



UNIVERSIDAD NACIONAL DEL LITORAL  
Facultad de Ingeniería Química

## Dimensión efectiva en escenarios de alta dimensionalidad

María Antonella Gieco

TESIS PRESENTADA COMO PARTE DE LOS REQUISITOS DE LA  
UNIVERSIDAD NACIONAL DEL LITORAL PARA LA OBTENCIÓN DEL GRADO DE

**Doctor en Matemática**

EN EL CAMPO DE: **Estadística**

INSTITUCIÓN DONDE SE REALIZÓ:

Instituto de Matemática Aplicada y Facultad de Ingeniería Química  
(CONICET-UNL)

DIRECTORES DE TESIS:

Dra. Liliana Forzani y Dr. Carlos Tolmasky

DEFENDIDA ANTE EL JURADO COMPUESTO POR:

XX

XX

XX

AÑO DE PRESENTACIÓN: 2018



---

# Índice General

---

<b>Resumen</b> .....	III
<b>Introducción</b> .....	V
<b>Notación</b> .....	XI
<b>Capítulo 1. Resultados conocidos en baja dimensión</b> .....	1
1.1 Test de Esfericidad.....	2
1.1.1 Distribución asintótica bajo la hipótesis nula.....	4
1.2 Test de Esfericidad Parcial.....	5
1.2.1 Distribución asintótica bajo la hipótesis nula.....	9
<b>Capítulo 2. Estadístico para esfericidad y esfericidad parcial en alta dimensión</b>	11
2.1 Resultados conocidos en alta dimensión.....	11
2.1.1 Modelos esféricos.....	12
2.1.2 Modelos spiked.....	16
2.2 Distribución asintótica del $\log(LRT_d)$ bajo la hipótesis $\mathcal{H}_d$ .....	22
2.3 Simulaciones para la distribución asintótica.....	26
<b>Capítulo 3. Test de esfericidad parcial en alta dimensión</b> .....	39
3.1 Test para determinar la dimensión $d$ del subespacio de los spikes.....	39
3.2 Simulaciones.....	41
3.2.1 Resultados de Passemier and Yao [36].....	46
<b>Capítulo 4. Potencia del test propuesto</b> .....	49

4.1	Distribución asintótica de $\log(LRT_d)$ cuando la verdadera dimensión es $d_1$ .....	49
4.2	Potencia del Test .....	50
4.3	Simulaciones .....	51
4.4	Algunos comentarios .....	55
<b>Capítulo 5.</b>	<b>Cociente de verosimilitudes penalizado</b> .....	<b>57</b>
5.1	Intuición para el estadístico penalizado .....	57
5.2	Versión penalizada del cociente de verosimilitudes .....	58
5.3	Comparación con el método de Kritchman–Nadler .....	60
<b>Conclusiones generales</b> .....	<b>69</b>	
<b>Apéndice A.</b>	<b>Resultados conocidos</b> .....	<b>73</b>
<b>Apéndice B.</b>	<b>Resultados auxiliares</b> .....	<b>77</b>
<b>Apéndice C.</b>	<b>Demostraciones del Capítulo 2</b> .....	<b>83</b>
C.1	Demostración de la Proposición 2.4 .....	83
C.2	Demostración de la Proposición 2.5 .....	91
<b>Apéndice D.</b>	<b>Demostraciones del Capítulo 3</b> .....	<b>99</b>
D.1	Demostración de la Proposición 3.1 .....	99
D.1.1	Caso $p < m$ .....	99
D.1.2	Caso $p > m + d + 3$ .....	102
D.1.3	Caso $y = 1$ .....	102
<b>Apéndice E.</b>	<b>Demostraciones del Capítulo 4</b> .....	<b>103</b>
E.1	Demostración de la Proposición 4.1 .....	103
E.2	Demostración de la Proposición 4.2 .....	106
<b>Apéndice F.</b>	<b>Demostraciones del Capítulo 5</b> .....	<b>111</b>
F.1	Demostración de la Proposición 5.1 .....	111
F.2	Demostración de la Proposición 5.3 .....	112
<b>Bibliografía</b> .....	<b>115</b>	

---

# Resumen

---

Los avances tecnológicos permiten actualmente medir una gran cantidad de variables asociadas a un fenómeno de interés y almacenar los datos en forma relativamente económica. Sin embargo, frecuentemente muchas de estas variables en realidad no contribuyen información relevante, sino que pueden tener un aporte redundante o introducir ruido que no contribuye a explicar el fenómeno bajo estudio. Además, en algunos contextos como la investigación biomédica, es cada vez más sencillo medir nuevas variables, pero resulta inviable aumentar la cantidad de individuos incluidos en tales estudios. Estas condiciones dan lugar a problemas en los que la cantidad de variables  $p$  puede ser del mismo orden que la cantidad de casos disponibles en la muestra,  $n$ , o incluso mucho mayor. Tales escenarios se denominan frecuentemente de alta y ultra-alta dimensionalidad, respectivamente. En este contexto, la varianza de las estimaciones adquiere un rol decisivo, de modo que buena parte de la teoría y de la metodología estadística en alta dimensión intenta cuantificar y mitigar la variabilidad de los estimadores obtenidos con muestras finitas.

Una opción común en estos escenarios es reducir la dimensionalidad del espacio de predictores, intentando no sacrificar, en la medida de lo posible, la cantidad de información disponible. La hipótesis subyacente, ampliamente aceptada, es que los datos usualmente presentan una *dimensión efectiva* menor que la observada o, dicho de otro modo, que la información relevante reside en un espacio de baja dimensión pero la observamos corrompida con ruido de alta dimensionalidad. De este modo, el problema de estimar correctamente tal dimensión efectiva reviste un gran interés teórico y práctico.

En escenarios de alta dimensionalidad, este problema se ha abordado recientemente combinando elementos de la teoría espectral de matrices aleatorias [51] junto con una generalización de la condición de esfericidad parcial de las matrices de covarianza, conocida como *modelo spike*, introducido por Johnstone [23], en el cual todos los autovalores poblacionales, con excepción de los primeros (mayores, fijos) son iguales. El comportamiento de los autovalores muestrales correspondientes a estos modelos ha sido ampliamente estudiado en la década pasada [7, 8, 38]. El problema de estimar cuántas componentes deben ser consideradas ha sido estudiado y resuelto en el caso tradicional de  $p$  fijo y  $n$  creciendo a infinito [33]. Sin embargo, en escenarios de alta dimensión, los estadísticos usuales ya no pueden aplicarse y la teoría estadística tradicional no resulta adecuada ni tampoco fácil de adaptar.

En este trabajo de tesis introducimos nuevos procedimientos para inferir la existencia de una estructura lineal subyacente en los datos y su dimensión efectiva  $d$ . La propuesta se basa en una nueva mirada al cociente de verosimilitudes para el test de esfericidad parcial. Una de las contribuciones específicas de esta tesis es estudiar la distribución asintótica del logaritmo de este estadístico cuando  $p, n \rightarrow \infty$  con  $p/n \rightarrow y \in (0, 1)$  lo que nos permite introducir un test secuencial para estimar la dimensión  $d$  del subespacio informativo en el modelo spike. También abordamos el problema de ultra-alta dimensión; es decir,  $p, n \rightarrow \infty$ ,  $p/n \rightarrow y \in [1, \infty)$ . En este caso, el estadístico de cociente de verosimilitudes no está definido. Proponemos entonces una nueva definición del estadístico adaptada a este caso, y estudiamos la distribución de este nuevo estadístico bajo la hipótesis nula. Finalmente, estudiamos la potencia de los tests propuestos y a partir de este análisis introducimos también una corrección que permite extender la definición del estimador basado en criterios de información propuesto anteriormente por Wax and Kailath [50] y Kritchman and Nadler [25]. Probamos su consistencia en escenarios de alta y ultra-alta dimensión. Este nuevo estimador nos permite testear la dimensión del subespacio spiked incluso para valores de  $p/n$  cercanos a 1, situación en la que otros enfoques resultan deficientes [25]. Además de las garantías teóricas, el procedimiento propuesto para estimar  $d$  logra resultados competitivos respecto a las técnicas del estado del arte.

---

# Introducción

---

Los avances tecnológicos permiten actualmente medir una gran cantidad de variables asociadas a un fenómeno de interés y almacenar los datos en forma relativamente económica. En ciencias biológicas, por ejemplo, es posible medir simultáneamente la expresión de miles de genes para estudiar su asociación con una característica fenotípica determinada y explorar con ello el origen de ciertas enfermedades. La expresión de cada gen es una variable en el problema estadístico asociado. De forma similar, es posible estudiar la carga genómica de un suelo derivada de su composición microbiológica para estudiar su influencia sobre el rendimiento de semillas y cultivos productivos. En el ámbito médico, es habitual registrar imágenes de alta resolución de diversas modalidades para facilitar el diagnóstico. Aquí, la intensidad en cada pixel de la imagen es una variable en el problema estadístico.

Fuera del contexto científico, el uso de internet, redes sociales, plataformas de entretenimiento en línea y comercio electrónico también generan abundantes datos en forma continua. La masividad de estos datos, en volumen y cantidad de variables, ayudan a conjeturar que puede extraerse de ellos patrones de regularidad subyacentes que ayuden a explicar diversos fenómenos, desde aquellos que se perciben relacionados con los datos, como las preferencias de consumo en sistemas de recomendación, como también otros indirectos como la predicción de variables socio-económicas o la prevalencia de enfermedades [32, 40].

Una característica compartida por ambos escenarios es que muchas de estas variables en realidad no contribuyen información nueva para explicar el fenómeno de interés, sino que pueden tener un aporte altamente redundante o directamente introducir ruido que no contribuye a explicar el fenómeno bajo estudio.

Además, en algunos contextos como la investigación biomédica, es cada vez más sencillo medir nuevas variables, pero resulta a menudo inviable aumentar la cantidad de individuos incluidos en tales estudios. Estas condiciones dan lugar a problemas en los que la cantidad de variables  $p$  puede ser del mismo orden que la cantidad de casos disponibles en la muestra,  $n$ , o incluso mucho mayor. Tales escenarios se denominan frecuentemente de alta y ultra-alta dimensionalidad, respectivamente. En este contexto, la varianza de las estimaciones adquiere un rol decisivo, de modo que buena parte de la teoría y de la metodología estadística en alta dimensionalidad intenta cuantificar y mitigar la variabilidad de los estimadores obtenidos con muestras finitas [48].

Desde el punto de vista de las aplicaciones, una opción frecuente es reducir la dimensionalidad de los datos como primer paso del análisis, ya sea seleccionando un subconjunto de las variables disponibles o escogiendo transformaciones que tienen por imagen un conjunto de menor dimensión. Esta *reducción de dimensiones* facilita el análisis exploratorio y constituye un medio rápido para poder aplicar análisis subsecuentes sobre las nuevas variables utilizando metodología estadística tradicional.

La hipótesis subyacente, ampliamente aceptada, es que los datos usualmente presentan una *dimensión efectiva* menor que la observada o, dicho de otro modo, que la información relevante reside en un espacio de baja dimensión pero la observamos corrompida con ruido de alta dimensionalidad. En este sentido, el *análisis de componentes principales* continúa siendo el enfoque más usado en aplicaciones para recuperar esa información relevante de baja dimensión. La solución consiste en proyectar los datos en las direcciones indicadas por los autovectores de la matriz de covarianza de los datos que corresponden a sus autovalores más grandes. Esta solución es la mejor aproximación afín de rango  $d < p$  en un sentido de mínimos cuadrados y es también la proyección  $d$ -dimensional que mejor preserva la variabilidad total en los datos [24]. A pesar de las diferentes mejoras y extensiones de la metodología a lo largo de los años (por ejemplo, variantes sparse, robustas o adaptadas a datos discretos), un problema mucho menos abordado en forma rigurosa es decidir cuál es esa dimensión efectiva, denotada aquí por  $d$ , que caracteriza los datos. La elección correcta de  $d$  es importante no sólo para no perder información relevante, sino también para modelar los datos de forma eficiente.

### **Estimación de la dimensión efectiva**

Una heurística usada con frecuencia para elegir la dimensión  $d$  se basa en graficar la fracción de varianza total capturada por las primeras  $k$  componentes principales a medida que  $k$  crece.

El gráfico se llama *scree plot* y la intención es capturar aquellas direcciones que contribuyen una fracción de varianza que es significativamente mayor que la del resto, lo que se traduce en un codo o quiebre abrupto en la gráfica. Con datos reales, no obstante, ese quiebre muchas veces no es apreciable y se recurre a tomar aquellas componentes que suman en conjunto una fracción pre-establecida de la variabilidad total. Variantes gráficas similares se describen por ejemplo en [17], sin que ninguna de ellas logre resultados completamente satisfactorios. Una consecuencia de esta elección es que pocas veces se presta atención a la naturaleza de la información que se desecha en las componentes descartadas.

Desde un punto de vista más riguroso, cuando  $p$  es fijo y la muestra disponible es arbitrariamente grande ( $n \rightarrow \infty$ ), la elección de la dimensión efectiva  $d$  puede basarse en pruebas estadísticas bien conocidas. El punto de partida para estos tests es observar que para que exista ese subespacio característico de menor dimensión, los datos no deben presentar simetría esférica. Esta condición queda expresada por la igualdad de todos los autovalores de la matriz de covarianzas. Así, para que la reducción dimensional sea posible, esta igualdad no debe abarcar los  $p$  autovalores sino a una cantidad  $(p - d) < p$  de ellos, en particular, aquellos más chicos. Por este motivo los procedimientos clásicos que intentan inferir  $d$  se conocen como *tests de esfericidad parcial*. Suponiendo que los datos tienen distribución normal, se conoce la distribución asintótica del logaritmo del estadístico correspondiente al cociente de verosimilitudes para testear la hipótesis nula de que la dimensión efectiva es  $d$  contra la alternativa que es mayor [33]. Esto permite construir tests secuenciales para inferir la dimensión, comenzando con la hipótesis  $d = 0$  y deteniendo el procedimiento cuando no puede rechazarse la hipótesis nula. Conocer la distribución asintótica del logaritmo del cociente de verosimilitudes también permite establecer estimadores de  $d$  basados en criterios de información [41, 42].

La extensión de estos resultados a contextos de alta dimensionalidad no es trivial. Por ejemplo, Kritchman and Nadler [25] mostraron que usando el estimador propuesto en [42] no es posible detectar las componentes significativas cuando su magnitud es similar a las que no aportan información relevante, lo que usualmente resulta en una subestimación de  $d$ , especialmente cuando  $n$  es chico (problema de baja potencia). Más recientemente, el problema de estimar la dimensión  $d$  en escenarios de alta dimensionalidad se ha abordado combinando elementos de la teoría espectral de matrices aleatorias [51] junto con una generalización de la condición de esfericidad parcial de las matrices de covarianza, conocida como *modelo spike* [23]. Brevemente, este modelo supone que las componentes del vector aleatorio son independientes, que  $p - d$  de ellas tienen una misma varianza  $\sigma^2$  y que las restantes tienen una varianza mayor  $\lambda_i > \sigma^2$ , para  $i = 1, \dots, d$ . Supongamos por un momento que los datos presentan una estructura lineal

subyacente de dimensión  $d = 1$ . Su matriz de covarianza bajo el modelo spike puede escribirse como  $\Sigma = \sigma^2 \mathbf{I}_p + \lambda \mathbf{v}\mathbf{v}^\top$ , donde  $\lambda > 0$ . Bajo condiciones de esfericidad ( $\lambda = 0$ ), la teoría de matrices aleatorias establece que los autovalores de la covarianza muestral  $\mathbf{S}_n$  tienen una distribución definida cuando  $p, n, \rightarrow \infty$  tales que  $p/n \rightarrow y \in (0, \infty)$ , conocida como distribución de Marchenko-Pastur [51]. Esta distribución (y su soporte) depende de  $\sigma^2$  y de la relación  $p/n$ . Por otra parte, cuando  $\lambda$  toma un valor grande, el espectro de  $\mathbf{S}_n$  muestra un autovalor claramente separado del resto, que si responden bien a lo predicho por la distribución de Marchenko-Pastur. Esta información es usada en algunos procedimientos para estimar  $d$ , por ejemplo los propuestos en [26, 39]. No obstante, estos procedimientos a menudo guardan componentes heurísticas y no estudian formalmente el escenario en el que tanto  $p$  como  $n$  crecen arbitrariamente pero guardando una relación fija  $p/n > 1$ , como es característico en los escenarios de ultra-alta dimensionalidad.

## Contribución

En este trabajo de tesis introducimos nuevos procedimientos para inferir la existencia de una estructura lineal subyacente en los datos y su dimensión efectiva  $d$ . La propuesta se basa en una nueva mirada al cociente de verosimilitudes para el test de esfericidad parcial. Una de las contribuciones específicas de esta tesis es estudiar la distribución asintótica del logaritmo de este estadístico cuando  $p, n \rightarrow \infty$  con  $p/n \rightarrow y \in (0, 1)$ . Esto nos permite introducir un test secuencial para estimar la dimensión  $d$  del subespacio informativo en el modelo spike, corregir el estimador basado en criterios de información propuesto anteriormente por Kritchman and Nadler [25] y probar su consistencia en escenarios de alta dimensión.

También abordamos el problema de ultra-alta dimensionalidad, con  $p, n \rightarrow \infty$  y  $p/n \rightarrow y \in [1, \infty)$ . En este contexto el estadístico basado en el cociente de verosimilitudes no está definido, pero motivamos una nueva definición intercambiando los roles de las filas y columnas de la matrix de datos. Estudiando su distribución asintótica, proponemos un test secuencial válido en el contexto de ultra-alta dimensión y extendemos el estimador basado en criterios de información a este nuevo escenario.

La reunión de estas contribuciones específicas nos permite unificar un test secuencial y un procedimiento para estimar  $d$  que, a diferencia de los resultados conocidos, son válidos en cualquier escenario de dimensionalidad. Además, el estudio de la potencia de estas pruebas muestran una ventaja competitiva respecto a los procedimientos conocidos.

## Publicaciones

La mayor parte de los resultados de esta tesis se incluyeron en un artículo publicado en el *Journal of Multivariate Analysis* (2017) [13]. Resultados parciales se publicaron en congresos de la especialidad, tanto en el ámbito nacional como internacional.

- “*Likelihood ratio test for partial sphericity in high and ultrahigh dimensions*”. *Journal of Multivariate Analysis*, Volumen 159, Julio 2017, pág. 18-38, ISSN 0047-259X, con L. Forzani y C. Tolmasky, <https://doi.org/10.1016/j.jmva.2017.04.001>.
- “*Likelihood ratio test for partial sphericity in high and ultra-high dimensions*”. 9th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2016). Universidad de Sevilla - España - 2016.
- “*Tests de esfericidad parcial en alta dimensión*”. LXV Reunión anual de comunicaciones científicas. Unión Matemática Argentina - Bahía Blanca - 2016.

## Organización general de la Tesis

La tesis esta organizada de la siguiente manera. En el Capítulo 1 presentamos resultados conocidos en el contexto clásico de baja dimensión de los datos. En el Capítulo 2 introducimos el estadístico para los tests de esfericidad y esfericidad parcial en alta y ultra-alta dimensión y estudiamos su distribución asintótica bajo la hipótesis nula. Además, presentamos estudios de simulación realizados para analizar el comportamiento de la distribución asintótica obtenida.

A partir de la distribución encontrada en el Capítulo 2, en el Capítulo 3 obtenemos el test de esfericidad y esfericidad parcial y en el Capítulo 4 estudiamos su potencia asintótica, para lo cual, previamente, estudiamos la distribución asintótica del estadístico bajo una alternativa específica. Incluimos además en cada uno de los capítulos una serie de simulaciones que nos permiten evaluar el desempeño del test y compararlo con otras técnicas ya existentes.

Finalmente, a partir del análisis realizado en el Capítulo 4, en el Capítulo 5 introducimos un nuevo estimador la dimensión  $d$  del subespacio spiked y mostramos su consistencia en escenarios de alta y ultra-alta dimensionalidad.

Las pruebas de los resultados de cada capítulo se ofrecen al final de la tesis, en apéndices separados.



---

# Notación

---

## Básica

$\mathbb{R}^p$	Espacio euclídeo de dimensión $p$ (de todos los vectores reales de $p \times 1$ ).
$\mathbb{R}^{p \times q}$	Conjunto de todas las matrices reales de dimensión $p \times q$ .
$\mathbf{A}^\top$	Traspuesta de la matriz $\mathbf{A}$ .
$\text{tr}(\mathbf{A})$	Traza de la matriz $\mathbf{A}$ .
$ \mathbf{A} $	Determinante de la matriz $\mathbf{A}$ .
$\text{vec}(\mathbf{A})$	Operador que vectoriza la matriz $\mathbf{A}$ apilando sus columnas.
$\mathbf{A} \otimes \mathbf{B}$	Producto Kronecker de las matrices $\mathbf{A}$ y $\mathbf{B}$ .
$\mathbf{A} > \mathbf{B}$	Indica el orden parcial entre las matrices definidas positivas $\mathbf{A}$ y $\mathbf{B}$ .
$e_i$	Vector canónico con un 1 en la $i$ -ésima posición.
$\mathbf{I}_p$	Matriz identidad de orden $p$ .
$\rho(\mathbf{A})$	Espectro de la matriz $\mathbf{A}$ .
$\lambda_r$	$r$ -ésimo mayor autovalor de la matriz de covarianza $\Sigma$ .
$l_r$	$r$ -ésimo mayor autovalor de la matriz de covarianza muestral $\mathbf{S}_n$ .
$\Gamma(\cdot)$	Función Gamma definida por $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ para $\text{Re}(z) > 0$ , donde $\text{Re}(z)$ es la parte real del número complejo $z$ .
$\Gamma_p(\cdot)$	Función Gamma multivariada (ver definición en (A.2)).
$\log(\cdot)$	Función logaritmo natural, esto es, de base $e$ .

## Estadística

$\Sigma$	Matriz de covarianza poblacional.
$\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$	Distribución normal multivariada en $\mathbb{R}^p$ con vector de medias $\boldsymbol{\mu}$ y matriz de covarianza $\Sigma$ .
$\mathcal{W}_p(m, \Sigma)$	Matriz aleatoria de $p \times p$ con distribución Wishart con $m$ grados de libertad y matriz de covarianza $\Sigma$ .
$\chi_f^2$	Distribución chi-cuadrado con $f$ grados de libertad.
$\xrightarrow{\text{c.s.}}$	Convergencia casi segura.
$\rightsquigarrow$	Convergencia en distribución.
$\Phi(\cdot)$	Función de distribución acumulada de la distribución normal estándar.
$M_X(\cdot)$	Función generadora de momentos de la variable aleatoria $X$ .
$\bar{\mathbf{x}}$	Vector promedio para la muestra $\mathbf{x}_1, \dots, \mathbf{x}_n$ .
$\mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$	Matriz de covarianza de la muestra $\mathbf{x}_1, \dots, \mathbf{x}_n$ . (Estimador insesgado de $\Sigma$ ).
$\hat{\Sigma} = \frac{n-1}{n} \mathbf{S}_n$	Estimador de máxima verosimilitud de $\Sigma$ .
$\mathbb{E}(\mathbf{X})$	Esperanza o valor esperado del vector aleatorio $\mathbf{X}$ .



# Resultados conocidos en baja dimensión

---

En muchas situaciones prácticas, al recoger información de un conjunto de datos, las observaciones se toman sobre un gran número de variables. A pesar de tener más información, esto no siempre resulta beneficioso para el análisis: si tomamos demasiadas variables, es muy probable que estén relacionadas o que midan lo mismo bajo distintos puntos de vista. Resulta natural entonces, analizar de qué manera se puede reducir el número de variables, sin sacrificar, en la medida de lo posible, la cantidad de información disponible.

Un enfoque adoptado comúnmente es asociar el concepto de mayor información con el de mayor variabilidad o varianza. Es decir, cuanto mayor es la variabilidad de los datos (varianza) se considera que existe mayor información. El Análisis Factorial y el Análisis de Componentes Principales (FA y PCA, respectivamente, por sus siglas en inglés) son ejemplos de técnicas de análisis multivariante de reducción de datos, basados en la búsqueda de direcciones que mejor expliquen la variabilidad común o la variabilidad total, respectivamente. En ambos métodos, para que tal reducción exista, es preciso que existan direcciones de mayor variabilidad. Recíprocamente, si en todas las variables existe la misma variabilidad (misma varianza) no tiene sentido aplicar PCA o FA, pues no hay reducción posible.

Por otra parte, en muchos sistemas naturales, los datos de alta dimensión a menudo están muy estructurados y, por lo tanto, pueden entenderse mejor en un subespacio de dimensión menor. Detectar dicha estructura ha sido el foco de estudios recientes en diversos campos aplicados. En particular, se ha dedicado especial atención al caso en que la covarianza de los datos puede escribirse como una perturbación finita de la matriz identidad. Esto significa que todos los

autovalores de  $\Sigma$  son iguales a 1, excepto por una cantidad finita de ellos, digamos  $k$ , que permanece fija en el análisis.

