2$ .

(b) Más aún, si  $\mathbf{X}_i \sim \mathcal{E}_p(\mathbf{0}, \boldsymbol{\Sigma}_i, \phi)$ , entonces  $\tilde{\boldsymbol{\Sigma}}_1 = \tilde{\boldsymbol{\Sigma}}_2 \iff \boldsymbol{\Sigma}_1 \propto \boldsymbol{\Sigma}_2$ .

Observemos que en la parte (b) del Corolario 4.2.4 no se necesita la existencia de los segundos momentos.

## 4.2 Tests basados en signos en el contexto de alta dimensión

En esta sección vamos a presentar algunos test basados en signos y sus resultados  $(n, p)$ -asintóticos. Las demostraciones de estos resultados incluyen ideas y resultados que tienen interés intrínseco.

Supongamos que tenemos una muestra  $\mathbf{X}_1, \dots, \mathbf{X}_n$  proveniente de una distribución con media cero y matriz de covarianza  $\Sigma$ . Una de las ventajas de usar tests de matrices de covarianza basados en signos es que, al quedarnos sólo con la información de la forma de  $\Sigma$ , los estadísticos considerados quedan independientes de la escala y por lo tanto los tests que sean válidos para distribuciones normales también lo van a ser para distribuciones elípticas. Otra de las ventajas de usar este tipo de tests es que al normalizar cada una de las observaciones se gana robustez. Cuando  $p$  es chico, quedarse sólo con los signos de las observaciones se ve evidenciado en una pérdida de eficiencia respecto a los test Gaussianos clásicos. Sin embargo, cuando  $p$  es grande, bajo ciertas hipótesis se ve que las observaciones tienden a concentrarse en una esfera y por lo tanto no se pierde demasiada información al restringirnos a los signos. Esta última afirmación puede consultarse en Hall *et al.* (2005).

#### 4.2.1 El test de Paindaveine y Verdebout (2013) para testear esfericidad

Supongamos que tenemos una muestra  $\mathbf{U}_1, \dots, \mathbf{U}_n$  donde  $\mathbf{U}_i$  son vectores en  $\mathcal{S}^{p-1}$ . Queremos testear

$$H_0 : \text{los vectores } \mathbf{U}_i \text{ son uniformes en la esfera} \quad \text{vs.} \quad H_1 : H_0 \text{ es falsa.}$$

El estadístico clásico para este tipo de hipótesis fue propuesto por Rayleigh (1919), quien definió

$$R_n = \frac{p}{n} \sum_{i,j=1}^n \mathbf{U}_i^T \mathbf{U}_j.$$

Es fácil ver que, bajo  $H_0$  y para  $p$  fijo,  $R_n$  tiene distribución asintótica  $\chi_p^2$ . Sin embargo, esta aproximación deja de ser útil cuando  $p$  crece junto con  $n$ . Paindaveine y Verdebout (2013) proponen el estadístico estandarizado

$$R_n^{St} = \frac{R_n - p}{\sqrt{2p}} = \frac{\sqrt{2p}}{n} \sum_{1 \leq i < j \leq n} \mathbf{U}_i^T \mathbf{U}_j. \quad (4.7)$$

Antes de probar la distribución  $(n, p)$ -asintótica necesitamos el siguiente lema.

**Lema 4.2.1.** Sean  $\mathbf{U}_1, \dots, \mathbf{U}_n$  una muestra aleatoria en  $\mathbb{R}^p$  proveniente de una distribución uniforme en  $\mathcal{S}^{p-1}$  y llamamos  $\rho_{ij} = \mathbf{U}_i^T \mathbf{U}_j$  con  $i < j$ . Entonces,

- (i)  $\mathbb{E}(\mathbf{U}_i) = 0$  y  $\text{VAR}(\mathbf{U}_i) = \mathbf{I}_p/p$
- (ii) Las  $n(n-1)/2$  variables aleatorias  $\rho_{ij}$  son independientes dos a dos.
- (iii)  $\rho_{ij}^2 \sim \beta(1/2, (p-1)/2)$  si  $i < j$ .

DEMOSTRACIÓN.

- (i) Indicaremos con  $u_{ik}$  al  $k$ -ésimo elemento de  $\mathbf{U}_i$ . La esperanza de  $\mathbf{U}_i$  es cero porque  $\mathbf{U}_i \sim -\mathbf{U}_i$ . Para hallar su varianza, notemos que si  $\mathbf{A}$  es ortogonal se tiene que  $\mathbf{A}\mathbf{U}_i \sim \mathbf{U}_i$ . Esto implica

que  $\mathbf{A}\text{VAR}(\mathbf{U}_i)\mathbf{A}^\top = \text{VAR}(\mathbf{U}_i)$  para toda  $\mathbf{A}$  ortogonal. Por lo tanto,  $\text{VAR}(\mathbf{U}_i)$  es una matriz diagonal. Además, como  $\mathbb{E}(\mathbf{U}_i) = \mathbf{0}$ ,

$$1 = \mathbb{E}\left(\sum_{k=1}^p u_{ik}^2\right) = \sum_{i=1}^p \text{VAR}(u_{ik}),$$

lo que conjuntamente con el hecho de que  $u_{ik} \sim u_{is}$ ,  $k \neq s$ , implica que  $\text{VAR}(u_{ik}) = 1/p$  para  $k = 1, \dots, p$ , de donde  $\text{VAR}(\mathbf{U}_i) = \mathbf{I}_p/p$ .

- (ii) Si los índices  $\{i, j, k, s\}$  son distintos dos a dos, entonces es trivial que  $\rho_{ij}$  es independiente de  $\rho_{ks}$ . Queda sólo por ver que  $\rho_{ij}$  es independiente de  $\rho_{ik}$  si los índices  $\{i, j, k\}$  son todos distintos. Como la distribución uniforme en  $\mathcal{S}^{p-1}$  es invariante bajo transformaciones ortogonales, se tiene que  $\rho_{ij} \sim \mathbf{e}_{1,p}^\top \mathbf{U}_j$  donde  $\mathbf{e}_{1,p}$  es el primer vector canónico de  $\mathbb{R}^p$ . Análogamente,  $\rho_{ik}$  se distribuye como  $\mathbf{e}_{1,p}^\top \mathbf{U}_k$ . De estas expresiones se ve que  $\rho_{ij}$  es independiente de  $\rho_{ik}$ , como queríamos ver.
- (iii) Como en (ii), tenemos que  $\rho_{ij}$  se distribuye como  $\mathbf{e}_{1,p}^\top \mathbf{U}_j$ . El resultado se obtiene entonces aplicando el Lema 2.2.3 con la matriz  $\mathbf{B} = \mathbf{e}_{1,p}\mathbf{e}_{1,p}^\top$  pues la distribución uniforme es esférica.

El siguiente teorema establece la distribución  $(n, p)$ -asintótica de  $R_n^{St}$ .

**Teorema 4.2.2.** Sean  $p_n \rightarrow \infty$  y  $\{\mathbf{U}_{ni} \in \mathbb{R}^{p_n}, i = 1, \dots, n, n = 1, \dots\}$  un arreglo triangular tal que para cada  $n$  fijo los vectores aleatorios  $\mathbf{U}_{n1}, \dots, \mathbf{U}_{nn}$  son i.i.d. uniformes en  $\mathcal{S}^{p_n-1}$ . Entonces, el estadístico  $R_n^{St}$  definido en (4.7) es asintóticamente normal estándar.

DEMOSTRACIÓN. Vamos a hacer uso del Teorema Central del Límite para Martingalas. Para eso, vamos a considerar la filtración  $\{\mathcal{F}_{n,\ell}\}$  donde  $\mathcal{F}_{n,\ell}$  está definida como la  $\sigma$ -álgebra generada por  $\mathbf{U}_{n1}, \dots, \mathbf{U}_{n\ell}$  y por  $\mathbf{U}_{rs}$  donde  $r < n$ ,  $s \leq r$ . Notemos que estas  $\sigma$ -álgebras están anidadas en la forma necesaria para usar el Teorema 2.5.1, es decir,  $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$  para todo  $n \geq 1$ ,  $1 \leq i \leq n$ . Para simplificar la notación llamemos  $\mathbb{E}_{n,\ell}(\mathbf{X}) = \mathbb{E}(\mathbf{X}|\mathcal{F}_{n,\ell})$  y  $D_{n\ell} = \mathbb{E}_{n,\ell}(R_n^{St}) - \mathbb{E}_{n,\ell-1}(R_n^{St})$ .

Es trivial ver que  $\{D_{n\ell}\}_{\ell=1,\dots,n}$  es una martingala diferencia respecto a  $\{\mathcal{F}_{n,\ell}\}_{\ell=1,\dots,n}$ . Por otro lado, usando que  $\mathbb{E}(\mathbf{U}_{ni}) = \mathbf{0}$  para todo  $n \in \mathbb{N}$ ,  $i = 1, \dots, n$ , se puede ver que

$$D_{n\ell} = \frac{\sqrt{2p_n}}{n} \sum_{i=1}^{\ell-1} \mathbf{U}_{ni}^\top \mathbf{U}_{n\ell}.$$

Es claro que  $|D_{n\ell}| \leq \sqrt{2p_n}(\ell-1)/n$  en casi todo punto, entonces  $D_{n\ell}$  tiene varianza finita. Debemos ver las siguientes dos condiciones para poder usar el Teorema Central del Límite para Martingalas:

- (i) Si llamamos  $\sigma_{n\ell}^2 = \mathbb{E}_{n,\ell-1}(D_{n\ell}^2)$  entonces

$$\sum_{\ell=1}^n \sigma_{n\ell}^2 \xrightarrow{p} 1. \quad (4.8)$$

- (ii) Para todo  $\epsilon > 0$  vale que

$$\sum_{\ell=1}^n \mathbb{E}_{n,\ell-1}((D_{n\ell})^2 \mathbb{I}(|D_{n\ell}| > \epsilon)) \xrightarrow{p} 0. \quad (4.9)$$

Veamos primero (4.8). Eliminamos la dependencia con  $n$  en las cuentas que siguen. Usando el Lema 4.2.1.(i), obtenemos

$$\sigma_{n\ell}^2 = \mathbb{E}_{\ell-1}(D_{n\ell}^2) = \mathbb{E}_{\ell-1}\left(\frac{2p_n}{n^2} \sum_{i,j=1}^{\ell-1} \mathbf{U}_i^T \mathbf{U}_\ell \mathbf{U}_\ell^T \mathbf{U}_j\right) = \frac{2p_n}{n^2} \sum_{i,j=1}^{\ell-1} \mathbf{U}_i^T \mathbb{E}(\mathbf{U}_\ell \mathbf{U}_\ell^T) \mathbf{U}_j = \frac{2}{n^2} \sum_{i,j=1}^{\ell-1} \rho_{ij}.$$

Por lo tanto, obtenemos fácilmente que  $\mathbb{E}(\sigma_{n\ell}^2) = 2(\ell-1)/n^2$ , de donde se obtiene que

$$\mathbb{E}\left(\sum_{\ell=1}^n \sigma_{n\ell}^2\right) = \frac{2}{n^2} \sum_{\ell=1}^n (\ell-1) = \frac{n-1}{n} \rightarrow 1.$$

La conclusión es entonces inmediata si vemos que  $\text{VAR}(\sum_{\ell=1}^n \sigma_{n\ell}^2) \rightarrow 0$ . Usando el Lema 4.2.1, tenemos que los  $\rho_{ij}$  con  $i < j$  son independientes y  $\text{VAR}(\rho_{ij}) = 1/p_n$ . Por lo tanto

$$\begin{aligned} \text{VAR}\left(\sum_{\ell=1}^n \sigma_{n\ell}^2\right) &= \frac{4}{n^4} \text{VAR}\left(\sum_{\ell=1}^n \sum_{i,j=1}^{\ell-1} \rho_{ij}\right) = \frac{4}{n^4} \text{VAR}\left(2 \sum_{\ell=1}^n \sum_{1 \leq i < j \leq \ell-1} \rho_{ij}\right) \\ &= \frac{16}{n^4} \text{VAR}\left(\sum_{1 \leq i < j \leq n} (n-j) \rho_{ij}\right) = \frac{16}{n^4 p_n} \sum_{j=2}^n (j-1)(n-j)^2 \\ &\leq \frac{16}{n^2 p_n} \sum_{j=1}^{n-1} j = \frac{8(n-1)}{n p_n} \rightarrow 0, \end{aligned}$$

como queríamos.

Veamos ahora (4.9). Primero, notemos que si tenemos una sucesión de variables aleatorias  $X_n$  no negativas en casi todo punto tal que  $\mathbb{E}(X_n) \rightarrow 0$ , entonces  $X_n \xrightarrow{p} 0$ . En efecto, por la desigualdad de Markov se tiene que  $\mathbb{P}(|X_n| > \epsilon) \leq \mathbb{E}(|X_n|)/\epsilon \rightarrow 0$ . De esta propiedad se deduce que alcanza probar  $\sum_{\ell=1}^n \mathbb{E}[(D_{n\ell})^2 \mathbb{I}(|D_{n\ell}| > \epsilon)] \xrightarrow{p} 0$ . Observemos que como  $\mathbb{E}(D_{n\ell}) = 0$ ,

$$\text{VAR}(D_{n\ell}) = \mathbb{E}(D_{n\ell}^2) = \mathbb{E}(\sigma_{n\ell}^2) = \frac{2(\ell-1)}{n^2}.$$

Usando las desigualdades de Cauchy-Schwarz y Chebyshev, se tiene que

$$\begin{aligned} \sum_{\ell=1}^n \mathbb{E}[D_{n\ell}^2 \mathbb{I}(|D_{n\ell}| > \epsilon)] &\leq \sum_{\ell=1}^n \sqrt{\mathbb{E}(D_{n\ell}^4)} \sqrt{\mathbb{P}(|D_{n\ell}| \geq \epsilon)} \\ &\leq \frac{1}{\epsilon} \sum_{\ell=1}^n \sqrt{\mathbb{E}(D_{n\ell}^4)} \sqrt{\text{VAR}(D_{n\ell})} \leq \frac{\sqrt{2}}{\epsilon n} \sum_{\ell=1}^n \sqrt{\mathbb{E}(D_{n\ell}^4)(\ell-1)}. \end{aligned} \quad (4.10)$$

Para acotar  $\mathbb{E}(D_{n\ell}^4)$ , notemos que

$$\begin{aligned} \mathbb{E}(D_{n\ell}^4) &= \frac{4p_n^2}{n^4} \mathbb{E}\left(\left(\sum_{i=1}^{\ell-1} \rho_{il}\right)^4\right) = \frac{4p_n^2}{n^4} \sum_{a,b,c,d=1}^{\ell-1} \mathbb{E}(\rho_{al}\rho_{bl}\rho_{cl}\rho_{dl}) \\ &= \frac{4p_n^2}{n^4} \left\{ \sum_{a=1}^{\ell-1} \mathbb{E}(\rho_{al}^4) + 3 \sum_{\substack{a,b=1 \\ a \neq b}}^{\ell-1} \mathbb{E}(\rho_{al}^2) \mathbb{E}(\rho_{bl}^2) \right\} = \frac{4p_n^2}{n^4} \{(\ell-1)\mathbb{E}(\rho_{al}^4) + 3(\ell-1)(\ell-2)\mathbb{E}(\rho_{al}^2)^2\}. \end{aligned}$$

Por otra parte, por la parte (iii) del Lema 4.2.1 tenemos que  $\mathbb{E}(\rho_{a\ell}^2) = 1/p_n$  y  $\mathbb{E}(\rho_{a\ell}^4) = 3/(p_n(p_n+2))$ . Por lo tanto,

$$\mathbb{E}(D_{n\ell}^4) = \frac{4p_n^2}{n^4} \left\{ \frac{3(\ell-1)}{p_n(p_n+2)} + \frac{3(\ell-1)(\ell-2)}{p_n^2} \right\} \leq \frac{24(\ell-1)^2}{n^4}. \quad (4.11)$$

De (4.10) y (4.11) obtenemos

$$\begin{aligned} \sum_{\ell=1}^n \mathbb{E}[D_{n\ell}^2 \mathbb{I}(|D_{n\ell}| > \epsilon)] &\leq \frac{\sqrt{2}}{\epsilon n} \sum_{\ell=1}^n \sqrt{\frac{24(\ell-1)^3}{n^4}} \\ &= \frac{\sqrt{48}}{\epsilon n^3} \sum_{\ell=1}^n (\ell-1)^{3/2}. \end{aligned}$$

Es fácil ver que  $\sum_{\ell=1}^n (\ell-1)^{3/2} = O(n^{5/2})$ , porque  $(1/n) \sum_{\ell=1}^n (\ell/n)^{3/2} \rightarrow \int_0^1 x^{3/2} dx$ . Entonces  $\sum_{\ell=1}^n \mathbb{E}[D_{n\ell}^2 \mathbb{I}(|D_{n\ell}| > \epsilon)] \in O(n^{-1/2})$  y por lo tanto tiende a cero, como queríamos ver.  $\square$

#### 4.2.2 Otros tests de esfericidad basados en signos para el contexto de alta dimensión

Existen otros tests para estudiar esfericidad que están basados en la matriz de covarianza de signos espaciales. Por ejemplo, Sirkia *et al.* (2009) proponen el estadístico

$$Q_S = p \operatorname{tr}^2 \left( \mathbf{V} - \frac{1}{p} \mathbf{I}_p \right),$$

donde  $\mathbf{V}$  es la matriz de covarianza de signos espaciales. Sirkia *et al.* (2009) muestran que si  $p$  está fijo entonces

$$Q_S \frac{n(p+2)}{2} \xrightarrow{D} \chi_{(p+2)(p-1)/2}^2.$$

Zou *et al.* (2013) extienden este resultado al contexto de alta dimensión. Intuitivamente, como  $\chi_p^2$  asintóticamente es una normal con media  $p$  y varianza  $2p$ , entonces es esperable que

$$\frac{Q_S - \mathbb{E}(Q_S)}{\sqrt{\operatorname{VAR}(Q_S)}} \xrightarrow{D} N(0, 1).$$

Se puede ver que, bajo algunas hipótesis, efectivamente vale esta convergencia y la demostración usa también el Teorema Central del Límite para Martingalas. Además, en este último trabajo proponen una modificación del estadístico  $Q_S$  para eliminar el sesgo que produce introducir un estimador de la media cuando ésta es desconocida, es decir, cuando en vez de usar los signos  $\mathbf{U}_i = s(\mathbf{X}_i)$  se usa  $\widehat{\mathbf{U}}_i = s(\mathbf{X}_i - \widehat{\boldsymbol{\mu}})$ .

## Capítulo 5

# Tests de igualdad de matrices de covarianza basados en signos

En este capítulo nos vamos a basar en el signo de las observaciones originales para proponer algunos tests para la igualdad de matrices de covarianza de signos espaciales entre dos grupos.

### 5.1 Un test basado en signos con $p$ fijo

#### 5.1.1 Algunos lemas previos y deducción del estadístico

Supongamos que tenemos dos muestras aleatorias  $\mathbf{X}_{11}, \dots, \mathbf{X}_{1n_1}$  y  $\mathbf{X}_{21}, \dots, \mathbf{X}_{2n_2}$ ,  $\mathbf{X}_{ij} \in \mathbb{R}^p$  y que  $\mathbf{X}_{ij} \sim \mathcal{E}_p(\mathbf{0}, \boldsymbol{\Sigma}_i, \phi)$ ,  $i = 1, 2$ . Si  $\mathbb{E}(\|\mathbf{X}_i\|^2) < \infty$  entonces entenderemos que  $\boldsymbol{\Sigma}_i$  es la matriz de covarianza de la población correspondiente a la muestra  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ . Indicaremos por  $\tilde{\boldsymbol{\Sigma}}_i$  a las matrices de covarianza de los signos espaciales para cada población, es decir,

$$\tilde{\boldsymbol{\Sigma}}_i = \mathbb{E} \left[ \frac{\mathbf{X}_{i1} \mathbf{X}_{i1}^T}{\|\mathbf{X}_{i1}\|^2} \right] = \mathbb{E} [s(\mathbf{X}_{i1})s(\mathbf{X}_{i1})^T]$$

donde  $s(\mathbf{X}) = \mathbf{X}/\|\mathbf{X}\|$ . Un estimador de  $\tilde{\boldsymbol{\Sigma}}_i$  está dado por

$$\mathbf{V}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{\mathbf{X}_{ij} \mathbf{X}_{ij}^T}{\|\mathbf{X}_{ij}\|^2}$$

que corresponde a la versión muestral de  $\tilde{\boldsymbol{\Sigma}}_i$ .

Las hipótesis a testear son

$$H_0 : \tilde{\boldsymbol{\Sigma}}_1 = \tilde{\boldsymbol{\Sigma}}_2 \quad vs. \quad H_1 : \tilde{\boldsymbol{\Sigma}}_1 \neq \tilde{\boldsymbol{\Sigma}}_2. \quad (5.1)$$

Si ambas poblaciones tienen segundo momento finito, entonces por el Corolario 4.2.4 las hipótesis (5.1) son equivalentes a

$$H_0 : \boldsymbol{\Sigma}_1 \propto \boldsymbol{\Sigma}_2 \quad vs. \quad H_1 : H_0 \text{ es falsa.}$$

Vamos a proponer un test para (5.1) que se basa en  $\mathbf{V}_1$  y  $\mathbf{V}_2$ , es decir, las matrices de covarianza de signos espaciales muestrales. Como estas matrices se basan en los signos de las observaciones originales, por el Lema 4.1.1 podemos suponer sin perder generalidad que  $\mathbf{X}_{ij} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ .

El siguiente lema muestra que bajo  $H_0$  los signos de ambas poblaciones son igualmente distribuidos.

**Lema 5.1.1.** *Bajo  $H_0$  vale que  $s(\mathbf{X}_{11}) \sim s(\mathbf{X}_{21})$ .*

DEMOSTRACIÓN. Supongamos sin perder generalidad que  $\mathbf{X}_{i1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ . Observemos que bajo  $H_0$ ,  $\mathbb{E}(s(\mathbf{X}_{11})s(\mathbf{X}_{11})^T) = \mathbb{E}(s(\mathbf{X}_{21})s(\mathbf{X}_{21})^T)$ , es decir, las matrices de covarianza de signos espaciales son iguales y como ya mencionamos, por el Corolario 4.2.3, las matrices  $\boldsymbol{\Sigma}_1$  y  $\boldsymbol{\Sigma}_2$  son proporcionales. Notemos que  $\mathbf{X}_{11}$  y  $\mathbf{X}_{21}$  son normales multivariadas independientes cuyas distribuciones sólo se diferencian en la escala de las matrices de covarianza. Por lo tanto, existe una constante  $a > 0$  tal que  $\mathbf{X}_{11} \sim a\mathbf{X}_{21}$ , de donde  $s(\mathbf{X}_{11}) \sim s(\mathbf{X}_{21})$  como queríamos ver.  $\square$

Teniendo todo esto en cuenta, vamos a construir un test para las hipótesis (5.1). Por el Teorema Central del Límite Multivariado, tenemos que

$$\sqrt{n_i}(\mathbf{V}_i - \tilde{\boldsymbol{\Sigma}}_i) \xrightarrow{D} N(0, \boldsymbol{\Phi}_i) \quad (5.2)$$

con  $\boldsymbol{\Phi}_i = \text{VAR}(\text{vec}(s(\mathbf{X}_{i1})s(\mathbf{X}_{i1})^T))$ , donde entendemos que para matrices aleatorias  $\mathbf{U}_n \in \mathbb{R}^{p \times p}$  la convergencia  $\mathbf{U}_n \xrightarrow{D} \mathbf{U} \sim N(0, \boldsymbol{\Phi})$  significa que  $\text{vec}(\mathbf{U}_n) \xrightarrow{D} N(0, \boldsymbol{\Phi})$ , con  $\boldsymbol{\Phi} \in \mathbb{R}^{p^2 \times p^2}$ . Es importante observar que como bajo  $H_0$  se tiene que  $s(\mathbf{X}_{11}) \sim s(\mathbf{X}_{21})$ , entonces  $\text{vec}(s(\mathbf{X}_{11})s(\mathbf{X}_{11})^T) \sim \text{vec}(s(\mathbf{X}_{21})s(\mathbf{X}_{21})^T)$  y por lo tanto, bajo  $H_0$ ,  $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_2$ .