Por el contrario, en muchos análisis estadísticos (por ejemplo ANOVA con medidas repetidas) una suposición típica es que todas las variables aleatorias no están correlacionadas y tienen la misma varianza. Esta situación es conocida en la literatura como esfericidad de la matriz de covarianza. Notar que la hipótesis de variables no correlacionadas y con igual varianza puede ser interpretada en términos de los autovalores de la matriz de covarianza  $\Sigma$ : la hipótesis es cierta si y solo si todos los autovalores de la matriz de covarianza son iguales. Por su parte, el caso anterior, en el cual solo un subconjunto de los autovalores son iguales, recibe el nombre de esfericidad parcial.

Por lo antes dicho, resulta relevante considerar tests que nos permitan decidir, a partir de la información recogida en la muestra, si tales suposiciones se satisfacen. En este capítulo presentaremos los tests llamados de esfericidad y esfericidad parcial, para el escenario clásico de  $p$  fijo y  $n \rightarrow \infty$  que servirán como punto de partida para la obtención de los tests equivalentes en alta dimensión.

## 1.1 Test de Esfericidad

Supongamos que tenemos una muestra  $\mathbf{x}_1, \dots, \mathbf{x}_n$  de vectores aleatorios de dimensión  $p$ , independientes e idénticamente distribuidos  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$  y queremos testear la hipótesis nula  $\mathcal{H}_0 : \Sigma = \sigma^2 \mathbf{I}_p$  contra la alternativa  $\mathcal{H}_1 : \Sigma \neq \sigma^2 \mathbf{I}_p$ , con  $\sigma$  desconocido. Esta hipótesis nula es llamada *hipótesis de esfericidad* pues cuando es cierta, los contornos de igual densidad en la distribución normal, son esferas.

El test más utilizado en esta situación es el test de cociente de verosimilitudes (LRT por sus siglas en inglés) derivado por Mauchly [31], y que presentamos en el siguiente teorema.

**Teorema 1.1.** (Muirhead [33, Teorema 8.3.2]) Sean  $\mathbf{x}_1, \dots, \mathbf{x}_n$  vectores aleatorios independientes e idénticamente distribuidos  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$  y definamos

$$\mathbf{A} = n\hat{\Sigma} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top.$$

El test de cociente de verosimilitudes de nivel  $\alpha$  para las hipótesis  $\mathcal{H}_0 : \Sigma = \sigma^2 \mathbf{I}_p$  versus  $\mathcal{H}_a : \Sigma \neq \sigma^2 \mathbf{I}_p$  donde  $\sigma$  es desconocido, rechaza  $\mathcal{H}_0$  si

$$\mathbf{V}_0 = \frac{|\mathbf{A}|}{\left(\frac{1}{p} \operatorname{tr} \mathbf{A}\right)^p} = \frac{|\widehat{\Sigma}|}{\left(\frac{1}{p} \operatorname{tr} \widehat{\Sigma}\right)^p} \leq k_\alpha, \quad (1.1)$$

donde  $k_\alpha$  se elige de modo tal que el tamaño del test sea  $\alpha$ .

**Demostración del Teorema 1.1:** Salvo constantes, la función de verosimilitud es

$$L(\boldsymbol{\mu}, \Sigma) = |\Sigma|^{-n/2} e^{\operatorname{tr}(-\frac{1}{2}\Sigma^{-1}\mathbf{A})} \exp\left[-\frac{1}{2}n(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \Sigma^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})\right].$$

Por otra parte, el estadístico de cociente de verosimilitudes es

$$\boldsymbol{\Lambda} = \frac{\max_{\boldsymbol{\mu} \in \mathbb{R}^p, \sigma > 0} L(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_p)}{\max_{\boldsymbol{\mu} \in \mathbb{R}^p, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma)}. \quad (1.2)$$

El denominador de (1.2) se maximiza en  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$  y  $\hat{\Sigma} = \frac{1}{n} \mathbf{A}$  y dicho máximo es

$$\begin{aligned} \max_{\boldsymbol{\mu} \in \mathbb{R}^p, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma) &= L(\bar{\mathbf{x}}, \widehat{\Sigma}) \\ &= n^{np/2} e^{-np/2} |\mathbf{A}|^{-n/2}. \end{aligned} \quad (1.3)$$

Por su parte, el numerador en (1.2) se maximiza en  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$  y  $\hat{\sigma}^2 = \frac{1}{np} \operatorname{tr} \mathbf{A}$  y el valor máximo es

$$\begin{aligned} \max_{\boldsymbol{\mu} \in \mathbb{R}^p, \sigma > 0} L(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_p) &= L(\bar{\mathbf{x}}, \frac{\operatorname{tr} \mathbf{A}}{np} \mathbf{I}_p) \\ &= \left(\frac{1}{np} \operatorname{tr} \mathbf{A}\right)^{-np/2} e^{-np/2}. \end{aligned} \quad (1.4)$$

Luego, usando (1.3) y (1.4) en (1.2) tenemos

$$\boldsymbol{\Lambda}^{2/n} = \frac{|\mathbf{A}|}{\left(\frac{1}{p} \operatorname{tr} \mathbf{A}\right)^p} = \mathbf{V}_0.$$

El test de cociente de verosimilitudes rechaza la hipótesis  $\mathcal{H}_0$  si el estadístico de cociente de verosimilitudes  $\boldsymbol{\Lambda}$  es pequeño, lo cual es equivalente a rechazar la hipótesis  $\mathcal{H}_0$  si  $\mathbf{V}_0$  es pequeño.

Ahora bien, para poder elegir la constante  $k_\alpha$  en el Teorema 1.1 necesitamos conocer la distribución de  $\mathbf{V}_0$  bajo la hipótesis nula. Excepto en algunos casos especiales, las distribuciones exactas son extremadamente complicadas, por lo que, en general, sólo se obtienen las distribuciones asintóticas. En la siguiente sección daremos la distribución asintótica, bajo la hipótesis nula, del estadístico de esfericidad cuando la dimensión  $p$  permanece fija y el tamaño de muestra  $n$  crece a infinito.

### 1.1.1 Distribución asintótica bajo la hipótesis nula

El siguiente lema da la distribución límite, para  $n \rightarrow \infty$  y  $p$  fijo, bajo la hipótesis nula, del estadístico de cociente de verosimilitudes para el test de esfericidad.

**Lema 1.2.** *Cuando la hipótesis nula  $\mathcal{H}_0 : \Sigma = \sigma^2 \mathbf{I}_p$  es cierta,*

$$Z \doteq -2 \frac{m}{n} \rho \log(\Lambda) \rightarrow \chi_f^2, \quad (1.5)$$

donde  $f = (p+2)(p-1)/2$  y  $\rho = 1 - (2p^2 + p + 2)/(6pm)$ .

**Demostración del Lema 1.2:** Dado que la distribución  $\chi^2$  está caracterizada por sus momentos, el lema es consecuencia de probar que

$$M_Z(t) \rightarrow M_{\chi_f^2}(t) \quad \text{para } t \text{ cercano a } 0,$$

donde hemos denotado con  $M_X(t)$  a la función generadora de momentos de una variable aleatoria  $X$ . (ver Teorema 30.2 de Billingsley [10] y comentarios.) Ahora bien,

$$\begin{aligned} M_Z(t) &= \mathbb{E}(e^{tZ}) \\ &= \mathbb{E}(e^{-2t \frac{m}{n} \rho \log \Lambda}) \\ &= \mathbb{E}\left(\mathbf{V}_0^{-m\rho t}\right), \end{aligned}$$

motivo por el cual, nos interesa estudiar los momentos del estadístico  $\mathbf{V}_0$ . Utilizando el Corolario 8.3.6 de Muirhead [33] (ver A.5), tenemos que

$$\begin{aligned} \mathbb{E}\left(\mathbf{V}_0^{-m\rho t}\right) &= p^{-pm\rho t} \frac{\Gamma\left(\frac{1}{2}pm\right)}{\Gamma\left(\frac{1}{2}pm(1-2\rho t)\right)} \frac{\Gamma_p\left(\frac{1}{2}m(1-2\rho t)\right)}{\Gamma_p\left(\frac{1}{2}m\right)} \\ &= K p^{-pm\rho t} \frac{\Gamma_p\left(\frac{1}{2}m(1-2\rho t)\right)}{\Gamma\left(\frac{1}{2}pm(1-2\rho t)\right)} \\ &= K p^{-pm\rho t} \frac{\prod_{k=1}^p \Gamma\left(\frac{1}{2}m(1-2\rho t) - \frac{1}{2}(k-1)\right)}{\Gamma\left(\frac{1}{2}pm(1-2\rho t)\right)}, \end{aligned}$$

donde  $K$  es una constante que no involucra a  $t$ . Utilizando los resultados de la Sección 8.2.4 de Muirhead [33] con  $p = 1$ ,  $q = p$ ,  $x_k = m/2$ ,  $y_j = pm/2$ ,  $\xi_k = -(k-1)/2$  y  $\eta_k = 0$  con  $k = 1, \dots, p$  tenemos que la función generadora de momentos de  $Z$  definida en (1.5) es igual a

$$M_Z(t) = (1-2t)^{-f/2} [1 + O(n^{-2})], \quad (1.6)$$

la cual converge a la función generadora de momentos de una variable aleatoria con distribución chi-cuadrado con  $f$  grados de libertad, que es de la forma  $(1-2t)^{-f/2}$ , como queríamos probar.

□

**Observación 1.3.** En el Teorema anterior, si en lugar de elegir  $\rho$  como lo hicimos consideramos  $\rho = 1$  entonces en (1.6) se obtiene

$$M_Z(t) = (1 - 2t)^{-\frac{1}{2}f} [1 + O(n^{-1})].$$

Es decir, multiplicar por  $\rho = 1 - (2p^2 + p + 2)/(6pm)$  a  $-2\frac{m}{n}\log(\mathbf{\Lambda})$  hace que sea mejor (esto es, que converja más rápido) la aproximación obtenida.

Para finalizar la prueba del Teorema 1.1, debemos determinar el valor de  $k_\alpha$ . Como consecuencia del Lema 1.2, eligiendo  $k_\alpha$  de modo tal que  $\frac{n}{2}\log(k_\alpha) = -2\rho\frac{m}{n}q_{f,\alpha}$  donde  $q_{f,\alpha}$  es el cuantil  $\alpha$  de la distribución  $\chi_f^2$ , el test definido en (1.1) tendrá nivel asintótico igual a  $\alpha$ , lo cual completa la demostración del teorema.

□

## 1.2 Test de Esfericidad Parcial

En la Sección 1.1 hemos presentado el test de cociente de verosimilitudes para esfericidad de la matrices de covarianzas, esto es, para testear la hipótesis nula de que todos los autovalores de la matriz de covarianzas  $\Sigma$  son iguales. Si no rechazamos esta hipótesis, en contexto de componentes principales, concluimos que todas las componentes tienen la misma varianza y por lo tanto aportan igual cantidad a la variabilidad total, de modo que no existe reducción dimensional posible transformando las variables originales a las componentes principales. Por el contrario, si rechazamos la hipótesis nula de esfericidad (o igualdad de los autovalores), aún podría ser posible, por ejemplo, que los menores  $p - 1$  autovalores poblacionales sean iguales. Si esto es cierto y su valor común es pequeño comparado con el valor del autovalor más grande, entonces la mayoría de la variabilidad la explica la primera componente principal, dando lugar a una reducción sustancial de la dimensión de los datos. Resulta así razonable considerar la hipótesis nula de que los  $p - 1$  autovalores más chicos son iguales. Si rechazamos esta hipótesis nula, entonces aún podemos testear si los  $p - 2$  autovalores menores son iguales y así sucesivamente.

Es decir, en la práctica, testeamos secuencialmente la hipótesis nula

$$\mathcal{H}_d : \lambda_{d+1} = \dots = \lambda_p$$

para  $d = 0, 1, \dots, p - 2$ , donde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$  son los autovalores de la matriz de covarianzas poblacional  $\Sigma$ . Cuando la hipótesis nula  $\mathcal{H}_d$  es cierta, decimos que la matriz de covarianzas  $\Sigma$  se corresponde a un escenario de esfericidad parcial.

En la Sección 1.1 vimos que el test de cociente de verosimilitudes para la hipótesis

$$\mathcal{H}_0 : \lambda_1 = \dots = \lambda_p$$

está basado en el estadístico

$$\mathbf{V}_0 = \frac{|\mathbf{A}|}{\left(\frac{1}{p} \operatorname{tr} \mathbf{A}\right)^p} = \frac{\prod_{i=1}^p l_i}{\left(\frac{1}{p} \sum_{i=1}^p l_i\right)^p},$$

donde  $l_1 > \dots > l_p$  son los autovalores de la matriz de covarianzas muestral  $\mathbf{S}_n = \frac{1}{m} \mathbf{A} = \frac{1}{m} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$  ( $m = n - 1$ ).

Cuando la hipótesis que queremos testear es la igualdad de un subconjunto de los autovalores de la matriz  $\Sigma$ , más precisamente el subconjunto de los  $p - d$  autovalores menores, para  $1 \leq d \leq p - 2$ , entonces el estadístico de cociente de verosimilitudes es similar a  $\mathbf{V}_0$  excepto que sólo aparecen aquellos autovalores muestrales que se corresponden con los autovalores poblacionales cuya igualdad queremos testear. Esto es lo que se prueba en el siguiente teorema.

**Teorema 1.4.** (Teorema 9.6.1, Muirhead, 2005) *Dada una muestra de tamaño  $n$  de una población con distribución  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ , el estadístico de cociente de verosimilitudes para la hipótesis*

$$\mathcal{H}_d : \lambda_{d+1} = \dots = \lambda_p \quad (= \sigma^2, \text{ desconocido})$$

es  $\Lambda_d = \mathbf{V}_d^{n/2}$ , donde

$$\mathbf{V}_d = \frac{\prod_{i=d+1}^p l_i}{\left(\frac{1}{p-d} \sum_{i=d+1}^p l_i\right)^p}.$$

**Demostración del Teorema 1.4:** Como lo hicimos en la prueba del Teorema 1.1, vamos a maximizar el cociente de verosimilitudes. Recordemos que, ignorando las constantes, la función de verosimilitud es

$$L(\boldsymbol{\mu}, \Sigma) = |\Sigma|^{-n/2} e^{-\frac{1}{2} \operatorname{tr}(\Sigma^{-1} \mathbf{A})} e^{-\frac{1}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})}.$$

Entonces el estadístico resulta

$$\Lambda_d = \frac{\max_{\mathcal{H}_d} L(\boldsymbol{\mu}, \Sigma)}{\max_{\boldsymbol{\mu} \in \mathbb{R}^p, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma)}. \quad (1.7)$$

Como antes, el denominador de (1.7) se maximiza en  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$  y  $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\mathbf{A}$  y es

$$\begin{aligned} \max_{\boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} > 0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= L(\bar{\mathbf{x}}, \hat{\boldsymbol{\Sigma}}) \\ &= n^{np/2} e^{-np/2} |\mathbf{A}|^{-n/2}. \end{aligned} \quad (1.8)$$

Para maximizar el numerador, comencemos tomando derivadas respecto de  $\boldsymbol{\mu}$ . Tenemos entonces que, para cada  $\boldsymbol{\Sigma}$ ,  $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  se maximiza tomando  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ . Sólo resta maximizar la función

$$g(\boldsymbol{\Sigma}) = -\frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{A}) \quad (= \log(L(\bar{\mathbf{x}}, \boldsymbol{\Sigma}))). \quad (1.9)$$

Sean  $\mathbf{H}$  y  $\mathbf{Q}$  matrices ortogonales de  $p \times p$  tales que  $\boldsymbol{\Sigma} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^\top$  y  $\mathbf{S}_n = \mathbf{Q}\mathbf{L}\mathbf{Q}^\top$ , donde  $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_d, \sigma^2, \dots, \sigma^2)$  (recordemos que estamos bajo la hipótesis nula) y  $\mathbf{L} = \text{diag}(l_1, \dots, l_p)$ . Reemplazando entonces en (1.9), tenemos

$$\begin{aligned} g(\boldsymbol{\Sigma}) &= -\frac{n}{2} \log |\boldsymbol{\Lambda}| - \frac{1}{2} \text{tr} [m\mathbf{H}\boldsymbol{\Lambda}^{-1}\mathbf{H}^\top\mathbf{Q}\mathbf{L}\mathbf{Q}^\top] \\ &= -\frac{n}{2} \sum_{i=1}^d \log(\lambda_i) - \frac{n}{2}(p-d) \log(\sigma^2) - \frac{m}{2} \text{tr} [\boldsymbol{\Lambda}^{-1}\mathbf{P}^\top\mathbf{L}\mathbf{P}], \end{aligned} \quad (1.10)$$

donde  $\mathbf{P} = \mathbf{Q}^\top\mathbf{H} \in O(p)$ .

Particionamos ahora  $\mathbf{P} = [\mathbf{P}_1 : \mathbf{P}_2]$  de modo tal que  $\mathbf{P}_1 \in \mathbb{R}^{p \times d}$  y  $\mathbf{P}_2 \in \mathbb{R}^{p \times (p-d)}$ , y escribimos  $\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{p-d} \end{pmatrix}$  con  $\boldsymbol{\Lambda}_1 = \text{diag}(\lambda_1, \dots, \lambda_d)$ . Luego,

$$\begin{aligned} \text{tr}(\boldsymbol{\Lambda}^{-1}\mathbf{P}^\top\mathbf{L}\mathbf{P}) &= \text{tr} \left( \begin{bmatrix} \boldsymbol{\Lambda}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \sigma^{-2} \mathbf{I}_{p-d} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1^\top \\ \mathbf{P}_2^\top \end{bmatrix} \mathbf{L} \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_2 \end{bmatrix} \right) \\ &= \text{tr}(\boldsymbol{\Lambda}_1^{-1}\mathbf{P}_1^\top\mathbf{L}\mathbf{P}_1) + \frac{1}{\sigma^2} \text{tr}(\mathbf{L}\mathbf{P}_2\mathbf{P}_2^\top) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^p l_i + \text{tr} \left( \left[ \boldsymbol{\Lambda}_1^{-1} - \frac{1}{\sigma^2} \mathbf{I}_d \right] \mathbf{P}_1^\top\mathbf{L}\mathbf{P}_1 \right), \end{aligned}$$

donde hemos usado que  $\mathbf{P}_2\mathbf{P}_2^\top = \mathbf{I}_p - \mathbf{P}_1\mathbf{P}_1^\top$ . Por lo tanto, si reemplazamos en (1.10) tenemos

$$g(\boldsymbol{\Sigma}) = -\frac{n}{2} \sum_{i=1}^d \log(\lambda_i) - \frac{n}{2}(p-d) \log(\sigma^2) - \frac{m}{2\sigma^2} \sum_{i=1}^p l_i + \frac{m}{2} \text{tr} \left( \left[ \frac{1}{\sigma^2} \mathbf{I}_d - \boldsymbol{\Lambda}_1^{-1} \right] \mathbf{P}_1^\top\mathbf{L}\mathbf{P}_1 \right). \quad (1.11)$$

Ahora bien, es una cuestión sencilla de mostrar que si  $\mathbf{U} = \text{diag}(u_1, \dots, u_d)$  con  $u_1 > \dots > u_d > 0$  y  $\mathbf{V} = \text{diag}(v_1, \dots, v_p)$  con  $v_1 > \dots > v_p > 0$ , entonces para todo  $\mathbf{P}_1 \in \mathbf{V}_{d,p}$ , la variedad de Stiefel de matrices semi-ortogonales de  $p \times d$  (es decir, tales que  $\mathbf{P}_1^\top\mathbf{P}_1 = \mathbf{I}_d$ ), se verifica que

$\text{tr}(\mathbf{U}\mathbf{P}_1^\top\mathbf{V}\mathbf{P}_1) \leq \sum_{i=1}^d u_i v_i$ , con la igualdad sólo si  $\mathbf{P}_1$  es alguna de las  $2^d$  matrices de la forma

$$\mathbf{P}_1 = \begin{pmatrix} \pm 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \pm 1 \\ \dots & \dots & \dots \\ 0 & & \end{pmatrix}. \quad (1.12)$$

Aplicando este resultado al último término de (1.11) con  $\mathbf{U} = \sigma^{-2}\mathbf{I}_d - \mathbf{\Lambda}_1^{-1}$  y  $\mathbf{V} = \mathbf{L}$ , se sigue que ese término se maximiza respecto de  $\mathbf{P}_1$  cuando  $\mathbf{P}_1$  tiene la forma (1.12), y el máximo valor es

$$\frac{m}{2\sigma^2} \sum_{i=1}^d l_i - \frac{m}{2} \sum_{i=1}^d \frac{l_i}{\lambda_i}. \quad (1.13)$$

Dado que  $\mathbf{P}$  es ortogonal, la función  $g(\mathbf{\Sigma})$  se maximiza con respecto a  $\mathbf{P}$  cuando  $\mathbf{P}$  es de la forma

$$\mathbf{P} = \begin{pmatrix} \pm 1 & \dots & 0 & \vdots & \vdots & 0 \\ \vdots & \ddots & \vdots & & & \\ 0 & \dots & \pm 1 & & & \\ \hline & & & 0 & & \mathbf{P}_{22} \end{pmatrix}, \quad (1.14)$$

para cualquier  $\mathbf{P}_{22} \in O(p-k)$ . Así,  $\hat{\mathbf{H}} = \mathbf{Q}\hat{\mathbf{P}}$  da una estimación de máxima verosimilitud para  $\mathbf{H}$ . Ahora, de (1.11) y (1.13),

$$\max_{\mathbf{P} \in O(p)} g(\mathbf{\Sigma}) = -\frac{n}{2} \sum_{i=1}^d \log(\lambda_i) - \frac{n}{2}(p-d) \log(\sigma^2) - \frac{m}{2\sigma^2} \sum_{i=d+1}^p l_i - \frac{m}{2} \sum_{i=1}^d \frac{l_i}{\lambda_i}. \quad (1.15)$$

Tomando derivadas vemos que los estimadores de máxima verosimilitud para  $\lambda_i$  y  $\sigma^2$  son,

$$\hat{\lambda}_i = \frac{m}{n} l_i \quad (i = 1, \dots, d)$$

y

$$\hat{\sigma}^2 = \frac{m}{(p-d)n} \sum_{i=d+1}^p l_i.$$

Sustituyendo en (1.15), tenemos

$$\max_{\mathcal{H}_d} L(\boldsymbol{\mu}, \mathbf{\Sigma}) = \left(\frac{m}{n}\right)^{-np/2} \left(\prod_{i=1}^d l_i\right)^{-n/2} \left(\frac{1}{p-d} \sum_{i=d+1}^p l_i\right)^{-n(p-d)/2} e^{-np/2}. \quad (1.16)$$

Luego, de (1.8) y (1.16) resulta que el estadístico de cociente de verosimilitudes para  $\mathcal{H}_d$  es

$$\begin{aligned}
\Lambda_d &= \frac{\max_{\mathcal{H}_d} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} > 0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \\
&= \frac{\left(\frac{m}{n}\right)^{-np/2} \left(\prod_{i=1}^d l_i\right)^{-n/2} \left(\frac{1}{p-d} \sum_{i=d+1}^p l_i\right)^{-n(p-d)/2} e^{-np/2}}{n^{np/2} e^{-np/2} |\mathbf{A}|^{-n/2}} \\
&= \frac{\left(\frac{m}{n}\right)^{-np/2} \left(\prod_{i=1}^d l_i\right)^{-n/2} \left(\frac{1}{p-d} \sum_{i=d+1}^p l_i\right)^{-n(p-d)/2} e^{-np/2}}{n^{np/2} e^{-np/2} m^{-np/2} |\mathbf{S}_n|^{-n/2}} \\
&= \left[ \frac{\prod_{i=d+1}^p l_i}{\left(\frac{1}{p-d} \sum_{i=d+1}^p l_i\right)^{p-d}} \right]^{n/2} = \mathbf{V}_d^{n/2},
\end{aligned}$$

lo que completa la prueba. □

Para poder obtener un test de nivel de significancia  $\alpha$ , otra vez necesitamos encontrar la distribución del estadístico que hemos encontrado bajo la hipótesis nula. Tal y como hicimos en la Sección 1.1, estudiaremos la distribución asintótica dado que hallar la distribución exacta es muy complicado.

### 1.2.1 Distribución asintótica bajo la hipótesis nula

La teoría general de tests de cociente de verosimilitudes muestra que, cuando  $n \rightarrow \infty$ , la distribución asintótica bajo la hipótesis nula  $\mathcal{H}_d$  de  $-n \log \mathbf{V}_d$  es  $\chi_{(q-2)(q+1)/2}^2$ , donde  $q = p - d$ . Bartlett [9] introduce una mejora a  $-n \log \mathbf{V}_d$  proponiendo el estadístico

$$- \left( m - d - \frac{2q^2 + q + 2}{6q} \right) \log \mathbf{V}_d \quad (q = p - d). \quad (1.17)$$

Este estadístico deberíamos compararlo con (1.5), al cual se reduce en el caso en que  $d = 0$ , i.e.,  $q = p$ . Un nuevo ajuste en el factor multiplicativo fue propuesto por Lawley [27] y luego mejorado por James [18], de donde tenemos el siguiente teorema para la distribución asintótica.