En (5.2), la matrices  $\boldsymbol{\Phi}_i$  tienen la forma

$$\begin{aligned} \boldsymbol{\Phi}_i &= \text{VAR}(\text{vec}(s(\mathbf{X}_{i1})s(\mathbf{X}_{i1})^T)) = \text{VAR}(s(\mathbf{X}_{i1}) \otimes s(\mathbf{X}_{i1})) \\ &= \mathbb{E}(s(\mathbf{X}_{i1})s(\mathbf{X}_{i1})^T \otimes s(\mathbf{X}_{i1})s(\mathbf{X}_{i1})^T) - \text{vec}(\tilde{\boldsymbol{\Sigma}}_i)\text{vec}(\tilde{\boldsymbol{\Sigma}}_i)^T, \end{aligned} \quad (5.3)$$

ya que  $\mathbb{E}(s(\mathbf{X}_{i1}) \otimes s(\mathbf{X}_{i1})) = \mathbb{E}(\text{vec}(s(\mathbf{X}_{i1})s(\mathbf{X}_{i1})^T)) = \text{vec}(\tilde{\boldsymbol{\Sigma}}_i)$ .

Usando la independencia entre las dos muestras tenemos que, si  $n_1/N \rightarrow \tau_1$ ,  $n_2/N \rightarrow \tau_2$  con  $N = n_1 + n_2$ ,

$$\sqrt{N}(\mathbf{V}_1 - \mathbf{V}_2 - (\tilde{\boldsymbol{\Sigma}}_1 - \tilde{\boldsymbol{\Sigma}}_2)) \xrightarrow{D} N(\mathbf{0}, \frac{1}{\tau_1}\boldsymbol{\Phi}_1 + \frac{1}{\tau_2}\boldsymbol{\Phi}_2).$$

Llamaremos  $\boldsymbol{\Upsilon} = (1/\tau_1)\boldsymbol{\Phi}_1 + (1/\tau_2)\boldsymbol{\Phi}_2$ . Observemos que bajo  $H_0$ ,  $\tilde{\boldsymbol{\Sigma}}_1 = \tilde{\boldsymbol{\Sigma}}_2$  y  $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_2 = \boldsymbol{\Phi}$ , por lo tanto

$$\sqrt{n}(\mathbf{V}_1 - \mathbf{V}_2) \xrightarrow{D} N\left(\mathbf{0}, \left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right)\boldsymbol{\Phi}\right). \quad (5.4)$$

Un estadístico para testear (5.1) es el cuadrado de la norma de Frobenius de la diferencia entre las matrices de covarianza de signos espaciales muestrales, es decir,  $\text{tr}((\mathbf{V}_1 - \mathbf{V}_2)^2) = \|\text{vec}(\mathbf{V}_1 - \mathbf{V}_2)\|^2$ . Es por eso que a los tests propuestos en esta sección los llamaremos Tests de Frobenius e indicaremos a los métodos **F1**, **F2**, etc.

Por la continuidad de la norma y usando (5.4), obtenemos que bajo  $H_0$ ,

$$N\text{tr}((\mathbf{V}_1 - \mathbf{V}_2)^2) \xrightarrow{D} \text{tr}(\mathbf{U}^2) = \|\text{vec}(\mathbf{U})\|^2, \quad (5.5)$$

donde  $\text{vec}(\mathbf{U}) \sim N(\mathbf{0}, ((1/\tau_1 + 1/\tau_2)\mathbf{\Phi}))$ ,  $\mathbf{U} \in \mathbb{R}^{p \times p}$ . El estadístico del test que vamos a considerar es entonces

$$T_N = N \text{tr}((\mathbf{V}_1 - \mathbf{V}_2)^2). \quad (5.6)$$

Para obtener la distribución asintótica de  $T_N$ , usaremos el siguiente resultado.

**Lema 5.1.2.** *Sea  $\mathbf{U} \in \mathbb{R}^{p \times p}$  tal que  $\text{vec}(\mathbf{U}) \sim N(\mathbf{0}, \mathbf{\Upsilon})$ . Sean  $\theta_1 \geq \theta_2 \geq \dots \theta_{p^2}$  los autovalores de  $\mathbf{\Upsilon}$ , entonces  $\text{tr}(\mathbf{U}^2) = \|\text{vec}(\mathbf{U})\|^2 \sim \sum_{j=1}^{p^2} \theta_j \xi_j$  donde  $\xi_j \sim \chi_1^2$  independientes.*

DEMOSTRACIÓN. Tenemos que  $\mathbf{\Upsilon} \in \mathbb{R}^{p^2 \times p^2}$  admite una descomposición espectral del tipo

$$\mathbf{\Upsilon} = \sum_{j=1}^{p^2} \theta_j \mathbf{t}_j \mathbf{t}_j^T = \mathbf{T} \mathbf{\Theta} \mathbf{T}^T$$

donde  $\mathbf{T}$  es tal que  $\mathbf{T} \mathbf{T}^T = \mathbf{T}^T \mathbf{T} = \mathbf{I}$ . Luego, si  $\mathbf{u} = \text{vec}(\mathbf{U})$ , tenemos que  $\mathbf{v} = \mathbf{T}^T \mathbf{u} \sim N(\mathbf{0}, \mathbf{\Theta})$ , de donde

$$\text{tr}(\mathbf{U}^2) = \|\mathbf{u}\|^2 = \|\mathbf{v}\|^2 = \sum_{j=1}^{p^2} v_j^2 = \sum_{j=1}^{p^2} \theta_j \left( \frac{v_j}{\theta_j^{1/2}} \right)^2 = \sum_{j=1}^{p^2} \theta_j z_j^2$$

y el resultado se obtiene del hecho que  $z_j \sim N(0, 1)$  independientes.  $\square$

Como consecuencia de este Lema, tenemos el siguiente resultado.

**Lema 5.1.3.** *Sean  $\mathbf{X}_{ij} \in \mathbb{R}^p$ ,  $1 \leq j \leq n_i$ ,  $i = 1, 2$  independientes tales que  $\mathbf{X}_{ij} \sim \mathcal{E}_p(\mathbf{0}, \mathbf{\Sigma}_i, \phi)$ . Supongamos que si  $n_1/N \rightarrow \tau_1$ ,  $n_2/N \rightarrow \tau_2$  con  $N = n_1 + n_2$  y  $0 < \tau_i < 1$ ,  $i = 1, 2$ . Indiquemos por  $\tilde{\mathbf{\Sigma}}_i = \mathbb{E}[s(\mathbf{X}_{i1})s(\mathbf{X}_{i1})^T]$  donde  $s(\mathbf{X}) = \mathbf{X}/\|\mathbf{X}\|$ ,  $\mathbf{V}_i = (1/n_i) \sum_{j=1}^{n_i} s(\mathbf{X}_{ij})s(\mathbf{X}_{ij})^T$  y  $T_N = N \text{tr}((\mathbf{V}_1 - \mathbf{V}_2)^2)$ . Sean  $\theta_j$  los autovalores de  $\mathbf{\Upsilon} = (1/\tau_1)\mathbf{\Phi}_1 + (1/\tau_2)\mathbf{\Phi}_2$  donde  $\mathbf{\Phi}_i$  está definida en (5.3). Luego, bajo  $H_0 : \tilde{\mathbf{\Sigma}}_1 = \tilde{\mathbf{\Sigma}}_2$ , tenemos que  $T_N \xrightarrow{D} \sum_{j=1}^{p^2} \theta_j \xi_j$  donde  $\xi_j \sim \chi_1^2$  independientes.*

Basado en el Lema 5.1.3, podemos definir el siguiente procedimiento para testear  $H_0 : \tilde{\mathbf{\Sigma}}_1 = \tilde{\mathbf{\Sigma}}_2$ .

### Método F1

**Paso 1** Dadas las muestras  $X_{i1}, \dots, X_{in_i}$ , consideramos  $\hat{\mathbf{\Phi}}_i$  estimadores consistentes de  $\mathbf{\Phi}_i$  para  $i = 1, 2$ . Definimos  $\hat{\mathbf{\Upsilon}} = (1/\hat{\tau}_1)\hat{\mathbf{\Phi}}_1 + (1/\hat{\tau}_2)\hat{\mathbf{\Phi}}_2$  con  $\hat{\tau}_i = n_i/(n_1 + n_2)$ .

**Paso 2** Sean  $\hat{\theta}_\ell$  los autovalores de  $\hat{\mathbf{\Upsilon}}$ .

**Paso 3** Genere  $Z_1^*, \dots, Z_{p^2}^*$  i.i.d. tales que  $Z_i^* \sim N(0, 1)$  y sea  $\mathcal{T}_N^* = \sum_{j=1}^{p^2} \hat{\theta}_j Z_j^{*2}$ .

**Paso 4** Repita el **Paso 3**  $M_{\text{BOOT}}$  veces, para obtener  $M_{\text{BOOT}}$  valores  $\mathcal{T}_{N,r}^*$  for  $1 \leq r \leq M_{\text{BOOT}}$ .

El percentil  $(1 - \alpha)$  de la distribución asintótica de  $T_N$  puede aproximarse por el percentil  $(1 - \alpha)$  de la distribución empírica de  $\mathcal{T}_{N,r}^*$  para  $1 \leq r \leq M_{\text{BOOT}}$ . El  $p$ -valor se estima por  $\hat{p} = s/M_{\text{BOOT}}$  donde  $s$  es la cantidad de  $\mathcal{T}_{N,r}^*$  mayores o iguales que  $T_N$ .

Un problema con este método es que si  $p$  es grande, es computacionalmente muy costoso hallar los autovalores de  $\mathbf{\Upsilon} \in \mathbb{R}^{p^2 \times p^2}$ .

Otra forma sería aplicar un bootstrap paramétrico como sigue, usando que si las dos muestras son elípticas, podemos generar las muestras normal a partir de una normal.

### Método F2

**Paso 1** Dada la muestra  $X_{i1}, \dots, X_{in_i}$  considere estimadores  $\widehat{\Sigma}_i$  de  $\Sigma_i$  consistentes, basados en  $\mathbf{X}_{ij}$ ,  $1 \leq j \leq n_i$ .

**Paso 2** Genere para  $i = 1, 2$ ,  $X_{i,j}^*$ ,  $1 \leq j \leq n_i$ , con distribución  $N(\mathbf{0}, \widehat{\Sigma}_i)$ .

**Paso 3** Defina  $\mathbf{V}_i^*$  como el estimador de signos asociado a la muestra  $X_{i,j}^*$ ,  $1 \leq j \leq n_i$  y  $T_N^* = N \text{tr} [(\mathbf{V}_1^* - \mathbf{V}_2^*)^2]$ .

**Paso 4** Repita los Pasos 3 y 4  $M_{\text{BOOT}}$  veces para obtener  $M_{\text{BOOT}}$  valores  $T_N^*$  que permiten construir un test bootstrap.

El  $p$ -valor del test es

$$p - \text{valor} = \frac{\#\{T_N^* \geq T_N\}}{M_{\text{BOOT}}}.$$

En lugar de estos procedimientos, usaremos otra forma de obtener la distribución de  $T_N$ . Primero supondremos que bajo  $H_0$ ,  $\widetilde{\Sigma} = \widetilde{\Sigma}_1 = \widetilde{\Sigma}_2$  es una matriz diagonal que llamaremos  $\Psi = \text{DIAG}(\psi_1, \dots, \psi_p)$ . Observemos que en ese caso, las matrices  $\Sigma_1$  y  $\Sigma_2$  también son diagonales porque tienen los mismos autovectores que  $\widetilde{\Sigma}$ .

#### 5.1.2 Test para el caso $\widetilde{\Sigma} = \widetilde{\Sigma}_1 = \widetilde{\Sigma}_2$ diagonal

Consideremos el siguiente escenario: sea  $\mathbf{X} = (X_1, \dots, X_p)^T \sim N(\mathbf{0}, \mathbf{\Lambda})$  con  $\mathbf{\Lambda} = \text{DIAG}(\lambda_1, \dots, \lambda_p)$  y sea  $\widetilde{\Sigma} = (\tilde{\sigma}_{ij}) = \mathbb{E}(s(\mathbf{X})s(\mathbf{X})^T) = \Psi = \text{DIAG}(\psi_1, \dots, \psi_p)$ . Queremos una expresión para  $\Phi = \text{VAR}(\text{vec}(s(\mathbf{X})s(\mathbf{X})^T)) = \text{VAR}(s(\mathbf{X}) \otimes s(\mathbf{X})) = \mathbb{E}(s(\mathbf{X})s(\mathbf{X})^T \otimes s(\mathbf{X})s(\mathbf{X})^T) - \text{vec}(\widetilde{\Sigma})\text{vec}(\widetilde{\Sigma})^T$ .

Llamamos  $\Phi_{ij,k\ell}$  al elemento de  $\Phi$  correspondiente a la covarianza entre las coordenadas  $(i, j)$  y  $(k, \ell)$  de la matriz aleatoria  $s(\mathbf{X})s(\mathbf{X})^T$ . Tenemos que

$$\begin{aligned} \Phi_{ij,k\ell} &= \text{COV} \left( (s(\mathbf{X})s(\mathbf{X})^T)_{ij}, (s(\mathbf{X})s(\mathbf{X})^T)_{k\ell} \right) = \text{COV} \left( \frac{X_i X_j}{\|\mathbf{X}\|^2}, \frac{X_k X_\ell}{\|\mathbf{X}\|^2} \right) \\ &= \mathbb{E} \left( \frac{X_i X_j X_k X_\ell}{\|\mathbf{X}\|^4} \right) - \mathbb{E} \left( \frac{X_i X_j}{\|\mathbf{X}\|^2} \right) \mathbb{E} \left( \frac{X_k X_\ell}{\|\mathbf{X}\|^2} \right) \\ &= \mathbb{E} \left( \frac{X_i X_j X_k X_\ell}{\|\mathbf{X}\|^4} \right) - \tilde{\sigma}_{ij} \tilde{\sigma}_{k\ell} = \mathbb{E} \left( \frac{X_i X_j X_k X_\ell}{\|\mathbf{X}\|^4} \right) - \psi_j \psi_k \delta_{ij} \delta_{k\ell}. \end{aligned}$$

Observemos que las componentes de  $\mathbf{X}$  son independientes entre sí, pues su distribución es normal multivariada con matriz de covarianza diagonal. Si alguno de los cuatro índices  $\{i, j, k, \ell\}$  fuese distinto a los otros tres, digamos  $i$ , por la simetría e independencia entre componentes de  $\mathbf{X}$  tendríamos que  $(X_1, \dots, X_p)$  tiene la misma distribución que  $(X_1, \dots, X_{i-1}, -X_i, X_{i+1}, \dots, X_p)$  de donde

$$\frac{X_i X_j X_k X_\ell}{\|\mathbf{X}\|^4} \sim -\frac{X_i X_j X_k X_\ell}{\|\mathbf{X}\|^4}.$$

Por lo tanto, el primer sumando es 0. Como el segundo es cero pues  $\delta_{ij} = 0$ , tenemos que  $\Phi_{ij,kl} = 0$ . Esto prueba que los únicos elementos de  $\Phi$  no nulos son los de la forma  $\Phi_{ii,kk}$ ,  $\Phi_{ij,ij}$  y  $\Phi_{ij,ji}$ . Observemos que  $\Phi_{ij,ij} = \Phi_{ij,ji}$  por la simetría de  $s(\mathbf{X})s(\mathbf{X})^T$ . Podemos escribir entonces para  $i \neq j$  (pero  $i$  y  $k$  no necesariamente distintos)

$$\begin{aligned}\Phi_{ii,kk} &= \mathbb{E} \left( \frac{X_i^2 X_k^2}{\|\mathbf{X}\|^4} \right) - \psi_i \psi_k = \mathbb{E} (s_i^2(\mathbf{X}) s_k^2(\mathbf{X})) - \psi_i \psi_k \\ \Phi_{ij,ij} &= \mathbb{E} \left( \frac{X_i^2 X_j^2}{\|\mathbf{X}\|^4} \right) = \mathbb{E} (s_i^2(\mathbf{X}) s_j^2(\mathbf{X})),\end{aligned}\tag{5.7}$$

donde  $s(\mathbf{X}) = \mathbf{X}/\|\mathbf{X}\| = (s_1(\mathbf{X}), \dots, s_p(\mathbf{X}))^T$ . Por lo tanto, en base a  $n$  observaciones  $\mathbf{X}_1, \dots, \mathbf{X}_n$ ,  $\mathbf{X}_\ell = (X_{\ell,1}, \dots, X_{\ell,p})^T$  la matriz  $\Phi$  puede estimarse por  $\hat{\Phi}$  donde

$$\begin{aligned}\hat{\Phi}_{ii,kk} &= \frac{1}{n} \sum_{\ell=1}^n \frac{X_{\ell,i}^2 X_{\ell,k}^2}{\|\mathbf{X}_\ell\|^4} - \hat{\psi}_i \hat{\psi}_k, & \hat{\psi}_i &= \frac{1}{n} \sum_{\ell=1}^n \frac{X_{\ell,i}^2}{\|\mathbf{X}_\ell\|^2} \\ &= \frac{1}{n} \sum_{\ell=1}^n s_i^2(\mathbf{X}_\ell) s_k^2(\mathbf{X}_\ell) - \hat{\psi}_i \hat{\psi}_k, & \hat{\psi}_i &= \frac{1}{n} \sum_{\ell=1}^n s_i^2(\mathbf{X}_\ell) \\ \hat{\Phi}_{ij,ij} &= \frac{1}{n} \sum_{\ell=1}^n \frac{X_{\ell,i}^2 X_{\ell,j}^2}{\|\mathbf{X}_\ell\|^4} = \frac{1}{n} \sum_{\ell=1}^n s_i^2(\mathbf{X}_\ell) s_j^2(\mathbf{X}_\ell).\end{aligned}\tag{5.8}$$

Indicaremos con  $\Upsilon_{ij,kl}$  a los elementos de  $\Upsilon$ . Recordemos que bajo  $H_0$ ,  $\Phi_1 = \Phi_2 = \Phi$  y tomemos  $\hat{\tau}_i = n_i/N$ . Entonces, podemos estimar los elementos de  $\Upsilon$  de la siguiente forma:

$$\begin{aligned}\hat{\Upsilon}_{ij,ij} &= \left( \frac{1}{\hat{\tau}_1} + \frac{1}{\hat{\tau}_2} \right) \hat{\Phi}_{ij,ij} = \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \left\{ \sum_{\ell=1}^{n_1} s_i^2(\mathbf{X}_{1\ell}) s_j^2(\mathbf{X}_{1\ell}) + \sum_{\ell=1}^{n_2} s_i^2(\mathbf{X}_{2\ell}) s_j^2(\mathbf{X}_{2\ell}) \right\} \\ \hat{\Upsilon}_{ii,kk} &= \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \left\{ \sum_{\ell=1}^{n_1} s_i^2(\mathbf{X}_{1\ell}) s_k^2(\mathbf{X}_{1\ell}) + \sum_{\ell=1}^{n_2} s_i^2(\mathbf{X}_{2\ell}) s_k^2(\mathbf{X}_{2\ell}) \right\} - \left( \frac{1}{\hat{\tau}_1} + \frac{1}{\hat{\tau}_2} \right) \hat{\psi}_i \hat{\psi}_k,\end{aligned}\tag{5.9}$$

donde

$$\hat{\psi}_i = \frac{1}{N} \left\{ \sum_{\ell=1}^{n_1} s_i^2(\mathbf{X}_{1\ell}) + \sum_{\ell=1}^{n_2} s_i^2(\mathbf{X}_{2\ell}) \right\}.\tag{5.10}$$

Veamos ahora el siguiente resultado.

**Lema 5.1.4.** Sean  $\mathbf{X}_{ij} \in \mathbb{R}^p$ ,  $1 \leq j \leq n_i$ ,  $i = 1, 2$  independientes tales que  $\mathbf{X}_{ij} \sim \mathcal{E}_p(\mathbf{0}, \Sigma_i, \phi)$ . Supongamos que si  $n_1/N \rightarrow \tau_1$ ,  $n_2/N \rightarrow \tau_2$  con  $N = n_1 + n_2$  y  $0 < \tau_i < 1$ ,  $i = 1, 2$ . Indiquemos por  $\tilde{\Sigma}_i = \mathbb{E}[s(\mathbf{X}_{i1})s(\mathbf{X}_{i1})^T]$  donde  $s(\mathbf{X}) = \mathbf{X}/\|\mathbf{X}\|$ ,  $\mathbf{V}_i = (1/n_i) \sum_{j=1}^{n_i} s(\mathbf{X}_{ij})s(\mathbf{X}_{ij})^T$  y  $T_N = N \text{tr}((\mathbf{V}_1 - \mathbf{V}_2)^2)$ . Supongamos además que bajo  $H_0 : \tilde{\Sigma}_1 = \tilde{\Sigma}_2 = \tilde{\Sigma}$  es una matriz diagonal que llamaremos  $\Psi = \text{DIAG}(\psi_1, \dots, \psi_p)$ . Sea  $\Upsilon = (1/\tau_1)\Phi_1 + (1/\tau_2)\Phi_2$  donde  $\Phi_i$  está definida en (5.3) y llamemos  $\Gamma = (\gamma_{ij})$  con  $\gamma_{ij} = \Upsilon_{ii,jj}$ . Luego, bajo  $H_0$ ,

$$T_N \xrightarrow{D} 2 \sum_{i < j} \zeta_{ij} \xi_{ij} + \sum_{i=1}^p \Delta_i \xi_i$$

donde  $\xi_i \sim \chi_1^2$  independientes entre sí e independientes de  $\xi_{ij}$ ,  $\xi_{ij} \sim \chi_1^2$  son independientes,  $\Delta_1, \dots, \Delta_p$  son los autovalores de  $\mathbf{\Gamma}$  y  $\zeta_{ij} = \Upsilon_{ij,ij}$ .

DEMOSTRACIÓN. Hemos visto que bajo  $H_0$ ,  $\sqrt{N}(\mathbf{V}_1 - \mathbf{V}_2)$  converge en distribución a una matriz aleatoria  $\mathbf{U} = (u_{ij}) \in \mathbb{R}^{p \times p}$  simétrica tal que  $\text{vec}(\mathbf{U}) \sim N(\mathbf{0}, (1/\tau_1)\mathbf{\Phi}_1 + (1/\tau_2)\mathbf{\Phi}_2)$  donde  $\mathbf{\Phi}_i$  está definida como en (5.3). Por lo tanto, la distribución asintótica de  $T_N$  es la distribución de  $\text{tr}(\mathbf{U}^2)$ . Recordemos que los elementos fuera de la diagonal del triángulo superior de  $\mathbf{U}$  tienen covarianza 0 con cualquiera de los otros  $p(p+1)/2 - 1$  y por lo tanto, son independientes. En cambio, no podemos afirmar que los elementos de la diagonal sean independientes entre sí. Como

$$\text{tr}(\mathbf{U}^2) = \sum_{i,j=1}^p u_{ij}^2 = 2 \sum_{j>i} u_{ij}^2 + \sum_{i=1}^p u_{ii}^2$$

y

$$u_{ij}^2 = \left( \frac{u_{ij}}{\sqrt{\Upsilon_{ij,ij}}} \right)^2 \Upsilon_{ij,ij} \sim \chi_1^2 \Upsilon_{ij,ij}$$

tenemos que

$$\text{tr}(\mathbf{U}^2) \sim 2 \sum_{i<j} \zeta_{ij} \xi_{ij} + \|\mathbf{Z}\|^2 \quad (5.11)$$

donde  $\zeta_{ij} = \Upsilon_{ij,ij}$ ,  $\xi_{ij} \sim \chi_1^2$  independientes e independientes de  $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{\Gamma})$ . Luego como  $\|\mathbf{Z}\|^2 = \|\beta \mathbf{Z}\|^2$  con  $\beta$  los autovectores de  $\mathbf{\Gamma} = \beta^T \mathbf{\Delta} \beta$ , tenemos que

$$\text{tr}(\mathbf{U}^2) \sim 2 \sum_{i<j} \zeta_{ij} \xi_{ij} + \sum_{i=1}^p \Delta_i \xi_i \quad (5.12)$$

donde  $\xi_i \sim \chi_1^2$  independientes entre sí e independientes de las  $\xi_{ij}$ .  $\square$

Notemos que pudimos expresar a la distribución asintótica de  $T_N$  como una combinación lineal de  $p(p+1)/2$  variables  $\chi_1^2$ . Esto no se contradice con lo que obtuvimos en el Lema 5.1.3, porque muchos de los autovalores de la matriz  $\mathbf{\Upsilon}$  son cero debido a que la matriz aleatoria  $\mathbf{U}$  de (5.5) es normal singular. Además, tenemos que  $\sum_{k=1}^p \Phi_{ii,kk} = 0$ , para todo  $1 \leq i \leq p$  pues  $\sum_{k=1}^p s_k^2(\mathbf{X}) = 1$  y por lo tanto,  $\mathbf{\Gamma}$  también es singular. Como computacionalmente es complicado generar una normal singular, para estimar los percentiles de la distribución límite de  $T_N$  es mejor basarse en la expresión (5.12) que en (5.11).