**Teorema 1.5.** (Teorema 9.6.2, Muirhead, 2005) *Cuando la hipótesis nula  $\mathcal{H}_d$  es cierta, la distribución límite cuando  $n \rightarrow \infty$  del estadístico*

$$\mathbf{P}_d = - \left[ m - d - \frac{2q^2 + q + 2}{6q} + \sum_{i=1}^d \frac{\bar{l}_q^2}{(l_i - \bar{l}_q)^2} \right] \log \mathbf{V}_d \quad (1.18)$$

es  $\chi_{(q+2)(q-1)/2}^2$ , donde  $l_i$  para  $i = 1, \dots, p$  son los autovalores de la matriz de covarianza muestral,  $\bar{l}_q = \frac{1}{q} \sum_{i=d+1}^p l_i$  y  $q = p - d$ .

**Idea de la prueba del Teorema 1.5:**

La prueba es análoga al caso del Teorema 1.2 de esfericidad; es decir, se prueba utilizando aproximaciones que

$$M_{\mathbf{P}_d}(t) \rightarrow M_{\chi_d^2}(t) \quad \text{para } t \text{ cercano a } 0,$$

cuando  $n \rightarrow \infty$ , con  $d = \frac{1}{2}(q+2)(q-1)$ .

□

En lo que sigue de la tesis, extenderemos los tests presentados en este capítulo al caso en que la dimensión  $p$  ya no permanece fija durante el análisis sino que crece a infinito junto con el tamaño  $n$  de la muestra. En particular, en el próximo capítulo, derivaremos la distribución asintótica del estadístico de esfericidad y esfericidad parcial. Como veremos, para hallar dicha distribución, en lugar de buscar la función generadora de momentos del estadístico (como en los casos presentados en el presente capítulo), descompondremos el  $\log(LRT)$  en distintos sumandos. A uno de ellos le calcularemos la función generadora de momentos y veremos que converge a la función generadora de momentos de la distribución  $\mathcal{N}(0, 1)$  y probaremos que los términos remanentes convergen a 0, de donde obtendremos el resultado.

---

# Estadístico para esfericidad y esfericidad parcial en alta dimensión

---

En el capítulo previo hemos repasado los tests de esfericidad y esfericidad parcial en el contexto de baja dimensionalidad de los datos. Como vimos, los tests para estas hipótesis, están basados en la matriz de covarianza muestral (más precisamente en los autovalores de la matriz de covarianza muestral). Esto representa un problema en alta dimensión, pues ya no es posible obtener estimadores consistentes de la matriz de covarianza poblacional. Además, en el caso de alta dimensión, es decir, cuando la dimensión  $p$  y el tamaño  $n$  de muestra ambos divergen a infinito, la aproximación  $\chi^2$  ya no es buena para la distribución del estadístico de cociente de verosimilitudes. Más aún, en escenarios de “ultra-alta-dimensión” ( $p > n$ ) el estadístico ni siquiera está definido.

En este capítulo probaremos que el estadístico dado por el logaritmo del cociente verosimilitudes  $\log LRT_d$  para el test de esfericidad parcial en el caso en que la dimensión  $p$  de los datos es comparable al tamaño  $n$  de la muestra, tiene, bajo la hipótesis nula, distribución asintóticamente normal. Más aún, para el caso en que  $p > n$  daremos una definición para el estadístico que nos permitirá extender ambos tests al caso de ultra-alta-dimensión.

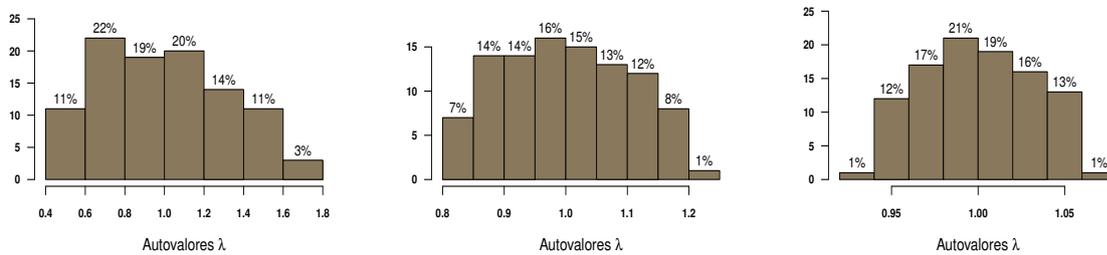
## 2.1 Resultados conocidos en alta dimensión

Durante el último medio siglo, se han dedicado una gran cantidad de esfuerzos a la comprensión del comportamiento de los autovalores y autovectores muestrales de matrices de covarianza. Dado que  $\mathbb{E}(\mathbf{S}_n) = \mathbf{\Sigma}$ , en caso de  $p$  fijo y  $n \rightarrow \infty$  la ley de los grandes números garantiza que  $\mathbf{S}_n \rightarrow \mathbf{\Sigma}$ . Desafortunadamente, cuando  $p$  es del mismo orden que  $n$  (o incluso mayor) la matriz

de covarianza de la muestra es un estimador notoriamente malo con estructura espectral (autovalores y autovectores) radicalmente diferente a la de la covarianza poblacional. Veamos esto más detenidamente en un par de ejemplos.

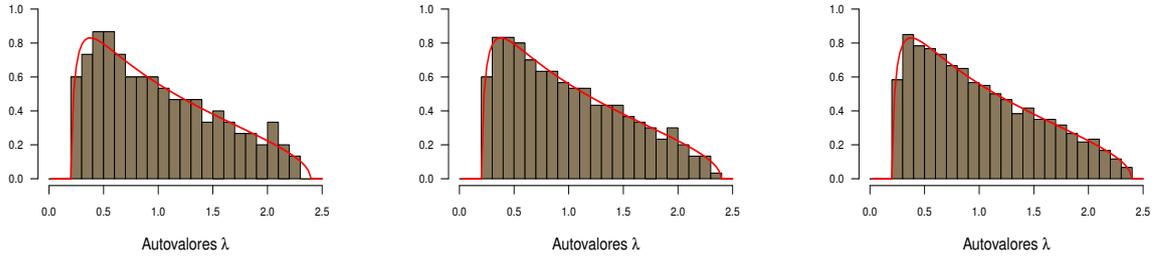
### 2.1.1 Modelos esféricos

Consideremos primero el caso  $p$  fijo. Supongamos que  $\Sigma = \mathbf{I}_p$  y tomemos una muestra  $\mathbf{x}_i$ , para  $i = 1, \dots, n$  de una población con distribución  $\mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ . Los histogramas en la Figura 2.1 muestran cómo los autovalores muestrales, para el caso  $p = 100$  y  $n$  creciendo, convergen a 1, valor común de los autovalores poblacionales.



**Figura 2.1.** Histograma de los autovalores correspondientes a la matriz de covarianza de una muestra de una población  $\mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ , para  $p = 100$ ,  $n = 1000, 10000, 100000$ ,  $\sigma^2 = 1$  (caso  $p$  fijo,  $n \rightarrow \infty$ ).

Sin embargo, cuando consideramos  $p$  variable, creciendo también a infinito junto con el tamaño de la muestra, los resultados son bien diferentes. Los histogramas en la Figura 2.2 corresponden a los autovalores de la matriz de covarianza muestral para  $p = 150$ ,  $n = 500$  (a la izquierda),  $p = 300$ ,  $n = 1000$  (centro) y  $p = 600$ ,  $n = 2000$  (a la derecha). Como podemos ver, los autovalores ya no se concentran alrededor de 1, sino que hay muchos autovalores mayores que 1 y algunos considerablemente más grandes que otros (distribución asimétrica con cola a la derecha). Es destacable además la distribución de los autovalores dentro de un intervalo aparentemente fijo (independientemente de los valores de  $p$  y  $n$  elegidos, siempre que el cociente entre ambos es 0.3). Es más, viendo estos gráficos uno podría llegar a sospechar que los datos tienen una estructura dimensional baja (es decir, pocos autovalores grandes y el resto más chicos), cuando en realidad la distribución de la que se extrajo la muestra es esférica. La línea roja es la distribución de los autovalores predicha por la distribución Marchenko-Pastur (2.1). Esta distribución límite fue primero establecida en 1967 por Marčenko and Pastur [30], motivo por el cual es referida actualmente como distribución Marchenko-Pastur. En su trabajo, ellos



**Figura 2.2.** Histograma de los autovalores correspondientes a la matriz de covarianza de una muestra de una población  $\mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ , para  $p = 150$ ,  $n = 500$  (izquierda),  $p = 300$ ,  $n = 1000$  (centro) y  $p = 600$ ,  $n = 2000$  (derecha),  $\sigma^2 = 1$  (caso  $p, n \rightarrow \infty$ ).

prueban que si la verdadera matriz de covarianza es la identidad (o un múltiplo de ella), cuando  $p$  y  $n$  ambos divergen a  $\infty$  con cociente fijo  $p/n = y < \infty$ , la distribución de los autovalores de  $\mathbf{S}_n$  (distribución espectral empírica, como los histogramas de la Figura 2.2), en el límite será

$$p_{y, \sigma^2}(x) = \begin{cases} \frac{1}{2\pi xy\sigma^2} \sqrt{(b-x)(x-a)}, & \text{si } a \leq x \leq b \\ 0, & \text{cualquier otro caso} \end{cases}, \quad (2.1)$$

con un punto masa adicional de valor  $1 - 1/y$  en el origen si  $y > 1$ , donde  $a = \sigma^2(1 - \sqrt{y})^2$  y  $b = \sigma^2(1 + \sqrt{y})^2$ . La constante  $y$  es la relación entre la dimensión y el tamaño de la muestra y  $\sigma^2$  el parámetro de escala.

A partir de este trabajo seminal de Marchenko y Pastur, muchas investigaciones se han dedicado a comprender las propiedades asintóticas de diversos estadísticos espectrales de matrices de covarianza muestrales en alta dimensión. En particular, a la luz de las aplicaciones a PCA, se ha mostrado gran interés en estudiar el comportamiento de los autovalores extremos de  $\mathbf{S}_n$ . Cuando la covarianza poblacional es la identidad, Geman [14] probó que, cuando  $p/n \rightarrow y$  para  $0 < y < \infty$ , el mayor autovalor de la matriz de covarianza muestral converge casi seguro (c.s.) a  $b$  (el extremo superior del soporte de la distribución Marchenko-Pastur), asumiendo una cierta condición de crecimiento en los momentos de la distribución subyacente. Más tarde, Yin et al. [53] y Bai et al. [4], respectivamente, probaron que la condición necesaria y suficiente para que el autovalor más grande de una matriz de covarianza muestral converja c.s. a un límite finito es que la distribución subyacente tenga cuarto momento finito y media cero, y que en tal caso el límite es  $b$ . Por otra parte, Bai and Yin [6], Silverstein [43] y Yin et al. [52] consideraron el límite c.s. del menor autovalor de una matriz de covarianza muestral, cuando  $y < 1$ , bajo la existencia del cuarto momento en la distribución subyacente, siendo dicho límite igual a  $a$ , el extremo inferior en el soporte de la distribución Marchenko-Pastur.

Observemos que si  $y = 1$  el resultado es trivialmente cierto (en tal caso  $a = 0$ ). Sin embargo, para el caso  $y > 1$ ,  $p > n$  y por lo tanto los  $p - n + 1$  autovalores más chicos de la matriz de covarianza muestral son iguales a 0. No obstante, si en tal caso consideramos el  $p - n + 2$  autovalor más chico, esto es el  $(n - 1)$ -ésimo mayor autovalor, el resultado sigue siendo válido. Por lo tanto, definiendo  $l_{\min} = \lambda_{\min\{p, n-1\}}(\mathbf{S}_n)$  se tiene que  $l_{\min} \rightarrow \sigma^2(1 - \sqrt{y})^2$  cuando  $p, n \rightarrow \infty$  tales que  $p/n \rightarrow y > 0$ .

Resumiendo, podemos decir que cuando la matriz de covarianza poblacional es la identidad y bajo el supuesto de que la relación  $p/n \rightarrow y$ , con  $0 < y < \infty$ , cuando  $p, n \rightarrow \infty$ , el autovalor más chico (o el  $l_{\min}$  definido antes) y el más grande de la matriz de covarianza muestral no convergen a sus contrapartes poblacionales sino que convergen a los límites inferior y superior del soporte de la distribución Marchenko-Pastur respectivamente (no hay autovalores “extraviados”, en el sentido que todos los autovalores muestrales convergen dentro del intervalo soporte de la densidad Marchenko-Pastur). Así, cuando existen algunos autovalores que se separan del resto, tendríamos información sobre la validez o no de la suposición de esfericidad en la matriz de covarianzas. De hecho, como veremos en la próxima sección, esta información ha sido utilizada para inferir la dimensión del subespacio de los autovalores spike (ver Sección 2.1.2) o, en el contexto del Capítulo 1, testear esfericidad parcial.

En relación al problema de determinar esfericidad en la matriz de covarianzas en escenarios de alta dimensionalidad, Bai et al. [5] desarrollan correcciones al test de cociente de máxima verosimilitudes tradicional de modo que resulte adecuado para testear si la covarianza poblacional es igual a la matriz identidad, bajo la restricción de que el cociente  $p/n \rightarrow y$ , siendo  $0 < y < 1$ . Por su parte, Jiang et al. [19] también consideran el problema de testear si la matriz de covarianza es la identidad pero lo abordan con un método diferente (el de las integrales de Selberg) lo que les permite extender el resultado de [5] al caso  $y = 1$ . (Observar que este test puede verse como un caso particular del test de esfericidad asumiendo que  $\sigma^2 = 1$  o más generalmente,  $\sigma^2$  conocido).

Srivastava [45] considera la hipótesis de esfericidad en contexto de ultra-alta-dimensión ( $p \gg n$ ). La idea clave en su trabajo es considerar la construcción del LRT intercambiando los roles de  $p$  y  $n$ . Esto es, bajo la hipótesis nula de esfericidad,  $m\mathbf{S}_n \sim \mathcal{W}_p(m, \sigma^2\mathbf{I}_p)$  ( $m = n - 1$ ) y por lo tanto  $\mathbf{W} = m\mathbf{S}_n = \mathbf{Y}^\top \mathbf{Y}$  con  $\mathbf{Y} \in \mathbb{R}^{m \times p}$  con entradas i.i.d. normales, con media 0 y varianza  $\sigma^2$ . Luego  $\widetilde{\mathbf{W}} = \mathbf{Y}\mathbf{Y}^\top \sim \mathcal{W}_m(p, \sigma^2\mathbf{I}_m)$  con  $p > m$  y así podemos obtener el test de cociente de verosimilitudes para  $\widetilde{\mathbf{W}}$  del mismo modo que fuera hecho para  $\mathbf{W}$  cuando  $p < m$ . La distribución asintótica del estadístico asociado es dada en términos de la distribución  $\chi^2$ . En Srivastava [44] el autor propone un test basado en estimadores consistentes de la traza de

potencias de  $\Sigma$ . Su test se basa en la primera y segunda media aritmética de los autovalores muestrales bajo la suposición de que  $n = \mathcal{O}(p^\delta)$ ,  $0 < \delta \leq 1$ .

Wang and Yao [49] proponen correcciones al test de cociente de verosimilitudes para la hipótesis de esfericidad bajo el régimen donde ambos,  $p, n \rightarrow \infty$  tales que  $p/n \rightarrow y \in (0, 1)$ . Asumiendo que  $(\mathbf{X}) = \{x_{ij}\}_{\substack{1 \leq i \leq p \\ 1 \leq j \leq n}}$  tiene entradas que satisfacen  $\mathbb{E}(x_{ij}) = 0$ ,  $\mathbb{E}(x_{ij}^2) = 1$  y  $\mathbb{E}(|x_{ij}|^4) < \infty$ , demuestran que el estadístico  $\mathcal{L}_n = \frac{2}{n} \log(L_n)$  con  $L_n = \left\{ \frac{|\mathbf{S}_n|}{\left[\frac{1}{p} \text{tr}(\mathbf{S}_n)\right]^p} \right\}^{n/2}$  tiene distribución asintóticamente normal. Más precisamente,

$$\mathcal{L}_n + (p - n) \log \left(1 - \frac{p}{n}\right) - p \rightsquigarrow \mathcal{N} \left( -\frac{\kappa - 1}{2} \log(1 - y) + \frac{1}{2} \beta y; -\kappa \log(1 - y) - \kappa y \right), \quad (2.2)$$

donde  $\kappa$  es igual a 2 si las  $\{x_{ij}\}$  son reales y es igual a 1 si son complejas y  $\beta = \mathbb{E}(|x_{ij}|^4) - 1 - \kappa$  es el coeficiente de curtosis. Notar que en el caso de  $\{x_{ij}\}$  reales y normales,  $\kappa = 2$  y  $\beta = 0$ , y por lo tanto (2.2) se reduce a

$$\mathcal{L}_n + (p - n) \log \left(1 - \frac{p}{n}\right) - p \rightsquigarrow \mathcal{N} \left( -\frac{1}{2} \log(1 - y); -2 \log(1 - y) - 2y \right),$$

que es lo mismo que lo obtenido en [5].

Li and Yao [28] consideran el test de esfericidad cuando la dimensión  $p$  de los datos es mucho mayor que el tamaño  $n$  de la muestra ( $p \gg n$ ). Para ello introducen un estadístico “quasi-LRT” definido como

$$\mathcal{L}_n = \frac{p}{n} \log \left[ \frac{\left( \frac{1}{n} \sum_{i=1}^n \tilde{l}_i \right)^n}{\prod_{i=1}^n \tilde{l}_i} \right],$$

donde  $\tilde{l}_i$ , para  $i = 1, \dots, n$ , son los autovalores de la matriz  $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$ , de  $n \times n$ , donde  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ , de  $n \times p$ , es la matriz de datos con  $n$  vectores aleatorios  $p$ -dimensionales independientes e idénticamente distribuídos tales que sus componentes  $x_{ij}$  (para  $1 \leq i \leq n$ ,  $1 \leq j \leq p$ ) satisfacen  $\mathbb{E}(x_{ij}) = 0$ ,  $\mathbb{E}(x_{ij}^2) = 1$  y  $\mathbb{E}(|x_{ij}|^4) = \nu_4 < \infty$ . Bajo estas hipótesis y asumiendo que el cociente  $p/n \rightarrow \infty$  cuando  $n \rightarrow \infty$ , prueban que, bajo la hipótesis nula de esfericidad,

$$\mathcal{L}_n - \frac{n}{2} - \frac{n^2}{6p} - \frac{\nu_4 - 2}{2} \rightsquigarrow \mathcal{N}(0, 1).$$

A partir de este resultado distribucional asintótico, obtienen un test para la hipótesis de esfericidad en el caso de ultra-alta-dimensión. Notar que la idea para la construcción del estadístico que denominan quasi-LRT es la misma que considera [45].

Jiang and Yang [20] estudian el test de esfericidad para la matriz de covarianza de poblaciones normalmente distribuídas y derivan el correspondiente teorema central del límite cuando la

dimensión de los datos y el tamaño de la muestra son comparables ( $p/n \rightarrow y \in (0, 1]$ ). Además comparan sus resultados con la tradicional aproximación chi-cuadrado, mostrando que ésta última puede fallar completamente cuando  $p$  crece a infinito junto con  $n$ . Formalmente, obtienen el siguiente teorema:

**Teorema 2.1.** *Sea  $m =: n - 1 > p$  para todo  $n \geq 3$  y  $\mathbf{V}$  definido por*

$$\mathbf{V} = \frac{|\mathbf{S}_n|}{\left[\frac{1}{p} \text{tr}(\mathbf{S}_n)\right]^p}.$$

*Supongamos que  $p/m \rightarrow y \in (0, 1]$ . Entonces, bajo la hipótesis nula  $\mathcal{H}_0 : \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$ ,  $(\log \mathbf{V} - \mu_m)/\sigma_m$  converge en distribución a una  $\mathcal{N}(0, 1)$  cuando  $n \rightarrow \infty$ , donde*

$$\begin{aligned} \mu_m &= -p - \left(m - p - \frac{1}{2}\right) \log \left(1 - \frac{p}{m}\right), \\ \sigma_m^2 &= -2 \left[\frac{p}{m} + \log \left(1 - \frac{p}{m}\right)\right]. \end{aligned}$$

Notar que el resultado obtenido en [20], cuando  $y \in (0, 1)$ , es el mismo que el de [49] para el caso de muestras normales y reales.

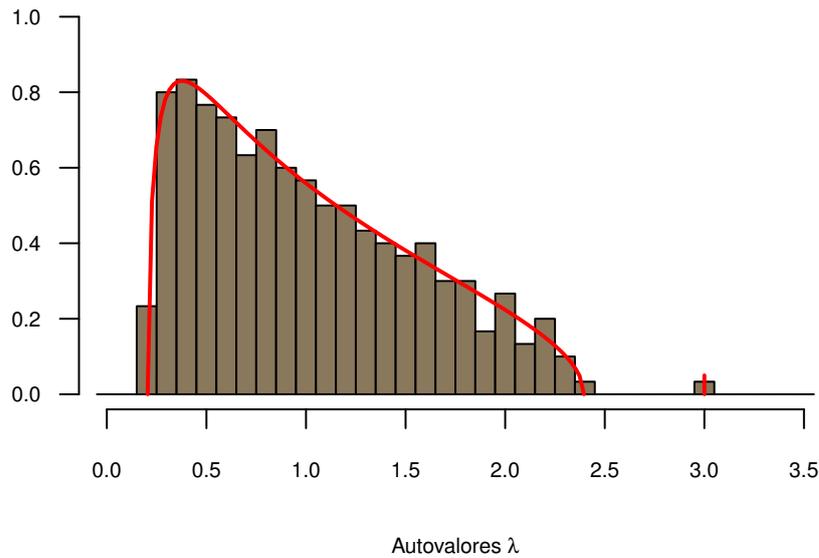
### 2.1.2 Modelos spiked

¿Qué pasa si realmente hay alguna estructura (lineal) de baja dimensión en los datos? En situaciones prácticas a menudo ocurre que hay uno o más autovalores grandes claramente separados del resto. En concreto, supongamos que la matriz de covarianza poblacional es una perturbación de rango  $k$ , con  $k \ll p$ , de la matriz identidad. Johnstone [23] llama a tal situación “modelo de covarianza spiked”, en el cual todos los autovalores poblacionales, con excepción de una cantidad fija y relativamente chica de ellos, son iguales. Notar que este problema, puesto en contexto con el capítulo anterior, refiere al modelo de esfericidad parcial. Dos cuestiones surgen inmediatamente para estos modelos: por un lado, cómo afecta esta pequeña perturbación al límite de los autovalores extremos de la matriz de covarianzas muestral en comparación con el caso nulo (i.e., cuando la covarianza poblacional es la identidad) y por otro lado, cómo determinar la dimensión de esta perturbación (*dimensión efectiva*). El comportamiento de los autovalores muestrales correspondientes a estos modelos ha sido ampliamente estudiado en la década pasada [3, 7, 8, 37, 38].

Para ilustrar esta situación, analicemos el ejemplo que mencionamos en la introducción, concentrándonos por un momento en el caso  $y < 1$ .

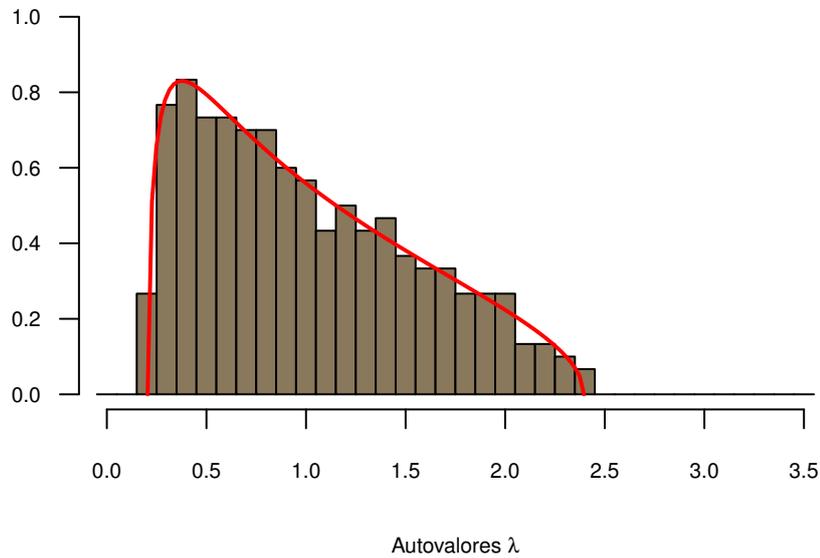
Supongamos que los datos presentan una estructura lineal subyacente de dimensión  $d = 1$ . En tal caso, su matriz de covarianza puede escribirse como  $\Sigma = \sigma^2 \mathbf{I}_p + \lambda \mathbf{v} \mathbf{v}^\top$ , para un vector unitario  $\mathbf{v}$  y  $\lambda > 0$ . Una forma de pensar este caso es que cada punto  $\mathbf{x}$  de los datos consiste de una parte de señal  $\sqrt{\lambda} g_0 \mathbf{v}$  donde  $g_0 \sim \mathcal{N}(0, 1)$  (i.e. un múltiplo gaussiano del vector fijo  $\sqrt{\lambda} \mathbf{v}$ ) más una parte de ruido  $\mathbf{g} \sim \mathcal{N}_p(\mathbf{0}, \sigma^2 \mathbf{I}_p)$  (independiente de  $g_0$ ). Luego  $\mathbf{x} = \mathbf{g} + \sqrt{\lambda} g_0 \mathbf{v} \sim \mathcal{N}_p(\mathbf{0}, \sigma^2 \mathbf{I}_p + \lambda \mathbf{v} \mathbf{v}^\top)$ .