Proponemos entonces implementar el test de la siguiente forma

### Método F3

**Paso 1** Dada la muestra  $X_{i1}, \dots, X_{in_i}$ , obtenga el estadístico  $T_N$ .

**Paso 2** Obtenga  $\Upsilon_{ii,jj}$  y  $\Upsilon_{ij,ij}$  como en (5.9).

**Paso 3** Sean  $\hat{\gamma}_{ij} = \hat{\Upsilon}_{ii,jj}$ ,  $1 \leq i, j \leq p$ ,  $\hat{\mathbf{\Gamma}} = (\hat{\gamma}_{ij})$ ,  $\hat{\Delta}_1, \dots, \hat{\Delta}_p$  los autovalores de  $\hat{\mathbf{\Gamma}}$  y  $\hat{\zeta}_{ij} = \hat{\Upsilon}_{ij,ij}$ ,  $1 \leq i < j \leq p$ . Por simplicidad, ordene a  $\hat{\zeta}_{ij}$  como  $\hat{\zeta}_1, \dots, \hat{\zeta}_q$  donde  $q = p(p-1)/2$  y sean  $\hat{\Delta}_1 \geq \dots \geq \hat{\Delta}_r > 0$  los autovalores distintos de cero.

**Paso 4** Genere  $W_1^*, \dots, W_q^*$  y  $Y_1^*, \dots, Y_r^*$  todas independientes tales que  $W_i^* \sim N(0, 1)$ ,  $Y_i^* \sim N(0, 1)$  donde  $q = p(p-1)/2$ . Defina  $\mathcal{T}_N^* = \sum_{i<j} \hat{\zeta}_j W_j^{*2} + \sum_{i=1}^r \hat{\Delta}_i Y_i^{*2}$ .

**Paso 5** Repita el **Paso 4**  $M_{\text{BOOT}}$  veces, para obtener  $M_{\text{BOOT}}$  valores  $\mathcal{T}_{N,s}^*$  for  $1 \leq s \leq M_{\text{BOOT}}$ .

El  $p$ -valor se estima por  $\hat{p} = s/M_{\text{BOOT}}$  donde  $s$  es la cantidad de  $\mathcal{T}_{N,r}^*$  mayores o iguales que  $T_N$ .

### 5.1.3 Test para el caso $\tilde{\Sigma} = \tilde{\Sigma}_1 = \tilde{\Sigma}_2$ no diagonal

El **Método F3** sólo puede implementarse cuando bajo  $H_0$ ,  $\tilde{\Sigma} = \tilde{\Sigma}_1 = \tilde{\Sigma}_2$  es una matriz diagonal que llamamos  $\Psi = \text{DIAG}(\psi_1, \dots, \psi_p)$ . Hemos visto en el Lema 4.1.1 que, para hallar la distribución asintótica de  $\mathbf{V}_i$  y luego la de  $T_N$ , podemos suponer sin pérdida de generalidad que  $\mathbf{X}_{ij} \sim N(\mathbf{0}, \Sigma_i)$ .

Sea  $\mathbf{X}_i$  tal que  $\mathbf{X}_i \sim N(\mathbf{0}, \Sigma_i)$  y  $\beta_i$  ortogonal tal que  $\Sigma_i = \beta_i^T \Lambda_i \beta_i$  con  $\Lambda_i = \text{DIAG}(\lambda_{i1}, \dots, \lambda_{ip})$ . Luego, si definimos  $\mathbf{Y}_i = \beta_i \mathbf{X}_i$  tenemos que  $\mathbf{Y}_i \sim N(\mathbf{0}, \Lambda_i)$ . Para reforzar la dependencia de la muestra considerada, indiquemos por  $V_i(\mathbf{X}) = (1/n_i) \sum_{j=1}^{n_i} s(\mathbf{X}_{ij})s(\mathbf{X}_{ij})^T$ ,  $T_N(\mathbf{X}) = N\text{tr}((\mathbf{V}_1(\mathbf{X}) - \mathbf{V}_2(\mathbf{X}))^2)$ ,

$$\begin{aligned}\Phi_i(\mathbf{X}) &= \text{VAR}(\text{vec}(s(\mathbf{X}_{i1})s(\mathbf{X}_{i1})^T)) \\ \Upsilon(\mathbf{X}) &= \frac{1}{\tau_1} \Phi_1(\mathbf{X}) + \frac{1}{\tau_2} \Phi_2(\mathbf{X}),\end{aligned}$$

y en forma análoga,  $V_i(\mathbf{Y})$ ,  $T_N(\mathbf{Y})$ ,  $\Phi_i(\mathbf{Y})$  y  $\Upsilon(\mathbf{Y})$ .

Por la invarianza de  $T_N$  bajo transformaciones ortogonales, tenemos que  $T_N(\mathbf{Y}) = T_N(\mathbf{X})$ . Luego, bajo  $H_0$ , la distribución asintótica de  $T_N(\mathbf{X})$  es la de  $T_N(\mathbf{Y})$  que por el Lema 5.1.4 es la distribución de

$$S = 2 \sum_{i < j} \zeta_{ij}(\mathbf{Y}) \xi_{ij} + \sum_{i=1}^p \Delta_i(\mathbf{Y}) \xi_i$$

donde  $\xi_1, \dots, \xi_p$  son i.i.d  $\xi_j \sim \chi_1^2$ ,  $\{\xi_j\}$  y  $\{\xi_{ij}\}_{i,j}$  son independientes y  $\xi_{ij} \sim \chi_1^2$  son independientes. Además  $\Delta_1(\mathbf{Y}), \dots, \Delta_p(\mathbf{Y})$  son los autovalores de  $\Gamma(\mathbf{Y}) = (\Upsilon_{ii,jj}(\mathbf{Y}))_{1 \leq i,j \leq p}$  y  $\zeta_{ij}(\mathbf{Y}) = \Upsilon_{ij,ij}(\mathbf{Y})$ . Por lo tanto, necesitamos tener una manera de estimar los elementos  $\Upsilon_{ii,jj}(\mathbf{Y})$  y  $\Upsilon_{ij,ij}(\mathbf{Y})$  a partir de las muestras  $\mathbf{X}_{ij} \in \mathbb{R}^p$ ,  $1 \leq j \leq n_i$ ,  $i = 1, 2$ . Mostraremos dos maneras de hacerlo.

Para la primer forma, observemos que  $\Upsilon(\mathbf{Y}) = (1/\tau_1) \Phi_1(\mathbf{Y}) + (1/\tau_2) \Phi_2(\mathbf{Y})$ , luego basta encontrar una expresión para  $\Phi_1(\mathbf{Y})$  que involucre a  $\mathbf{X}$ . Observemos que como  $\mathbf{Y}_i = \beta_i \mathbf{X}_i$  y  $\beta_i$  es ortogonal, entonces  $s(\mathbf{Y}_i) = \beta_i s(\mathbf{X}_i)$ . Por lo tanto,  $s(\mathbf{Y}_i)s(\mathbf{Y}_i)^T = \beta_i s(\mathbf{X}_i) s(\mathbf{X}_i) \beta_i^T$  de donde  $\text{vec}(s(\mathbf{Y}_{i1})s(\mathbf{Y}_{i1})^T) = (\beta_i \otimes \beta_i) \text{vec}(s(\mathbf{X}_{i1})s(\mathbf{X}_{i1})^T)$ , lo que implica que

$$\Phi_i(\mathbf{Y}) = (\beta_i \otimes \beta_i) \Phi_i(\mathbf{X}) (\beta_i^T \otimes \beta_i^T).$$

Bajo  $H_0$ ,  $\beta_1 = \beta_2 = \beta$ , luego podemos encontrar un estimador de  $\beta$ ,  $\hat{\beta}$ , usando por ejemplo los autovectores de

$$\mathbf{V} = \hat{\tau}_1 \mathbf{V}_1 + \hat{\tau}_2 \mathbf{V}_2. \quad (5.13)$$

Por lo tanto, un estimador de  $\Upsilon(\mathbf{Y})$  está dado por

$$\hat{\Upsilon}(\mathbf{Y}) = (\hat{\beta} \otimes \hat{\beta}) \hat{\Upsilon}(\mathbf{X}) (\hat{\beta}^T \otimes \hat{\beta}^T), \quad (5.14)$$

donde para  $\hat{\tau}_i = n_i/(n_1 + n_2)$

$$\hat{\mathbf{Y}}(\mathbf{X}) = \left( \frac{1}{\hat{\tau}_1} + \frac{1}{\hat{\tau}_2} \right) \hat{\Phi} \quad (5.15)$$

$$\begin{aligned} \hat{\Phi}_{ij,k\ell} &= \frac{1}{N} \sum_{r=1}^2 \sum_{m=1}^{n_r} s_i(\mathbf{X}_{rm}) s_j(\mathbf{X}_{rm}) s_k(\mathbf{X}_{rm}) s_\ell(\mathbf{X}_{rm}) \\ &\quad - \left( \frac{1}{N} \sum_{r=1}^2 \sum_{m=1}^{n_r} s_i(\mathbf{X}_{rm}) s_j(\mathbf{X}_{rm}) \right) \left( \frac{1}{N} \sum_{r=1}^2 \sum_{m=1}^{n_r} s_k(\mathbf{X}_{rm}) s_\ell(\mathbf{X}_{rm}) \right). \end{aligned} \quad (5.16)$$

Observemos que solo necesitamos calcular los elementos  $\hat{\Upsilon}_{ij,ij}(\mathbf{Y})$  y  $\hat{\Upsilon}_{ii,kk}(\mathbf{Y})$ . Por lo tanto, en el caso general, tenemos el siguiente procedimiento

#### Método F4

**Paso 1** Dadas las muestras  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ , obtener el estadístico  $T_N$  y  $\hat{\beta}$  como los autovectores de  $\hat{\tau}_1 \mathbf{V}_1 + \hat{\tau}_2 \mathbf{V}_2$ .

**Paso 2** Sean  $\hat{\Upsilon}(\mathbf{Y})$  y  $\hat{\Upsilon}(\mathbf{X})$  definidos en (5.14) y (5.15), respectivamente.

**Paso 3** Sean  $\hat{\gamma}_{ij} = \hat{\Upsilon}_{ii,jj}(\mathbf{Y})$ ,  $1 \leq i, j \leq p$ ,  $\hat{\Gamma} = (\hat{\gamma}_{ij})$ ,  $\hat{\Delta}_1, \dots, \hat{\Delta}_p$  los autovalores de  $\hat{\Gamma}$  y  $\hat{\zeta}_{ij} = \hat{\Upsilon}_{ij,ij}(\mathbf{Y})$ ,  $1 \leq i < j \leq p$ . Por simplicidad, ordene a  $\hat{\zeta}_{ij}$  como  $\hat{\zeta}_1, \dots, \hat{\zeta}_q$  donde  $q = p(p-1)/2$  y sean  $\hat{\Delta}_1 \geq \dots \geq \hat{\Delta}_r > 0$  los autovalores distintos de cero.

**Paso 4** Genere  $W_1^*, \dots, W_q^*$  y  $Y_1^*, \dots, Y_r^*$  todas independientes tales que  $W_i^* \sim N(0, 1)$ ,  $Y_i^* \sim N(0, 1)$  donde  $q = p(p-1)/2$ . Defina  $\mathcal{T}_N^* = \sum_{i < j} \hat{\zeta}_j W_j^{*2} + \sum_{i=1}^r \hat{\Delta}_i Y_i^{*2}$ .

**Paso 5** Repita el **Paso 4**  $M_{\text{BOOT}}$  veces, para obtener  $M_{\text{BOOT}}$  valores  $\mathcal{T}_{N,s}^*$  for  $1 \leq s \leq M_{\text{BOOT}}$ .

El  $p$ -valor se estima por  $\hat{p} = s/M_{\text{BOOT}}$  donde  $s$  es la cantidad de  $\mathcal{T}_{N,r}^*$  mayores o iguales que  $T_N$ .

Veamos ahora un último método que involucra otra forma de estimar a los elementos  $\Upsilon_{ii,jj}(\mathbf{Y})$  y  $\Upsilon_{ij,ij}(\mathbf{Y})$ . Recordemos que bajo  $H_0$ ,  $\text{VAR}(\mathbf{Y}_i) = \mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \mathbf{\Lambda}$  y esta matriz es diagonal. Usando lo visto en la Sección 5.1.2, sabemos que los únicos elementos de  $\Upsilon(\mathbf{Y})$  que no son necesariamente cero son los de la forma  $\Upsilon_{ii,jj}(\mathbf{Y})$ ,  $\Upsilon_{ij,ij}(\mathbf{Y})$  y  $\Upsilon_{ij,ji}(\mathbf{Y})$ . En el **Método F3** hicimos uso de este hecho debido a que sólomente estimamos los elementos de  $\Upsilon$  con esta forma en vez de estimar la matriz  $\Upsilon$  completa.

Recordemos que bajo  $H_0$ ,  $\beta_1 = \beta_2 = \beta$ . Consideremos un estimador consistente  $\hat{\beta}$  de  $\beta$  y definamos  $\hat{\mathbf{Y}}_{ij} = \hat{\beta} \mathbf{X}_{ij}$ ,  $i = 1, 2$ ,  $j = 1, \dots, n_i$ . Luego,  $\hat{\mathbf{Y}}_{ij} \xrightarrow{p} \mathbf{Y}_{ij} = \beta \mathbf{X}_{ij} \sim N(\mathbf{0}, \mathbf{\Lambda})$ . Por lo tanto, proponemos el siguiente método.

#### Método F5

**Paso 1** Dadas las muestras  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ , defina  $\mathbf{V}$  como en (5.13) y sea  $\hat{\beta}$  la matriz de autovectores de  $\mathbf{V}$ , tal que  $\mathbf{V} = \hat{\beta}^T \hat{\mathbf{\Lambda}} \hat{\beta}$  con  $\hat{\mathbf{\Lambda}}$  diagonal.

**Paso 2** Llamamos  $\hat{\mathbf{Y}}_{ij} = \hat{\beta} \mathbf{X}_{ij}$ .

**Paso 3** Aplicar el **Método F3** usando las muestras  $\hat{\mathbf{Y}}_{i1}, \dots, \hat{\mathbf{Y}}_{in_i}$ .

## 5.2 Una heurística para el caso $(n, p)$ –asintótico

En el capítulo anterior vimos el ejemplo de test basado en signos para testear esfericidad y demostramos que distribución  $(n, p)$ –asintótica del estadístico es  $N(0, 1)$ . Además este estadístico se obtuvo estandarizando al que se usaba en el caso clásico  $n$ –asintótico. Ahora, con esta misma idea, proponemos un test para testear igualdad de matrices de covarianza para dos grupos en el contexto de alta dimensión. Daremos primero una propuesta para el caso en que bajo  $H_0$  la matriz  $\tilde{\Sigma} = \tilde{\Sigma}_1 = \tilde{\Sigma}_2$  es diagonal y haremos otra para cuando dicha matriz no sea diagonal.

### 5.2.1 Test para el caso $\tilde{\Sigma} = \tilde{\Sigma}_1 = \tilde{\Sigma}_2$ diagonal

Supongamos primero, como en la sección anterior, que las matrices  $\tilde{\Sigma}_1$  y  $\tilde{\Sigma}_2$  son diagonales. Veamos en este caso cuáles son la esperanza y la varianza asintóticas del estadístico  $T_N$  de la sección anterior.

**Lema 5.2.1** *Sea  $T_N$  el estadístico definido en (5.6). Entonces, bajo  $H_0 : \tilde{\Sigma}_1 = \tilde{\Sigma}_2 = \tilde{\Sigma}$  y suponiendo que  $\tilde{\Sigma} = \text{DIAG}(\psi_1, \dots, \psi_p)$ , se tiene que*

$$\begin{aligned} \mathbb{E}(T_N) &\longrightarrow \text{tr}(\Upsilon) = \left( \frac{1}{\tau_1} + \frac{1}{\tau_2} \right) \left( 1 - \sum_{i=1}^p \psi_i^2 \right) =: E \\ \text{VAR}(T_N) &\longrightarrow 2 \left( 4 \left( \sum_{i>j} \Upsilon_{ij,ij}^2 \right) + \sum_{i=1}^p \Delta_i^2 \right) =: V, \end{aligned}$$

donde  $\Upsilon = (1/\tau_1)\Phi_1 + (1/\tau_2)\Phi_2$  con  $\Phi_1$  y  $\Phi_2$  definidas en (5.3) y  $\Delta_i$ ,  $i = 1, \dots, p$  los autovalores de la matriz  $\Gamma = (\gamma_{ij})$  donde  $\gamma_{ij} = \Upsilon_{ii,jj}$ .

DEMOSTRACIÓN. Por la observación hecha en (5.12) tenemos que

$$\text{tr}(\mathbf{U}^2) \sim 2 \sum_{i>j} \Upsilon_{ij,ij} \xi_{ij} + \sum_{i=1}^p \Delta_i \xi_i$$

donde  $\xi_{ij}$  y  $\xi_i$  son variables aleatorias i.i.d. distribuídas como una  $\chi_1^2$ . Entonces, tenemos que

$$\text{VAR}(\text{tr}(\mathbf{U}^2)) = 4 \text{VAR} \left( \sum_{i>j} \Upsilon_{ij,ij} \xi_{ij} \right) + \text{VAR} \left( \sum_{i=1}^p \Delta_i \xi_i \right) = 4 \left( \sum_{i>j} 2\Upsilon_{ij,ij}^2 \right) + \sum_{i=1}^p 2\Delta_i^2.$$

Como  $T_N \xrightarrow{D} \text{tr}(\mathbf{U}^2)$ , la segunda afirmación queda demostrada. Para ver la primera, notemos que

$$\begin{aligned} \mathbb{E}(\text{tr}(\mathbf{U}^2)) &= 2 \mathbb{E} \left( \sum_{i>j} \Upsilon_{ij,ij} \xi_{ij} \right) + \mathbb{E} \left( \sum_{i=1}^p \Delta_i \xi_i \right) = 2 \sum_{i>j} \Upsilon_{ij,ij} + \sum_{i=1}^p \Delta_i = 2 \sum_{i>j} \Upsilon_{ij,ij} + \text{tr}(\Gamma) \\ &= 2 \sum_{i>j} \Upsilon_{ij,ij} + \sum_{i=1}^p \Upsilon_{ii,ii} = \sum_{i,j=1}^p \Upsilon_{ij,ij} = \text{tr}(\Upsilon). \end{aligned}$$

Además, usando las expresiones obtenidas en (5.7) y usando que  $\tilde{\Sigma}_1 = \tilde{\Sigma}_2 = \text{DIAG}(\psi_1, \dots, \psi_p)$ , tenemos que

$$\begin{aligned} \text{tr}(\mathbf{Y}) &= \frac{1}{\tau_1} \sum_{i,j=1}^p (\Phi_1)_{ij,ij} + \frac{1}{\tau_2} \sum_{i,j=1}^p (\Phi_2)_{ij,ij} \\ &= \frac{1}{\tau_1} \left( \sum_{i,j=1}^p \mathbb{E}(s_i^2(\mathbf{X}_{11})s_j^2(\mathbf{X}_{11})) - \sum_{i=1}^p \psi_i^2 \right) + \frac{1}{\tau_2} \left( \sum_{i,j=1}^p \mathbb{E}(s_i^2(\mathbf{X}_{21})s_j^2(\mathbf{X}_{21})) - \sum_{i=1}^p \psi_i^2 \right) \\ &= \frac{1}{\tau_1} \left( 1 - \sum_{i=1}^p \psi_i^2 \right) + \frac{1}{\tau_2} \left( 1 - \sum_{i=1}^p \psi_i^2 \right) = \left( \frac{1}{\tau_1} + \frac{1}{\tau_2} \right) \left( 1 - \sum_{i=1}^p \psi_i^2 \right), \end{aligned}$$

como queríamos ver.  $\square$

Podemos estimar a la esperanza y varianza asintótica de  $T_N$  (definidas como  $E$  y  $V$  respectivamente) de la siguiente forma:

$$\begin{aligned} \hat{E} &= \left( \frac{1}{\hat{\tau}_1} + \frac{1}{\hat{\tau}_2} \right) \left( 1 - \sum_{i=1}^p \hat{\psi}_i^2 \right) \\ \hat{V} &= 2 \left( 4 \left( \sum_{i>j} \hat{\Upsilon}_{ij,ij}^2 \right) + \sum_{i=1}^p \hat{\Delta}_i^2 \right) \end{aligned} \tag{5.17}$$

donde  $\hat{\tau}_i = n_i/N$ ,  $\hat{\Upsilon}_{ij,ij}$  y  $\hat{\Upsilon}_{ii,kk}$  definidos en (5.9),  $\hat{\psi}_i$  definido en (5.10) y  $\hat{\Delta}_i$  los autovalores de la matriz  $\hat{\Gamma} = (\hat{\gamma}_{ij})$  con  $\hat{\gamma}_{ij} = \hat{\Upsilon}_{ii,jj}$ .

De esta forma, para el caso con  $\tilde{\Sigma}_1$  y  $\tilde{\Sigma}_2$  diagonales proponemos el estadístico

$$D_{Np} = \frac{T_N - \hat{E}}{\sqrt{\hat{V}}} \tag{5.18}$$

y por lo tanto tenemos el siguiente método.

### Método H1

**Paso 1** Dadas las muestras  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ , calcule  $\hat{E}$  y  $\hat{V}$  como en (5.17).

**Paso 2** Calcule el estadístico  $D_{Np}$  definido en (5.18).

El  $p$ -valor se estima como el percentil de una  $N(0, 1)$  correspondiente al valor que toma  $D_{Np}$ .

En este caso, el estadístico  $D_{Np}$  no es invariante por rotaciones. Por eso, para el caso en que  $\tilde{\Sigma}_1$  y  $\tilde{\Sigma}_2$  no sean necesariamente diagonales proponemos otro método.

### 5.2.2 Test para el caso $\tilde{\Sigma} = \tilde{\Sigma}_1 = \tilde{\Sigma}_2$ no diagonal

En primer lugar, siguiendo la idea del **Método F5** de la Subsección 5.1.3, proponemos el siguiente método.

## Método H2

**Paso 1** Dadas las muestras  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ , defina  $\mathbf{V}$  como en (5.13) y sea  $\hat{\boldsymbol{\beta}}$  la matriz de autovectores de  $\mathbf{V}$ , tal que  $\mathbf{V} = \hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{\beta}}$  con  $\hat{\boldsymbol{\Lambda}}$  diagonal.

**Paso 2** Llamamos  $\hat{\mathbf{Y}}_{ij} = \hat{\boldsymbol{\beta}} \mathbf{X}_{ij}$ .

**Paso 3** Aplicar el **Método H1** usando las muestras  $\hat{\mathbf{Y}}_{i1}, \dots, \hat{\mathbf{Y}}_{in_i}$ .

Veamos ahora un último método. Para eso, notemos que por el Lema 5.1.3, tenemos que  $T_N \xrightarrow{D} \sum_{i=1}^{p^2} \theta_i \xi_i$  con  $\xi_i \sim \chi_1^2$  independientes, donde  $\theta_1, \dots, \theta_{p^2}$  son los autovalores de  $\boldsymbol{\Upsilon}$ . Entonces, como en el Lema 5.2.1, es fácil ver que  $\mathbb{E}(T_N) \rightarrow \sum_{i=1}^{p^2} \theta_i = \text{tr}(\boldsymbol{\Upsilon})$  y  $\text{VAR}(T_N) \rightarrow 2 \sum_{i=1}^{p^2} \theta_i^2 = 2 \text{tr}(\boldsymbol{\Upsilon}^2)$ . Luego,

$$\tilde{D}_{Np} = \frac{T_N - \text{tr}(\hat{\boldsymbol{\Upsilon}})}{\sqrt{2 \text{tr}(\hat{\boldsymbol{\Upsilon}}^2)}}, \quad (5.19)$$

donde  $\hat{\boldsymbol{\Upsilon}}$  es un estimador de  $\boldsymbol{\Upsilon}$ , provee un estadístico de interés en el caso  $(n, p)$ -asintótico. Un posible estimador, basado en la expresión (5.3) sería

$$\begin{aligned} \hat{\boldsymbol{\Upsilon}} &= \left( \frac{1}{\hat{\tau}_1} + \frac{1}{\hat{\tau}_2} \right) \hat{\boldsymbol{\Phi}} \\ &= \left( \frac{1}{\hat{\tau}_1} + \frac{1}{\hat{\tau}_2} \right) \frac{1}{N} \left\{ \sum_{j=1}^{n_1} s(\mathbf{X}_{1j}) s(\mathbf{X}_{1j})^T \otimes s(\mathbf{X}_{1j}) s(\mathbf{X}_{1j})^T + \sum_{j=1}^{n_2} s(\mathbf{X}_{2j}) s(\mathbf{X}_{2j})^T \otimes s(\mathbf{X}_{2j}) s(\mathbf{X}_{2j})^T \right\} \\ &\quad - \left( \frac{1}{\hat{\tau}_1} + \frac{1}{\hat{\tau}_2} \right) \text{vec}(\mathbf{V}) \text{vec}(\mathbf{V})^T, \end{aligned} \quad (5.20)$$

donde  $\mathbf{V}$  es un estimador de  $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Sigma}}_1 = \tilde{\boldsymbol{\Sigma}}_2$ , por ejemplo,

$$\mathbf{V} = \frac{1}{N} \left\{ \sum_{j=1}^{n_1} s(\mathbf{X}_{1j}) s(\mathbf{X}_{1j})^T + \sum_{j=1}^{n_2} s(\mathbf{X}_{2j}) s(\mathbf{X}_{2j})^T \right\}. \quad (5.21)$$

La siguiente proposición muestra una forma alternativa de calcular  $\text{tr}(\hat{\boldsymbol{\Upsilon}})$  y  $\text{tr}(\hat{\boldsymbol{\Upsilon}}^2)$ .