Una pregunta natural que surge es si esta perturbación de rango 1 puede detectarse a partir de la matriz de covarianza muestral  $\mathbf{S}_n$ . Para alcanzar cierta intuición, grafiquemos algunos ejemplos concretos. El siguiente histograma corresponde a los autovalores de  $\mathbf{S}_n$  para  $p = 300$ ,  $n = 1000$ ,  $\sigma^2 = 1$ ,  $\mathbf{v} = \mathbf{e}_1$  es el primer elemento de la base canónica de  $\mathbb{R}^p$  y  $\lambda = 1.5$ . (Tenemos entonces que los autovalores poblacionales son todos iguales a 1 excepto el mayor que es 2.5).



**Figura 2.3.** Histograma de los autovalores de la matriz de covarianza muestral cuando  $\Sigma = \sigma^2 \mathbf{I}_p + \lambda \mathbf{v} \mathbf{v}^\top$ , con  $p = 300$ ,  $n = 1000$ ,  $\sigma^2 = 1$ ,  $\lambda = 1.5$  y  $\mathbf{v} = \mathbf{e}_1$

La Figura 2.3 sugiere que un autovalor de  $\mathbf{S}_n$  que “sale” del soporte de la distribución de Marchenko-Pastur (a continuación, veremos la ubicación límite de ese autovalor. El segmento rojo en el gráfico corresponde a la estimación de la misma). Vale la pena notar que el autovalor más grandes de  $\Sigma$  es  $1 + \lambda = 2.5$ , mientras que el mayor autovalor de  $\mathbf{S}_n$  parece considerablemente más grande que eso.



**Figura 2.4.** Histograma de los autovalores de la matriz de covarianza muestral cuando  $\Sigma = \sigma^2 \mathbf{I}_p + \lambda \mathbf{v}\mathbf{v}^\top$  con  $p = 300$ ,  $n = 1000$ ,  $\sigma^2 = 1$ ,  $\lambda = 0.5$  y  $\mathbf{v} = \mathbf{e}_1$

Repitamos el experimento pero tomando  $\lambda = 0.5$ , en cuyo caso el mayor autovalor de  $\Sigma$  es  $1 + \lambda = 1.5$  (ver Figura 2.4). Como podemos ver, parecería que para  $\lambda = 0.5$ , la distribución de los autovalores es la misma que para el caso en que  $\Sigma = \mathbf{I}_p$ .

Esto motiva la siguiente pregunta: ¿Para qué valores de  $\lambda$  esperamos ver un autovalor de  $\mathbf{S}_n$  saliendo del soporte de la distribución de Marchenko-Pastur, y cuál es el valor límite que esperamos que tome?

El comportamiento del mayor autovalor en el caso de vectores aleatorios complejos con distribución normal fue estudiado por Baik et al. [8]. En este trabajo, los autores prueban que existe un valor crítico por debajo del cual no esperamos ver ningún cambio en la distribución de los autovalores, respecto del caso nulo (i.e.  $\Sigma = \mathbf{I}_p$ ), y por encima del cual esperamos que existan autovalores que se salgan del soporte. Este fenómeno se conoce como *transición de fase BBP* (llamado así justamente por Baik, Ben Arous y Pécché). En pocas palabras, si los autovalores spike están cerca de uno (o, más generalmente, de  $\sigma^2$ ), entonces sus versiones muestrales se comportarán aproximadamente de la misma manera que si la covarianza verdadera fuera la identidad ( $\sigma^2 \mathbf{I}_p$  respectivamente). Sin embargo, cuando los autovalores spike son mayores que el umbral  $(1 + \sqrt{y})$  ( $\sigma^2(1 + \sqrt{y})$  respectivamente), las contrapartes muestrales tienen un comportamiento asintótico diferente. En [7], los autores consideran el modelo spiked con vectores

aleatorios más generales: complejos o reales y no necesariamente gaussianos. Para establecer los límites casi seguros de los autovalores extremos de la muestra, también encuentran que estos límites dependen de los valores críticos  $\sigma^2(1 + \sqrt{y})$  y  $\sigma^2(1 - \sqrt{y})$  por arriba y por abajo, respectivamente. Por ejemplo, supongamos que hay  $M$  autovalores en la matriz de covarianza poblacional mayores que el umbral  $\sigma^2(1 + \sqrt{y})$ , entonces prueban que los  $M$  autovalores más grandes de la matriz de covarianza muestral tendrán sus límites (c.s.) por encima del extremo superior del soporte de la distribución Marchenko-Pastur. Más precisamente, para  $\lambda \neq 1$  sea la función  $g(\lambda) = \lambda + \frac{y\sigma^2\lambda}{\lambda - \sigma^2}$ . Entonces prueban que para cada  $k \in \{1, \dots, M\}$ ,  $l_k \xrightarrow{c.s.} g(\lambda_k)$ , donde  $l_k$ , para  $k = 1, \dots, p$  son los autovalores de  $\mathbf{S}_n$  mientras que  $\lambda_k$ , para  $k = 1, \dots, p$  son sus contrapartes poblacionales. También prueban que para todo  $1 \leq i \leq L$ , para un rango  $L$  prefijado,  $l_{M+i} \xrightarrow{c.s.} \sigma^2(1 + \sqrt{y})^2$ . En forma análoga, si existen  $M_1$  autovalores poblacionales menores que el valor crítico inferior  $\sigma^2(1 - \sqrt{y})$ , los  $M_1$  autovalores más chicos de la covarianza muestral tienen su límite por debajo del extremo inferior del soporte de la Marchenko-Pastur.

Para los casos  $y > 1$  e  $y = 1$  se proponen resultados análogos. Vale la pena observar que en el caso  $y > 1$  tenemos que  $p > n$  y por lo tanto los  $p - n$  autovalores más chicos son iguales a 0 (recordar que la distribución Marchenko-Pastur presenta un punto masa de valor  $1 - 1/y$  en el origen cuando  $y > 1$ ). En tal caso tendremos que  $l_n, \dots, l_p = 0$  mientras que  $l_{n-1} \rightarrow \sigma^2(1 - \sqrt{y})^2$ . Por lo tanto, a diferencia del caso  $y < 1$ , los autovalores poblacionales más chicos no afectan a la distribución espectral empírica de  $\mathbf{S}_n$ .

En cuanto a la segunda cuestión, sobre cómo determinar el número de spikes (o, equivalentemente, testear esfericidad parcial en la matriz de covarianzas, lo que se corresponde con la existencia de una estructura o señal de baja dimensión en los datos), no existe consenso en cómo proceder en forma sistemática. En el campo de las aplicaciones, un enfoque común consiste en verificar cuántas de las variables transformadas explican una gran parte de la varianza en los datos y se presta poca atención (si es que hay alguna) a la naturaleza de lo que se descarta. Una excepción a este enfoque simplista se presenta en Plerou et al. [39], en el cual los autores comparan la distribución espectral empírica con la distribución típica en el caso nulo (i.e.  $\mathbf{\Sigma} = \mathbf{I}_p$ ). Un enfoque similar se ofrece en Laloux et al. [26], en donde los autores buscan estimar el valor de  $\sigma^2$  que mejor se ajusta a la distribución Marchenko-Pastur, y dicho valor determina automáticamente cuántos y cuáles son los autovalores spike (es decir, si observamos nuevamente la Figura 2.3, la idea es encontrar el valor de  $\sigma^2$  de modo de ajustar lo mejor posible la curva roja. Los autovalores que quedan fuera del soporte de distribución, corresponden a los spikes).

Otros métodos propuestos están basados principalmente en la aplicación de criterios de la teoría de la información para selección de modelos, introducidos por Akaike [1, 2] (AIC -Criterio

de Información de Akaike-), Schwarz [42] (BIC -Criterio de Información Bayesiano-) y Rissanen [41] (MDL -Longitud de Descripción Mínima-). A grandes rasgos, la idea es tomar la cantidad de factores que minimiza los criterios AIC o BIC/MDL (ver Wax and Kailath [50]). Sin embargo, estos métodos se basan en expansiones asintóticas para muestras de gran tamaño y pueden no funcionar bien cuando la dimensión de los datos es grande en comparación con el tamaño de muestra.

Más recientemente, el problema de estimar la dimensión  $d$  en escenarios de alta dimensionalidad se ha abordado combinando elementos de la teoría espectral de matrices aleatorias, como Onatski [34] o Harding et al. [16] en economía, y Kritchman and Nadler [25] en procesamiento de señales matriciales o literatura de quimiometría. Estos últimos en particular, consideran el estimador MDL desarrollado en [50] y muestran que con MDL no es posible detectar las componentes significativas cuando su magnitud es similar a las que no aportan información relevante, lo que usualmente resulta en una subestimación de la dimensión del subespacio de los autovalores spike, especialmente cuando  $n$  es chico. Luego proponen un nuevo estimador que mejora la tasa de detección. Sin embargo, solo prueban la consistencia de su estimador en el escenario en que  $p$  es fijo y  $n \rightarrow \infty$ .

Basados en los resultados de [3] y [38] (que, como mencionamos anteriormente, dan las distribuciones límite de los autovalores extremos de la matriz de covarianza proveniente de un modelo poblacional spiked) y bajo la suposición de que  $p/n \rightarrow y > 0$  cuando  $p$  y  $n$  crecen a infinito, Passemier and Yao [36] presentan un estimador del número de autovalores spike considerando el espaciado o diferencia entre dos autovalores muestrales consecutivos. La idea es que tal espaciado es menor entre aquellos autovalores que corresponden a la parte de ruido. Así el estimador que proponen se basa en detectar cuándo esa diferencia se vuelve pequeña. Formalmente, supongamos que existen  $q_0$  autovalores spike y definamos con  $\delta_{n,j} = l_j - l_{j+1}$ , para  $j \geq 1$ , a la diferencia entre dos autovalores consecutivos ( $l_k$  para  $k = 1, \dots, p$  son los  $p$  autovalores de la matriz de covarianza muestral en orden decreciente). Aplicando los resultados de [7], es fácil ver que para  $j > q_0$   $\delta_{n,j} \rightarrow 0$  mientras que si  $j \leq q_0$  la diferencia  $\delta_{n,j}$  tiende a un número positivo si todos los autovalores spike son distintos. Entonces el estimador de  $q_0$  propuesto es

$$\hat{q}_n = \min\{j \in \{1, \dots, s\} : \delta_{n,j+1} < d_n\}, \quad (2.3)$$

donde  $s > q_0$  es un número fijo lo suficientemente grande y  $d_n$  es un umbral (que debe determinarse). En la práctica  $s$  debe pensarse como una cota superior preliminar sobre el posible número de spikes.

Bajo la suposición de que todos los autovalores spike son simples y mayores que el valor crítico  $\sigma^2(1 + \sqrt{y})$ , prueban la consistencia de su estimador en el caso en que  $\sigma^2$  es conocido, como veremos en el siguiente teorema. Previamente enunciaremos un supuesto necesario sobre la distribución subyacente. Para ello, consideremos  $\mathbf{X} = \mathbf{E}\mathbf{V}^{1/2}\mathbf{Y}$ , donde  $\mathbf{Y} \in \mathbb{R}^p$  es un vector aleatorio con media  $\mathbf{0}$  con componentes i.i.d.,  $\mathbf{E}$  es una matriz ortogonal y

$$\mathbf{V} = \text{cov}(\mathbf{X}) = \sigma^2 \begin{pmatrix} \boldsymbol{\Sigma}_{q_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-q_0} \end{pmatrix},$$

donde  $\boldsymbol{\Sigma}_{q_0}$  tiene  $q_0$  autovalores  $\alpha_j$ , con  $j = 1, \dots, q_0$ , no nulos, distintos de 1 y simples. Se tiene entonces que el espectro de  $\mathbf{V}$  es  $\rho(\mathbf{V}) = \sigma^2(\underbrace{\alpha_1, \dots, \alpha_{q_0}}_{q_0}, \underbrace{1, \dots, 1}_{p-q_0})$ .

**Supuesto 2.2.** Las componentes  $y_i$  del vector  $\mathbf{Y}$  tienen distribución simétrica y decaimiento sub-exponencial. Es decir, existen constantes  $C$  y  $C'$  tales que para todo  $t \geq C'$

$$\mathbb{P}(|y_i|^C) \leq e^{-t}.$$

**Teorema 2.3.** (Teorema 1, Passemier and Yao [36]) Sean  $\mathbf{x}_1, \dots, \mathbf{x}_n$   $n$  copias i.i.d. de  $\mathbf{X} = \mathbf{E}\mathbf{V}^{1/2}\mathbf{Y}$ , donde  $\mathbf{Y} \in \mathbb{R}^p$  es un vector aleatorio con media  $\mathbf{0}$  y componentes i.i.d. que satisfacen el supuesto 2.2, con  $\mathbf{E}$  y  $\mathbf{V}$  como las definimos arriba.

Supongamos además que los  $q_0$  autovalores distintos de  $\boldsymbol{\Sigma}_{q_0}$  satisfacen  $\alpha_1 > \dots > \alpha_{q_0} > (1 + \sqrt{y})$  ( $p/n \rightarrow y > 0$  cuando  $p, n \rightarrow \infty$ ).

Sea  $(d_n)_{n \geq 0}$  una sucesión de números reales tales que  $d_n \rightarrow 0$  y  $n^{2/3}d_n \rightarrow \infty$ . Entonces el estimador  $\hat{q}_n$  es fuertemente consistente, i.e.  $\hat{q}_n \xrightarrow{\text{c.s.}} q_0$  cuando  $n \rightarrow \infty$ .

Cuando todos los autovalores spike (en la población) son distintos, las diferencias entre los autovalores spike en la muestra tienden a una constante positiva, mientras que si existen (al menos) dos autovalores spike (poblacionales) iguales, dicha diferencia tenderá a cero. Este hecho genera cierta ambigüedad con las diferencias correspondientes a los autovalores de la parte no spike (ruido) puesto que allí las diferencias también tienden a cero. En Passemier and Yao [35] los autores extienden los resultados de [36] al caso en que existan autovalores spike con multiplicidad mayor que 1. La idea es que la convergencia de los espaciamientos  $\delta_{n,j}$  para  $j > q_0$  (ruido) es más rápida ( $\mathcal{O}_P(n^{-2/3})$ ) que la de  $\delta_{n,j}$  para spikes múltiples ( $\mathcal{O}_P(n^{-1/2})$ ). Esta es la característica clave que utilizan los autores para modificar el estimador (2.3) a la situación de autovalores spike repetidos con la elección de un nuevo umbral  $d_n$ ; a saber, le añaden la condición  $d_n = o(n^{-1/2})$ .

En lo que resta del capítulo, presentaremos los fundamentos teóricos que nos permitirán definir un test de cociente de verosimilitudes para determinar la dimensión del subespacio spiked cualquiera sea la relación entre  $p$  y  $n$  (esto es,  $p \leq n$  o  $p > n$ ).

## 2.2 Distribución asintótica del $\log(LRT_d)$ bajo la hipótesis $\mathcal{H}_d$

Recordemos que, para  $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , con  $\mathbf{X} \in \mathbb{R}^p$ , el *test de esfericidad* está dado por

$$\mathcal{H}_0 : \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p \quad \text{vs} \quad \mathcal{H}_a : \boldsymbol{\Sigma} \neq \sigma^2 \mathbf{I}_p,$$

con  $\sigma$  desconocido.

En el Capítulo 1 vimos cuál es el estadístico apropiado para este test y su distribución asintótica en el caso tradicional en que la dimensión de los datos  $p$  es fija. Para el caso de alta dimensión, como mencionamos en la sección precedente, Jiang and Yang [20] mostraron que el test tradicional (con  $p$  fijo) puede fallar completamente. Para dar respuesta a este problema, en su Teorema 1 (ver 2.1), ellos encuentran la distribución asintótica del logaritmo del cociente de verosimilitud para esfericidad de la matriz de covarianza cuando tanto  $p$  como  $n$  crecen de modo tal que  $p < n - 1$  y el cociente  $p/n \rightarrow y \in (0, 1]$ . Basados en ese resultado, ellos determinan la región de rechazo para que el test tenga nivel de significancia asintótico igual a un valor  $\alpha$  dado.

En el Capítulo 1 vimos también que, en caso de que la hipótesis nula  $\mathcal{H}_0$  sea rechazada, procedemos a testear secuencialmente la hipótesis nula

$$\mathcal{H}_d : \lambda_{d+1} = \dots = \lambda_p, \tag{2.4}$$

para  $d = 0, \dots, p-2$ , donde  $\lambda_1, \dots, \lambda_p$  son los autovalores de  $\boldsymbol{\Sigma}$ . Como en el caso de esfericidad, este tipo de tests han sido ampliamente estudiados en la literatura multivariada para el caso de  $p$  fijo y  $n$  creciendo a infinito (ver Sección 1.2). Recordemos que este test, llamado de esfericidad parcial, se basa en el estadístico

$$LRT_d = \frac{l_{d+1} \cdots l_p}{\left( \frac{1}{p-d} \sum_{i=d+1}^p l_i \right)^{p-d}}, \tag{2.5}$$

donde  $l_i$  son los autovalores (en orden decreciente) de  $\mathbf{S}_n$  (la matriz de covarianza muestral). Recordemos también que en su trabajo Lawley [27] (luego mejorado por James [18]) prueba que la distribución asintótica de (2.5) bajo  $\mathcal{H}_d$  es

$$-\rho \log(LRT_d) \rightarrow \chi_{(p-d+2)(p-d-1)/2}^2, \tag{2.6}$$

para  $\rho = m - d - \frac{2(p-d)^2 + (p-d) + 2}{6(p-d)} + \sum_{i=1}^d \frac{\hat{\sigma}^2}{(l_i - \hat{\sigma}^2)^2}$  cuando  $n$  crece con  $p$  fijo, donde  $\hat{\sigma}^2 = \sum_{i=d+1}^p l_i / (p-d)$ .

Como fue señalado anteriormente, el estadístico de cociente de verosimilitud solo existe si  $p \leq n - 1 \doteq m$ . Cuando  $p > m$ ,  $p - m$  de los autovalores muestrales  $\{l_i\}$  son iguales a 0 y por lo tanto el LRT es idénticamente nulo. Sin embargo, como lo hiciera [45], aún podemos construir un test del siguiente modo: para  $p > m$ , bajo  $\mathcal{H}_0$ ,  $\mathbf{W} = m\hat{\Sigma} \sim W_p(m, \sigma^2 \mathbf{I}_p)$  y por lo tanto  $\mathbf{W} \stackrel{d}{=} \mathbf{Y}^T \mathbf{Y}$  con  $\mathbf{Y} \in \mathbb{R}^{m \times p}$  cuyos elementos son normales independientes con media 0 y varianza  $\sigma^2$ . En consecuencia,  $\widetilde{\mathbf{W}} = \mathbf{Y} \mathbf{Y}^T \sim W_m(p, \sigma^2 I_m)$  con  $p > m$  y entonces podemos considerar construir el estadístico de cociente de verosimilitudes para  $\widetilde{\mathbf{W}}$  como lo hicimos antes para  $\mathbf{W}$  en el caso de  $p < m$ . Este artilugio cobra pleno sentido si tenemos en cuenta que los autovalores de la matriz  $\widetilde{\mathbf{W}}$  (de  $m \times m$ ) son los mismos que los autovalores no nulos de la matriz  $\mathbf{W}$  (de  $p \times p$ ). Obtenemos entonces el estadístico de cociente de verosimilitudes para  $\widetilde{\mathbf{W}}$  como

$$LRT_0^* = \frac{l_1 \dots l_m}{\left(\frac{1}{m} \sum_{i=1}^m l_i\right)^m}.$$

Usando [20, Teorema 1] obtenemos la distribución asintótica del  $\log LRT_0^*$  con  $m$  y  $p$  intercambiando sus roles. A partir de este resultado, podemos obtener un test de nivel asintótico  $\alpha$  para el problema de esfericidad de la matriz de covarianza en el caso de  $p > m$  y  $p/m \rightarrow y > 1$ .

Esto nos motiva a definir el estadístico para el test de esfericidad parcial, en el caso de  $p > m$ , del siguiente modo

$$LRT_d^* = \frac{l_{d+1} \dots l_m}{\left(\frac{1}{m-d} \sum_{i=d+1}^m l_i\right)^{m-d}}. \quad (2.7)$$

Con el objetivo de derivar un test para esfericidad parcial que tenga nivel de significancia asintótico igual a  $\alpha$  para el caso de  $p$  y  $m$  creciendo, y con  $p < m$  or  $p > m + d + 3$  (la necesidad de imponer la restricción sobre  $p$  de ser mayor que  $m + d + 3$  será explicado más adelante, en la demostración de la proposición correspondiente), primero necesitamos encontrar la distribución asintótica de  $LRT_d$  y  $LRT_d^*$  respectivamente. En las siguientes dos proposiciones derivamos la distribución asintótica cuando  $p, m$  crecen a infinito y  $p/m \rightarrow y$  con  $y \in (0, \infty)$ .

En el resto del trabajo, asumiremos la siguiente Condición  $Q_0$ :

**Condición  $Q_0$**  : existe  $q_0 \ll \min(p, m)$  independiente de  $p$  y  $m$  tal que  $\mathcal{H}_d$  definida en (2.4) es cierta para  $d \leq q_0$ .

Es decir, asumiremos que  $\mathcal{H}_d$  es cierta para algún  $d \leq q_0$ . Esta cota  $q_0$  (equivalente a la cota  $s$  del trabajo de [36]) es una cota superior preliminar al número posible de spikes. Por simplicidad en la notación, denotaremos al estadístico simplemente por  $LRT_d$  en ambos casos (de  $p/m < 1$  y  $p/m > 1$ ).

**Proposición 2.4.** *Sea  $\mathbf{A} = m\mathbf{S}_n \sim W_p(m, \boldsymbol{\Sigma})$ , con  $p < m$ . Bajo la hipótesis nula de que la verdadera dimensión es  $d \geq 0$  fija, i.e.,*

$$\mathcal{H}_d : \lambda_{d+1} = \cdots = \lambda_p,$$

la distribución asintótica de  $LRT_d$  definido en (2.5) (cuando  $m$  y  $p$  crecen y  $p/m \rightarrow y \in (0; 1]$ ) está dada por

$$\frac{\log(LRT_d) - \mu_{m,p,d}}{\sigma_{m,p,d}} \rightarrow \mathcal{N}(0, 1),$$

donde

- $\mu_{m,p,d} = \tilde{\mu}_{m,p} + \log A_{m,p,d} + \log B_{m,p,d}$ ,
- $\sigma_{m,p,d}^2 = -2 \left[ \frac{p-d}{m} + \log \left( 1 - \frac{p}{m} \right) \right]$ ,

con

$$\begin{aligned} \tilde{\mu}_{m,p} &= -p - \left(m - p - \frac{1}{2}\right) \log \left( 1 - \frac{p}{m} \right), \\ A_{m,p,d} &= \frac{\prod_{i=1}^h \lambda_i^{s_i}}{\prod_{i=1}^d l_i}, \\ B_{m,p,d} &= \left( 1 + \frac{\sum_{i=1}^d l_i - \sum_{i=1}^h s_i \lambda_i}{\sum_{i=d+1}^p l_i} \right)^{p-d}. \end{aligned}$$

**Proposición 2.5.** *Sea  $\mathbf{A} = m\mathbf{S}_n \sim W_p(m, \boldsymbol{\Sigma})$ , con  $p > m + d + 3$ . Bajo la hipótesis nula de que la verdadera dimensión es  $d \geq 0$  fija, i.e.,*

$$\mathcal{H}_d : \lambda_{d+1} = \cdots = \lambda_p,$$

la distribución asintótica de  $LRT_d$  definido en (2.7) (cuando  $m$  y  $p$  crecen y  $p/m \rightarrow y \geq 1$ ) está dada por

$$\frac{\log(LRT_d) - \mu_{m,p,d}}{\sigma_{m,p,d}} \rightarrow \mathcal{N}(0, 1),$$

donde

- $\mu_{m,p,d} = \tilde{\mu}_{m,p,d} + \log B_{m,p,d} + \log C_{m,p,d} + \log D_{m,p,d}$ ,

$$\blacksquare \sigma_{m,p,d}^2 = -2 \left[ \frac{m}{p-d} + \log \left( 1 - \frac{m}{p-d} \right) \right],$$

con

$$\begin{aligned} \tilde{\mu}_{m,p,d} &= -m - \left( p - d - m - \frac{1}{2} \right) \log \left( 1 - \frac{m}{p-d} \right), \\ B_{m,p,d} &= \left( 1 + \frac{\sum_{i=1}^d l_i - \sum_{i=1}^h s_i \lambda_i}{\sum_{i=d+1}^m l_i} \right)^{m-d} \left( \frac{m-d}{m} \right)^{m-d}, \\ C_{m,p,d} &= \left[ \frac{\sigma^2(p-d)}{m} \right]^d, \\ D_{m,p,d} &= \frac{\prod_{i=1}^h \left( 1 + \frac{\lambda_i}{\sigma^2} \frac{m}{p-d-m-1} \right)^{s_i}}{\prod_{i=1}^d l_i}. \end{aligned}$$

**Observación 2.6.** La Proposición 2.4 contiene el test de esfericidad estudiado por Jiang and Yang [20] para  $m > p$ . Más aún, si  $p > m$ , bajo la hipótesis nula de esfericidad, se obtiene que el nuevo estadístico definido tiene la misma distribución dada en Jiang and Yang [20], excepto que  $p$  y  $m$  intercambian sus roles. Así la distribución asintótica del  $LRT_0$  para  $p > m$  bajo la hipótesis nula de que la dimensión es  $d = 0$  se sigue del resultado en Jiang and Yang [20], intercambiando los roles de  $m$  y  $p$ .

**Observación 2.7.** Notemos que  $\mu_{m,p,d}$  es una variable aleatoria que, además de los autovalores muestrales (es decir, los autovalores de  $\widehat{\Sigma}$ ), depende de los verdaderos valores (i.e. poblacionales) de  $\sigma^2$ ,  $\lambda_i$ ,  $i = 1, \dots, d$ . Luego, para poder utilizar la distribución asintótica de las Proposiciones 2.4 y 2.5 para obtener un test que nos permita determinar la dimensión  $d$  del subespacio de los spikes, necesitamos reemplazar esos valores por estimaciones apropiadas. Al parámetro  $\sigma^2$  podemos reemplazarlo por su estimación consistente  $\hat{\sigma}^2 = \sum_{i=d+1}^p l_i / (p-d)$  (Ver Lema B.2). En cambio, para los  $\lambda_i$  la situación es algo más complicada. Como mencionamos en la sección previa, se sabe que, en el límite, las estimaciones de los autovalores spike experimentan una transición de fase, [8]. De hecho, recordemos que, si  $\lambda > \sigma^2(1 + \sqrt{y})$ , entonces

$$l \rightarrow \lambda \left( 1 + \frac{y\sigma^2}{\lambda - \sigma^2} \right), \quad (2.8)$$

mientras que para los autovalores  $\lambda$  que se encuentran dentro del rango  $(\sigma^2, \sigma^2(1 + \sqrt{y})]$ , el límite es  $\sigma^2(1 + \sqrt{y})^2$ , haciendo que los mismos resulten “invisibles”, esto es, indistinguibles de los autovalores de la parte de ruido, ya que los autovalores muestrales correspondientes a un número fijo de autovalores en la parte de ruido convergen al mismo límite  $\sigma^2(1 + \sqrt{y})^2$ , [7] (ver también [35, 36]). Por lo tanto, no podemos reemplazar directamente  $\lambda_i$  por  $l_i$  en  $\mu_{m,p,d}$  ya que, aun cuando  $\lambda_i$  es mayor que el umbral  $\sigma^2(1 + \sqrt{y})$ , los  $l_i$  son estimadores sesgados de  $\lambda_i$ .

Sin embargo, por (2.8) conocemos exactamente cual es el sesgo. Así, para sustituir los  $\lambda_i$  en la expresión de  $\mu_{m,p,d}$ , podemos utilizar la ecuación sugerida por (2.8):

$$l_i = \lambda_i \left( 1 + \frac{p}{m} \frac{\hat{\sigma}^2}{\lambda_i - \hat{\sigma}^2} \right) \quad (2.9)$$

para obtener un estimador consistente para  $\lambda_i \geq \sigma^2(1 + \sqrt{y})$ , a saber:

$$\tilde{\lambda}_i = \frac{1}{2} \left\{ l_i + \hat{\sigma}^2 - \hat{\sigma}^2 \frac{p}{m} + \sqrt{-4l_i \hat{\sigma}^2 + \left( l_i + \hat{\sigma}^2 - \hat{\sigma}^2 \frac{p}{m} \right)^2} \right\}. \quad (2.10)$$

Aquí,  $\tilde{\lambda}_i$  es el valor de  $\lambda_i$  obtenido como solución de la ecuación (2.9). Notar que hemos descartado una solución dado que es una solución extraña para (2.9). En el Lema B.5 probamos la consistencia del estimador  $\tilde{\lambda}_i$  para el caso en que el correspondiente autovalor poblacional  $\lambda_i$  sea mayor que el umbral y también probamos que, en caso de que  $\lambda_i \in (\sigma^2, \sigma^2(1 + \sqrt{y})]$ , entonces  $\tilde{\lambda}_i \rightarrow \sigma^2(1 + \sqrt{y})$ .

En el límite, el discriminante de (2.10) será no-negativo si y solo si  $\lambda_i \geq \sigma^2(1 + \sqrt{y})$ . Ahora bien, la versión muestral del discriminante puede ser negativo cuando los autovalores poblacionales están cerca del umbral (o son menores que el umbral). En tal caso, consideraremos  $\tilde{\lambda}_i = \hat{\sigma}^2(1 + \sqrt{p/m})$  que es el valor de  $\tilde{\lambda}_i$  para  $l_i$  que anula el discriminante. Reemplazando  $\lambda_i$  en  $\mu_{m,p,d}$  por  $\tilde{\lambda}_i$  y  $\sigma^2$  por  $\hat{\sigma}^2$  tenemos una nueva aproximación para la distribución asintótica, bajo la suposición de que todos los autovalores spike son mayores que el umbral, que podemos utilizar para llevar a cabo el test. Por otro lado, si uno o más autovalores spike son menores que el umbral, esta nueva aproximación de la distribución asintótica nos dará un test con nivel de significancia no mayor que  $\alpha$ . (Ver Observación 3.3 en el Capítulo 3).

Resumiendo, tenemos el siguiente Corolario:

**Corolario 2.8.** *Bajo las hipótesis de la Proposición 2.4 o 2.5, si todos los autovalores spike  $\lambda_i$ ,  $i = 1, \dots, d$  son mayores que el umbral  $\sigma^2(1 + \sqrt{y})$ ,*

$$\frac{\log LRT_d - \hat{\mu}_{m,p,d}}{\sigma_{m,p,d}} \rightarrow \mathcal{N}(0, 1),$$

donde  $\hat{\mu}_{m,p,d}$  se obtiene de  $\mu_{m,p,d}$  reemplazando  $\lambda_i$  por  $\tilde{\lambda}_i$  definida en (2.10) y  $\sigma^2$  por  $\hat{\sigma}^2$ .

### 2.3 Simulaciones para la distribución asintótica

En esta Sección presentamos una serie de simulaciones realizadas con el objetivo de estudiar el comportamiento de la distribución asintótica obtenida en las Proposiciones 2.4 y 2.5, como

así también de la aproximación asintótica de la distribución dada en el Corolario 2.8. Vamos a recordar que estas dos distribuciones son asintóticamente equivalente cuando todos los autovalores en el subespacio spiked son mayores que el umbral  $\sigma^2(1 + \sqrt{y})$ , mientras que si existen autovalores spike en el rango  $(\sigma^2, \sigma^2(1 + \sqrt{y})]$ , la distribución dada en el Corolario 2.8 es sólo una aproximación.

Este estudio lo realizamos siguiendo los dos escenarios utilizados en [36]. En su trabajo ellos consideran dos modelos con subespacio spike de dimensiones  $d = 4$  y  $d = 5$  y autovalores spike  $(7, 6, 5, 4)$  y  $(259.72, 17.97, 11.04, 7.88, 4.82)$ , respectivamente. A partir de ahora y en adelante, los llamaremos **Modelo 1** y **Modelo 2** respectivamente. Para cada uno de los modelos, consideramos ocho valores diferentes para  $y$ , a saber:  $0.1, 0.3, 0.6, 0.9$  para el caso  $p < m$ , y  $1/0.1 = 10, 1/0.3 = 3.33, 1/0.6 = 1.67, 1/0.9 = 1.11$ , para el caso  $p > m$ . (Es válido destacar que [36] sólo trabajaron con los casos  $0.3$  y  $0.6$ ). En ambos modelos  $\sigma^2$  lo consideramos igual a  $1$ . Notemos que en estos escenarios, todos los autovalores spike son mayores que el umbral  $\sigma^2(1 + \sqrt{y})$ . Finalmente, en todas las simulaciones suponemos que conforme  $p$  y  $m$  crecen, su cociente permanece constante e igual a su límite  $y$ .

En las Figuras 2.5 y 2.7 podemos observar la distribución correspondiente a la Proposición 2.4 (curva roja) y la aproximación dada en el Corolario 2.8 (curva azul) para cada uno de los valores elegidos anteriormente para el cociente  $p/m$  cuando  $p < m$  (esto es,  $p/m = 0.1, 0.3, 0.6, 0.9$ ) para los Modelos 1 y 2 respectivamente. Como podemos ver, el comportamiento de la distribución asintótica mejora cuando  $p$  y  $m$  crecen en los cuatro casos:  $p/m = 0.1$ ,  $p/m = 0.3$ ,  $p/m = 0.6$  y  $p/m = 0.9$ . Sin embargo, la aproximación no es buena cuando  $y$  es cercano a  $0$  y  $p$  y  $m$  son chicos (ver primera línea de las Figuras 2.5 y 2.7).

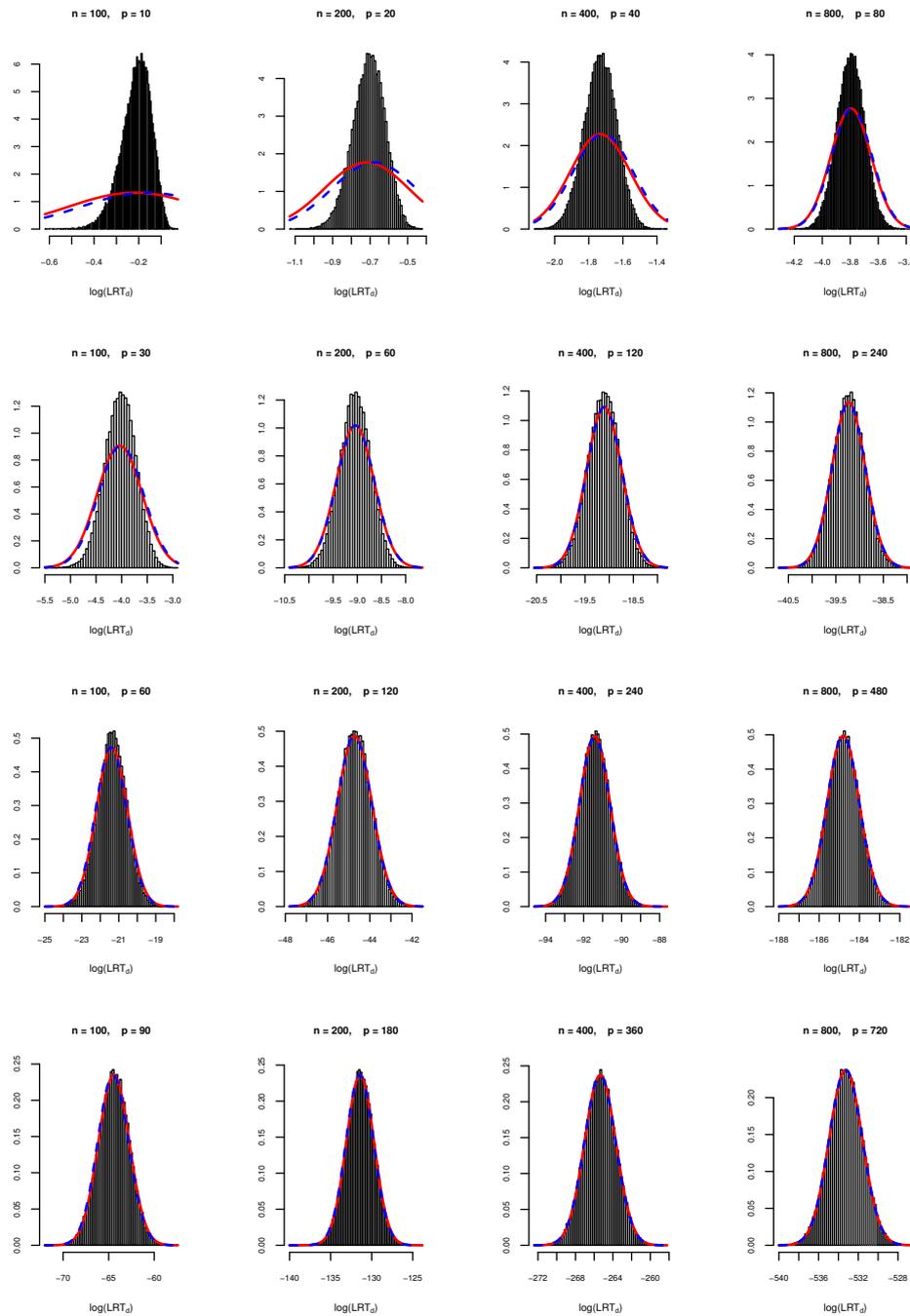
Para el caso  $p > m$ , según podemos observar en las Figura 2.6 y 2.8, los resultados son esencialmente los mismos. Sin embargo, la tasa de convergencia a la verdadera distribución es menor a medida que el cociente  $p/m$  crece. Además, en este caso, las aproximaciones no son buenas cuando  $y$  tiende a infinito y  $p$  y  $m$  son chicos (ver primera línea de las Figuras 2.6 y 2.8).

Las Figuras 2.9 y 2.11 corresponden a modelos spiked con  $d = 2$ ,  $p/m = 0.3$  y  $p/m = 0.6$  respectivamente, con  $p = 9, 21, 30, 72$ , y varios valores para  $(\lambda_1, \lambda_2)$ , mostrando el comportamiento por encima y por debajo del umbral  $\sigma^2(1 + \sqrt{y})$ . Más específicamente, para los valores de  $p$  y  $n$  mencionados, consideramos los siguientes tres casos:

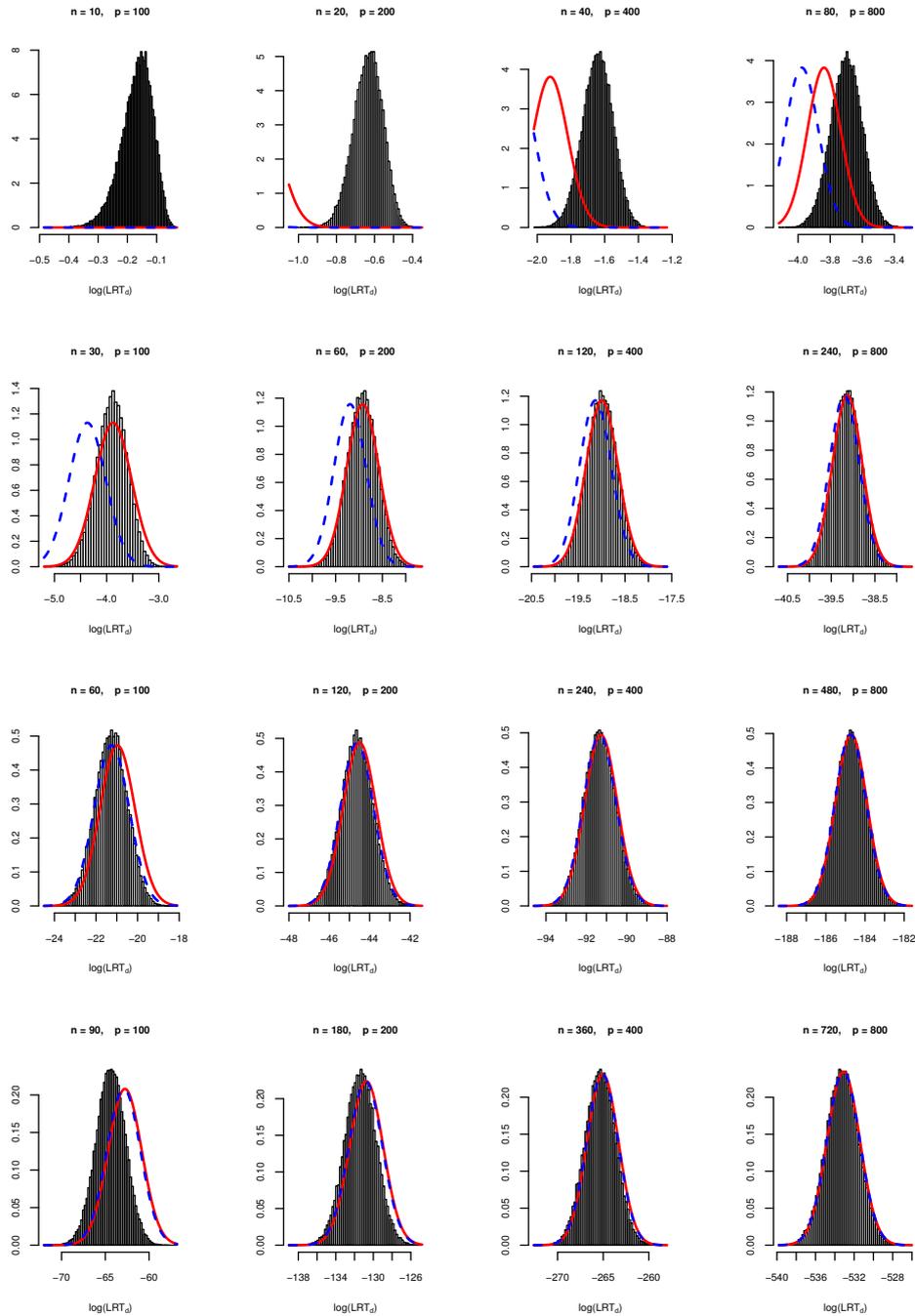
- a)  $(\lambda_1, \lambda_2) = \left(1 + \frac{1}{2}\sqrt{\frac{p}{n}}, 1 + \frac{1}{4}\sqrt{\frac{p}{n}}\right)$ , los dos autovalores menores que el umbral,
- b)  $(\lambda_1, \lambda_2) = \left(10, 1 + \frac{1}{2}\sqrt{\frac{p}{n}}\right)$ , un autovalor por encima del umbral y el otro por debajo,
- c)  $(\lambda_1, \lambda_2) = (10, 2)$ , los dos autovalores mayores que el umbral.

Las Figuras 2.10 y 2.12 corresponden al mismo modelo pero intercambiando los roles de  $p$  y  $m$ . El comportamiento de ambas, la distribución asintótica dada en las Proposiciones 2.4 y 2.5 (curva roja) y la aproximación dada por el Corolario 2.8 (curva azul), mejora cuando  $p$  y  $m$  crecen, como es esperable cuando los autovalores spike están alejados del ruido. El comportamiento es similar para  $p/m = .3$ ,  $p/m = .6$ ,  $p/m = 1/.3$  y  $p/m = 1/.6$ .

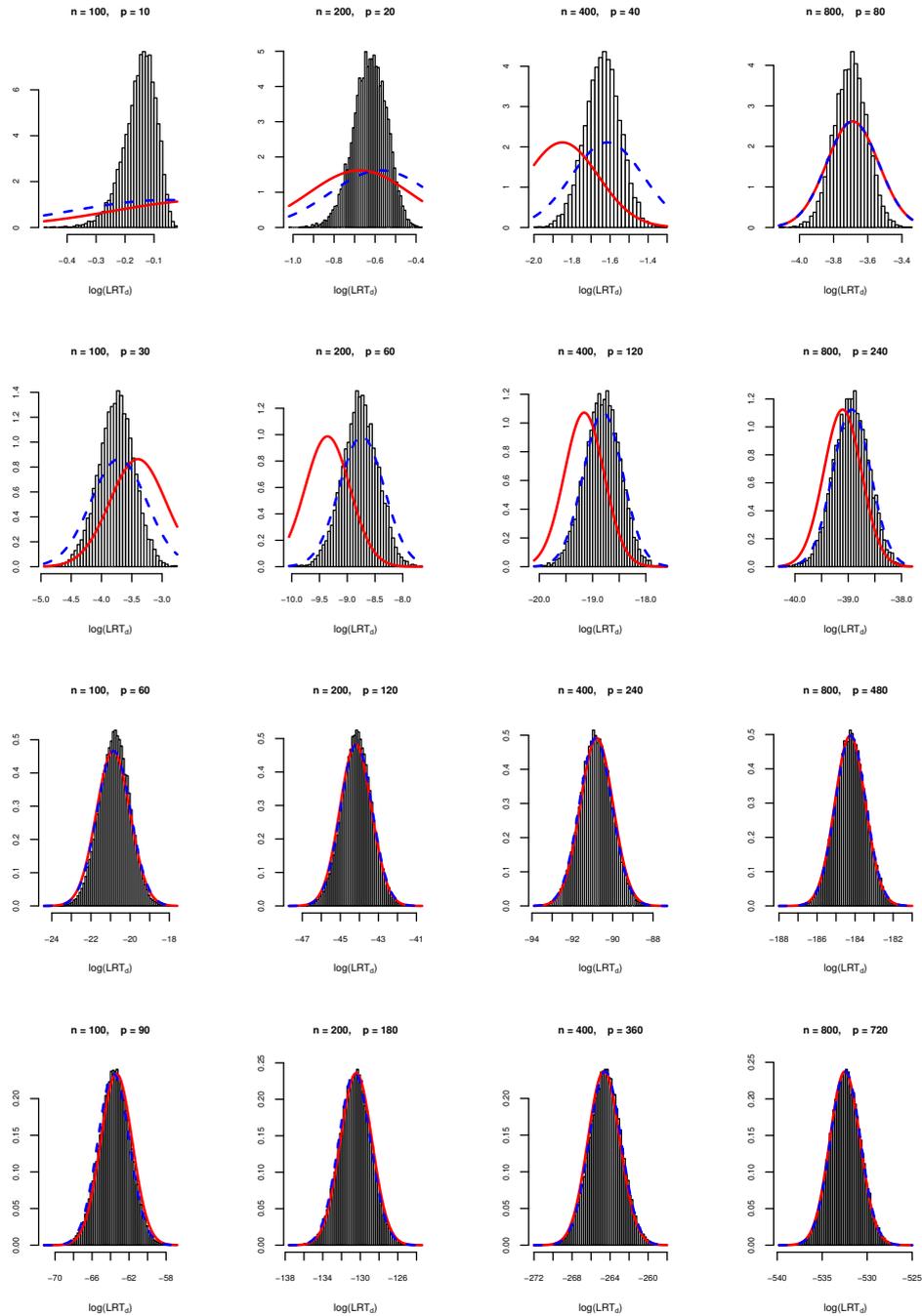
Finalmente, las Figuras 2.13 y 2.14 muestran cómo la clásica aproximación  $\chi^2$  para  $p$  fijo dada en (2.6) es cada vez peor a medida que  $p$  se hace grande en relación a  $n$  (pero manteniéndose siempre  $p < n$ ) cuando  $d = 1$  para dos caso de autovalor spike: mayor que el umbral (Figura 2.13) y menor que el umbral (Figura 2.14). Esta misma situación ya fue observada para el caso  $d = 0$  en Jiang and Yang [20].



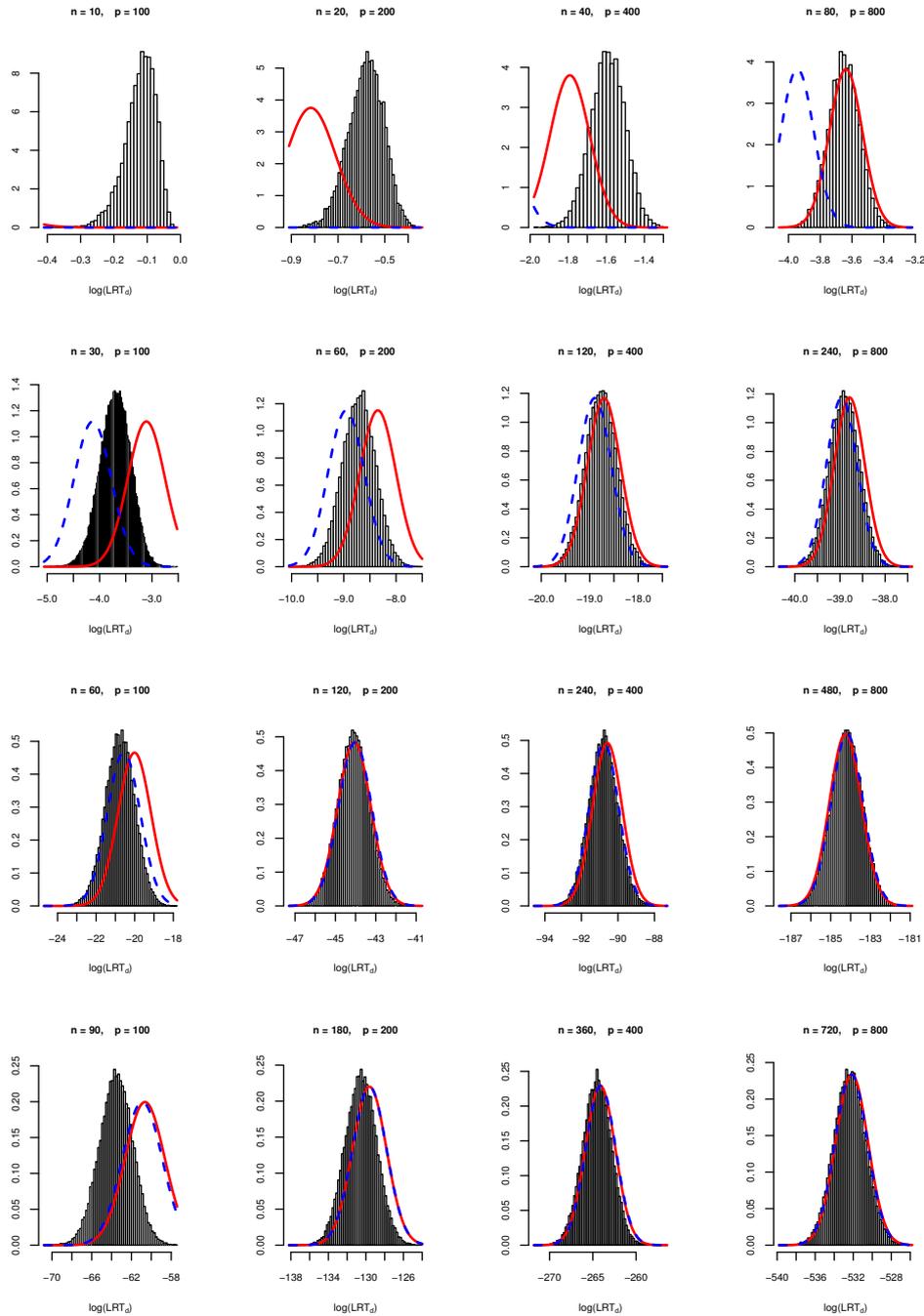
**Figura 2.5.** Simulación con  $\sigma^2 = 1$ ,  $n = 100, 200, 300, 400$ , para  $d = 4$ ,  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (7, 6, 5, 4)$ ,  $p/n = .1, .3, .6, .9$  de la primera fila a la cuarta respectivamente. La curva roja es la distribución asintótica de  $\log LRT_d$  dada en la Proposición 2.4 y la curva azul la aproximación dada en el Corolario 2.8.