**Proposición 5.2.2.** *Sea  $\hat{\boldsymbol{\Upsilon}}$  definida en (5.20). Para simplificar notación renombramos a los vectores aleatorios  $s(\mathbf{X}_{11}), \dots, s(\mathbf{X}_{1n_1})$  y  $s(\mathbf{X}_{21}), \dots, s(\mathbf{X}_{2n_2})$  como  $\mathbf{Y}_1, \dots, \mathbf{Y}_N$  respectivamente, donde  $N = n_1 + n_2$ . Entonces,*

(a) *Sea  $\mathbf{V} = (v_{ij})$  definida en (5.21). Entonces,*

$$\text{tr}(\hat{\boldsymbol{\Upsilon}}) = \left( \frac{1}{\hat{\tau}_1} + \frac{1}{\hat{\tau}_2} \right) \left( 1 - \sum_{i,j=1}^p v_{ij}^2 \right).$$

(b) Indicamos como  $Y_{qi}$  al  $i$ -ésimo elemento de  $\mathbf{Y}_q$ . Llamamos  $\alpha(\mathbf{Y}_q, \mathbf{Y}_t) = \sum_{i=1}^p Y_{qi}Y_{ti}$ , es decir, al producto escalar entre  $\mathbf{Y}_q$  y  $\mathbf{Y}_t$ . Entonces,

$$\text{tr}(\widehat{\mathbf{Y}}^2) = \left( \frac{1}{\widehat{\tau}_1} + \frac{1}{\widehat{\tau}_2} \right)^2 \left\{ \frac{1}{N^2} \sum_{t,q=1}^N \alpha^4(\mathbf{Y}_q, \mathbf{Y}_t) - \frac{2}{N^3} \sum_{q=1}^N \left[ \sum_{t=1}^N \alpha^2(\mathbf{Y}_q, \mathbf{Y}_t) \right]^2 + \frac{1}{N^4} \left[ \sum_{t,q=1}^N \alpha^2(\mathbf{Y}_q, \mathbf{Y}_t) \right]^2 \right\}.$$

DEMOSTRACIÓN. Para demostrar (a), observemos que

$$\text{tr}(\widehat{\mathbf{Y}}) = \left( \frac{1}{\widehat{\tau}_1} + \frac{1}{\widehat{\tau}_2} \right) \left( \frac{1}{N} \sum_{q=1}^N \text{tr}(\mathbf{Y}_q \mathbf{Y}_q^T \otimes \mathbf{Y}_q \mathbf{Y}_q^T) - \text{tr}(\text{vec}(\mathbf{V}) \text{vec}(\mathbf{V})^T) \right).$$

Si fijamos un  $1 \leq q \leq N$ , entonces es fácil ver que

$$\text{tr}(\mathbf{Y}_q \mathbf{Y}_q^T \otimes \mathbf{Y}_q \mathbf{Y}_q^T) = \text{tr}^2(\mathbf{Y}_q \mathbf{Y}_q^T) = \left[ \sum_{i=1}^p Y_{qi}^2 \right]^2 = 1.$$

Por lo tanto,  $\text{tr}(\widehat{\mathbf{Y}}) = (1/\widehat{\tau}_1 + 1/\widehat{\tau}_2) (1 - \text{tr}(\text{vec}(\mathbf{V}) \text{vec}(\mathbf{V})^T))$  y el resultado se obtiene al notar que  $\text{tr}(\text{vec}(\mathbf{V}) \text{vec}(\mathbf{V})^T) = \sum_{i,j=1}^p v_{ij}^2$ .

Probemos ahora la parte (b). Llamemos  $c = 1/\widehat{\tau}_1 + 1/\widehat{\tau}_2$ . Notemos que

$$\frac{1}{c} \widehat{\Upsilon}_{ij,kl} = \frac{1}{N} \sum_{q=1}^N Y_{qi} Y_{qj} Y_{qk} Y_{ql} - \frac{1}{N^2} \left( \sum_{q=1}^N Y_{qi} Y_{qj} \right) \left( \sum_{q=1}^N Y_{qk} Y_{ql} \right).$$

Entonces,

$$\frac{1}{c^2} \widehat{\Upsilon}_{ij,kl}^2 = \rho_{1,ijkl} - 2\rho_{2,ijkl} + \rho_{3,ijkl} \quad (5.22)$$

donde

$$\begin{aligned} \rho_{1,ijkl} &= \frac{1}{N^2} \sum_{q,t=1}^N Y_{qi} Y_{qj} Y_{qk} Y_{ql} Y_{ti} Y_{tj} Y_{tk} Y_{tl}, \\ \rho_{2,ijkl} &= \frac{1}{N^3} \left( \sum_{q=1}^N Y_{qi} Y_{qj} Y_{qk} Y_{ql} \right) \left( \sum_{q=1}^N Y_{qi} Y_{qj} \right) \left( \sum_{q=1}^N Y_{qk} Y_{ql} \right), \\ \rho_{3,ijkl} &= \frac{1}{N^4} \left( \sum_{q,t=1}^N Y_{qi} Y_{qj} Y_{ti} Y_{tj} \right) \left( \sum_{q,t=1}^N Y_{qk} Y_{ql} Y_{tk} Y_{tl} \right). \end{aligned}$$

Por otra parte,  $\text{tr}(\widehat{\mathbf{Y}}^2) = \sum_{i,j,k,\ell=1}^p \widehat{\mathbf{Y}}_{ij,kl}^2$  y cálculos inmediatos muestran que

$$\begin{aligned} \sum_{i,j,k,\ell}^p \rho_{1,ijkl} &= \frac{1}{N^2} \sum_{q,t=1}^N \alpha^4(\mathbf{Y}_q, \mathbf{Y}_t), \\ \sum_{i,j,k,\ell}^p \rho_{2,ijkl} &= \frac{1}{N^3} \sum_{q,t,r=1}^N \alpha^2(\mathbf{Y}_q, \mathbf{Y}_t) \alpha^2(\mathbf{Y}_q, \mathbf{Y}_r) = \frac{1}{N^3} \sum_{q=1}^N \left[ \sum_{t=1}^N \alpha^2(\mathbf{Y}_q, \mathbf{Y}_t) \right]^2, \\ \sum_{i,j,k,\ell}^p \rho_{3,ijkl} &= \frac{1}{N^4} \sum_{q,t,r,v=1}^N \alpha^2(\mathbf{Y}_q, \mathbf{Y}_t) \alpha^2(\mathbf{Y}_r, \mathbf{Y}_v) = \frac{1}{N^4} \left[ \sum_{t,q=1}^N \alpha^2(\mathbf{Y}_q, \mathbf{Y}_t) \right]^2. \end{aligned}$$

Combinando esto con (5.22), se obtiene el resultado deseado.  $\square$

La Proposición 5.2.2 muestra que para el cálculo de  $\text{tr}(\widehat{\mathbf{Y}})$  y  $\text{tr}(\widehat{\mathbf{Y}}^2)$  alcanza con tener un estimador  $\mathbf{V}$  de  $\widetilde{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{\Sigma}}_1 = \widetilde{\boldsymbol{\Sigma}}_2$  y calcular los  $(n_1 + n_2)(n_1 + n_2 + 1)/2$  productos escalares entre cada par de vectores en el conjunto  $\{\mathbf{X}_{ij}, i = 1, 2, j = 1, \dots, n_i\}$ . Computacionalmente, esto es considerablemente más eficiente que calcular la matriz  $\widehat{\mathbf{Y}}$  cuando  $p$  es grande.

Resumiendo, tenemos el siguiente método.

### Método H3

**Paso 1** Si  $p$  es chico, siga el **Paso 1a**. Si  $p$  es grande, siga el **Paso 1b**.

**Paso 1a** Dadas las muestras  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ , defina  $\widehat{\mathbf{Y}}$  como en (5.20). Calcule  $\text{tr}(\widehat{\mathbf{Y}})$  y  $\text{tr}(\widehat{\mathbf{Y}}^2)$ .

**Paso 1b** Dadas las muestras  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ , calcule  $\text{tr}(\widehat{\mathbf{Y}})$  y  $\text{tr}(\widehat{\mathbf{Y}}^2)$  a partir de la Proposición 5.2.2.

**Paso 2** Calcule el estadístico  $\widetilde{D}_{Np}$  definido en (5.19).

Como antes, el  $p$ -valor se estima como el percentil de una  $N(0, 1)$  correspondiente al valor que toma  $\widetilde{D}_{Np}$ .

## 5.3 Test de permutaciones

Por último, proponemos un test para testear igualdad de matrices de covarianza para dos grupos basándonos en el estadístico del test definido en la Sección 5.1.

Recordemos que este estadístico era

$$T_N = N \text{tr}((\mathbf{V}_1 - \mathbf{V}_2)^2) \quad (5.23)$$

donde  $\mathbf{V}_i$  son las matrices de covarianza de signos espaciales para cara grupo. Es decir, podemos pensar que dicho estadístico mide la distancia entre ambas matrices. Llamamos, como antes,  $n_i$  al tamaño del grupo  $i$  para  $i = 1, 2$ .

Supongamos que  $\mathbf{X}_1, \dots, \mathbf{X}_{n_1}$  son las observaciones de la primer población y  $\mathbf{X}_{n_1+1}, \dots, \mathbf{X}_{n_1+n_2}$  son las observaciones de la segunda población. Primero, permutamos los índices  $(1, \dots, n_1, n_1 + 1, \dots, n_1 + n_2)$  obteniendo  $(\pi(1), \dots, \pi(n_1), \pi(n_1 + 1), \dots, \pi(n_1 + n_2))$ . Se generan dos nuevas muestras de la siguiente forma: las observaciones  $\mathbf{X}_{\pi(1)}, \dots, \mathbf{X}_{\pi(n_1)}$  se asignan a la primer muestra y las observaciones  $\mathbf{X}_{\pi(n_1+1)}, \dots, \mathbf{X}_{\pi(n_1+n_2)}$  a la segunda. Luego, calculamos el estadístico  $T_N^*$  a partir de estas dos nuevas muestras. Bajo  $H_0$ , el estadístico  $T_N^*$  se distribuye igual que  $T_N$ . Esto se debe a que mezclar las dos muestras bajo  $H_0$  no representa un cambio en la distribución. En cambio, bajo  $H_1$  las distribuciones de estos dos estadísticos no son la misma y se espera que el valor de  $T_N$  sea mayor que el de  $T_N^*$  porque luego de mezclar las dos muestras se pierde la separación entre ambas. Por lo tanto, se espera una menor diferencia entre las nuevas matrices de covarianza de signos espaciales  $\mathbf{V}_1^*$  y  $\mathbf{V}_2^*$  respecto de las originales  $\mathbf{V}_1$  y  $\mathbf{V}_2$ .

Entonces planteamos el siguiente test:

**Paso 1** Dada la muestra  $X_{i1}, \dots, X_{in_i}$ ,  $i = 1, 2$ , obtenemos el estadístico  $T_N$  definido como en (5.6).

**Paso 2** Permutamos la muestra  $X_{i1}, \dots, X_{in_i}$  y así obtenemos dos muestras nuevas  $X_{i1}^*, \dots, X_{in_i}^*$  de igual tamaño que las originales.

**Paso 3** Obtenemos el estadístico  $T_N^*$  basándonos en las muestras  $X_{i1}^*, \dots, X_{in_i}^*$ .

**Paso 4** Se repite el **Paso 2** y **Paso 3**  $M_{\text{PERM}}$  veces, para obtener  $M_{\text{PERM}}$  valores  $T_{N,r}^*$   $1 \leq r \leq M_{\text{PERM}}$ .

El  $p$ -valor del test se estima como  $\hat{p} = s/M_{\text{PERM}}$  donde  $s$  es la cantidad de  $T_{N,r}^*$  mayores o iguales que  $T_n$ .

# Capítulo 6

## Simulación

En este capítulo resumiremos los resultados de un estudio de simulación diseñado para analizar el rendimiento de algunos tests para igualdad de matrices de covarianza. En los tests para los cuales se tiene un resultado asintótico, nos interesa estudiar la velocidad de convergencia. Por otro lado, tenemos el test descrito en la Sección 5.2 cuyo desarrollo y aproximación normal son heurísticos en esta etapa. En este caso, la simulación ayuda a ver su eficacia (empírica) y así conjeturar algún posible resultado asintótico.

Cuando  $H_0$  es cierta, nos referiremos con  $\Sigma$  a la matriz de covarianza común para los dos grupos.

Este capítulo lo organizaremos de la siguiente forma: en la Sección 6.1 daremos algunas condiciones generales en las que se realizó la simulación. Luego, dividiremos la simulación en dos secciones: en la Sección 6.2 presentamos los resultados de la simulación para los tests en donde se asume que bajo  $H_0$  la matriz  $\Sigma$  es diagonal. En la Sección 6.3 realizamos el estudio de simulación para los tests que no asumen que bajo  $H_0$  dicha matriz sea diagonal.

### 6.1 Condiciones de la simulación

En lo que sigue, llamamos  $n_1$  y  $n_2$  al tamaño de cada una de las dos muestras y a  $p$  a la dimensión de cada uno de los vectores de las muestras. Tenemos que  $\mathbf{X}_{11}, \dots, \mathbf{X}_{1n_1}$  y  $\mathbf{X}_{21}, \dots, \mathbf{X}_{2n_2}$  son dos muestras aleatorias y llamemos  $\mathbf{X}_1$  y  $\mathbf{X}_2$  a vectores aleatorios tales que  $\mathbf{X}_{ij} \sim \mathbf{X}_i$ . Para los modelos y parámetros de esta simulación, nos basamos en la simulación hecha en Schott (2007).

En las simulaciones hechas, tomamos siempre  $n_1 = n_2 = n$ . Tanto a  $n$  como a  $p$  lo tomamos variando en el conjunto  $\{4, 8, 16, 32, 64, 128\}$  y realizamos 1000 repliaciones para cada posible valor de  $n$  y  $p$  en ese conjunto. El nivel de significación nominal de los tests lo fijamos como 0.05.

En todos los casos consideramos que  $\mathbf{X}_1$  tiene matriz de dispersión  $\mathbf{I}_p$ , que coincide con la covarianza si ésta existe. Por otro lado,  $\mathbf{X}_2$  tiene matriz de dispersión  $\mathbf{D}_\delta$  donde  $\mathbf{D}_\delta$  es una matriz diagonal cuyos elementos de la forma  $(i, i)$  son iguales a  $\delta$  si  $i$  es múltiplo de 4 e igual a 1 si no lo es. Supondremos siempre que el parámetro de posición es conocido e igual a  $\mathbf{0}$ . Tenemos entonces un modelo distinto para cada elección de  $\delta$ . Al calcular la proporción de rechazos cuando  $\delta = 1$ ,

estamos estimando el nivel de significación del test y en cualquier otro caso se está estimando la potencia.

Indicamos por  $t_{p,f}(\mathbf{0}, \Sigma)$  a la distribución  $t$ -multivariada en  $\mathbb{R}^p$  centrada en  $\mathbf{0}$ , con  $f$  grados de libertad y matriz de dispersión  $\Sigma$  y decimos que  $\mathbf{X} \sim \epsilon_{p,\lambda}(\mathbf{0}, \Sigma)$  si  $\mathbf{X} = Z\Sigma^{1/2}\mathbf{U}$  donde  $Z \sim \epsilon(\lambda)$  y  $\mathbf{U}$  uniforme en la esfera, con  $\mathbf{U}$  y  $Z$  independientes. Consideramos cuatro modelos para la distribución de las variables aleatorias  $\mathbf{X}_1$  y  $\mathbf{X}_2$ .

- (A) Ambas distribuciones son normales, es decir,  $\mathbf{X}_1 \sim N(\mathbf{0}, \mathbf{I}_p)$  y  $\mathbf{X}_2 \sim N(\mathbf{0}, \mathbf{D}_\delta)$ .
- (B) Ambas distribuciones son  $t$ -multivariadas con 1 grado de libertad, es decir,  $\mathbf{X}_1 \sim t_{p,1}(\mathbf{0}, \mathbf{I}_p)$  y  $\mathbf{X}_2 \sim t_{p,1}(\mathbf{0}, \mathbf{D}_\delta)$ .
- (C) Ambas distribuciones son  $t$ -multivariadas con  $p + 1$  grados de libertad, es decir,  $\mathbf{X}_1 \sim t_{p,p+1}(\mathbf{0}, \mathbf{I}_p)$  y  $\mathbf{X}_2 \sim t_{p,p+1}(\mathbf{0}, \mathbf{D}_\delta)$ .
- (D) Ambas distribuciones son exponenciales multivariadas con ambos valores de  $\lambda$  igual a 1, es decir,  $\mathbf{X}_1 \sim \epsilon_{p,1}(\mathbf{0}, \mathbf{I}_p)$  y  $\mathbf{X}_2 \sim \epsilon_{p,1}(\mathbf{0}, \mathbf{D}_\delta)$ .

## 6.2 Simulación en el caso $\Sigma$ diagonal

En primer lugar estudiaremos los resultados de una simulación para varios tests presentados en esta tesis que asumen que, bajo  $H_0$ , la matriz común de covarianza para los dos grupos es diagonal. Recordemos que hay varios tests para los cuales no se pierde generalidad al asumir esto y, por lo tanto, se incluyen en esta sección.

Más específicamente, consideramos los siguientes tests:

1. **Test de cociente de máxima verosimilitud** definido en la Sección 3.1 y basado en el estadístico que indicamos por  $T_{n,MV}$ . Solamente se simula en los casos en que cada  $n$  sea mayor o igual a  $p$  ya que en otro caso no tiene validez.
2. **Test de cociente de máxima verosimilitud adaptado al contexto  $(n, p)$ -asintótico** definido en la Subsección 3.2.1. Para este caso sólomente simulamos en los casos en que  $n$  sea estrictamente mayor que  $p$ . Denotamos con  $T_{n,MVNP}$  al estadístico de este test.
3. **Test de Schott** definido en la Subsección 3.2.2. Indicamos como  $T_{n,SCH}$  al estadístico del test.
4. **Test Frobenius basado en signos** definido en el **Método F3** de la Subsección 5.1.2 y basado en el estadístico que indicamos por  $T_{n,FROB}$ . Para estimar los percentiles de la distribución límite usamos  $M_{BOOT} = 2000$ .
5. **Test heurístico  $(n, p)$ -asintótico basado en signos** definido en el **Método H1** de la Subsección 5.2.1. Indicamos por  $T_{n,NP}$  a su estadístico.
6. **Test de permutaciones** definido en la Sección 5.3. Usamos 400 permutaciones para cada replicación (es decir,  $M_{PERM} = 400$ ). El estadístico del test se denota por  $T_{n,PERM}$ .

En todas las tablas y gráficos los tests se identifican con su estadístico. En el análisis de nivel y potencia también nos referiremos a los tests de esta misma forma. Para los tests basados en los estadísticos  $T_{n,MV}$ ,  $T_{n,MVNP}$  y  $T_{n,SCH}$  consideraremos los cuatro modelos (A), (B), (C) y (D), mientras que para los tests basados en  $T_{n,FROB}$ ,  $T_{n,NP}$  y  $T_{n,PERM}$  consideraremos solo el modelo (A). Esto se debe a que los cuatro modelos corresponden a distribuciones elípticas y por el Lema 4.1.1, podemos suponer sin perder generalidad que las muestras provienen de una distribución normal.

En la Tabla 6.1 se muestran los niveles de significación estimados bajo el modelo (A) para cada uno de los seis tests considerados. En las Tablas 6.2 a 6.6 se exhiben, bajo este mismo modelo, las frecuencias de rechazo obtenidas para cada test. Los valores de  $\delta$  que consideramos para evaluar potencia son  $\{0.5, 1.5, 2, 2.5, 3\}$ . Por último, en las Tablas 6.7 a 6.9 se presentan los niveles de significación estimados bajo los modelos (B), (C) y (D) para los Tests  $T_{n,MV}$ ,  $T_{n,MVNP}$  y  $T_{n,SCH}$ .

### 6.2.1 Observaciones sobre los niveles de significación estimados

Para el análisis de los resultados de la simulación, haremos primero un estudio de los niveles de significación estimados para cada test. Como varios de los resultados teóricos que tenemos son  $n$  o  $(n, p)$ -asintóticos, es importante analizar, para ciertos  $n$  y  $p$  fijos, si las frecuencias de rechazo obtenidas son lejanas al nivel nominal 0.05. Supongamos que hacemos  $q$  replicaciones de un test que tiene nivel  $\alpha$  desconocido. Queremos testear  $\tilde{H}_0 : \alpha = \alpha_0$  vs.  $\tilde{H}_1 : \alpha \neq \alpha_0$ . Llamamos  $\hat{p}$  a la frecuencia de rechazo de las  $q$  replicaciones y  $z_\gamma$  al percentil  $\gamma$  de una normal estándar. Usando la aproximación normal para la distribución binomial, es fácil ver que un test de nivel asintótico  $\gamma$  tiene como región de aceptación  $\hat{p} \in \left[ \alpha_0 - z_{\gamma/2} \sqrt{\alpha_0(1-\alpha_0)/q}, \alpha_0 + z_{\gamma/2} \sqrt{\alpha_0(1-\alpha_0)/q} \right]$ . En nuestra simulación tenemos  $q = 1000$ ,  $\alpha_0 = 0.05$  y tomaremos  $\gamma = 0.01$ . Por lo tanto, se obtiene la región de aceptación  $\hat{p} \in [0.0322, 0.6775]$ . Es decir, si para un test, un  $n$  y un  $p$  fijos obtenemos que la frecuencia de rechazo está fuera de este intervalo, podemos afirmar con nivel asintótico 1% que la probabilidad de error de tipo I del test para los  $n$  y  $p$  elegidos es distinta de 0.05. Si la frecuencia de rechazo de un test excede al extremo superior de este intervalo, se dice que el test es liberal y si esta frecuencia es menor al extremo inferior del intervalo, se dice que el test es conservador.

En primer lugar, en las Tablas 6.7 a 6.9 se muestran las frecuencias de rechazo obtenidas bajo  $H_0$  para los tests  $T_{n,MV}$ ,  $T_{n,MVNP}$  y  $T_{n,SCH}$  en cada uno de los modelos (B), (C) y (D). Recordemos que los resultados asintóticos para estos dos tests asumen la normalidad de las observaciones. Observemos que dichas frecuencias son mucho mayores al nominal 0.05 y por lo tanto los tests  $T_{n,MV}$ ,  $T_{n,MVNP}$  y  $T_{n,SCH}$  no son informativos en estos tres modelos. Notemos que esta diferencia se intensifica en los modelos (B) y (D), debido a que la distribución que asume el modelo (C) es más cercana a una distribución normal.

En la Tabla 6.1 se muestran los niveles de significación estimados para el modelo (A). En este modelo, el test  $T_{n,MV}$ , aplicable solo cuando  $n \geq p$ , tiene un nivel de significación mucho mayor a 0.05 salvo cuando  $n$  es mucho mayor que  $p$ . Esto implica que este test no es informativo cuando  $n$  y  $p$  son cercanos. Por otro lado, se puede ver que test  $T_{n,MVNP}$  es liberal cuando  $p$  es chico. De todas formas, se observa que dicho test adapta a  $T_{n,MV}$  a contextos de alta dimensión, ya que los niveles de significación estimados se acercan mucho más al nivel nominal.

El test  $T_{n,SCH}$  es liberal cuando  $p$  es chico, con una frecuencia de rechazo mayor a 0.85 en algunos casos. Por otra parte, el test  $T_{n,NP}$  también resulta liberal cuando  $p$  es chico, pero con frecuencias

de rechazo menores que las obtenidas en el test  $T_{n,\text{SCH}}$ . En esta misma tabla se puede ver que los tests  $T_{\text{NP}}$  y  $T_{n,\text{FROB}}$  son conservadores cuando  $p$  es mucho mayor que  $n$ . Las frecuencias de rechazo de estos dos tests se acerca al nivel nominal 0.05 cuando tanto  $n$  como  $p$  son grandes.

Respecto al test  $T_{n,\text{PERM}}$ , bajo  $H_0$  se esperaría que el nivel estimado sea siempre cercano al nominal 0.05. Sin embargo, si se miran los casos en que  $n = 4$ , el nivel estimado es siempre menor a 0.05. Esto se debe a que, dadas las muestras  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ ,  $i = 1, 2$ , hay a lo sumo  $\binom{8}{4}/2 = 35$  valores que podrían tomar los  $T_{N,r}^*$ . En efecto, estos valores quedan determinados por cuáles observaciones pertenecen a cada grupo (sin importar el orden) y se divide por dos por la simetría entre los dos grupos (es decir, da lo mismo que las observaciones  $a, b, c, d$  pertenezcan todas al primer grupo o todas al segundo). Notemos que al estimar el  $p$ -valor del test, contamos cuántos de los  $T_{N,r}^*$  son mayores o iguales que  $T_N$  y luego dividimos por el total de permutaciones. Al estar incluyendo los  $T_{N,r}^*$  que son iguales a  $T_N$ , disminuye la probabilidad de que dicho  $p$ -valor sea menor a 0.05. Esto se traduce en una menor frecuencia de rechazo, tal como se ve en la Tabla 6.1. Cuando  $n$  toma valores más grandes no se observa esta diferencia con el nivel nominal 0.05 porque son muchos más los valores posibles para las variables  $T_{N,r}^*$ .

### 6.2.2 Observaciones sobre las potencias estimadas

En esta sección estudiaremos las frecuencias de rechazo bajo hipótesis alternativas. Los resultados se muestran en las Tablas 6.2 a 6.6.

Vamos a estudiar la potencia estimada sólomente bajo el modelo (A), debido a que ya vimos que los tests  $T_{n,\text{MV}}$  y  $T_{n,\text{SCH}}$  no son informativos en el resto de los modelos.

Recordemos que el test  $T_{n,\text{MV}}$  únicamente es informativo en el caso en que  $n$  sea grande y  $p$  sea chico. En este caso, se puede ver que el test tiene una potencia alta, lo que indica que es el uso de este test es una buena alternativa en el contexto de observaciones normales y tamaños de muestra mucho mayores a la dimensión.

El test  $T_{n,\text{SCH}}$  tiene una potencia baja cuando  $n$  es chico. Sin embargo, esta potencia aumenta rápidamente cuando  $n$  crece y para valores de  $\delta$  mayores a 1, resulta mayor que la obtenida para el test  $T_{n,\text{MV}}$ . Recordemos que este test es liberal para algunas elecciones de  $n$  y de  $p$  y esto puede ser en parte lo que explique una alta frecuencia de rechazo para hipótesis alternativas.

Por otra parte, el test  $T_{n,\text{MVNP}}$  tiene una potencia bastante superior a la de  $T_{n,\text{SCH}}$  cuando  $\delta = 0.5$ . Sin embargo, en los casos en que  $\delta > 1$  este orden se invierte y cuando  $p$  es grande,  $T_{n,\text{MVNP}}$  pasa a tener una potencia baja en relación con los otros cinco tests considerados.

Si comparamos las potencias de los tests  $T_{n,\text{FROB}}$ ,  $T_{n,\text{NP}}$  y  $T_{n,\text{PERM}}$ , se observa que el test  $T_{n,\text{NP}}$  tiene mayor potencia que los otros dos en el caso en que  $p$  sea chico y  $n$  sea grande. Por el contrario, cuando  $n$  es grande y  $p$  es chico, se puede ver que el test  $T_{n,\text{PERM}}$  es el que tiene mayor potencia. Para el resto de los casos las potencias de los tres tests tienen comportamientos similares.

Las potencias para los tests que proponemos,  $T_{n,\text{FROB}}$ ,  $T_{n,\text{NP}}$  y  $T_{n,\text{PERM}}$ , convergen más lentamente a 1 en comparación con el test  $T_{n,\text{SCH}}$ . La ventaja de los tests propuestos, como ya vimos, es que son informativos para cualquier distribución elíptica mientras que el test  $T_{n,\text{SCH}}$  sólomente es informativo en el caso normal.

En la Figura 6.1 se pueden comparar las frecuencias de rechazo obtenidas para los tests  $T_{n,\text{SCH}}$ ,

$T_{n,\text{FROB}}$ ,  $T_{n,\text{NP}}$  y  $T_{n,\text{PERM}}$  para cada  $\delta$  en el conjunto  $\{0.5, 1, 1.5, 2, 2.5, 3\}$ . Recordemos que  $\delta = 1$  corresponde a la hipótesis nula. En cada uno de los cuatro gráficos se fija un valor de  $p$  y se muestran las potencias estimadas para cada uno de los cuatro tests cuando  $n = 8$ ,  $n = 32$  y  $n = 128$ . Los gráficos para los valores  $p = 8$  y  $p = 32$  no se presentan por ser muy similares a otros de los gráficos sí presentados.

Por último, en la Figura 6.10 se estudia empíricamente la distribución asintótica del estadístico  $T_{n,\text{NP}}$  cuando  $(n, p) \rightarrow \infty$ . Para cada  $n$  y  $p$  en el conjunto  $\{4, 8, 16, 32, 64, 128\}$ , se obtuvieron 1000 valores del estadístico  $T_{n,\text{NP}}$  y se construye la densidad empírica correspondiente. Luego, comparamos gráficamente esta densidad empírica con la que corresponde a una normal estándar. En la figura se muestran los 36 gráficos obtenidos, ordenados por  $n$  y por  $p$ . Se puede ver cuando  $n$  y  $p$  son grandes ambas densidades son muy similares. Este hecho sugiere que  $T_{n,\text{NP}} \xrightarrow{D} N(0, 1)$  cuando  $(n, p) \rightarrow \infty$ . Dicha convergencia se analizará en estudios posteriores.

$p =$	4						8						16					
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64
$T_{n,MV}$	0.709	0.235	0.109	0.071	0.066	0.039	NA	0.985	0.439	0.168	0.087	0.070	NA	NA	1.000	0.835	0.345	0.161
$T_{n,MVNP}$	NA	0.087	0.084	0.088	0.061	NA	NA	NA	0.080	0.071	0.065	0.070	NA	NA	NA	0.071	0.062	0.078
$T_{n,SCH}$	0.059	0.075	0.083	0.091	0.089	0.066	0.051	0.067	0.061	0.072	0.066	0.071	0.053	0.051	0.066	0.060	0.058	0.081
$T_{n,FROB}$	0.035	0.051	0.052	0.039	0.057	0.051	0.029	0.040	0.045	0.048	0.036	0.041	0.015	0.020	0.042	0.044	0.048	0.054
$T_{n,NP}$	0.066	0.087	0.075	0.059	0.074	0.067	0.051	0.049	0.057	0.062	0.046	0.050	0.024	0.031	0.052	0.052	0.050	0.063
$T_{n,PERM}$	0.031	0.069	0.057	0.038	0.059	0.048	0.041	0.053	0.051	0.052	0.035	0.038	0.028	0.048	0.056	0.049	0.047	0.057
$p =$			32						64						128			
$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,MV}$	NA	NA	NA	1.000	1.000	0.733	NA	NA	NA	NA	1.000	1.000	NA	NA	NA	NA	NA	1.000
$T_{n,MVNP}$	NA	NA	NA	NA	0.060	0.050	NA	NA	NA	NA	NA	0.049	NA	NA	NA	NA	NA	NA
$T_{n,SCH}$	0.037	0.061	0.056	0.048	0.050	0.061	0.039	0.061	0.057	0.061	0.047	0.053	0.047	0.054	0.053	0.044	0.054	0.071
$T_{n,FROB}$	0.009	0.023	0.027	0.035	0.042	0.044	0.012	0.020	0.030	0.044	0.039	0.047	0.012	0.016	0.032	0.034	0.047	0.066
$T_{n,NP}$	0.019	0.031	0.030	0.036	0.048	0.046	0.017	0.025	0.032	0.046	0.041	0.051	0.016	0.023	0.037	0.037	0.048	0.063
$T_{n,PERM}$	0.029	0.053	0.044	0.040	0.051	0.049	0.041	0.061	0.043	0.052	0.044	0.049	0.036	0.042	0.056	0.049	0.046	0.063

Tabla 6.1: Niveles de significación estimados para modelo (A)

$p =$	4						8						16					
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64
$T_{n,MV}$	0.710	0.273	0.170	0.233	0.416	0.769	NA	0.987	0.546	0.402	0.537	0.861	NA	NA	1.000	0.935	0.844	0.970
$T_{n,MVNP}$	NA	0.114	0.141	0.257	0.468	0.824	NA	NA	0.124	0.226	0.476	0.862	NA	NA	NA	0.176	0.460	0.920
$T_{n,SCH}$	0.069	0.095	0.119	0.169	0.314	0.601	0.050	0.085	0.108	0.155	0.295	0.637	0.055	0.061	0.106	0.160	0.303	0.698
$T_{n,FROB}$	0.037	0.056	0.067	0.106	0.178	0.340	0.030	0.043	0.064	0.094	0.169	0.430	0.017	0.027	0.058	0.103	0.197	0.530
$T_{n,NP}$	0.077	0.091	0.091	0.139	0.210	0.379	0.057	0.066	0.079	0.111	0.203	0.473	0.030	0.041	0.072	0.126	0.226	0.555
$T_{n,PERM}$	0.039	0.068	0.066	0.107	0.177	0.325	0.037	0.064	0.071	0.096	0.172	0.429	0.034	0.055	0.080	0.117	0.211	0.533
$p =$			32						64						128			
$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,MV}$	NA	NA	NA	1.000	1.000	1.000	NA	NA	NA	NA	1.000	1.000	NA	NA	NA	NA	NA	1.000
$T_{n,MVNP}$	NA	NA	NA	NA	0.362	0.901	NA	NA	NA	NA	NA	0.800	NA	NA	NA	NA	NA	NA
$T_{n,SCH}$	0.038	0.080	0.096	0.144	0.309	0.679	0.033	0.070	0.092	0.149	0.307	0.720	0.032	0.055	0.088	0.139	0.297	0.708
$T_{n,FROB}$	0.015	0.033	0.054	0.086	0.213	0.550	0.014	0.024	0.055	0.103	0.227	0.575	0.011	0.020	0.064	0.092	0.219	0.588
$T_{n,NP}$	0.020	0.045	0.063	0.097	0.228	0.571	0.020	0.033	0.063	0.111	0.237	0.589	0.015	0.025	0.067	0.096	0.233	0.591
$T_{n,PERM}$	0.026	0.066	0.075	0.104	0.236	0.553	0.039	0.066	0.088	0.118	0.249	0.583	0.035	0.052	0.092	0.118	0.238	0.597

Tabla 6.2: Potencia estimada para el modelo (A) con  $\delta = 0.5$

$p =$	4							8							16						
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128		
$T_{n,MV}$	0.716	0.235	0.144	0.125	0.169	0.253	0.253	NA	0.988	0.487	0.227	0.217	0.370	NA	NA	1.000	0.881	0.534	0.518		
$T_{n,MVNP}$	NA	0.100	0.119	0.141	0.212	0.334	0.334	NA	NA	0.097	0.112	0.179	0.375	NA	NA	NA	0.098	0.154	0.341		
$T_{n,SCH}$	0.062	0.100	0.124	0.153	0.260	0.445	0.445	0.055	0.077	0.107	0.150	0.258	0.515	0.053	0.073	0.090	0.139	0.263	0.504		
$T_{n,FROB}$	0.040	0.064	0.065	0.070	0.115	0.150	0.150	0.036	0.045	0.061	0.072	0.093	0.192	0.014	0.027	0.050	0.081	0.107	0.270		
$T_{n,NP}$	0.071	0.090	0.090	0.089	0.142	0.196	0.196	0.055	0.067	0.069	0.101	0.111	0.228	0.029	0.043	0.059	0.095	0.126	0.294		
$T_{n,PERM}$	0.029	0.069	0.078	0.073	0.118	0.147	0.147	0.046	0.067	0.068	0.081	0.092	0.200	0.031	0.048	0.066	0.090	0.115	0.272		
$p =$																					
$n =$	4	8	16	32	64	128	128	4	8	16	32	64	128	4	8	16	32	64	128		
$T_{n,MV}$	NA	NA	NA	1.000	1.000	0.940	0.940	NA	NA	NA	NA	1.000	1.000	NA	NA	NA	NA	NA	1.000		
$T_{n,MVNP}$	NA	NA	NA	NA	0.127	0.286	0.286	NA	NA	NA	NA	NA	0.238	NA	NA	NA	NA	NA	NA		
$T_{n,SCH}$	0.043	0.071	0.071	0.123	0.239	0.538	0.538	0.040	0.067	0.081	0.118	0.233	0.537	0.044	0.055	0.077	0.119	0.219	0.550		
$T_{n,FROB}$	0.012	0.024	0.039	0.060	0.126	0.281	0.281	0.011	0.025	0.037	0.077	0.122	0.293	0.013	0.023	0.044	0.077	0.127	0.306		
$T_{n,NP}$	0.019	0.038	0.042	0.067	0.133	0.292	0.292	0.016	0.033	0.041	0.082	0.126	0.302	0.017	0.023	0.047	0.076	0.133	0.307		
$T_{n,PERM}$	0.033	0.059	0.061	0.073	0.135	0.289	0.289	0.043	0.063	0.062	0.092	0.129	0.303	0.038	0.052	0.068	0.095	0.146	0.316		

Tabla 6.3: Potencia estimada para el modelo (A) con  $\delta = 1.5$

$p =$	4							8							16						
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128		
$T_{n,MV}$	0.719	0.262	0.208	0.227	0.429	0.779	0.779	NA	0.990	0.562	0.403	0.540	0.890	NA	NA	1.000	0.933	0.855	0.965		
$T_{n,MVNP}$	NA	0.126	0.172	0.253	0.503	0.841	0.841	NA	NA	0.131	0.222	0.484	0.891	NA	NA	NA	0.187	0.483	0.925		
$T_{n,SCH}$	0.070	0.147	0.244	0.385	0.683	0.941	0.941	0.075	0.139	0.215	0.418	0.737	0.974	0.062	0.126	0.231	0.424	0.820	0.992		
$T_{n,FROB}$	0.043	0.072	0.098	0.116	0.251	0.471	0.471	0.035	0.061	0.084	0.158	0.313	0.696	0.019	0.040	0.091	0.182	0.403	0.831		
$T_{n,NP}$	0.080	0.102	0.121	0.159	0.300	0.523	0.523	0.057	0.091	0.113	0.191	0.357	0.733	0.029	0.058	0.115	0.199	0.435	0.849		
$T_{n,PERM}$	0.041	0.078	0.104	0.123	0.248	0.470	0.470	0.050	0.086	0.098	0.169	0.322	0.688	0.034	0.069	0.119	0.191	0.417	0.836		
$p =$																					
$n =$	4	8	16	32	64	128	128	4	8	16	32	64	128	4	8	16	32	64	128		
$T_{n,MV}$	NA	NA	NA	1.000	1.000	1.000	1.000	NA	NA	NA	NA	1.000	1.000	NA	NA	NA	NA	NA	1.000		
$T_{n,MVNP}$	NA	NA	NA	NA	0.354	0.894	0.894	NA	NA	NA	NA	NA	0.819	NA	NA	NA	NA	NA	NA		
$T_{n,SCH}$	0.054	0.103	0.191	0.448	0.841	0.999	0.999	0.044	0.105	0.183	0.438	0.882	0.999	0.038	0.076	0.173	0.419	0.900	1.000		
$T_{n,FROB}$	0.013	0.033	0.061	0.173	0.471	0.903	0.903	0.012	0.041	0.072	0.208	0.515	0.952	0.014	0.028	0.077	0.190	0.531	0.970		
$T_{n,NP}$	0.022	0.050	0.079	0.190	0.493	0.917	0.917	0.020	0.052	0.087	0.230	0.533	0.958	0.021	0.035	0.086	0.206	0.541	0.972		
$T_{n,PERM}$	0.039	0.078	0.100	0.197	0.481	0.908	0.908	0.050	0.089	0.109	0.241	0.538	0.957	0.051	0.065	0.114	0.218	0.547	0.968		

Tabla 6.4: Potencia estimada para el modelo (A) con  $\delta = 2$

$p =$	4						8						16						
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,MV}$	0.736	0.291	0.273	0.381	0.698	0.968	0.968	NA	0.993	0.633	0.586	0.833	0.997	NA	NA	1.000	0.974	0.978	1.000
$T_{n,MVNP}$	NA	0.142	0.239	0.412	0.750	0.981	0.981	NA	NA	0.177	0.397	0.797	0.997	NA	NA	NA	0.314	0.814	0.999
$T_{n,SCH}$	0.093	0.197	0.390	0.644	0.903	0.997	0.997	0.098	0.222	0.382	0.724	0.980	1.000	0.075	0.194	0.411	0.787	0.995	1.000
$T_{n,FROB}$	0.050	0.080	0.130	0.211	0.445	0.803	0.803	0.041	0.080	0.145	0.291	0.614	0.965	0.019	0.060	0.153	0.358	0.789	0.990
$T_{n,NP}$	0.087	0.127	0.166	0.274	0.501	0.835	0.835	0.062	0.112	0.176	0.336	0.659	0.973	0.037	0.086	0.191	0.397	0.805	0.991
$T_{n,PERM}$	0.050	0.095	0.133	0.214	0.453	0.798	0.798	0.057	0.102	0.164	0.300	0.615	0.962	0.038	0.099	0.189	0.383	0.794	0.990
$p =$			32							64						128			
$n =$	4	8	16	32	64	128	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,MV}$	NA	NA	NA	1.000	1.000	1.000	1.000	NA	NA	NA	NA	1.000	1.000	NA	NA	NA	NA	NA	1.000
$T_{n,MVNP}$	NA	NA	NA	NA	0.682	0.998	0.998	NA	NA	NA	NA	NA	0.998	NA	NA	NA	NA	NA	NA
$T_{n,SCH}$	0.070	0.176	0.405	0.837	0.998	1.000	1.000	0.052	0.167	0.412	0.847	1.000	1.000	0.032	0.101	0.369	0.856	1.000	1.000
$T_{n,FROB}$	0.014	0.050	0.132	0.395	0.852	0.999	0.999	0.015	0.051	0.142	0.414	0.898	1.000	0.017	0.037	0.144	0.410	0.935	1.000
$T_{n,NP}$	0.032	0.074	0.163	0.431	0.860	0.999	0.999	0.022	0.067	0.166	0.437	0.907	1.000	0.025	0.048	0.157	0.432	0.940	1.000
$T_{n,PERM}$	0.053	0.110	0.180	0.448	0.848	0.999	0.999	0.058	0.112	0.192	0.459	0.901	1.000	0.054	0.089	0.212	0.464	0.944	1.000

Tabla 6.5: Potencia estimada para el modelo (A) con  $\delta = 2.5$

$p =$	4						8						16						
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,MV}$	0.744	0.329	0.348	0.542	0.864	0.999	0.999	NA	0.994	0.706	0.730	0.962	1.000	NA	NA	1.000	0.993	0.998	1.000
$T_{n,MVNP}$	NA	0.155	0.304	0.570	0.886	0.999	0.999	NA	NA	0.235	0.560	0.949	1.000	NA	NA	NA	0.476	0.958	1.000
$T_{n,SCH}$	0.117	0.279	0.513	0.817	0.987	1.000	1.000	0.122	0.313	0.571	0.900	0.997	1.000	0.099	0.294	0.638	0.958	1.000	1.000
$T_{n,FROB}$	0.059	0.101	0.174	0.327	0.631	0.952	0.952	0.044	0.103	0.202	0.442	0.852	0.998	0.022	0.083	0.230	0.575	0.960	1.000
$T_{n,NP}$	0.096	0.145	0.222	0.390	0.685	0.968	0.968	0.069	0.134	0.255	0.494	0.876	0.998	0.050	0.110	0.278	0.620	0.966	1.000
$T_{n,PERM}$	0.055	0.114	0.185	0.322	0.637	0.953	0.953	0.060	0.123	0.230	0.457	0.846	0.998	0.039	0.126	0.277	0.604	0.963	1.000
$p =$			32							64						128			
$n =$	4	8	16	32	64	128	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,MV}$	NA	NA	NA	1.000	1.000	1.000	1.000	NA	NA	NA	NA	1.000	1.000	NA	NA	NA	NA	NA	1.000
$T_{n,MVNP}$	NA	NA	NA	NA	0.905	1.000	1.000	NA	NA	NA	NA	NA	1.000	NA	NA	NA	NA	NA	NA
$T_{n,SCH}$	0.096	0.281	0.661	0.977	1.000	1.000	1.000	0.054	0.238	0.651	0.987	1.000	1.000	0.029	0.160	0.641	0.992	1.000	1.000
$T_{n,FROB}$	0.024	0.074	0.241	0.629	0.982	1.000	1.000	0.019	0.066	0.227	0.675	0.994	1.000	0.017	0.052	0.236	0.701	0.995	1.000
$T_{n,NP}$	0.039	0.099	0.270	0.663	0.983	1.000	1.000	0.027	0.094	0.251	0.694	0.995	1.000	0.027	0.067	0.267	0.720	0.995	1.000
$T_{n,PERM}$	0.060	0.136	0.291	0.672	0.985	1.000	1.000	0.070	0.143	0.305	0.707	0.993	1.000	0.064	0.127	0.318	0.734	0.995	1.000

Tabla 6.6: Potencia estimada para el modelo (A) con  $\delta = 3$

$p =$	4				8				16									
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128					
$T_{n,MV}$	0.980	0.974	0.990	0.997	0.999	1.000	1.000	NA	1.000	1.000	1.000	1.000	1.000					
$T_{n,MVNP}$	NA	0.940	0.990	0.997	0.999	1.000	1.000	NA	1.000	1.000	1.000	1.000	1.000					
$T_{n,SCH}$	0.753	0.915	0.973	0.992	0.999	1.000	1.000	0.853	0.975	0.999	1.000	1.000	1.000					
$p =$	32																	
$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,MV}$	NA	NA	NA	1.000	1.000	1.000	NA	NA	NA	NA	1.000	1.000	NA	NA	NA	NA	NA	1.000
$T_{n,MVNP}$	NA	NA	NA	NA	1.000	1.000	NA	NA	NA	NA	NA	1.000	NA	NA	NA	NA	NA	NA
$T_{n,SCH}$	0.985	1.000	1.000	1.000	1.000	1.000	0.991	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000

Tabla 6.7: Niveles de significación estimados para el modelo (B) (distribuciones  $t_{p,1}(\mathbf{0}, \mathbf{I}_p)$ )

$p =$	4				8				16									
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128					
$T_{n,MV}$	0.794	0.468	0.401	0.469	0.536	0.606	0.606	NA	0.993	0.703	0.501	0.458	0.440					
$T_{n,MVNP}$	NA	0.261	0.365	0.492	0.588	0.672	0.672	NA	NA	0.222	0.338	0.405	0.449					
$T_{n,SCH}$	0.189	0.329	0.411	0.521	0.607	0.680	0.680	0.183	0.265	0.354	0.429	0.459	0.504					
$p =$	32																	
$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,MV}$	NA	NA	NA	1.000	1.000	0.936	NA	NA	NA	NA	NA	1.000	1.000	NA	NA	NA	NA	1.000
$T_{n,MVNP}$	NA	NA	NA	NA	NA	0.215	0.265	NA	NA	NA	NA	NA	0.202	NA	NA	NA	NA	NA
$T_{n,SCH}$	0.148	0.208	0.256	0.295	0.296	0.329	0.134	0.187	0.236	0.258	0.285	0.285	0.305	0.130	0.178	0.223	0.253	0.232

Tabla 6.8: Niveles de significación estimados para el modelo (C) (distribuciones  $t_{p,p+1}(\mathbf{0}, \mathbf{I}_p)$ )