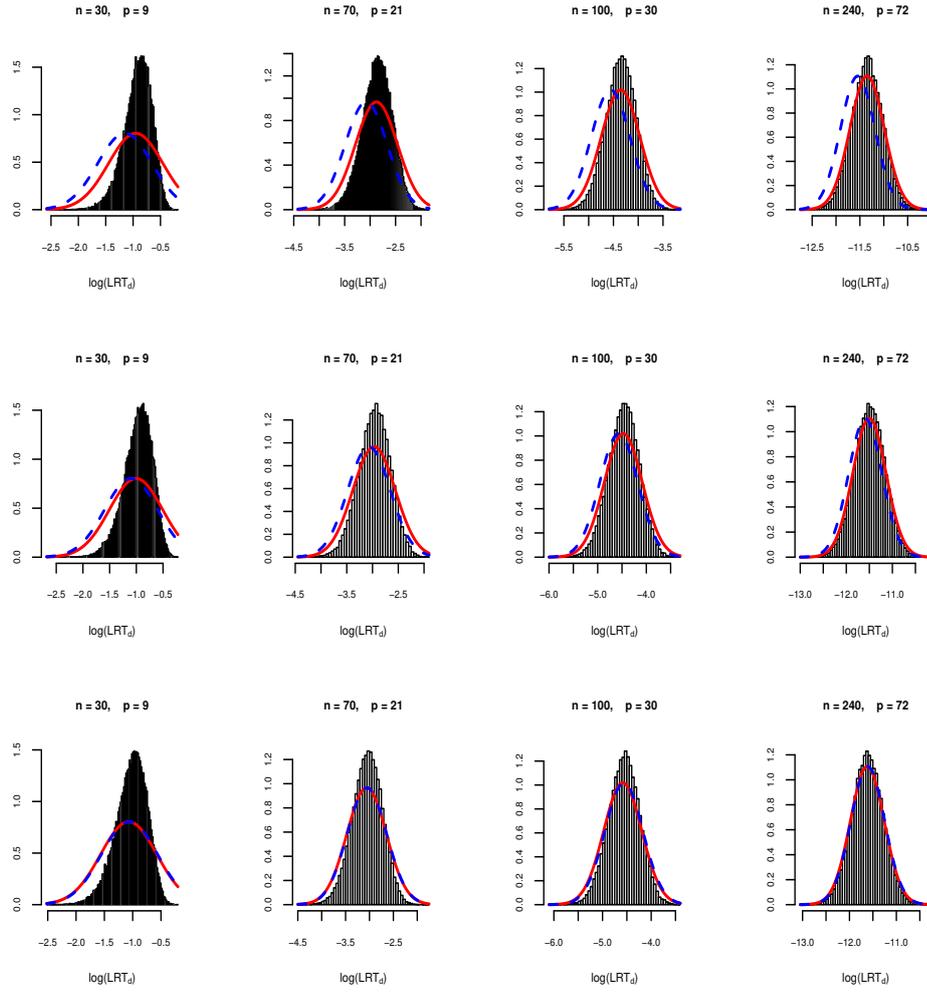
**Figura 2.6.** Simulación con  $\sigma^2 = 1$ ,  $p = 100, 200, 300, 400$ , para  $d = 4$ ,  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (7, 6, 5, 4)$ ,  $p/n = 1/.1, 1/.3, 1/.6, 1/.9$  de la primera fila a la cuarta respectivamente. La curva roja es la distribución asintótica de  $\log LRT_d$  dada en la Proposición 2.5 y la curva azul la aproximación dada en el Corolario 2.8.



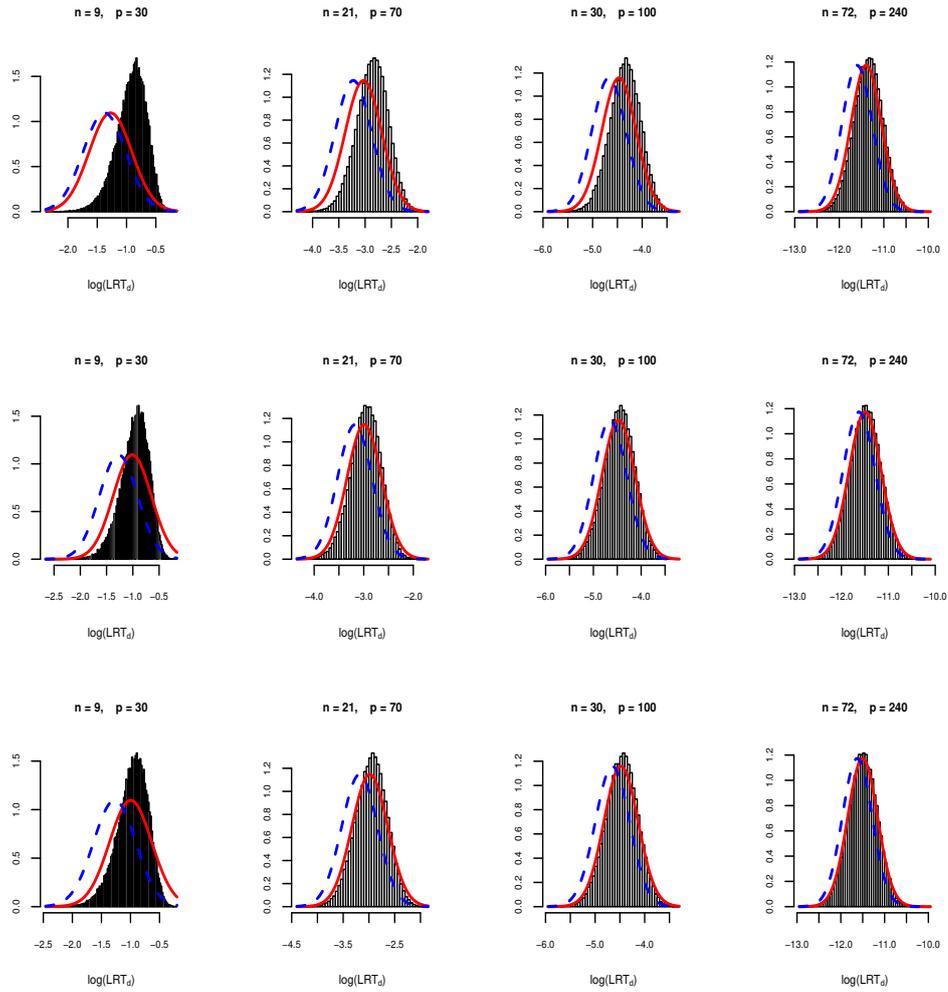
**Figura 2.7.** Simulación con  $\sigma^2 = 1$ ,  $n = 100, 200, 300, 400$ , para  $d = 5$ ,  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = (259.72, 17.97, 11.04, 7.88, 4.82)$ ,  $p/n = .1, .3, .6, .9$  de la primera fila a la cuarta respectivamente. La curva roja es la distribución asintótica de  $\log LRT_d$  dada en la Proposición 2.4 y la curva azul la aproximación dada en el Corolario 2.8.



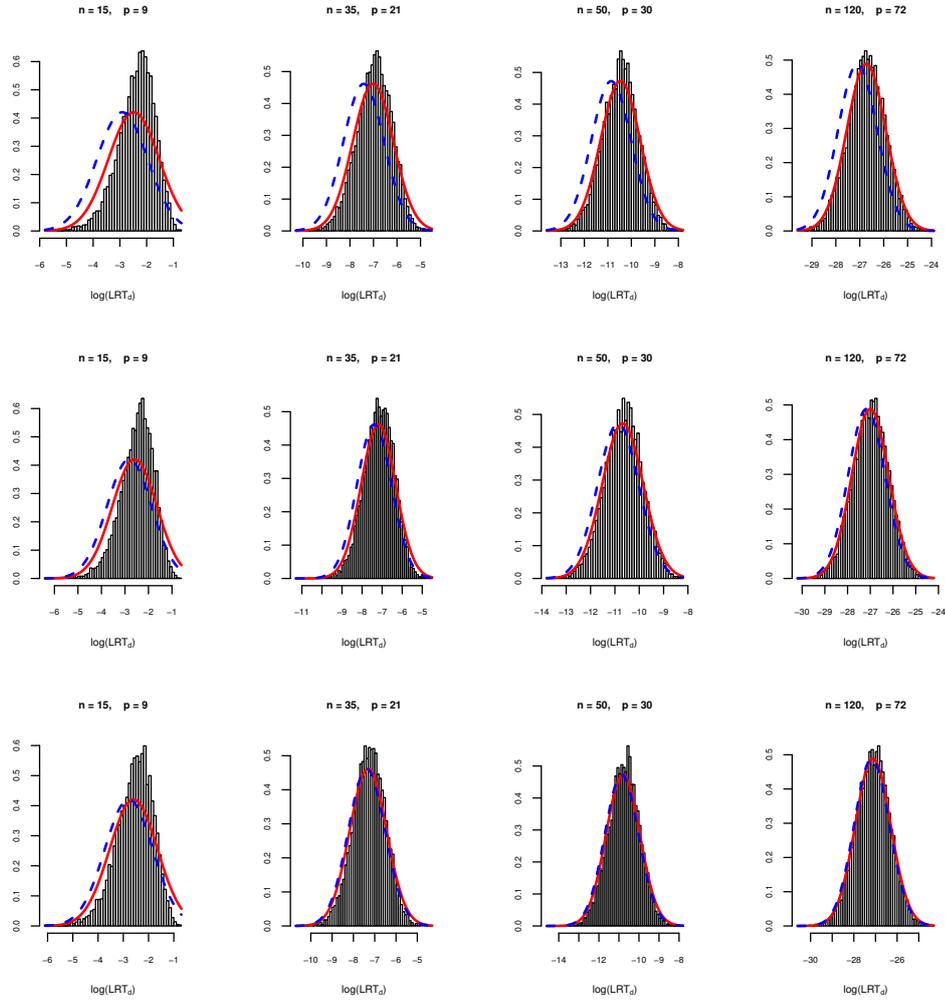
**Figura 2.8.** Simulación con  $\sigma^2 = 1$ ,  $p = 100, 200, 300, 400$ , para  $d = 5$ ,  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = (259.72, 17.97, 11.04, 7.88, 4.82)$ ,  $p/n = 1/.1, 1/.3, 1/.6, 1/.9$  de la primera fila a la cuarta respectivamente. La curva roja es la distribución asintótica de  $\log LRT_d$  dada en la Proposición 2.5 y la curva azul la aproximación dada en el Corolario 2.8.



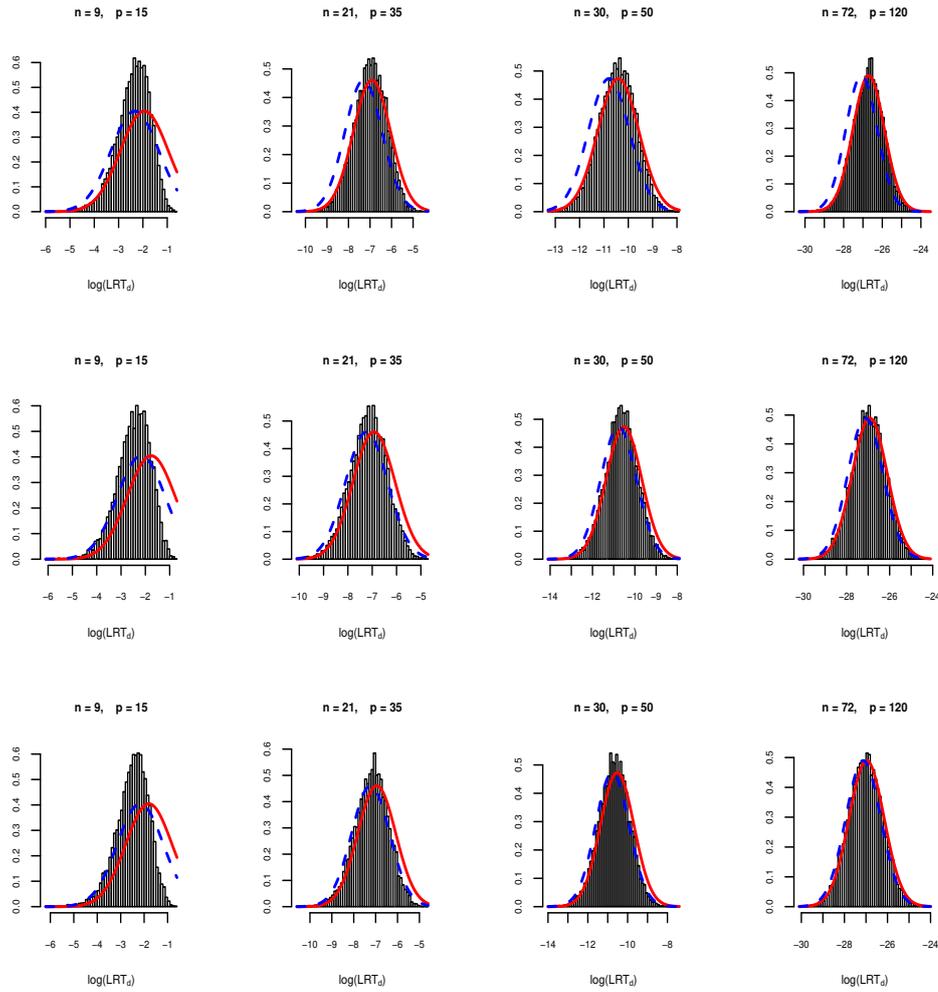
**Figura 2.9.** Simulación con  $\sigma^2 = 1$ ,  $d = 2$ ,  $p = 9, 21, 30, 72$  y  $p/n = .3$ . Primera fila  $(\lambda_1, \lambda_2) = (1 + 1/2\sqrt{p/n}, 1 + 1/4\sqrt{p/n})$ , segunda fila  $(\lambda_1, \lambda_2) = (10, 1 + 1/2\sqrt{p/n})$  y tercera fila  $(\lambda_1, \lambda_2) = (10, 2)$ . La curva roja es la distribución asintótica de  $\log LRT_d$  dada en la Proposición 2.4 y la curva azul la aproximación dada en el Corolario 2.8.



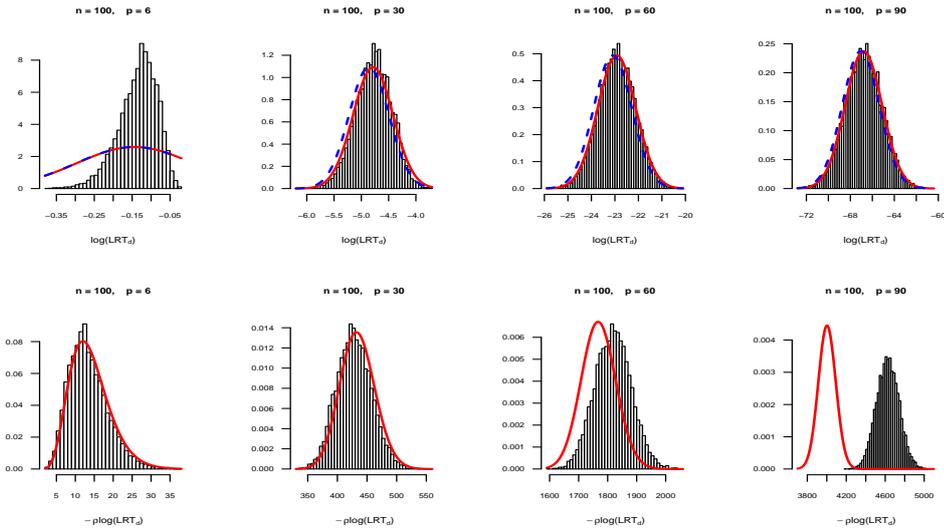
**Figura 2.10.** Simulación con  $\sigma^2 = 1$ ,  $d = 2$ ,  $n = 9, 21, 30, 72$  y  $p/n = 1/3$ . Primera fila  $(\lambda_1, \lambda_2) = (1 + 1/2\sqrt{p/n}, 1 + 1/4\sqrt{p/n})$ , segunda fila  $(\lambda_1, \lambda_2) = (10, 1 + 1/2\sqrt{p/n})$  y tercera fila  $(\lambda_1, \lambda_2) = (10, 2)$ . La curva roja es la distribución asintótica de  $\log LRT_d$  dada en la Proposición 2.5 y la curva azul la aproximación dada en el Corolario 2.8.



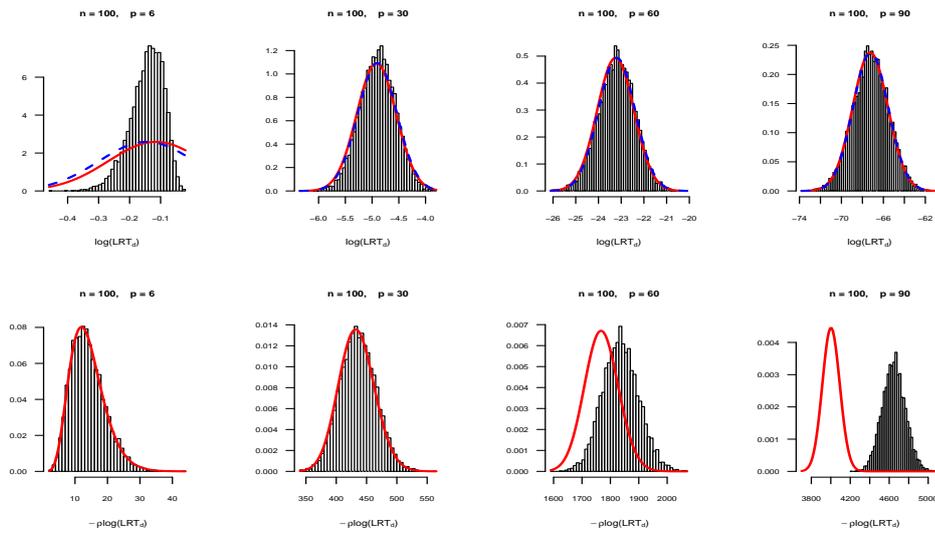
**Figura 2.11.** Simulación con  $\sigma^2 = 1$ ,  $d = 2$ ,  $p = 9, 21, 30, 72$  y  $p/n = .6$ . Primera fila  $(\lambda_1, \lambda_2) = (1 + 1/2\sqrt{p/n}, 1 + 1/4\sqrt{p/n})$ , segunda fila  $(\lambda_1, \lambda_2) = (10, 1 + 1/2\sqrt{p/n})$  y tercera fila  $(\lambda_1, \lambda_2) = (10, 2)$ . La curva roja es la distribución asintótica de  $\log LRT_d$  dada en la Proposición 2.4 y la curva azul la aproximación dada en el Corolario 2.8.



**Figura 2.12.** Simulación con  $\sigma^2 = 1$ ,  $d = 2$ ,  $n = 9, 21, 30, 72$  y  $p/n = 1/6$ . Primera fila  $(\lambda_1, \lambda_2) = (1 + 1/2\sqrt{p/n}, 1 + 1/4\sqrt{p/n})$ , segunda fila  $(\lambda_1, \lambda_2) = (10, 1 + 1/2\sqrt{p/n})$  y tercera fila  $(\lambda_1, \lambda_2) = (10, 2)$ . La curva roja es la distribución asintótica de  $\log LRT_d$  dada en la Proposición 2.5 y la curva azul la aproximación dada en el Corolario 2.8.



**Figura 2.13.** Simulación con  $\sigma^2 = 1$ ,  $d = 1$ ,  $\lambda_1 = 40$ ,  $n = 100$  y  $p = 6, 30, 60, 90$ . En la primera fila la curva roja es la distribución asintótica de  $\log LRT_d$  dada en la Proposición 2.4 y la curva azul la aproximación dada en el Corolario 2.8. La segunda fila es la aproximación  $\chi^2$  para  $p$  fijo dada en (2.6).



**Figura 2.14.** Simulación con  $\sigma^2 = 1$ ,  $d = 1$ ,  $\lambda_1 = 1.4$ ,  $n = 100$  y  $p = 6, 30, 60, 90$ . En la primera fila la curva roja es la distribución asintótica de  $\log LRT_d$  dada en la Proposición 2.4 y la curva azul la aproximación dada en el Corolario 2.8. La segunda fila es la aproximación  $\chi^2$  para  $p$  fijo dada en (2.6).



# Test de esfericidad parcial en alta dimensión

## 3.1 Test para determinar la dimensión $d$ del subespacio de los spikes

Como mencionamos en el Capítulo 2, el test para inferir la dimensión  $d$  del subespacio de los autovalores spike, será hecho en forma secuencial, tal como se hace en el caso clásico de  $p$  fijo. El procedimiento comienza considerando la hipótesis nula  $\Sigma = \sigma^2 \mathbf{I}_p$ . Si esta hipótesis es rechazada, entonces procedemos a testear si los  $p - 1$  autovalores más chicos son iguales, y así sucesivamente. Esto es, testeamos secuencialmente la hipótesis nula para cada  $d \in \{0, \dots, q_0\}$ , cuando la condición  $Q_0$  es cierta. Paramos cuando el test falla en rechazar la hipótesis nula y escogemos ese valor como el valor estimado para la dimensión  $d$ .

Para que el test tenga nivel de significancia  $\alpha$ , la región de rechazo será el conjunto  $\{d : LRT_d < c_\alpha\}$  donde  $c_\alpha$  se elige de modo que  $\mathbb{P}_{\mathcal{H}_d}\{LRT_d < c_\alpha\} = \alpha$ . Usando el Corolario 2.8 podemos obtener un test de nivel asintótico no mayor a  $\alpha$  considerando la región de rechazo

$$\left\{ d : \frac{\log LRT_d - \hat{\mu}_{m,p,d}}{\sigma_{m,p,d}} < z_\alpha \right\}, \quad (3.1)$$

donde  $\hat{\mu}_{m,p,d}$  fue definida en el Corolario 2.8,  $\sigma_{m,p,d}$  fue definida en las Proposiciones 2.4 y 2.5, según corresponda, y  $z_\alpha$  es el cuantil de orden  $1 - \alpha$  (i.e. el cuantil que acumula  $1 - \alpha$ ) de la distribución normal estándar. Este test tendrá nivel de significancia asintótico igual a  $\alpha$  cuando todos los autovalores (poblacionales) spike sean mayores que el umbral  $\sigma^2(1 + \sqrt{y})$ . Por otro lado, si uno o más autovalores son menores que el umbral, el test con región de rechazo (3.1) tendrá nivel de significancia no mayor que  $\alpha$  y por lo tanto será un test conservativo. Resumiendo, tenemos la siguiente proposición:

**Proposición 3.1.** *El test para las hipótesis*

$\mathcal{H}_d$  : El subespacio spike tiene dimensión  $d$

vs

$\mathcal{H}_a$  : El subespacio spike tiene dimensión mayor que  $d$

con región de rechazo definida en (3.1) tiene nivel de significancia asintótico, cuando  $p/m \rightarrow y > 0$ , igual a  $\Phi(z_\alpha + L)$ , con

$$L = \begin{cases} \frac{1}{\sqrt{-2[y + \log(1-y)]}} \sum_{i \in J_2} s_i \left[ \frac{\lambda_i}{\sigma^2} - (1 + \sqrt{y}) - \log \frac{\lambda_i}{\sigma^2(1 + \sqrt{y})} \right] & \text{si } p < m, \quad y \in (0, 1); \\ \frac{1}{\sqrt{-2\left[\frac{1}{y} + \log(1 - \frac{1}{y})\right]}} \sum_{i \in J_2} s_i \left\{ \frac{\lambda_i}{y\sigma^2} - \frac{1 + \sqrt{y}}{y} - \log \frac{\sigma^2(y-1) + \lambda_i}{\sqrt{y}\sigma^2(\sqrt{y} + 1)} \right\} & \text{si } p > m + d + 3, y > 1; \\ 0 & \text{si } y = 1, \end{cases}$$

donde  $J_1 = \{i \leq h : \lambda_i > \sigma^2(1 + \sqrt{y})\}$ ,  $J_2 = \{i \leq h : \lambda_i \in (\sigma^2, \sigma^2(1 + \sqrt{y})]\}$ ,  $\Phi$  es la función de distribución acumulada de la distribución  $\mathcal{N}(0, 1)$  y  $z_\alpha$  es su cuantil de orden  $1 - \alpha$ .