$p =$	4				8				16									
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128					
$T_{n,MV}$	0.977	0.909	0.874	0.894	0.918	0.905	0.905	NA	1.000	1.000	0.999	1.000	1.000					
$T_{n,MVNP}$	NA	0.821	0.853	0.908	0.932	0.934	0.934	NA	NA	0.996	0.999	1.000	1.000					
$T_{n,SCH}$	0.572	0.730	0.802	0.870	0.916	0.935	0.935	0.762	0.926	0.977	0.996	0.998	1.000					
$p =$	32																	
$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,MV}$	NA	NA	NA	1.000	1.000	1.000	NA	NA	NA	NA	NA	1.000	1.000	NA	NA	NA	NA	1.000
$T_{n,MVNP}$	NA	NA	NA	NA	NA	1.000	NA	NA	NA	NA	NA	NA	1.000	NA	NA	NA	NA	NA
$T_{n,SCH}$	0.982	1.000	1.000	1.000	1.000	1.000	0.994	1.000	1.000	1.000	1.000	1.000	0.990	1.000	1.000	1.000	1.000	1.000

Tabla 6.9: Niveles de significación estimados para el modelo (D) (distribuciones  $\epsilon_{p,1}(\mathbf{0}, \mathbf{I}_p)$ )

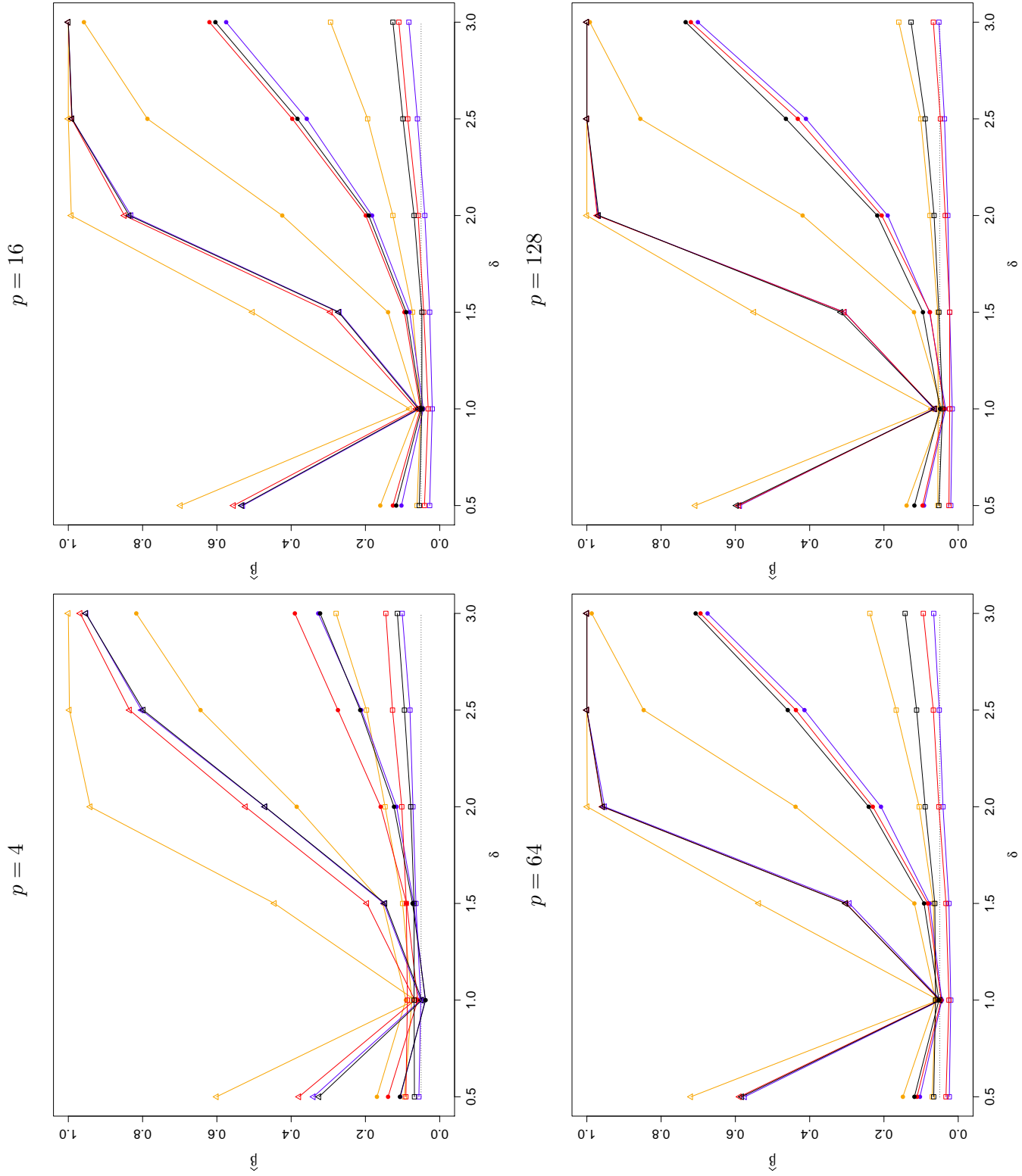


Figura 6.1: Frecuencia de rechazo,  $\hat{\beta}$ , cuando  $p = 4, 16, 64$  y  $128$  de los tests basados en los estadísticos  $T_{n, SCH}$ ,  $T_{n, PERM}$ ,  $T_{n, FROB}$  y  $T_{n, NP}$  en líneas naranja, negro, azul y rojo, respectivamente. En triángulos, círculos y cuadrados se indican la potencias estimadas para  $n = 128, 32$  y  $8$ , respectivamente.

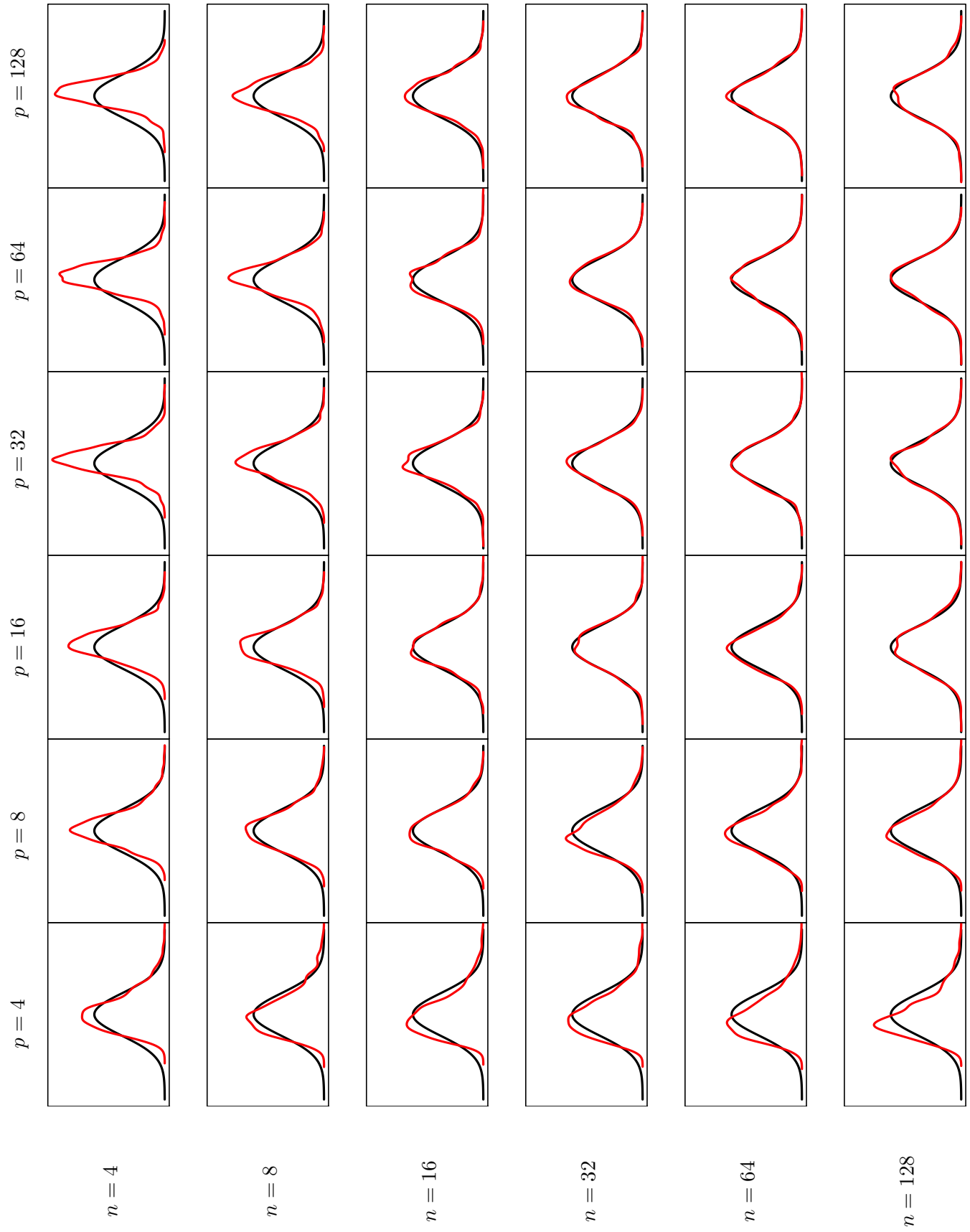


Tabla 6.10: Estimadores de la densidad del estadístico  $T_{n, NP}$ , bajo la hipótesis nula, en rojo y densidad normal estándar en negro.

## 6.3 Simulación en el caso $\Sigma$ no diagonal

En las Secciones 5.1 y 5.2 presentamos algunos tests para los que sí se pierde generalidad al asumir que bajo  $H_0$  la matriz de covarianza común entre los grupos es diagonal. Es por eso que se debieron definir nuevos métodos para el caso general. En esta sección haremos un breve análisis de una simulación realizada para estos casos.

Consideraremos los siguientes tres tests:

1. **Test Frobenius basado en signos** definido en el **Método F5** de la Subsección 5.1.3. Indicaremos al estadístico de este test como  $T_{n,\text{FROB2}}$ .
2. **Test heurístico  $(n, p)$ –asintótico basado en signos** definido en el **Método H2** de la Subsección 5.2.2. Al estadístico de este test lo llamaremos  $T_{n,\text{NP2}}$ .
3. **Test heurístico  $(n, p)$ –asintótico basado en signos** definido en el **Método H3** de la Subsección 5.2.2. Al estadístico de este test lo llamaremos  $T_{n,\text{NP3}}$ .

En el análisis de nivel y potencia y en las tablas nos referiremos al test con el nombre de su estadístico.

Para cada uno de ellos únicamente simulamos bajo el modelo (A). Esto se debe a que se basan en los signos de las observaciones originales y, suponiendo elipticidad, no se pierde generalidad al asumir normalidad. Las condiciones de la simulación son las mismas que para la Sección 6.2 y están detalladas en la Sección 6.1. En este caso, estudiaremos la potencia sólo en el caso  $\delta = 2$ .

### 6.3.1 Observaciones sobre los niveles de significación estimados

En la Tabla 6.11 se muestran las frecuencias de rechazo obtenidas bajo  $H_0$  para los tres tests de esta subsección. En esta tabla se observa que los tests  $T_{n,\text{NP2}}$  y  $T_{n,\text{NP3}}$  son liberales cuando  $p$  es chico. Recordemos que el mismo problema tenían los tests  $T_{n,\text{SCH}}$  y  $T_{n,\text{NP}}$  analizados en la Sección 6.2.

Lo que más llama la atención de esta tabla es que cuando  $p$  crece, las frecuencias de rechazo para los tres tests disminuyen considerablemente. De hecho, cuando  $p$  es grande y  $n$  es chico, esta frecuencia se hace cero. Esta situación se vuelve más extrema en el test  $T_{\text{NP3}}$ . Es decir, estos tests son claramente conservadores cuando  $p$  es grande.

Recordemos que en el test  $T_{n,\text{NP3}}$  se estima la matriz completa  $\Upsilon \in \mathbb{R}^{p^2 \times p^2}$ . Lo que puede estar ocurriendo es que cuando  $p$  es grande se necesita un  $n$  mucho mayor a  $p$  para obtener una buena estimación de los  $p^4$  elementos de  $\Upsilon$ . En la Subsección 5.1.3 se propuso el **Método F4** que también necesita una estimación de todos los elementos de  $\Upsilon$ . Por este motivo y por su complejidad computacional, no se incluyó en esta simulación.

Por otra parte, la idea de los tests  $T_{n,\text{NP2}}$  y  $T_{n,\text{FROB2}}$  es rotar las observaciones originales (o equivalentemente, rotar sus signos) de forma tal que podamos usar los tests respectivos para el caso  $\Sigma$  diagonal. Para eso, se necesita tener un estimador  $\hat{\beta}$  de la matriz de autovectores  $\beta$  de  $\Sigma$ . Cuando  $p$  es mayor a  $n$ , se obtiene una mala estimación para dicha matriz  $\beta$ . Por eso, para estos

casos, las frecuencias de rechazo obtenidas son cercanas a cero, es decir, se alejan mucho del nivel nominal 0.05.

### 6.3.2 Observaciones sobre las potencias estimadas

Por último, haremos un breve análisis de las potencias estimadas cuando  $\delta = 2$ . Estos resultados se muestran en la Tabla 6.12.

Vimos que para ciertos valores de  $n$  y  $p$  los niveles estimados de algunos test fueron cero. Este hecho trae como consecuencia una baja potencia para estos casos. Efectivamente, para el test  $T_{n, NP3}$  las potencias estimadas para  $n \leq p$  resultan siempre iguales o cercanas a cero. Esto quiere decir que se necesita que  $n$  sea mucho mayor a  $p$  para que el test resulte potente.

Esta situación mejora en los tests  $T_{n, NP2}$  y  $T_{n, FROB2}$ . Por ejemplo, se puede ver que en los casos en que  $n = p$  la potencia de estos tests aumenta y se acerca a 1 cuando  $n = p = 128$  mientras que las frecuencias de rechazo para el test  $T_{n, NP3}$  son casi siempre cero en estos casos.

La potencia de  $T_{n, NP2}$  es levemente mayor a la de  $T_{n, FROB2}$  en la mayoría de los casos. Sin embargo, al comparar estos dos tests es importante tener en cuenta que el test  $T_{n, NP2}$  es más liberal que  $T_{n, FROB2}$ , como puede observarse en la Tabla 6.11. Las potencias de estos dos tests resultan siempre menores que las obtenidas para los tests  $T_{n, NP}$  y  $T_{n, FROB}$  estudiadas en la Sección 6.2. Esta diferencia es casi nula cuando  $n$  es grande en relación con  $p$  y se intensifica en los casos en que  $n < p$ .

$p =$	4				8				16									
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128					
$T_{n,FROB2}$	0.038	0.054	0.052	0.039	0.059	0.045	0.045	0.015	0.030	0.037	0.044	0.036	0.038					
$T_{n,NP2}$	0.071	0.093	0.077	0.058	0.075	0.068	0.068	0.046	0.047	0.054	0.058	0.046	0.049					
$T_{n,NP3}$	0.037	0.070	0.069	0.053	0.072	0.067	0.067	0.007	0.017	0.026	0.041	0.034	0.042					
$p =$			32							64			128					
$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,FROB2}$	0.000	0.000	0.004	0.010	0.030	0.036	0.000	0.000	0.001	0.004	0.015	0.028	0.000	0.000	0.000	0.000	0.004	0.025
$T_{n,NP2}$	0.000	0.001	0.008	0.017	0.032	0.039	0.000	0.000	0.002	0.005	0.016	0.030	0.000	0.000	0.000	0.000	0.004	0.027
$T_{n,NP3}$	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Tabla 6.11: Niveles de significación estimados bajo el modelo (A) para los tests de la Sección 6.3

$p =$	4				8				16									
	$n =$	4	8	16	32	64	128	4	8	16	32	64	128					
$T_{n,FROB2}$	0.045	0.072	0.100	0.117	0.251	0.472	0.021	0.052	0.072	0.156	0.310	0.697	0.001	0.006	0.052	0.152	0.376	0.825
$T_{n,NP2}$	0.091	0.104	0.123	0.161	0.300	0.523	0.052	0.080	0.102	0.182	0.353	0.730	0.007	0.021	0.071	0.179	0.409	0.843
$T_{n,NP3}$	0.050	0.080	0.114	0.147	0.295	0.519	0.010	0.025	0.056	0.142	0.319	0.711	0.000	0.000	0.005	0.045	0.258	0.775
$p =$			32							64					128			
$n =$	4	8	16	32	64	128	4	8	16	32	64	128	4	8	16	32	64	128
$T_{n,FROB2}$	0.000	0.000	0.010	0.087	0.402	0.890	0.000	0.000	0.002	0.044	0.331	0.928	0.000	0.000	0.000	0.002	0.171	0.907
$T_{n,NP2}$	0.001	0.002	0.014	0.103	0.426	0.899	0.000	0.000	0.005	0.049	0.354	0.932	0.000	0.000	0.000	0.002	0.187	0.912
$T_{n,NP3}$	0.000	0.000	0.000	0.000	0.053	0.661	0.000	0.000	0.000	0.000	0.000	0.158	0.000	0.000	0.000	0.000	0.000	0.000

Tabla 6.12: Potencia estimada bajo el modelo (A) y  $\delta = 2$  para los tests de la Sección 6.3

# Capítulo 7

## Código

### 7.1 auxiliares.R

```
library(matrixcalc)
library(mvtnorm)

#####
#FUNCIONES AUXILIARES
#####

#####
#Obtiene una muestra aleatoria de una N(0,Sigma)
#####
obtenerMuestraNormal = function(n,p,sigma,...){
  return(rmvnorm(n,rep(0,p),sigma))
}

#####
#Obtiene una muestra aleatoria de una T de Student multivariada con
#covarianza sigma y df grados de libertad
#####
obtenerMuestrat = function(n,p,sigma,df,...){
  return(rmvt(n,sigma = sigma, df = df))
}

#####
#Obtiene una muestra aleatoria de una exponencial multivariada con parametro lambda
#parametro de centralidad cero y matriz de covarianza sqrtSigma %*% sqrtSigma \trasp
#####
obtenerMuestraExponencialMultivariada = function(n,p,sqrtsigma,lambda,...){
  unif = rmvnorm(n,rep(0,p),diag(p))
  exp = rexp(n,lambda)
  unif = t(apply(unif,1,normalizar))
  esferica = exp * unif
  return(esferica %*% t(sqrtsigma))
}

#####
#Devuelve la raiz cuadrada de una matriz diagonal.
#####
obtenerRaizCuadradaDeMatrizDiagonal = function(matriz){
  return (diag(sqrt(diag(matriz))))
}
```

```

}

#####
#Normaliza un vector respecto a la norma euclidea (obtiene un signo)
#####
normalizar = function(x){
  norma = sqrt(sum(x^2))
  return(x/norma)
}

#####
#Obtiene la traza de una matriz
#####
traza = function(matriz){sum(diag(matriz))}

#####
#Obtiene una lista de matrices identidad, donde el tamaño de la i-esima es ps[i]
#####
obtenerListaDeMatricesIdentidad = function(ps){
  lp = length(ps)
  ret = list()
  for(i in 1:lp){
    ret[[i]] = diag(ps[i])
  }
  return(ret)
}

#####
#Obtiene una lista de matrices diagonales, donde la i-esima tiene tamaño ps[i] y los ←
  elementos de
#la diagonal son bloques (1,1,1,delta), completando con 1's si fuera necesario.
#####
obtenerListaDeMatricesDiagonalesParaPotencia = function(ps,listadeltas){
  lp = length(ps)
  ret = list()
  for(i in 1:lp){
    nbloques = ps[i] %% 4
    sobrantes = ps[i] - nbloques*4
    ret[[i]] = diag(c(rep(c(1,1,1,listadeltas[i]),nbloques),rep(1,sobrantes)))
  }
  return(ret)
}

#####
#Calcula la covarianza sin centrar
#####
covarianza = function(x){
  n=nrow(x)
  media = apply(x,2,mean)
  covar =cov(x)*(n-1)/n
  covarianza=covar+ media%*%t(media)
  return(covarianza)
}

#####
# Devuelve la proporción de elementos del vector "observaciones" que son menores que "x"
#####
obtenerAcumuladaParaEmpirica = function(x,observaciones){
  len = length(observaciones)
  return (mean (1*(observaciones < x)))
}

#####
# Genera el nombre para el directorio y archivo del output

```

```
#####
obtenerNombreDeDirectorioYArchivoParaOutput = function(funcionParaLasMuestras ,n1 ,n2 ,p ,←
  nombreDelTest ,delta = 1 ,df1 = NULL ,df2 = NULL ,lambda1 = NULL ,lambda2 = NULL ){
  if(identical(funcionParaLasMuestras ,obtenerMuestraNormal)){directorio="normal"}
  if(identical(funcionParaLasMuestras ,obtenerMuestrat)){directorio=paste("t_df1_" ,df1 , "←
    df2_" ,df2 ,sep="") }
  if(identical(funcionParaLasMuestras ,obtenerMuestraExponencialMultivariada)){
    directorio=paste("exp_lambda1_" ,lambda1 , "_lambda2_" ,lambda2 ,sep="") }
  if(!(file.exists(directorio))){
    dir.create(directorio)
  }

  if(delta==1) {
    archivo= paste(nombreDelTest , "_H0_n1_" ,n1 , "_n2_" ,n2 , "_p_" ,p , ".txt" ,sep="") }
  if(delta!=1){
    archivo= paste(nombreDelTest , "_H1_delta_" ,delta , "_n1_" ,n1 , "_n2_" ,n2 , "_p_" ,p , ".txt" ,sep←
      ="" )}

  return(list(directorio = directorio , archivo = archivo))
}

```

## 7.2 TCMV.R

```
source("auxiliares.R")

#####
# TEST DE COCIENTE DE MAXIMA VEROSIMILITUD
# Se incluye el test T_{n,MV} y T_{n,MVNP}
#####

#####
#Obtenemos el valor del estadistico transformado (-2*log) de maxima verosimilitud.
#Lo obtengo asi para evitar problemas numericos (el estadistico sin transformar es ←
  demasiado grande).
#####
obtenerEstadisticoMV = function(n1 ,n2 ,p ,muestra1 ,muestra2){
  W1 = covarianza(muestra1)
  W2 = covarianza(muestra2)
  n = n1 + n2
  W = (W1 * n1 + W2 * n2) / n
  #Notar que las otras constantes del TCMV estan absorbidas al caclular los determinantes
  #de las matrices W1,W2 y W
  estad = n * log(det(W)) - n1 * log(det(W1)) - n2 * log(det(W2))
  return(estad)
}

#####
#Obtenemos el valor de mu_n del Corolario 3.2.2.
#####
obtenerMu = function(n1 ,n2 ,p){
  n = n1 + n2
  t1 = n * (2*n - 2*p - 1) * log(1 - p/n)
  t2 = n1 * (2*n1 - 2*p - 1) * log(1 - p/n1)
  t3 = n2 * (2*n2 - 2*p - 1) * log(1 - p/n2)
  return((1/4)*(t1 - t2 - t3))
}

#####
#Obtenemos el valor de sigmaSq_n del Corolario 3.2.2.

```

```

#####
obtenerSigmaSq = function(n1,n2,p){
  n = n1 + n2
  t1 = log(1 - p/n)
  t2 = ((n1/n)^2) * log(1 - p/n1)
  t3 = ((n2/n)^2) * log(1 - p/n2)
  return((1/2)*(t1 - t2 - t3))
}

#####
#Obtenemos el valor del estadístico del Corolario 3.2.2.
#####
obtenerEstadisticoMVnp= function(n,estadisticoMV,mu,sigmaSq){
  num = (-1/2)* estadisticoMV -mu
  denom = n * sqrt(sigmaSq)
  return(num/denom)
}

#####
#Corre una vez el test de CMV (n,p)-asintotico definido en
#el Corolario 3.2.2. (ver especificaciones de "correrTestCMV")
#####
correrUnTCMVnp = function(n1,n2,p,muestra1,muestra2,mu,sigmaSq){
  estadisticoMV = obtenerEstadisticoMV(n1,n2,p,muestra1,muestra2)
  estadisticoMVnp = obtenerEstadisticoMVnp(n1 + n1,estadisticoMV,mu,sigmaSq)
  pvalor = pnorm(estadisticoMVnp)
  return(list(pval = pvalor, estad = estadisticoMVnp))
}

#####
#Corre una vez el test de CMV (ver especificaciones de "correrTestCMV")
#####
correrUnTCMV = function(n1,n2,p,muestra1,muestra2){
  estad = obtenerEstadisticoMV(n1,n2,p,muestra1,muestra2)
  df = p* (p+1)/2
  pval = 1 - pchisq(estad,df)
  return(list(pval = pval, estad = estad))
}

#####
#cantidad es la cantidad de simulaciones a realizar
#n1 y n2 son los tamanos de las muestras
#p es la dimension de cada elemento de las muestras
#sigma 1 y sigma2 son las matrices de covarianza para cada una de las muestras
#nivel es el nivel del test
#df1 y df2 son los grados de libertad a usar para cada grupo si se usan las distribuciones↔
T de Student multivariadas
#lambda1 y lambda2 son los parametros lambda a utilizar para cada grupo si se usan las ↔
distribuciones exponencial multivariada.
#Si versionNp = T, se hace el test del Corolario 3.2.2. En caso contrario, se hace el test↔
clasico de maxima verosimilitud.
#Devuelve los estadisticos obtenidos, los pvalores y la proporcion de rechazos
#
#Funciones para muestras disponibles:
#obtenerMuestraNormal parametros: n,p,sigma
#obtenerMuestrat parametros: n,p,sigma,df
#obtenerMuestraExponencialMultivariada parametros:n,p,sqrtsigma,lambda
#####
correrTestCMV = function(cantidad,n1,n2,sigma1,sigma2,delta=1,funcionParaLasMuestras,↔
  nivel = 0.05,df1 = NULL, df2 = NULL, lambda1 = NULL, lambda2 = NULL, versionNp = F↔
  ,...){
  p = length(sigma1[,])

  #Nombre de los archivos de salida
  if(versionNp){

```

```

output = obtenerNombreDeDirectorioYArchivoParaOutput(funcionParaLasMuestras = ←
  funcionParaLasMuestras, n1 = n1, n2 = n2, p = p, nombreDelTest = "TCMV_NP", delta ←
  = delta, df1 = df1, df2 = df2, lambda1 = lambda1, lambda2 = lambda2)
} else {
  output = obtenerNombreDeDirectorioYArchivoParaOutput(funcionParaLasMuestras = ←
    funcionParaLasMuestras, n1 = n1, n2 = n2, p = p, nombreDelTest = "TCMV", delta = ←
    delta, df1 = df1, df2 = df2, lambda1 = lambda1, lambda2 = lambda2)
}
directorio = output$directorio
archivo = output$archivo
nombre = paste(directorio, "\\ ", archivo, sep = " ")

pvalores=estadisticos= rep(0,cantidad)
rechazo = 0
#Obtenemos la raiz cuadrada de las matrices sigmal y sigma2
sqrtsigma1 = obtenerRaizCuadradaDeMatrizDiagonal(sigmal)
sqrtsigma2 = obtenerRaizCuadradaDeMatrizDiagonal(sigma2)

#Obtenemos las constantes mu y sigmaSq
mu = obtenerMu(n1,n2,p)
sigmaSq = obtenerSigmaSq(n1,n2,p)

for (irep in 1:cantidad){
  set.seed(1234 + irep)
  if(versionNp){
    print(c("TCMV_NP irep=", irep, ",n1:", n1, ",n2:",n2,"p:",p, "delta:", delta, "df1:"←
      , df1, "lambda1:", lambda1))
  } else {
    print(c("TCMV irep=", irep, ",n1:", n1, ",n2:",n2,"p:",p, "delta:", delta, "df1:", ←
      df1, "lambda1:", lambda1))
  }
  #Obtenemos las muestras
  muestra1 = funcionParaLasMuestras(n = n1,p = p,sigma = sigmal,df = df1, sqrtsigma = ←
    sqrtsigma1,lambda = lambda1)
  muestra2 = funcionParaLasMuestras(n = n2,p = p,sigma = sigma2,df = df2, sqrtsigma = ←
    sqrtsigma2,lambda = lambda2)
  #Obtenemos los resultados del test
  if(versionNp){
    resultados = correrUnTCMVnp(n1,n2,p,muestra1,muestra2,mu,sigmaSq)
  } else {
    resultados = correrUnTCMV(n1,n2,p,muestra1,muestra2)
  }
  pvalores[irep] = resultados$pval
  estadisticos[irep] = resultados$estad
  if(resultados$pval < nivel){
    rechazo = rechazo + 1
  }

  salida<-cbind(irep, pvalores[irep],estadisticos[irep],rechazo)
  largo.sal=length(salida)
  # Guardo en cada iteracion
  write(t(salida),file=nombre,ncol=largo.sal,append=T)
}
rechazo = rechazo / cantidad
return(list(rechazo = rechazo, pvalores = pvalores,estadisticos = estadisticos))
}

```

### 7.3 schott.R

```

source("auxiliares.R")

#####
# TEST DE SCHOTT (SUBSECCION 3.2.2)
# Se incluye el test T_{n,SCH}
#####

#####
# Devuelve el estadístico de Schott.
# n1 y n2 son los tamaños de las muestras
# s1 y s2 son las matrices de covarianza sin centrar
#####
obtenerEstadisticoSchott = function(n1,n2,s1,s2){
  n = n1 + n2
  s = (n1 * s1 + n2 * s2)/n
  num = traza(s1 %*% s1) * (1 - ((n1 - 2)/((n1 + 2)*(n1 - 1)))
  num = num + traza(s2 %*% s2) * (1 - ((n2 - 2)/((n2 + 2)*(n2 - 1)))
  num = num - 2* traza(s1 %*% s2)
  num = num - (traza(s1)^2) * (n1/((n1 + 2)*(n1 - 1)))
  num = num - (traza(s2)^2) * (n2/((n2 + 2)*(n2 - 1)))
  a = (n^2/((n+2)*(n-1))) * (traza(s %*% s) - ((traza(s)^2)/n))
  denom = 2* (n/(n1*n2)) * a
  return (num/denom)
}

#####
#cantidad es la cantidad de simulaciones a hacer
#n es el tamaño de muestra para cada grupo
#p es la dimensión de cada observación
#sigma1 es la matriz de covarianza para obtener la m.a. del primer grupo
#sigma2 es la matriz de covarianza para obtener la m.a. del segundo grupo
#delta es el valor de los autovalores que diferencian a sigma1 y sigma2. Ver Sección 6.1
#nivel es el percentil de la distribución límite que se toma como valor crítico (si el ←
estadístico observado da mayor, rechazamos, si no, no rechazamos)
#funcionParaLasMuestras es la función a utilizar para obtener las muestras aleatorias
#df1 y df2 son los grados de libertad a usar para cada grupo si se usan las distribuciones ←
T de Student multivariadas
#lambda1 y lambda2 son los parámetros lambda a utilizar para cada grupo si se usan las ←
distribuciones exponencial multivariada.
#Devuelve los estadísticos obtenidos, los pvalores, los cuantiles correspondientes al
#nivel (valor crítico) y la proporción de rechazos.
#
#Funciones para muestras disponibles:
#obtenerMuestraNormal parametros: n,p,sigma
#obtenerMuestrat parametros: n,p,sigma,df
#obtenerMuestraExponencialMultivariada parametros:n,p,sqrtsigma,lambda
#####

correrTestSchott = function(cantidad,n1,n2,sigma1,sigma2,delta=1,funcionParaLasMuestras, ←
  nivel = 0.05,df1 = NULL, df2 = NULL, lambda1 = NULL, lambda2 = NULL,...){
  p = length(sigma1[1,])
  #Nombre de los archivos de salida
  output = obtenerNombreDeDirectorioYArchivoParaOutput(funcionParaLasMuestras = ←
    funcionParaLasMuestras, n1 = n1, n2 = n2, p = p, nombreDelTest = "SCHOTT", delta = ←
    delta, df1 = df1, df2 = df2, lambda1 = lambda1, lambda2 = lambda2)
  directorio = output$directorio
  archivo = output$archivo
  nombre = paste(directorio,"\\",archivo,sep = "")

  rechazo = 0
  estadisticos = pvalores = rep(0,cantidad)
  #Obtenemos la raíz cuadrada de las matrices sigma1 y sigma2
  sqrtsigma1 = obtenerRaizCuadradaDeMatrizDiagonal(sigma1)
  sqrtsigma2 = obtenerRaizCuadradaDeMatrizDiagonal(sigma2)

```

```

for(irep in 1:cantidad){
  set.seed(1234 + irep)
  print(c("SCHOTT irep=", irep, ",n1:", n1, ",n2:",n2,"p:",p, "delta:", delta, "df1:", ←
    df1, "lambda1:", lambda1))
  muestra1 = funcionParaLasMuestras(n = n1,p = p,sigma = sigma1,df = df1, sqrtsigma = ←
    sqrtsigma1,lambda = lambda1)
  muestra2 = funcionParaLasMuestras(n = n2,p = p,sigma = sigma2,df = df2, sqrtsigma = ←
    sqrtsigma2,lambda = lambda2)
  s1 = covarianza(muestra1)
  s2 = covarianza(muestra2)
  estadistico = obtenerEstadisticoSchott(n1,n2,s1,s2)
  pvalor = 1 - pnorm(estadistico)
  if(pvalor < nivel){
    rechazo = rechazo + 1
  }
  estadisticos[irep] = estadistico
  pvalores[irep] = pvalor
  salida<-cbind(irep, pvalores[irep], estadisticos[irep],rechazo)
  largo.sal=length(salida)
  #Guardo en cada iteracion
  write(t(salida),file=nombre,ncol=largo.sal,append=T)
}
rechazo = rechazo / cantidad
return(list(pvalores = pvalores, estadisticos = estadisticos, rechazo = rechazo))
}

```

## 7.4 frobenius.R

```

source("auxiliares.R")

#####
# TEST DE FROBENIUS (SECCION 5.1)
# Se incluye el test  $T_{\{n,FROB\}}$  y  $T_{\{n,FROB2\}}$ 
#####

#####
#Obtiene el estadistico dadas las dos matrices de covarianza muestrales y el tamano de la ←
muestra original
#####
obtenerEstadistico = function(n,mat1,mat2){
  diferencia = mat1 - mat2
  norma =frobenius.norm(diferencia)
  norma2=n*norma^2
  return (norma2)
}

#####
#Estima  $E(s_{-i}(x)^2 * s_{-k}(x)^2)$  para todo  $k \geq i$ . Devuelve una matriz triangular superior ←
cuyo elemento (i,k) con  $k \geq i$  es la estimacion de
# $E(s_{-i}(x)^2 * s_{-k}(x)^2)$ . La muestra que recibe como parametro se supone ya normalizada.
#####
estimarEsperanzasDeProductosDeComponentesAlCuadrado = function(muestra){
  cuadrados = muestra^2
  n = nrow(muestra)
  p= ncol(muestra)

  cruzadas = matrix(rep(0,p^2),p,p)

```

```

for(i in 1:p){
  for(k in i:p){
    cruzadas[i,k]=mean(cuadrados[,i] * cuadrados[,k])
  }
}
return(cruzadas)
}

#####
#En lo que sigue, se llama U a la matriz limite de  $\sqrt{N}$  ( $V_1 - V_2$ ),
#donde  $V_i$  es la matriz de covarianza muestral de signos espaciales de la muestra i.
#
#Para las funciones que siguen:
#
#cruzadas1 y cruzadas2 son las matrices cuyo elemento (i,k) estima, para cada
#una de las dos muestras, a  $E(s_i(x)^2 * s_k(x)^2)$  ( $k \geq i$ )
#
#simples1 y simples2 son vectores cuya coordenada i estima, para cada una de las
#dos muestras, a  $E(s_i(x)^2)$ 
#
#tau_1 y tau_2 son las proporciones muestrales para cada grupo, es decir,  $n_1/n$  y  $n_2/n \leftarrow$ 
#respectivamente
#####

#####
#Estima el valor de  $cov(U_{ii}, U_{kk})$ ,  $k \geq i$ .
#####
estimarCovarianzaEntreElementosDeLaDiagonal = function(n1,n2,cruzadas1,cruzadas2,simples1←
,simples2,i,k){
  term1 = (1/n1 + 1/n2)*(n1* cruzadas1[i,k] + n2*cruzadas2[i,k])
  term2 = (1/(n1 + n2))*(1/n1 + 1/n2)*(n1*simples1[i] + n2*simples2[i])*(n1*simples1[k] + ←
n2*simples2[k])
  return(term1 - term2)
}

#####
#Estima el valor de  $var(U_{ij})$ ,  $j > i$ .
#####
estimarVarianzaParaElementosNoDiagonales = function(n1,n2,cruzadas1,cruzadas2,i,j){
  return((1/n1 + 1/n2)*(n1* cruzadas1[i,j] + n2*cruzadas2[i,j]))
}

#####
#Para lo que sigue, recordemos que la distribucion de  $tr(U^2)$  se escribe como
#  $2 \sum a_i y_i + ||Z||^2$  donde  $y_i$  son i.i.d.  $\chi^2_{-1}$ , Z es  $N(0, \Gamma)$  independiente
#de las  $y_i$ , el elemento (i,j) de Gamma es  $cov(U_{ii}, U_{jj})$  y los  $a_i$  son  $var(U_{ij}) \leftarrow$ 
#para  $j > i$ .
#Ver la seccion 5.1 de la tesis para una deducccion de esta cuenta.
#####

#####
#Obtiene el vector que corresponde a las estimaciones de las varianzas de los elementos
#no diagonales, para ser usado como los coeficientes estimados ( $\hat{a}_i$ )
#de la combinacion lineal de  $\chi^2_{-1}$  de la distribucion de  $tr(U^2)$ 
#####
obtenerVectorDeCombinacionLinealDeChisq = function(n1,n2,cruzadas1,cruzadas2){
  p = ncol(cruzadas1)
  ret = rep(0,(p-1)*p/2)
  index = 0
  for(i in 1:(p-1)){
    for(j in (i+1):p){
      index = index + 1
      ret[index] = estimarVarianzaParaElementosNoDiagonales(n1,n2,cruzadas1,cruzadas2,i,j)
    }
  }
}

```

```

return(ret)
}

#####
#Obtiene una matriz cuyo elemento (i,j) es la covarianza muestral entre U_{ii} y U_{jj}.
#Dicha matriz se usa para estimar \Gamma, es decir, la matriz de covarianza del vector Z
#que interviene en la distribucion limite de tr(U^2).
#####
obtenerMatrizDeCovarianzaDeLosElementosDiagonales = function(n1,n2,cruzadas1,cruzadas2, ←
  simples1,simples2){
  p = length(cruzadas1[1,])
  ret = matrix(rep(0,p^2),p,p)
  for(i in 1:p){
    for(j in i:p){
      valor = estimarCovarianzaEntreElementosDeLaDiagonal(n1,n2,cruzadas1,cruzadas2, ←
        simples1,simples2,i,j)
      ret[i,j] = valor
      ret[j,i] = valor
    }
  }
  return(ret)
}

#####
#Estimamos los percentiles de una distrubcion W obtenida como una combinacion lineal de ←
chi cuadrado
#con un grado de libertad mas la norma al cuadrado de una normal multivariada Z.
#Aca usamos que la norma al cuadrado de una normal multivariada se distribuye como
#una combinacion lineal de p chi^2_1 cuyos coeficientes son los autovalores de la
#matriz de covarianza de normal.
#percentiles es un vector que indica que percentiles se quieren estimar
#coeficientes es un vector que corresponde a los coeficientes de la comb. lineal de chi^2_ ←
1
#cova es la matriz de covarianza de la normal multivariada Z.
#simul.size es la cantidad se iteraciones para estimar los percentiles
#estad es un numero al que se le calcula el p-valor correspondiente a una hipotesis nula H ←
_0
#donde H_0 supone que estad proviene de una distribucion T.
#####
percentil.limite= function(percentiles ,coeficientes ,cova,p,simul.size = 2000,estad){
  #Obtenemos los autovalores de cova
  autovalores=eigen(cova)$values
  p=length(autovalores)
  numeroDeSumandosIndep = length(coeficientes)
  #Obtenemos las realizaciones
  realizaciones = rep(0,simul.size)
  for (iboot in 1:simul.size){
    normalesAleatorias = rnorm(numeroDeSumandosIndep)
    randomNormal = rnorm(p)
    realizaciones[iboot] = 2 * sum(normalesAleatorias^2 *coeficientes) + sum(randomNormal ←
^2*autovalores)
  }
  #Hallamos el pvalor estimado
  pvalor=mean(1*(estad<=realizaciones))
  return(list(cuantil=quantile(realizaciones,percentiles),pvalor=pvalor))
}

#####
#cov1 y cov2 son las matrices de covarianza de signos espaciales muestrales
#Devuelve el poolado entre estas dos matrices
#####
obtenerEstimadorDeSigma = function(n1,n2,cov1,cov2){
  return((n1*cov1 + n2*cov2)/(n1 + n2))
}

```

```

#####
#Devuelve la matriz de autovectores de una matriz simetrica
#Se utiliza como el estimador de la matriz beta del metodo F5
#####
obtenerEstimadorDeBeta = function(estSigma){
  return(eigen(estSigma,symmetric = T)$vectors)
}

#####
#cantidad es la cantidad de simulaciones a hacer
#n es el tamano de muestra para cada grupo
#p es la dimension de cada observacion
#sigma1 es la matriz de covarianza para obtener la m.a. del primer grupo
#sigma2 es la matriz de covarianza para obtener la m.a. del segundo grupo
#delta es el valor de los autovalores que diferencian a sigma1 y sigma2. Ver Seccion 6.1
#nivel es el percentil de la distribucion limite que se toma como valor critico (si el ←
estadistico observado da mayor, rechazamos, si no, no rechazamos)
#funcionParaLasMuestras es la funcion a utilizar para obtener las muestras aleatorias
#df1 y df2 son los grados de libertad a usar para cada grupo si se usan las distribuciones ←
T de Student multivariadas
#lambda1 y lambda2 son los parametros lambda a utilizar para cada grupo si se usan las ←
distribuciones exponencial multivariada.
#Si diagonal = T, usa el metodo F3. Si diagonal = F, usa el metodo F5.
#Devuelve los estadisticos obtenidos, los pvalores, los cuantiles correspondientes al
#nivel (valor critico) y la proporcion de rechazos.
#
#Funciones para muestras disponibles:
#obtenerMuestraNormal parametros: n,p,sigma
#obtenerMuestrat parametros: n,p,sigma,df
#obtenerMuestraExponencialMultivariada parametros:n,p,sqrtsigma,lambda
#####

correrTestFrobenius = function(cantidad,n1,n2,sigma1,sigma2,delta=1, ←
  funcionParaLasMuestras,nivel = 0.05, df1 = NULL, df2 = NULL,lambda1 = NULL, lambda2 = ←
  NULL, diagonal = T,...){
  p = length(sigma1[1,])

  #Nombre de los archivos de salida
  if(diagonal){
    output = obtenerNombreDeDirectorioYArchivoParaOutput(funcionParaLasMuestras = ←
      funcionParaLasMuestras, n1 = n1, n2 = n2, p = p, nombreDelTest = "FROB", delta = ←
      delta, df1 = df1, df2 = df2, lambda1 = lambda1, lambda2 = lambda2)
  }else{
    output = obtenerNombreDeDirectorioYArchivoParaOutput(funcionParaLasMuestras = ←
      funcionParaLasMuestras, n1 = n1, n2 = n2, p = p, nombreDelTest = "FROB2", delta = ←
      delta, df1 = df1, df2 = df2, lambda1 = lambda1, lambda2 = lambda2)
  }
  directorio = output$directorio
  archivo = output$archivo
  nombre = paste(directorio,"\\",archivo,sep = "")

  rechazo = 0
  pvalores=cuantiles=estadisticos=rep(0,cantidad)
  n = n1 + n2
  tau1 = n1/n
  tau2 = n2/n
  #Obtenemos las raices cuadradas de las matrices sigma.
  sqrtsigma1 = obtenerRaizCuadradaDeMatrizDiagonal(sigma1)
  sqrtsigma2 = obtenerRaizCuadradaDeMatrizDiagonal(sigma2)

  for(irep in 1:cantidad){
    set.seed(1234 + irep)
    if(diagonal){
      print(c("FROB irep=", irep, ",n1:", n1, ",n2:",n2,"p:",p, "delta:", delta, "df1:", ←
        df1, "lambda1:", lambda1))
    }
  }
}

```

```

} else {
  print(c("FROB2 irep=", irep, ",n1:", n1, ",n2:", n2, "p:", p, "delta:", delta, "df1:", ←
        df1, "lambda1:", lambda1))
}
#Obtengo las muestras
muestra1 = funcionParaLasMuestras(n = n1, p = p, sigma = sigma1, df = df1, sqrtsigma = ←
  sqrtsigma1, lambda = lambda1)
muestra2 = funcionParaLasMuestras(n = n2, p = p, sigma = sigma2, df = df2, sqrtsigma = ←
  sqrtsigma2, lambda = lambda2)
#Obtengo los signos de las muestras
s1 = apply(muestra1, 1, normalizar)
s1 = t(s1)
s2 = apply(muestra2, 1, normalizar)
s2 = t(s2)
#Estimamos las matrices de covarianza de los signos
cov1 = covarianza(s1)
cov2 = covarianza(s2)

#Rotamos las observaciones en el caso no diagonal
if(!diagonal){
  estSigma = obtenerEstimadorDeSigma(n1, n2, cov1, cov2)
  estBeta = obtenerEstimadorDeBeta(estSigma)
  s1 = s1 %*% estBeta
  s2 = s2 %*% estBeta
  cov1 = covarianza(s1)
  cov2 = covarianza(s2)
}

#Obtenemos el valor del estadístico
estadistico = obtenerEstadistico(n, cov1, cov2)
#Obtenemos las estimaciones
cruzadas1 = estimarEsperanzasDeProductosDeComponentesAlCuadrado(s1)
cruzadas2 = estimarEsperanzasDeProductosDeComponentesAlCuadrado(s2)
simples1 = diag(cov1)
simples2 = diag(cov2)
#Obtenemos el vector correspondiente a la combinación lineal de chisq de la ←
distribucion limite
vector.CL.Chisq = obtenerVectorDeCombinacionLinealDeChisq(n1, n2, cruzadas1, cruzadas2)
#Obtenemos la matriz de covarianza correspondiente a la cual se le toma la norma al ←
cuadrado en la distribucion limite
matriz.cov.diag = obtenerMatrizDeCovarianzaDeLosElementosDiagonales(n1, n2, cruzadas1, ←
  cruzadas2, simples1, simples2)
#Estimamos el percentil de la distribucion limite
valorCritico = percentil.limite(percentiles = 1-nivel, coeficientes=vector.CL.Chisq, ←
  cova=matriz.cov.diag, p, estad= estadistico)
#Obtenemos la proporcion de rechazos a H_0
if( estadistico > valorCritico$cuantil){
  rechazo = rechazo + 1
}

pvalores[irep] = valorCritico$pvalor
cuantiles[irep]=valorCritico$cuantil
estadisticos[irep]= estadistico
salida<-cbind(irep, pvalores[irep], estadisticos[irep], cuantiles[irep], rechazo)
largo.sal=length(salida)
# Guardo en cada iteracion
write(t(salida), file=nombre, ncol=largo.sal, append=T)
}

rechazo = rechazo/cantidad
return(list(rechazo=rechazo, pvalores=pvalores, cuantil=cuantiles, estadisticos=←
  estadisticos))
}

```

## 7.5 np.R

```

source("frobenius.R")

#####
# TEST (n,p)-ASINTOTICO (SECCION 5.2)
# Se incluyen los tests T_{n,NP} y T_{n,NP2}
#####

#####
#Obtiene el estadistico del test (n,p).
#n es la suma de los tamanos de las muestras
#cov1 y cov2 son las matrices de covarianza de signos espaciales para cada muestra
#tau_1 y tau_2 son las proporciones muestrales para cada grupo, es decir, n1/n y n2/n ←
respectivamente
#varTn es la estimacion de la varianza del estadistico Tn (usado en el test de Frobenius)
#Ver la seccion 4.2 de la tesis para una deduccion de esta cuenta.
#####
obtenerEstadisticoTestnp = function(n, cov1, cov2, tau1, tau2, varTn){
  ret = obtenerEstadistico(n, cov1, cov2)
  print(c("ret", ret))
  media = 1/tau1 + 1/tau2 - ((1/tau1) * sum(diag(cov1)^2) + (1/tau2) * sum(diag(cov2)^2))
  print(c("media:", media))
  print(c("var", varTn))
  return ((ret - media) / sqrt(varTn))
}

#####
#Obtiene un valor estimado de la varianza de Tn
#cruzadas1 y cruzadas2 son las mismas matrices usadas en el test de Frobenius
#simples1 y simples2 son los mismos vectores usados en el test de Frobenius
#tau_1 y tau_2 son las proporciones muestrales para cada grupo, es decir, n1/n y n2/n ←
respectivamente.
#Ver la seccion 4.2 de la tesis para una deduccion de esta cuenta.
#####
obtenerVarianzaEstimadaDeTn = function(n1, n2, cruzadas1, cruzadas2, simples1, simples2){
  p = length(cruzadas1[1,])
  ret = 0
  for(i in 1:(p-1)){
    for(j in (i+1):p){
      #sumamos 4 veces cada varianza estimada de elementos no diagonales de U.
      ret = ret + 4 * (estimarVarianzaParaElementosNoDiagonales(n1, n2, cruzadas1, cruzadas2, i←
, j)^2)
    }
  }
  #Obtenemos la matriz de covarianza \hat{\Gamma} de los elementos diagonales de U
  #y sumamos los cuadrados de sus autovalores.
  covDiagonal = obtenerMatrizDeCovarianzaDeLosElementosDiagonales(n1, n2, cruzadas1, ←
cruzadas2, simples1, simples2)
  ret = ret + sum(eigen(covDiagonal)$values^2)
  return(2 * ret)
}

#####
#cantidad es la cantidad de simulaciones a hacer
#n es el tamaño de muestra para cada grupo
#p es la dimension de cada observacion
#sigma1 es la matriz de covarianza para obtener la m.a. del primer grupo
#sigma2 es la matriz de covarianza para obtener la m.a. del segundo grupo
#delta es el valor de los autovalores que diferencian a sigma1 y sigma2. Ver Seccion 6.1
#nivel es el percentil de la distribucion limite que se toma como valor critico (si el ←
estadistico observado da mayor, rechazamos, si no, no rechazamos)