**Observación 3.2.** Consideramos la función  $f(x) = x - \log(x)$ . Es fácil ver que  $f$  es estrictamente creciente si  $x > 1$ .

1. Por lo tanto, por un lado tenemos que  $\frac{\lambda}{\sigma^2} - (1 + \sqrt{y}) - \log \frac{\lambda}{\sigma^2(1 + \sqrt{y})} = f\left(\frac{\lambda}{\sigma^2}\right) - f(\sigma^2(1 + \sqrt{y})) < 0$  si  $\lambda \in J_2$ .
2. Por otro lado, si definimos

$$\begin{aligned} g(\lambda) &= \frac{\lambda}{y\sigma^2} - \frac{1 + \sqrt{y}}{y} - \log \frac{\sigma^2(y-1) + \lambda}{\sqrt{y}\sigma^2(\sqrt{y} + 1)} \\ &= \frac{\lambda}{y\sigma^2} + 1 - \frac{1}{y} - \log \left( \frac{\lambda}{y\sigma^2} + 1 - \frac{1}{y} \right) - 1 + \frac{1}{y} - \frac{1 + \sqrt{y}}{y} - \log \frac{\sqrt{y}}{1 + \sqrt{y}} \\ &= f\left(\frac{\lambda}{y\sigma^2} + 1 - \frac{1}{y}\right) - 1 - \frac{1}{\sqrt{y}} - \log \frac{\sqrt{y}}{1 + \sqrt{y}}, \end{aligned}$$

claramente  $g$  es creciente donde  $f$  es creciente. Entonces, si  $\lambda \in J_2$ ,  $1 = \frac{1}{y} + 1 - \frac{1}{y} < \frac{\lambda}{\sigma^2 y} + 1 - \frac{1}{y}$  y así la función  $f$  es creciente y por lo tanto también lo es la  $g$ . Además se verifica que  $g(\sigma^2(1 + \sqrt{y})) = 0$  y por lo tanto  $\frac{\lambda}{y\sigma^2} - \frac{1 + \sqrt{y}}{y} - \log \frac{\sigma^2(y-1) + \lambda}{\sqrt{y}\sigma^2(\sqrt{y} + 1)} < 0$  si  $\lambda \in J_2$ .

**Observación 3.3.** Si  $J_2$  es vacío, es decir todos los autovalores spike son mayores que el umbral, entonces  $L = 0$  y el test definido por la Proposición 3.1 tendrá nivel de significancia asintótico  $\alpha$ . Caso contrario, si existen autovalores spike en el rango  $(\sigma^2, \sigma^2(1 + \sqrt{y})]$ , entonces

por la observación anterior,  $L < 0$  y en tal caso el test es conservativo, esto es, tendrá nivel de significancia asintótico menor a  $\alpha$ .

## 3.2 Simulaciones

En esta sección presentamos distintas simulaciones llevadas a cabo para mostrar el desempeño del test para determinar la dimensión  $d$  del subespacio spike. Como lo hicimos en las simulaciones del Capítulo 2, consideraremos los modelos propuestos en [36] para evaluar nuestros resultados así y poder compararlos con los reportados allí. Recordemos que, en su trabajo, los autores consideran dos escenarios con subespacios spike de dimensión  $d = 4$  y  $d = 5$  y autovalores spike  $(7, 6, 5, 4)$  y  $(259.72, 17.97, 11.04, 7.88, 4.82)$ , respectivamente **Modelo 1** y **Modelo 2**. En ambos casos consideran  $\sigma^2 = 1$ . En ambos escenarios, la relación entre  $p$  y  $m$  utilizada por ellos son  $p/m = 0.3$  and  $p/m = 0.6$ . En nuestro caso, como lo hicimos antes, agregamos las relaciones  $p/m = 0.9, 1/0.9, 1/0.6, 1/0.3$  para contemplar las situaciones de  $p > m$  y  $p \approx m$ . En todas las simulaciones suponemos que, conforme  $p$  y  $m$  crecen, su cociente permanece constante e igual a su límite  $y$ .

Los resultados presentados en las Tablas 3.1–3.16 corresponden a la proporción de veces, sobre 1000 réplicas, que el test elige la dimensión  $d$ . Cabe mencionar, que la varianza  $\sigma^2$  no se supone conocida, sino que la estimamos con el promedio de los  $p - d$  autovalores muestrales más pequeños (que en la población son todos iguales a  $\sigma^2$ ). En la Sección 3.2.1 presentamos los resultados reportados en Passemier and Yao [36]. Las Tablas 3.2–3.3 y 3.10–3.11 deben ser comparadas con las Tablas 4–6 en [36] (reproducidas en las Tablas 3.17–3.20 de la sección mencionada). Como podemos ver, nuestro método es muy competitivo cuando  $p/m = 0.1$  y  $p/m = 0.3$  pero su desempeño decae para cocientes más grandes. El problema es que, aún cuando la distribución asintótica bajo la hipótesis nula  $\mathcal{H}_d$  es casi perfecta, como observamos anteriormente, el test subestima la dimensión del subespacio spike. El test del cociente de verosimiludes elige, cerca del 95 % de las veces, ya sea la dimensión verdadera o un valor que es menor que la dimensión verdadera.

Claramente, si estuviéramos testeando la hipótesis  $\mathcal{H}_d$  : verdadera dimensión es  $d$  vs  $\mathcal{H}_a$  : verdadera dimensión es mayor que  $d$ , el test del cociente de verosimiludes no rechazará la hipótesis nula el 95 % de la veces como esperamos. Sin embargo, dado que el test es secuencial, el problema es que podría detenerse en una dimensión que es menor a la verdadera. Como veremos en el próximo capítulo, esto se debe a que la potencia del test decrece cuando el cociente  $p/m$

crece a 1 (en el caso  $p < m$ ) o decrece a 1 (en el caso  $p > m$ ), fenómeno ya advertido por Jiang and Yang [20] en su Tabla 1, para  $p < m$  y  $p/m$  creciendo a 1.

Para dar respuesta a este problema de subestimación de la dimensión, presentaremos en el Capítulo 5 una potencial solución basada en el estudio más detallado del comportamiento del estadístico de cociente de verosimilitudes bajo una hipótesis alternativa específica.

TABLA 3.1

*Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.1$ .*

$(p, n)$	Distribución de $\hat{d}$								
	1	2	3	4	5	6	7	8	9
(10,100)	0.000	0.000	0.013	<b>0.987</b>	0.000	0.000	0.000	0.000	0.000
(20,200)	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000
(40,400)	0.000	0.000	0.000	<b>0.997</b>	0.003	0.000	0.000	0.000	0.000
(80,800)	0.000	0.000	0.000	<b>0.991</b>	0.006	0.003	0.000	0.000	0.000

TABLA 3.2

*Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.3$ .*

$(p, n)$	Distribución de $\hat{d}$								
	1	2	3	4	5	6	7	8	9
(30,100)	0.000	0.000	0.110	<b>0.885</b>	0.003	0.002	0.000	0.000	0.000
(60,200)	0.000	0.000	0.032	<b>0.940</b>	0.020	0.007	0.001	0.000	0.000
(120,400)	0.000	0.000	0.006	<b>0.952</b>	0.031	0.007	0.003	0.000	0.001
(240,800)	0.000	0.000	0.001	<b>0.953</b>	0.026	0.014	0.004	0.001	0.001

TABLA 3.3

*Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.6$ .*

$(p, n)$	Distribución de $\hat{d}$								
	1	2	3	4	5	6	7	8	9
(60,100)	0.000	0.036	0.522	<b>0.409</b>	0.022	0.007	0.002	0.002	0.000
(120,200)	0.000	0.006	0.460	<b>0.496</b>	0.021	0.010	0.004	0.003	0.000
(240,400)	0.000	0.002	0.436	<b>0.525</b>	0.023	0.007	0.003	0.003	0.001
(480,800)	0.000	0.002	0.384	<b>0.563</b>	0.021	0.014	0.010	0.004	0.001

TABLA 3.4

*Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.9$ .*

$(p, n)$	Distribución de $\hat{d}$								
	1	2	3	4	5	6	7	8	9
(90,100)	0.040	0.382	0.413	<b>0.127</b>	0.019	0.011	0.003	0.004	0.000
(180,200)	0.021	0.279	0.487	<b>0.170</b>	0.013	0.014	0.006	0.007	0.002
(360,400)	0.012	0.270	0.482	<b>0.193</b>	0.015	0.012	0.006	0.006	0.002
(720,800)	0.007	0.258	0.485	<b>0.209</b>	0.014	0.007	0.003	0.009	0.003

TABLA 3.5

Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.9 = 1.11$ .

$(p, n)$	Distribución de $\hat{d}$								
	1	2	3	4	5	6	7	8	9
(100,90)	0.077	0.373	0.313	<b>0.098</b>	0.000	0.000	0.000	0.000	0.000
(200,180)	0.033	0.367	0.386	<b>0.148</b>	0.009	0.002	0.000	0.000	0.000
(400,360)	0.035	0.329	0.423	<b>0.140</b>	0.018	0.013	0.005	0.004	0.001
(800,720)	0.024	0.327	0.444	<b>0.147</b>	0.017	0.008	0.005	0.005	0.005

TABLA 3.6

Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.6 = 1.67$ .

$(p, n)$	Distribución de $\hat{d}$								
	1	2	3	4	5	6	7	8	9
(100,60)	0.059	0.414	0.365	<b>0.116</b>	0.026	0.012	0.002	0.000	0.002
(200,120)	0.020	0.336	0.462	<b>0.137</b>	0.022	0.010	0.007	0.003	0.001
(400,240)	0.012	0.254	0.518	<b>0.183</b>	0.015	0.008	0.005	0.003	0.001
(800,480)	0.004	0.228	0.525	<b>0.191</b>	0.025	0.012	0.007	0.003	0.003

TABLA 3.7

Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.3 = 3.33$ .

$(p, n)$	Distribución de $\hat{d}$								
	1	2	3	4	5	6	7	8	9
(100,30)	0.296	0.501	0.178	<b>0.021</b>	0.001	0.000	0.000	0.000	0.000
(200,60)	0.131	0.515	0.290	<b>0.053</b>	0.011	0.000	0.000	0.000	0.000
(400,120)	0.076	0.448	0.353	<b>0.098</b>	0.020	0.005	0.000	0.000	0.000
(800,240)	0.042	0.409	0.413	<b>0.111</b>	0.018	0.006	0.001	0.000	0.000

TABLA 3.8

Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.1 = 10$ .

$(p, n)$	Distribución de $\hat{d}$								
	1	2	3	4	5	6	7	8	9
(100,10)	0.000	0.000	0.000	<b>0.000</b>	0.000	0.000	0.000	0.000	0.000
(200,20)	0.000	0.000	0.000	<b>0.000</b>	0.000	0.000	0.000	0.000	0.000
(400,40)	0.000	0.000	0.000	<b>0.000</b>	0.000	0.000	0.000	0.000	0.000
(800,80)	0.000	0.000	0.000	<b>0.000</b>	0.000	0.000	0.000	0.000	0.000

TABLA 3.9

Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.1$ .

$(p, n)$	Distribución de $\hat{d}$									
	1	2	3	4	5	6	7	8	9	10
(10,100)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000
(20,200)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000
(40,400)	0.000	0.000	0.000	0.000	<b>0.999</b>	0.001	0.000	0.000	0.000	0.000
(80,800)	0.000	0.000	0.000	0.000	<b>0.991</b>	0.007	0.002	0.000	0.000	0.000

TABLA 3.10

Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.3$ .

$(p, n)$	Distribución de $\hat{d}$									
	1	2	3	4	5	6	7	8	9	10
(30,100)	0.000	0.000	0.000	0.011	<b>0.982</b>	0.004	0.003	0.000	0.000	0.000
(60,200)	0.000	0.000	0.000	0.002	<b>0.972</b>	0.017	0.008	0.001	0.000	0.000
(120,400)	0.000	0.000	0.000	0.000	<b>0.976</b>	0.014	0.007	0.002	0.001	0.000
(240,800)	0.000	0.000	0.000	0.000	<b>0.966</b>	0.020	0.006	0.007	0.000	0.000

TABLA 3.11

Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.6$ .

$(p, n)$	Distribución de $\hat{d}$									
	1	2	3	4	5	6	7	8	9	10
(60,100)	0.000	0.000	0.000	0.285	<b>0.700</b>	0.006	0.006	0.003	0.000	0.000
(120,200)	0.000	0.000	0.000	0.184	<b>0.782</b>	0.021	0.009	0.004	0.000	0.000
(240,400)	0.000	0.000	0.000	0.146	<b>0.810</b>	0.027	0.008	0.003	0.005	0.000
(480,800)	0.000	0.000	0.000	0.130	<b>0.801</b>	0.039	0.016	0.008	0.003	0.002

TABLA 3.12

Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.9$ .

$(p, n)$	Distribución de $\hat{d}$									
	1	2	3	4	5	6	7	8	9	10
(90,100)	0.000	0.000	0.048	0.637	<b>0.254</b>	0.030	0.015	0.006	0.005	0.004
(180,200)	0.000	0.000	0.018	0.629	<b>0.307</b>	0.018	0.011	0.005	0.003	0.004
(360,400)	0.000	0.000	0.011	0.655	<b>0.302</b>	0.007	0.010	0.004	0.005	0.004
(720,800)	0.000	0.000	0.011	0.610	<b>0.316</b>	0.019	0.021	0.004	0.008	0.002

TABLA 3.13

Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.9 = 1.11$ .

$(p, n)$	Distribución de $\hat{d}$									
	1	2	3	4	5	6	7	8	9	10
(100,90)	0.000	0.000	0.047	0.465	<b>0.220</b>	0.000	0.000	0.000	0.000	0.000
(200,180)	0.000	0.000	0.034	0.567	<b>0.278</b>	0.004	0.000	0.000	0.000	0.000
(400,360)	0.000	0.000	0.043	0.617	<b>0.265</b>	0.021	0.004	0.002	0.001	0.001
(800,720)	0.000	0.000	0.017	0.657	<b>0.264</b>	0.019	0.011	0.006	0.004	0.003

TABLA 3.14

Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.6 = 1.67$ .

$(p, n)$	Distribución de $\hat{d}$									
	1	2	3	4	5	6	7	8	9	10
(100,60)	0.000	0.000	0.045	0.640	<b>0.262</b>	0.027	0.016	0.006	0.002	0.000
(200,120)	0.000	0.000	0.010	0.623	<b>0.312</b>	0.022	0.012	0.014	0.003	0.003
(400,240)	0.000	0.000	0.005	0.631	<b>0.322</b>	0.019	0.012	0.005	0.001	0.002
(800,480)	0.000	0.000	0.006	0.607	<b>0.341</b>	0.018	0.022	0.002	0.002	0.002

TABLA 3.15

*Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.3 = 3.33$ .*

$(p, n)$	Distribución de $\hat{d}$									
	1	2	3	4	5	6	7	8	9	10
(100,30)	0.000	0.004	0.407	0.546	<b>0.043</b>	0.000	0.000	0.000	0.000	0.000
(200,60)	0.000	0.000	0.150	0.730	<b>0.113</b>	0.006	0.001	0.000	0.000	0.000
(400,120)	0.000	0.000	0.069	0.730	<b>0.176</b>	0.018	0.006	0.001	0.000	0.000
(800,240)	0.000	0.000	0.036	0.690	<b>0.255</b>	0.014	0.005	0.000	0.000	0.000

TABLA 3.16

*Distribución empírica de  $\hat{d}$  sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.1 = 10$ .*

$(p, n)$	Distribución de $\hat{d}$									
	1	2	3	4	5	6	7	8	9	10
(100,10)	0.125	0.820	0.052	0.000	<b>0.000</b>	0.000	0.000	0.000	0.000	0.000
(200,20)	0.005	0.490	0.504	0.001	<b>0.000</b>	0.000	0.000	0.000	0.000	0.000
(400,40)	0.000	0.139	0.791	0.070	<b>0.000</b>	0.000	0.000	0.000	0.000	0.000
(800,80)	0.000	0.025	0.696	0.273	<b>0.006</b>	0.000	0.000	0.000	0.000	0.000

### 3.2.1 Resultados de Passemier and Yao [36]

La Tablas 3.2–3.3 y 3.10–3.11 deben ser comparadas con las Tablas 4-6 en [36]. Para facilitar la lectura en la presente secci3n reproducimos los resultados reportados en [36]. Adem3s, a modo de resumen, las Figuras 3.1a y 3.1b muestran la frecuencia de  $\hat{d} = d$  y  $\hat{q} = d$  para los **Modelo 1** y **Modelo 2**, respectivamente.

TABLA 3.17

*Distribuci3n empírica de  $\hat{q}$  sobre 1000 repeticiones para el Modelo 1 con  $p/n = 0.3$ .  
Método de Passemier and Yao [36].*

$(p, n)$	Distribuci3n de $\hat{q}$					
	1	2	3	4	5	6
(30, 100)	0.078	0.091	0.0	<b>0.658</b>	0.094	0.0
(60, 200)	0.038	0.026	0.0	<b>0.805</b>	0.098	0.001
(120, 400)	0.008	0.004	0.0	<b>0.878</b>	0.091	0.0
(240, 800)	0.001	0.001	0.0	<b>0.907</b>	0.088	0.0

TABLA 3.18

*Distribuci3n empírica de  $\hat{q}$  sobre 1000 repeticiones para el Modelo 1 con  $p/n = 0.6$ .  
Método de Passemier and Yao [36].*

$(p, n)$	Distribuci3n de $\hat{q}$					
	1	2	3	4	5	6
(60, 100)	0.081	0.096	0.001	<b>0.674</b>	0.079	0.0
(120, 200)	0.034	0.036	0.0	<b>0.806</b>	0.089	0.001
(240, 400)	0.015	0.007	0.0	<b>0.892</b>	0.071	0.0
(480, 800)	0.003	0.0	0.0	<b>0.926</b>	0.069	0.0

TABLA 3.19

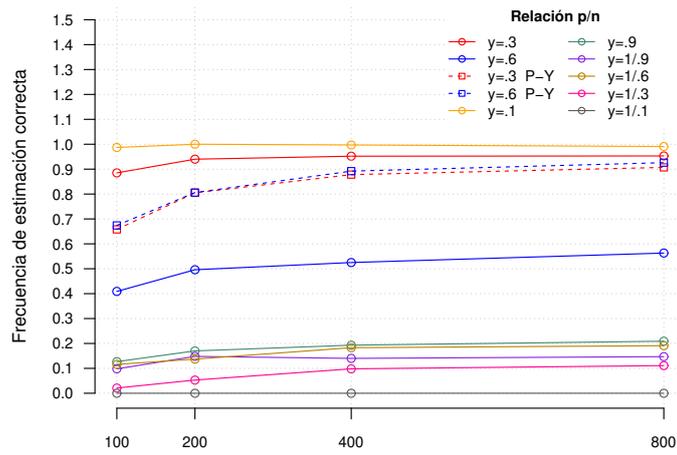
*Distribuci3n empírica de  $\hat{q}$  sobre 1000 repeticiones para el Modelo 2 con  $p/n = 0.3$ .  
Método de Passemier and Yao [36].*

$(p, n)$	Distribuci3n de $\hat{q}$						
	1	2	3	4	5	6	7
(30, 100)	0.003	0.015	0.008	0.0	<b>0.849</b>	0.0	0.125
(60, 200)	0.0	0.001	0.0	0.0	<b>0.89</b>	0.107	0.002
(120, 400)	0.0	0.001	0.0	0.0	<b>0.927</b>	0.072	0.0
(240, 800)	0.0	0.0	0.0	0.0	<b>0.916</b>	0.084	0.0

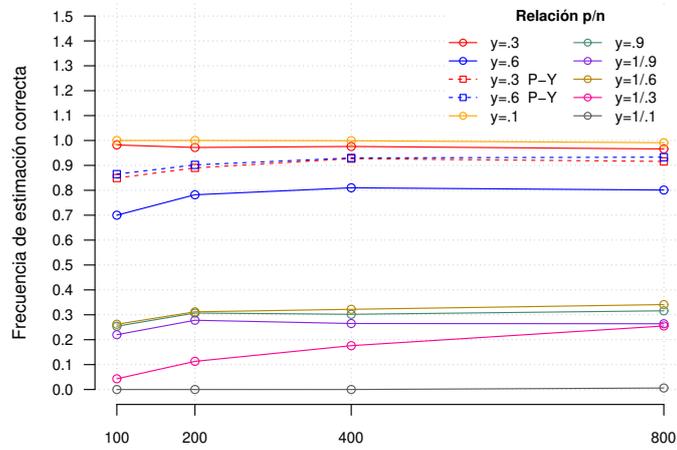
TABLA 3.20

*Distribuci3n empírica de  $\hat{q}$  sobre 1000 repeticiones para el Modelo 2 con  $p/n = 0.6$ .  
Método de Passemier and Yao [36].*

$(p, n)$	Distribuci3n de $\hat{q}$						
	1	2	3	4	5	6	7
(60, 100)	0.001	0.009	0.004	0.0	<b>0.865</b>	0.122	0.002
(120, 200)	0.0	0.0	0.001	0.0	<b>0.902</b>	0.097	0.0
(240, 400)	0.0	0.0	0.0	0.0	<b>0.93</b>	0.07	0.0
(480, 800)	0.0	0.0	0.0	0.0	<b>0.933</b>	0.067	0.0



(a) Modelo 1 ( $d = 4$ )



(b) Modelo 2 ( $d = 5$ )

**Figura 3.1.** Frecuencia de estimación correcta sobre 1000 repeticiones independientes para el método de Passemier and Yao [36] (línea de puntos) y el test propuesto en este capítulo.



## Potencia del test propuesto

En los capítulos anteriores nos hemos centrado en el estudio de la distribución del logaritmo del estadístico de cociente de verosimilitudes  $\log(LRT_d)$  bajo la hipótesis nula  $\mathcal{H}_d$  y, a partir de ella, hemos desarrollado un test para determinar la dimensión  $d$  del subespacio spiked. Con el objetivo de alcanzar mayor entendimiento sobre el comportamiento asintótico del estadístico, nos gustaría conocer la distribución del mismo cuando no estamos considerando el número correcto (esto es, verdadero) de spikes. Dicho análisis nos permitirá determinar cuál es la potencia del test considerado.

A lo largo de este capítulo asumiremos que la verdadera dimensión del subespacio spiked es  $d_1$  y que el estadístico en consideración es  $LRT_d$  con  $d$  menor o mayor que  $d_1$ .

### 4.1 Distribución asintótica de $\log(LRT_d)$ cuando la verdadera dimensión es $d_1$

**Proposición 4.1.** *La distribución asintótica del cociente de verosimilitudes logarítmico  $\log LRT_d$  para  $p/m \rightarrow y$  cuando el modelo spike tiene dimensión verdadera  $d_1$ , con  $d \neq d_1 \leq q_0$ , está dada por*

$$\frac{\log(LRT_d) - \mu_{m,p,d,d_1}}{\sigma_{m,p,d_1}} \rightarrow \mathcal{N}(0, 1),$$

donde

a) para  $p < m$ ,  $\mu_{m,p,d,d_1}$  es igual a  $\mu_{m,p,d_1}$  más

$$\begin{cases} \sum_{i=h_0+1}^{h_1} s_i (\log k_i - k_i + 1) & \text{si } d \leq d_1, \\ (d - d_1) [(1 + \sqrt{y})^2 - \log(1 + \sqrt{y})^2 - 1] & \text{si } d_1 \leq d \leq q_0, \end{cases} \quad (4.1)$$

b) para  $p > m + d_1 + 3$ ,  $\mu_{m,p,d,d_1}$  es igual a  $\mu_{m,p,d_1}$  más

$$\begin{cases} \sum_{i=h_0+1}^{h_1} s_i \left( \log \frac{k_i}{y} - \frac{k_i}{y} + 1 \right) & \text{si } d \leq d_1, \\ (d - d_1) \left[ \frac{(1 + \sqrt{y})^2}{y} - \log \frac{(1 + \sqrt{y})^2}{y} - 1 \right] & \text{si } d_1 \leq d \leq q_0. \end{cases} \quad (4.2)$$

En todas partes,  $\mu_{m,p,d_1}$  se define como en las Proposiciones 2.4 y 2.5 según corresponda, pero reemplazando  $d$  por  $d_1$ , con  $k_i = \lambda_i / \sigma^2 \{1 + y\sigma^2 / (\lambda_i - \sigma^2)\}$  si  $i \in J_{1,k}$  y  $k_i = (1 + \sqrt{y})^2$  si  $i \in J_{2,k}$  o cuando  $\lambda_i = \sigma^2$  donde para  $k = 0, 1$ ,  $J_{1,k} = \{i \leq h_k : \lambda_i > \sigma^2(1 + \sqrt{y})\}$ ,  $J_{2,k} = \{i \leq h_k : \lambda_i \in (\sigma^2, \sigma^2(1 + \sqrt{y})]\}$ , y  $h_0, h_1$  son tales que  $d = s_1 + \dots + s_{h_0}$  y  $d_1 = s_1 + \dots + s_{h_1}$ .