```

```

#funcionParaLasMuestras es la funcion a utilizar para obtener las muestras aleatorias
#df1 y df2 son los grados de libertad a usar para cada grupo si se usan las distribuciones ←
  T de Student multivariadas
#lambda1 y lambda2 son los parametros lambda a utilizar para cada grupo si se usan las ←
  distribuciones exponencial multivariada.
#Si diagonal = T, usa el metodo H1. Si diagonal = F, usa el metodo H2.
#Devuelve los estadisticos obtenidos, los pvalores, los cuantiles correspondientes al
#nivel (valor critico) y la proporcion de rechazos.
#
#Funciones para muestras disponibles:
#obtenerMuestraNormal   parametros: n,p,sigma
#obtenerMuestrat       parametros: n,p,sigma,df
#obtenerMuestraExponencialMultivariada parametros:n,p,sqrtsigma,lambda
#####

correrTestnp = function(cantidad,n1,n2,sigma1,sigma2,delta=1,funcionParaLasMuestras,nivel←
  = 0.05,df1= NULL, df2 = NULL, lambda1 = NULL, lambda2 = NULL,diagonal = T,...){
  p = length(sigma1[,])

  #Nombre de los archivos de salida
  if(diagonal){
    output = obtenerNombreDeDirectorioYArchivoParaOutput(funcionParaLasMuestras = ←
      funcionParaLasMuestras, n1 = n1, n2 = n2, p = p, nombreDelTest = "FROB_NP", delta ←
      = delta, df1 = df1, df2 = df2, lambda1 = lambda1, lambda2 = lambda2)
  }else{
    output = obtenerNombreDeDirectorioYArchivoParaOutput(funcionParaLasMuestras = ←
      funcionParaLasMuestras, n1 = n1, n2 = n2, p = p, nombreDelTest = "FROB_NP3", delta←
      = delta, df1 = df1, df2 = df2, lambda1 = lambda1, lambda2 = lambda2)
  }
  directorio = output$directorio
  archivo = output$archivo
  nombre = paste(directorio,"\\",archivo,sep = "")

  estadisticos = pvalores = rep(0,cantidad)
  rechazo = 0
  n = n1 + n2
  tau1 = n1 /n
  tau2 = n2 /n
  #Obtengo la raiz cuadrada de las matrices sigma1 y sigma2
  sqrtsigma1 = obtenerRaizCuadradaDeMatrizDiagonal(sigma1)
  sqrtsigma2 = obtenerRaizCuadradaDeMatrizDiagonal(sigma2)

  for(irep in 1:cantidad){
    set.seed(1234 + irep)
    if(diagonal){
      print(c("FROB_NP irep=", irep, ",n1:", n1, ",n2:",n2,"p:",p, "delta:", delta, "df1:"←
        , df1, "lambda1:", lambda1))
    }else{
      print(c("FROB_NP3 irep=", irep, ",n1:", n1, ",n2:",n2,"p:",p, "delta:", delta, "df1:"←
        , df1, "lambda1:", lambda1))
    }
    muestra1 = funcionParaLasMuestras(n = n1,p = p,sigma = sigma1,df = df1, sqrtsigma = ←
      sqrtsigma1,lambda = lambda1)
    muestra2 = funcionParaLasMuestras(n = n2,p = p,sigma = sigma2,df = df2, sqrtsigma = ←
      sqrtsigma2,lambda = lambda2)
    #Obtengo los signos de las muestras
    s1 =apply(muestra1,1,normalizar)
    s1=t(s1)
    s2= apply(muestra2,1,normalizar)
    s2= t(s2)
    #Obtengo las matrices de covarianza de signos espaciales
    cov1 = covarianza(s1)
    cov2 = covarianza(s2)

    #Rotamos las observaciones en el caso no diagonal

```

```

if(!diagonal){
  estSigma = obtenerEstimadorDeSigma(n1,n2,cov1,cov2)
  estBeta = obtenerEstimadorDeBeta(estSigma)
  s1 = s1 %*% estBeta
  s2 = s2 %*% estBeta
  cov1 = covarianza(s1)
  cov2 = covarianza(s2)
}
#Obtengo las estimaciones auxiliares
cruzadas1 = estimarEsperanzasDeProductosDeComponentesAlCuadrado(s1)
cruzadas2 = estimarEsperanzasDeProductosDeComponentesAlCuadrado(s2)
simples1 = diag(cov1)
simples2 = diag(cov2)
#Estimo la varianza de Tn (el estadistico del test Frobenius)
varTn = obtenerVarianzaEstimadaDeTn(n1,n2,cruzadas1,cruzadas2,simples1,simples2)
#Obtengo el estadistico del test (n,p) y el pvalor
estadistico = obtenerEstadisticoTestnp(n,cov1,cov2,tau1,tau2,varTn)
pvalor = 1 - pnorm(estadistico)
if(pvalor < nivel){
  rechazo = rechazo + 1
}

estadisticos[irep] = estadistico
pvalores[irep] = pvalor
salida<-cbind(irep, pvalores[irep],estadisticos[irep],rechazo)
largo.sal=length(salida)
# Guardo en cada iteracion
write(t(salida),file=nombre,ncol=largo.sal,append=T)
}
rechazo = rechazo / cantidad
return(list(pvalores = pvalores, estadisticos = estadisticos,rechazo = rechazo))
}

```

## 7.6 np3.R

```

source("frobenius.R")

#####
# TEST (n,p)-ASINTOTICO (SUBSECCION 5.2.2, METODO H3)
# Se incluye el test T- $\{n, NP3\}$ 
#####

#####
#Devuelve el estadistico del test
#####
obtenerEstadisticoTestnp3 = function(n,cov1,cov2,espTn,varTn){
  ret = obtenerEstadistico(n,cov1,cov2)
  print(c("ret",ret))
  print(c("media:", espTn))
  print(c("var", varTn))
  return ((ret - espTn) / sqrt(varTn))
}

#####
#Devuelve un estimador de la matriz Upsilon definida en la Seccion 5.1
#####
obtenerEstimadorDeUpsilon = function(muestra1,muestra2,tau1,tau2,estSigma){
  N = dim(muestra1)[1] + dim(muestra2)[1]
  p = dim(muestra1)[2]

```

```

term1 = matrix(rep(0,p^4),p^2,p^2)
muestras = rbind(muestral,muestra2)
for(i in 1:N){
  prodmat = muestras[i,] %*% t(muestras[i,])
  subterm = prodmat %x% prodmat
  term1 = term1 + subterm
}
term1 = term1/N
vecestSigma = vec(estSigma)
term2 = vecestSigma %*% t(vecestSigma)
dim(term2)
return((1/tau1 + 1/tau2)*(term1 - term2))
}

#####
#Devuelve un estimador de la esperanza del estadistico del test T_{n,FROB}
#####
obtenerEstimadorDeLaEsperanza = function(tau1,tau2,estSigma){
  return((1/tau1 + 1/tau2)*(1 - sum(estSigma^2)))
}

#####
#Devuelve el producto escalar entre dos vectores
#####
productoEscalar = function(vec1, vec2){
  return(sum(vec1*vec2))
}

#####
#Devuelve una matriz cuyo elemento (i,j) es el producto escalar entre la fila i-esima y j-
esima de la matriz muestra.
#####
obtenerMatrizDeProductosEscalares = function(muestra){
  N = dim(muestra)[1]
  ret = matrix(rep(0,N^2),N,N)
  for(i in 1:N){
    for(j in i:N){
      valor = productoEscalar(muestra[i,], muestra[j,])
      ret[i,j] = valor
      ret[j,i] = valor
    }
  }
  return(ret)
}

#####
#Devuelve un estimador de la varianza del estadistico del test T_{n,FROB}
#####
obtenerEstimadorDeLaVarianza = function(tau1,tau2,muestra){
  matrizDeEscalares = obtenerMatrizDeProductosEscalares(muestra)
  N = dim(muestra)[1]
  term1 = 0
  for(i in 1:N){
    for(j in 1:N){
      term1 = term1 + matrizDeEscalares[i,j]^4
    }
  }
  term1 = term1/(N^2)
  term2 = 0
  for(i in 1:N){
    parcial = 0
    for(j in 1:N){
      parcial = parcial + matrizDeEscalares[i,j]^2
    }
  }
  term2 = term2 + parcial^2
}

```

```

}
term2 = term2/(N^3)
term3 = 0
for(i in 1:N){
  for(j in 1:N){
    term3 = term3 + matrizDeEscalares[i,j]^2
  }
}
term3 = (term3^2) / (N^4)
return(2* ((1/taul + 1/tau2)^2) * (term1 - 2* term2 + term3))
}

#####
#cantidad es la cantidad de simulaciones a realizar
#n1 y n2 son los tamanos de las muestras
#p es la dimension de cada elemento de las muestras
#sigma 1 y sigma2 son las matrices de covarianza para cada una de las muestras
#nivel es el nivel del test
#df1 y df2 son los grados de libertad a usar para cada grupo si se usan las distribuciones ←
T de Student multivariadas
#lambda1 y lambda2 son los parametros lambda a utilizar para cada grupo si se usan las ←
distribuciones exponencial multivariada.
#Si ver1 = T se usa la version que estima Upsilon. En el otro caso se usa la version 2 que ←
usa la Proposicion 5.2.2.
#Devuelve los estadisticos obtenidos, los pvalores y la proporcion de rechazos
#
#Funciones para muestras disponibles:
#obtenerMuestraNormal parametros: n,p,sigma
#obtenerMuestrat parametros: n,p,sigma,df
#obtenerMuestraExponencialMultivariada parametros:n,p,sqrtsigma,lambda
#####

correrTestnp2v1 = function(cantidad,n1,n2,sigma1,sigma2,delta=1, funcionParaLasMuestras, ←
  nivel = 0.05,df1= NULL, df2 = NULL, lambda1 = NULL, lambda2 = NULL, ←
  numeroDePermutaciones = 1,ver1 = F){
  p = length(sigma1[1,])

  #Nombre de los archivos de salida
  output = obtenerNombreDeDirectorioYArchivoParaOutput(funcionParaLasMuestras = ←
    funcionParaLasMuestras, n1 = n1, n2 = n2, p = p, nombreDelTest = "TCMV", delta = ←
    delta, df1 = df1, df2 = df2, lambda1 = lambda1, lambda2 = lambda2)
  directorio = output$directorio
  archivo = output$archivo
  nombre = paste(directorio,"\\",archivo,sep = "")

  estadisticos = pvalores = rep(0,cantidad)
  rechazo = 0
  n = n1 + n2
  tau1 = n1 /n
  tau2 = n2 /n
  #Obtengo la raiz cuadrada de las matrices sigma1 y sigma2
  sqrtsigma1 = obtenerRaizCuadradaDeMatrizDiagonal(sigma1)
  sqrtsigma2 = obtenerRaizCuadradaDeMatrizDiagonal(sigma2)

  for(irep in 1:cantidad){
    set.seed(1234 + irep)
    print(c("FROB_NP3 irep=", irep, ",n1:", n1, ",n2:",n2, "p:",p, "delta:", delta, "df1:", ←
      df1, "lambda1:", lambda1))

    #Obtenemos las muestras
    muestral = funcionParaLasMuestras(n = n1,p = p,sigma = sigma1,df = df1, sqrtsigma = ←
      sqrtsigma1,lambda = lambda1)

```

```

muestra2 = funcionParaLasMuestras(n = n2,p = p,sigma = sigma2,df = df2, sqrtsigma = ←
  sqrtsigma2,lambda = lambda2)
#Obtengo los signos de las muestras
s1 =apply(muestra1,1,normalizar)
s1=t(s1)
s2= apply(muestra2,1,normalizar)
s2= t(s2)
s = rbind(s1,s2)

#Obtengo las matrices de covarianza de signos espaciales
cov1 = covarianza(s1)
cov2 = covarianza(s2)
estSigma = obtenerEstimadorDeSigma(n1,n2,cov1,cov2)
if(ver1){
  #Estimo Upsilon
  estUpsilon = obtenerEstimadorDeUpsilon(s1,s2,tau1,tau2,estSigma)
  #Estimo la esperanza y varianza
  esp = traza(estUpsilon)
  var = 2*sum(estUpsilon^2)
}else{
  #Estimo la esperanza y varianza
  esp = obtenerEstimadorDeLaEsperanza(tau1,tau2,estSigma)
  var = obtenerEstimadorDeLaVarianza(tau1,tau2,s)
}
estadistico = obtenerEstadisticoTestnp3(n,cov1,cov2,esp,var)

#Obtengo el p-valor
pvalor = 1 - pnorm(estadistico)
if(pvalor < nivel){
  rechazo = rechazo + 1
}
estadisticos[irep] = estadistico
pvalores[irep] = pvalor
salida<-cbind(irep, pvalores[irep],estadisticos[irep],rechazo)
largo.sal=length(salida)
#Guardo en cada iteracion
write(t(salida),file=nombre,ncol=largo.sal,append=T)
}
rechazo = rechazo / cantidad
return(list(pvalores = pvalores, estadisticos = estadisticos,rechazo = rechazo))
}

```

## 7.7 permutaciones.R

```

source("frobenius.R")
#####
# TEST DE PERMUTACIONES (SECCION 5.3)
# Se incluye el test T_{n,PERM}
#####

#####
# Devuelve dos muestras que se obtienen permutando los elementos de las dos muestras ←
  originales
#####
obtenerMuestrasPermutadas = function(n1,n2,p,muestra1,muestra2){
  n = n1 + n2
  #Pego las dos muestras
  bind = rbind(muestra1,muestra2)

```

```

#Mezclo las filas
bind = bind[sample(1:n),]
return(list(muestraNueva1 = bind[1:n1,], muestraNueva2 = bind[(n1+1):n,]))
}

#####
#Obtiene un vector donde cada elemento es el estadístico del test luego de permutar
#los elementos de las dos muestras. Las muestras ya se suponen normalizadas.
#cantidad es la cantidad de permutaciones a hacer
#n1 y n2 son el tamaño de cada muestra
#p es la dimensión de cada elemento de cada muestra
#muestra1 y muestra2 son las muestras ya normalizadas
#####
obtenerVectorDeEstadisticos.permutaciones = function(numeroDePermutaciones, n1, n2, p, ←
  muestra1, muestra2){
  n = n1 + n2
  ret = rep(0, numeroDePermutaciones)
  for(i in 1:numeroDePermutaciones){
    muestrasNuevas = obtenerMuestrasPermutadas(n1, n2, p, muestra1, muestra2)
    muestraNueva1 = muestrasNuevas$muestraNueva1
    muestraNueva2 = muestrasNuevas$muestraNueva2
    covNueva1 = covarianza(muestraNueva1)
    covNueva2 = covarianza(muestraNueva2)
    ret[i] = obtenerEstadistico(n, covNueva1, covNueva2)
  }
  return(ret)
}

#####
# Devuelve el pvalor, el estadístico y el cuantil de un test de permutaciones
# (ver especificaciones de la función "correrSimulacionTestDePermutaciones").
# Las muestras NO se suponen todavía normalizadas
#####
obtenerpValorYEstadisticoDelTestdePermutaciones = function(numeroDePermutaciones, n1, n2, p, ←
  muestra1, muestra2, nivel){
  #Normalizamos cada muestra
  s1 = apply(muestra1, 1, normalizar)
  s2 = apply(muestra2, 1, normalizar)
  s1 = t(s1)
  s2 = t(s2)
  n = n1 + n2
  #Obtenemos los estadísticos obtenido luego de las permutaciones
  vectorDeEstadisticosPermutados = obtenerVectorDeEstadisticos.permutaciones(←
    numeroDePermutaciones, n1, n2, p, s1, s2)

  #Obtenemos el estadístico para las muestras sin permutar
  cov1 = covarianza(s1)
  cov2 = covarianza(s2)
  estadistico = obtenerEstadistico(n, cov1, cov2)
  return(list(pval = 1 - obtenerAcumuladaParaEmpirica(estadistico, ←
    vectorDeEstadisticosPermutados), estadistico = estadistico, cuantil = quantile(←
    vectorDeEstadisticosPermutados, 1 - nivel)))
}

#####
#cantidad es la cantidad de simulaciones a hacer
#n es el tamaño de muestra para cada grupo
#p es la dimensión de cada observación
#sigma1 es la matriz de covarianza para obtener la m.a. del primer grupo
#sigma2 es la matriz de covarianza para obtener la m.a. del segundo grupo
#delta es el valor de los autovalores que diferencian a sigma1 y sigma2. Ver Sección 6.1
#nivel es el percentil de la distribución límite que se toma como valor crítico (si el ←
  estadístico observado da mayor, rechazamos, si no, no rechazamos)
#funcionParaLasMuestras es la función a utilizar para obtener las muestras aleatorias

```

```

#df1 y df2 son los grados de libertad a usar para cada grupo si se usan las distribuciones ←
  T de Student multivariadas
#lambda1 y lambda2 son los parametros lambda a utilizar para cada grupo si se usan las ←
  distribuciones exponencial multivariada.
#numeroDePermutaciones es la cantidad de permutaciones a realizar dentro de cada iteracion ←
.
#Devuelve los estadisticos obtenidos, los pvalores, los cuantiles correspondientes al
#nivel (valor critico) y la proporcion de rechazos.
#
#Funciones para muestras disponibles:
#obtenerMuestraNormal  parametros: n,p,sigma
#obtenerMuestrat  parametros: n,p,sigma,df
#obtenerMuestraExponencialMultivariada  parametros:n,p,sqrtsigma ,lambda
#####

correrTestDePermutaciones = function(cantidad,n1,n2,sigma1,sigma2,delta=1, ←
  funcionParaLasMuestras ,nivel = 0.05, df1 = NULL, df2 = NULL, lambda1 = NULL, lambda2 = ←
  NULL ,numeroDePermutaciones = 400){
  p = length(sigma1[1,])

  #Nombre de los archivos de salida
  output = obtenerNombreDeDirectorioYArchivoParaOutput(funcionParaLasMuestras = ←
    funcionParaLasMuestras , n1 = n1, n2 = n2, p = p, nombreDelTest = "PERM" , delta = ←
    delta, df1 = df1, df2 = df2, lambda1 = lambda1, lambda2 = lambda2)
  directorio = output$directorio
  archivo = output$archivo
  nombre = paste(directorio,"\\",archivo,sep = "")

  rechazo = 0
  pvalores = estadisticos = cuantiles = rep(0,cantidad)
  #Obtenemos la raiz cuadrada de las matrices sigma1 y sigma2
  sqrtsigma1 = obtenerRaizCuadradaDeMatrizDiagonal(sigma1)
  sqrtsigma2 = obtenerRaizCuadradaDeMatrizDiagonal(sigma2)

  for(irep in 1: cantidad){
    set.seed(1234 + irep)
    print(c("PERM irep=", irep, ",n1:" , n1, ",n2:" ,n2,"p:" ,p, "delta:" , delta, "df1:" , df1 ←
      , "lambda1:" , lambda1))
    #Obtenemos las muestras
    muestra1 = funcionParaLasMuestras(n = n1,p = p,sigma = sigma1,df = df1, sqrtsigma = ←
      sqrtsigma1,lambda = lambda1)
    muestra2 = funcionParaLasMuestras(n = n2,p = p,sigma = sigma2,df = df2, sqrtsigma = ←
      sqrtsigma2,lambda = lambda2)
    #Obtenemos el resultado del test
    resultadoDelTest = obtenerpValorYEstadisticoDelTestdePermutaciones( ←
      numeroDePermutaciones ,n1 ,n2 ,p ,muestra1 ,muestra2 ,nivel)
    pvalores[irep] = resultadoDelTest$pval
    estadisticos[irep] = resultadoDelTest$estadistico
    cuantiles[irep] = resultadoDelTest$cuantil

    if(resultadoDelTest$pval < nivel){
      rechazo = rechazo + 1
    }
    salida<-cbind(irep, pvalores[irep],estadisticos[irep],cuantiles[irep],rechazo)
    largo.sal=length(salida)
    # Guardo en cada iteracion
    write(t(salida),file=nombre,ncol=largo.sal,append=T)
  }
  rechazo = rechazo/cantidad
  return(list(rechazo = rechazo, pvalores = pvalores,estadisticos = estadisticos, ←
    cuantiles = cuantiles))
}

```

# Bibliografía

- [1] Aramaki, E., Maskawa, S. y Morita, M. (2011) Twitter Catches The Flu: Detecting Influenza Epidemics using Twitter.
- [2] Barnes, E. (1899) The theory of the gamma function. *Messenger Math.* **29**, 64-128.
- [3] Cai, T., Liu, W., Xia, Y. (2013) Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association* **108**, 265-277.
- [4] Durrett, R. (2010) Probability: Theory and Examples. *Cambridge U. Press*.
- [5] Fan, J., Han, F. y Liu, H. (2013) Challenges of Big Data Analysis. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1568-1576.
- [6] Fujikoshi, Y., Ulyanov, V. y Shimizu, R. (2011). Multivariate statistics: High-dimensional and large-sample approximations. *John Wiley and Sons*.
- [7] Hall, P., Heyde, C.C. (1980) Martingale Limit Theory and Its Application. *Academic Press*.
- [8] Hall, P., Marron, J.S., Neeman, A. (2005) Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B* **67 (3)** 427-444.
- [9] Jiang, T., Yang F. (2013) Central Limit Theorems for Classical Likelihood Ratio Tests for High-Dimensional Normal Distributions. *Annals of Statistics* **41** 2029-2074.
- [10] Li, J., Chen, S. (2012) Two Sample Tests for High Dimensional Covariance Matrices. *Ann. Statist.* **40 (2)** 908-940.
- [11] Ledoit, O., Wolf, M. (2002) Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30 (4)**, 1081-1102.
- [12] Lukacs, E. (1970) Characteristic Functions. *Charles Griffin & Co, London*.
- [13] Maygar, A., Tyler, D. (2014) The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions. *Biometrika* **101 Issue 3**, p673.
- [14] Muirhead, R. (1982) Aspects of Multivariate Statistical Theory. *John Wiley and Sons*.
- [15] Paindaveine, D., Verdebout, T. (2013) Universal Asymptotics for High-Dimensional Sign Tests.
- [16] Rao, C. (1973) Linear Statistical Inference and its Applications. *John Wiley and Sons*.
- [17] Rayleigh, L. (1919). On the problem of random vibrations and random ights in one, two and three dimensions. *Phil. Mag.* **37** 321-346.

- [18] Schott, J. (2007) A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics & Data Analysis* **51** 6535-6542.
- [19] Seber, G.A.F (1984) Multivariate Observations. *John Wiley and Sons*.
- [20] Sirkia, S., Taskinen, S., Oja, H., Tyler, D. (2009) Tests and estimates of shape based on spatial signs and ranks. *J. Nonparametr. Stat.*, **21** (2), 155-176.
- [21] Srivastava, M. (2005) Some tests concerning the covariance matrix in high-dimensional data. *J. Japan Statist. Soc. Vol.* **35** (2) 251-272.
- [22] Srivastava, M., Yanagihara H. (2010) Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis* **101** 1319-1320.
- [23] Wakaki, H. (2007) Error bounds for highdimensional Edgeworth expansions for some tests on covariance matrices. *Hiroshima Statistical Research Group Technical Report, (07-04)*.
- [24] Zou, C., Peng, L., Feng, L., Wang, Z. (2013). Multivariate sign-based high-dimensional tests for sphericity. *Biometrika*, *ast040*.