A partir de la Proposición 4.1 conocemos la distribución de  $\log(LRT_d)$  cuando la verdadera dimensión en realidad es  $d_1 \neq d$ . Sin embargo, notemos que dado que el test secuencialmente testea si la dimensión es  $d$  versus si es mayor que  $d$ , para explorar la potencia del test derivado en el Capítulo 3, vamos a considerar la hipótesis alternativa de que la verdadera dimensión es  $d_1 > d$ .

## 4.2 Potencia del Test

La siguiente proposición nos da la potencia asintótica del test de cociente de verosimilitudes para testear las hipótesis

$$\begin{aligned} \mathcal{H}_d &: \text{El subespacio spike tiene dimensión } d \\ \mathcal{H}_a &: \text{El subespacio spike tiene dimensión mayor que } d, \end{aligned} \quad (4.3)$$

para el caso de la hipótesis alternativa específica de que el verdadero modelo es spiked de dimensión  $d_1 > d$  y que todos los autovalores spike son mayores que el umbral.

**Proposición 4.2.** *La potencia asintótica del test de cociente de verosimilitudes correspondiente a las hipótesis (4.3) para la hipótesis alternativa específica de que el verdadero modelo es spiked de dimensión  $d_1$ , con  $d_1 \in (d, q_0]$ , y que todos los autovalores spike son mayores que el umbral, esta dada por:*

1. Caso  $p < m$ :

$$\psi(d_1) = \Phi \left( \frac{\sum_{i=h_0+1}^{h_1} s_i \left\{ \frac{\lambda_i}{\sigma^2} - \log\left(\frac{\lambda_i}{\sigma^2}\right) - 1 \right\} + z_\alpha \sigma_{m,p,d}}{\sigma_{m,p,d_1}} \right). \quad (4.4)$$

2. Caso  $p > m + d_1 + 3$ :

$$\psi(d_1) = \Phi \left( \frac{\sum_{i=h_0+1}^{h_1} s_i \left\{ \frac{\lambda_i}{y\sigma^2} - \frac{1}{y} - \log\left(1 - \frac{1}{y} + \frac{\lambda_i}{y\sigma^2}\right) \right\} + z_\alpha \sigma_{m,p,d}}{\sigma_{m,p,d_1}} \right), \quad (4.5)$$

donde  $\Phi$  es la distribución acumulada de la normal estándar y  $z_\alpha$  es su cuantil de orden  $1 - \alpha$ .

## 4.3 Simulaciones

Continuando con las simulaciones presentadas en las secciones previas, en esta sección mostramos, para cada uno de los escenarios considerados, los resultados obtenidos en cuanto a la proporción de rechazos, entre 1000 repeticiones, de que la dimensión del subespacio spiked es igual a  $d$  para  $d = 1, 2, 3$  en el caso del **Modelo 1** y  $d = 1, 2, 3, 4$  para el **Modelo 2**. Es decir, estudiamos la probabilidad de rechazar la hipótesis nula para cada uno de los valores por debajo de la verdadera dimensión. Recordemos que para el **Modelo 1** la dimensión del subespacio de los spikes es  $d_1 = 4$ , con autovalores spike iguales a  $(7, 6, 5, 4)$ , mientras que en el **Modelo 2** la dimensión del subespacio de los spikes es  $d_1 = 5$ , con autovalores spike iguales a  $(259.72, 17.97, 11.04, 7.88, 4.82)$ .

TABLA 4.1

Proporción de rechazos sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.1$ .

$(p, n)$	$d$		
	1	2	3
(10,100)	1.000	1.000	0.987
(20,200)	1.000	1.000	1.000
(40,400)	1.000	1.000	1.000
(80,800)	1.000	1.000	1.000

TABLA 4.2

Proporción de rechazos sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.3$ .

$(p, n)$	$d$		
	1	2	3
(30,100)	1.000	1.000	0.890
(60,200)	1.000	1.000	0.968
(120,400)	1.000	1.000	0.994
(240,800)	1.000	1.000	0.999

TABLA 4.3

Proporción de rechazos sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.6$ .

$(p, n)$	$d$		
	1	2	3
(60,100)	1.000	0.964	0.442
(120,200)	1.000	0.994	0.534
(240,400)	1.000	0.998	0.562
(480,800)	1.000	0.998	0.614

TABLA 4.4

Proporción de rechazos sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.9$ .

$(p, n)$	$d$		
	1	2	3
(90,100)	0.960	0.578	0.165
(180,200)	0.979	0.700	0.213
(360,400)	0.988	0.718	0.236
(720,800)	0.993	0.735	0.250

TABLA 4.5

Proporción de rechazos sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.9 = 1.11$ .

$(p, n)$	$d$		
	1	2	3
(100,90)	0.784	0.411	0.098
(200,180)	0.912	0.545	0.159
(400,360)	0.933	0.604	0.181
(800,720)	0.971	0.644	0.200

TABLA 4.6

Proporción de rechazos sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.6 = 1.67$ .

$(p, n)$	$d$		
	1	2	3
(100,60)	0.940	0.526	0.161
(200,120)	0.980	0.644	0.182
(400,240)	0.988	0.734	0.216
(800,480)	0.996	0.768	0.243

TABLA 4.7

Proporción de rechazos sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.3 = 3.33$ .

$(p, n)$	$d$		
	1	2	3
(100,30)	0.701	0.200	0.022
(200,60)	0.869	0.354	0.064
(400,120)	0.924	0.476	0.123
(800,240)	0.958	0.549	0.136

TABLA 4.8

Proporción de rechazos sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.1$ .

$(p, n)$	$d$		
	1	2	3
(100,10)	1.000	1.000	1.000
(200,20)	1.000	1.000	1.000
(400,40)	1.000	1.000	1.000
(800,80)	1.000	1.000	1.000

TABLA 4.9

Proporción de rechazos sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.1$ .

$(p, n)$	$d$			
	1	2	3	4
(10,100)	1.000	1.000	1.000	1.000
(20,200)	1.000	1.000	1.000	1.000
(40,400)	1.000	1.000	1.000	1.000
(80,800)	1.000	1.000	1.000	1.000

TABLA 4.10

Proporción de rechazos sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.3$ .

$(p, n)$	$d$			
	1	2	3	4
(30,100)	1.000	1.000	1.000	0.989
(60,200)	1.000	1.000	1.000	0.998
(120,400)	1.000	1.000	1.000	1.000
(240,800)	1.000	1.000	1.000	1.000

TABLA 4.11

Proporción de rechazos sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.6$ .

$(p, n)$	$d$			
	1	2	3	4
(60,100)	1.000	1.000	1.000	0.715
(120,200)	1.000	1.000	1.000	0.816
(240,400)	1.000	1.000	1.000	0.854
(480,800)	1.000	1.000	1.000	0.870

TABLA 4.12

Proporción de rechazos sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.9$ .

$(p, n)$	$d$			
	1	2	3	4
(90,100)	1.000	1.000	0.952	0.315
(180,200)	1.000	1.000	0.982	0.353
(360,400)	1.000	1.000	0.989	0.334
(720,800)	1.000	1.000	0.989	0.379

TABLA 4.13

Proporción de rechazos sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.9 = 1.11$ .

$(p, n)$	$d$			
	1	2	3	4
(100,90)	0.732	0.732	0.685	0.220
(200,180)	0.883	0.883	0.849	0.282
(400,360)	0.954	0.954	0.911	0.294
(800,720)	0.996	0.996	0.979	0.322

TABLA 4.14

Proporción de rechazos sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.6 = 1.67$ .

$(p, n)$	$d$			
	1	2	3	4
(100,60)	0.999	0.999	0.954	0.314
(200,120)	1.000	1.000	0.990	0.367
(400,240)	1.000	1.000	0.995	0.364
(800,480)	1.000	1.000	0.994	0.387

TABLA 4.15

Proporción de rechazos sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.3 = 3.33$ .

$(p, n)$	$d$			
	1	2	3	4
(100,30)	1.000	0.996	0.589	0.043
(200,60)	1.000	1.000	0.850	0.120
(400,120)	1.000	1.000	0.931	0.201
(800,240)	1.000	1.000	0.964	0.274

TABLA 4.16

Proporción de rechazos sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.1 = 10$ .

$(p, n)$	$d$			
	1	2	3	4
(100,10)	0.872	0.052	0.000	0.000
(200,20)	0.995	0.505	0.001	0.000
(400,40)	1.000	0.861	0.070	0.000
(800,80)	1.000	0.975	0.279	0.006

## 4.4 Algunos comentarios

En primer lugar notemos que para  $y$  fijo, cuando  $m$  y  $p$  crecen,  $\sigma_{m,p,d}/\sigma_{m,p,d_1} \rightarrow 1$  cuando  $p/m \rightarrow y$ . Ahora bien, a medida que más grande son los autovalores  $\lambda_i$  con  $i \in \{h_0 + 1, \dots, h_1\}$ , mayor es  $\sum_{i=h_0+1}^{h_1} s_i \{\lambda_i/\sigma^2 - \log(\lambda_i/\sigma^2) - 1\} > 0$  y  $\sum_{i=h_0+1}^{h_1} s_i [\lambda_i/(y\sigma^2) - 1/y - \log\{1 - 1/y + \lambda_i/(y\sigma^2)\}] > 0$ . En consecuencia, mayores valores de  $\lambda_i$  implican una mayor potencia. Más aún  $\lambda_i \rightarrow \infty$  para  $i \in \{h_0 + 1, \dots, h_1\}$  implica que la potencia tienda a 1. Por otro lado, para  $\lambda_i$  fijo, con  $i \in \{h_0 + 1, \dots, h_1\}$  e  $y \rightarrow 1$ , tenemos que  $\sigma_{m,p,d_1} \rightarrow \infty$ ,  $\sigma_{m,p,d}/\sigma_{m,p,d_1} \rightarrow 1$  y, por lo tanto, la potencia decrece a  $\alpha$ . Para  $y \rightarrow 0$ ,  $\sigma_{m,p,d_1} \rightarrow 0$ ,  $\sigma_{m,p,d}/\sigma_{m,p,d_1} \rightarrow 1$  y in tal caso, la potencia tiende a 1.

A la luz de la Proposición 4.2 y la explicación dada arriba, los resultados obtenidos en el Capítulo 3 no deberían sorprendernos. En aquellas simulaciones vimos que el test de cociente de verosimilitudes subestima la verdadera dimensión cuando el límite de  $p/m$  es cercano a 1 (por arriba o por debajo). Esto se confirma con el comportamiento de la potencia del test cuando  $y \approx 1$ . La consecuencia de esto es que el test secuencial se detiene antes de lo que se supone debería hacerlo. A pesar de esto, sabemos cuál es el comportamiento del estadístico como función de la hipótesis nula (gracias a la Proposición 4.1). Precisamente, el conocer tal comportamiento nos permitirá modificar el estadístico penalizando el número de spikes elegidos, como estudiaremos en el próximo capítulo.



---

# Cociente de verosimilitudes penalizado

---

El test que hemos desarrollado en los capítulos previos es secuencial y por lo tanto elige como estimación para la dimensión  $d$  del subespacio spiked el primer valor para el cual no rechazamos la hipótesis nula. Dado que la potencia del test decrece cuando  $p/n$  crece o decrece a 1, el test subestima la verdadera dimensión cuando  $p/n$  esta cerca de 1 o bien cuando los autovalores spike son pequeños (esto es, cuando están cerca o son menores que el umbral  $\sigma^2(1 + \sqrt{y})$ ).

Con el objetivo de solucionar este problema de subestimación de la dimensión, vamos a estudiar el comportamiento del estadístico  $\log(LRT_d)$  como una función de la dimensión  $d$  cuando la verdadera dimensión es  $d_1$ , para cualquier valor de  $d_1$ .

En la Proposición 5.1 veremos que la tasa de crecimiento de la media asintótica del logaritmo del cociente de verosimilitudes cambia al pasar de  $d < d_1$  a  $d > d_1$ . Esto nos permitirá definir un nuevo estimador consistente de la verdadera dimensión  $d_1$  del subespacio spiked a través de criterios de teoría de la información, como lo hicieron [50] para el caso de  $p$  fijo.

## 5.1 Intuición para el estadístico penalizado

Antes de dar la definición de la versión penalizada del logaritmo del estadístico de cociente de verosimilitudes, veamos un ejemplo que nos ayude a alcanzar cierta intuición. Para ello consideremos el **Modelo 1** (recordemos que para este modelo la verdadera dimensión es  $d_1 = 4$ , con autovalores spike  $\lambda = (7, 6, 5, 4)$  y  $\sigma^2 = 1$ .) En la Figura 5.1a graficamos la función  $\log(LRT_d)$ . Como podemos observar,  $\log(LRT_d)$  es creciente como función de la dimensión  $d$ . Sin embargo, la tasa de crecimiento es diferente según sea  $d < d_1$  o  $d > d_1$ . La idea entonces es

restarle “algo” al  $\log(LRT_d)$  de modo tal que obtengamos una función que resulte creciente para  $d < d_1$  y decreciente para  $d > d_1$ . Si logramos esto, la estimación de la verdadera dimensión  $d_1$  estará dada por el máximo de la función mencionada.

Para determinar ese “algo” que necesitamos restar al  $\log(LRT_d)$ , en la Proposición 5.1 estudiamos la media asintótica del  $\log(LRT_d)$  cuando la verdadera dimensión es  $d_1$ . Y es ahí que vemos que la media del estadístico es igual a un término constante (como función de  $d$ ) más otro término que es diferente según sea  $d < d_1$  o  $d > d_1$ . Además, vemos que si a la media le restamos el término diferente correspondiente a la zona en que  $d > d_1$  más un  $\epsilon > 0$ , entonces obtenemos el comportamiento que buscamos (es decir, la media menos ese término presenta un máximo en  $d = d_1$ ). La aparición del  $\epsilon$  intuitivamente tiene que ver con que, si sólo le restáramos el término en cuestión, obtendríamos una función que luego de  $d_1$  sería constante (pues, básicamente, le estamos restando su crecimiento) y no decreciente como queremos para que resulte más fácil de detectar computacionalmente.

En la Figura 5.1b graficamos la función  $\log(LRT_d) - d \{h[\sigma^2(1 + \sqrt{y}) + \epsilon]\}$  (esto es, el  $\log(LRT_d)$  más el término penalizante). Podemos observar allí que la función resultante presenta un máximo global en la verdadera dimensión  $d_1$ .

## 5.2 Versión penalizada del cociente de verosimilitudes

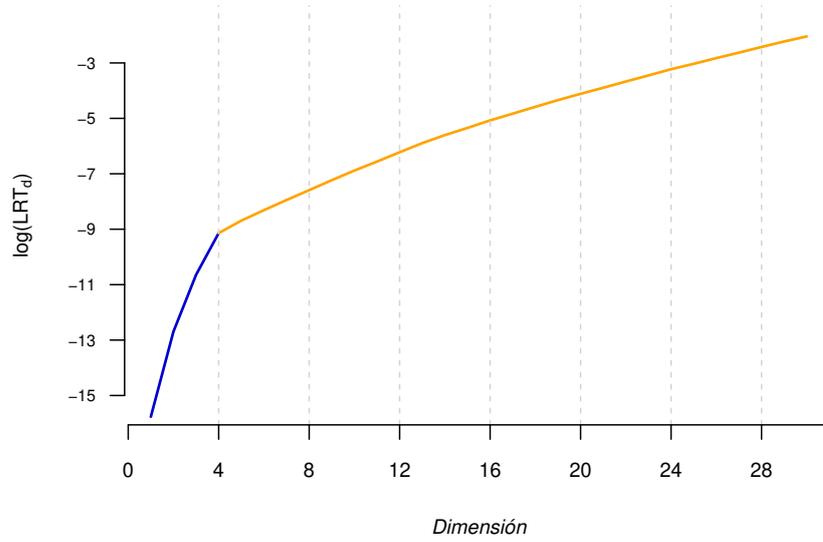
**Proposición 5.1.** *Supongamos que el subespacio spiked tiene dimensión verdadera  $d_1$ . Dada  $\mu_{m,p,d,d_1}$  definida en la Proposición 4.1 y  $\epsilon \geq 0$ , consideremos, para  $\tilde{y} = \max(1, y)$ , la función  $g(d) = \mu_{m,p,d,d_1} - (d - d_1) \{h[\sigma^2(1 + \sqrt{y})] + \epsilon\}$  con*

$$h(\lambda) = \frac{\lambda}{\tilde{y}\sigma^2} \left(1 + \frac{y\sigma^2}{\lambda - \sigma^2}\right) - \log \left[ \frac{\lambda}{\tilde{y}\sigma^2} \left(1 + \frac{y\sigma^2}{\lambda - \sigma^2}\right) \right] - 1. \quad (5.1)$$

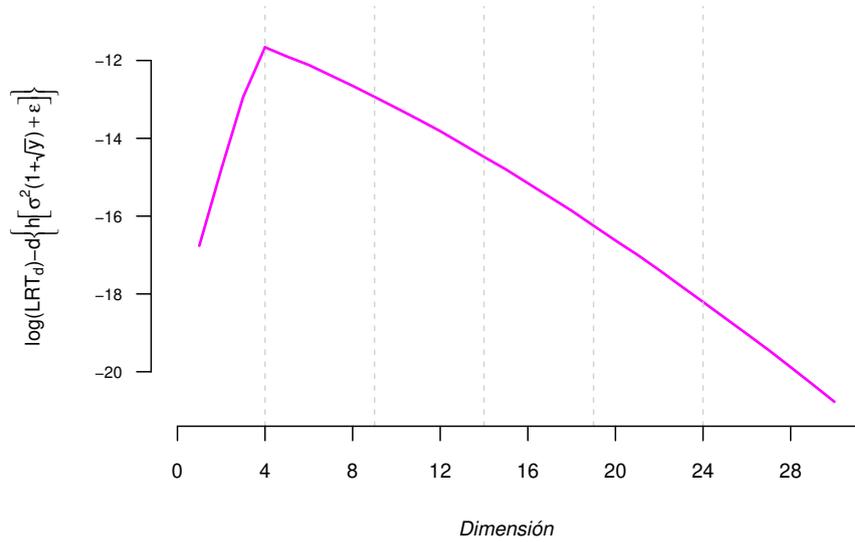
*Sea  $\lambda^* > \sigma^2(1 + \sqrt{y})$  tal que  $\epsilon = h(\lambda^*) - h[\sigma^2(1 + \sqrt{y})]$ . Entonces, si todos los autovalores spike son mayores que  $\lambda^*$ , la función  $g$  es creciente para  $d < d_1$ , decreciente para  $d \geq d_1$  y, por lo tanto, tiene un máximo global en  $d = d_1$ .*

**Corolario 5.2.** *Consideremos la función  $g(d)$  definida en la Proposición 5.1. Entonces, para  $\epsilon = 0$ , la función  $g$  es creciente para  $d < d_1$  y constante para  $d \geq d_1$ . Más aún, si todos los autovalores spike son mayores que el umbral  $\sigma^2(1 + \sqrt{y})$ , entonces  $g$  es estrictamente creciente para  $d \leq d_1$ .*

La Proposición 5.1 nos describe el comportamiento de la media de la distribución de los distintos estadísticos utilizados en el test secuencial. En la población, si  $\epsilon > 0$  y todos los



(a)  $\log(LRT_d)$  como función de la dimensión  $d$ . En la parte azul ( $d < d_1$ ), la media asintótica del  $\log(LRT_d)$  es igual a  $\mu_{m,p,d_1} + \sum_{i=h_0+1}^{h_1} s_i \left( \log \frac{k_i}{\tilde{y}} - \frac{k_i}{\tilde{y}} + 1 \right)$ , mientras que en la parte anaranjada ( $d > d_1$ ) es igual a  $\mu_{m,p,d_1} + (d - d_1) \left[ \frac{(1 + \sqrt{y})^2}{\tilde{y}} - \log \frac{(1 + \sqrt{y})^2}{\tilde{y}} - 1 \right]$ .



(b)  $\log(LRT_d)$  penalizado como función de la dimensión  $d$ .

**Figura 5.1.** Comportamiento de  $\log(LRT_d)$  y de la versión penalizada como función de la dimensión  $d$  para el Modelo 1, cuando  $y = 0.3$ ,  $p = 60$  y  $n = 200$ .

autovalores spike son mayores que  $\lambda^*$ , entonces la media de la distribución del estadístico del test más un término penalizante presenta un máximo global en  $d = d_1$ . Inspirados por este resultado, definiremos un nuevo estimador ( $\hat{d}_\epsilon$ ) para la verdadera dimensión. Para  $\epsilon \geq 0$  fijo, sea

$$\hat{d}_\epsilon = \min \left\{ \arg \max_{0 \leq j \leq q_0} \{ \log(LRT_j) - j \{ h[\sigma^2(1 + \sqrt{y})] + \epsilon \} \} \right\}, \quad (5.2)$$

donde  $q_0$  es una cota superior para  $d$  (Condición  $Q_0$ ).

Ahora, dado que la función  $h(\lambda)$  es estrictamente creciente cuando  $\lambda > \sigma^2(1 + \sqrt{y})$  (ver (F.2)), existirá un valor  $\lambda^*$  tal que  $\epsilon = h(\lambda^*) - h\{\sigma^2(1 + \sqrt{y})\}$ . Este nuevo estimador, aunque se perderá los autovalores posicionados entre el umbral y  $\lambda^*$ , con probabilidad alta, detectará todos los autovalores mayores que  $\lambda^*$  cuando  $m, p \rightarrow \infty$ . Como una consecuencia de la Proposición 5.1, tenemos:

**Proposición 5.3.** *Supongamos que la dimensión verdadera del subespacio spiked es  $d_1$ . Si todos los autovalores spike son mayores que  $\lambda^*$ , entonces  $\hat{d}_\epsilon$  definido en (5.2) es un estimador consistente de  $d_1$ , es decir que*

$$\mathbb{P}(\hat{d}_\epsilon = d_1) \rightarrow 1 \quad \text{cuando } p, m \rightarrow \infty, \text{ con } \frac{p}{m} \rightarrow y > 0.$$

**Observación 5.4.** *Como lo discutimos arriba, la elección de  $\epsilon$  determinará cuáles autovalores spike no podrán ser detectados. El método se perderá, probablemente, todo autovalor  $\lambda$  que sea menor que  $\lambda^*$ , donde  $\lambda^*$  satisface  $\epsilon = h(\lambda^*) - h\{\sigma^2(1 + \sqrt{y})\}$ . Por lo tanto, la estrategia para la elección de  $\epsilon$  deberá ser elegir primero un valor  $\lambda^*$  ligeramente mayor que  $\sigma^2(1 + \sqrt{y})$  para luego definir  $\epsilon$  como la mencionada diferencia. Vale la pena remarcar que, como lo dice la proposición, este método es consistente siempre y cuando hayamos elegido un valor para  $\lambda^*$  que se encuentre a la izquierda de todos los autovalores spike que son mayores que el umbral  $\sigma^2(1 + \sqrt{y})$ . En otro caso, si el valor elegido para  $\lambda^*$  resulta ser mayor que algunos de los autovalores relevantes, esos autovalores muy probablemente no serán detectados y, por lo tanto, la dimensión estimada será menor que la verdadera  $d_1$ .*

### 5.3 Comparación con el método de Kritchman–Nadler

En procedimiento definido en (5.2) nos proporciona otra forma de estimar la verdadera dimensión del subespacio spike. Para ilustrar este estimador replicaremos las simulaciones ya

presentadas en el Capítulo 3. Las Tablas 5.1–5.16 muestran los resultados para los casos correspondientes a las Tablas 3.1–3.16. Como podemos observar, los resultados mejoran notablemente.

TABLA 5.1

*Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.1$ .*

$(p, n)$	Distribución de $\hat{d}$ penalizado y de K–N							
	3		4		5		6	
	HD–MDL	K–N	HD–MDL	K–N	HD–MDL	K–N	HD–MDL	K–N
(10,100)	0.000	0.000	<b>0.999</b>	<b>0.963</b>	0.001	0.037	0.000	0.000
(20,200)	0.000	0.000	<b>1.000</b>	<b>0.982</b>	0.001	0.018	0.000	0.000
(30,300)	0.000	0.000	<b>1.000</b>	<b>0.988</b>	0.000	0.012	0.000	0.000
(40,400)	0.000	0.000	<b>1.000</b>	<b>0.990</b>	0.001	0.009	0.000	0.001

TABLA 5.2

*Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.3$ .*

$(p, n)$	Distribución de $\hat{d}$ penalizado y de K–N							
	3		4		5		6	
	HD–MDL	K–N	HD–MDL	K–N	HD–MDL	K–N	HD–MDL	K–N
(30,100)	0.000	0.000	<b>0.999</b>	<b>0.979</b>	0.021	0.001	0.000	0.000
(60,200)	0.000	0.000	<b>0.999</b>	<b>0.992</b>	0.008	0.001	0.000	0.000
(120,400)	0.000	0.000	<b>1.000</b>	<b>0.995</b>	0.005	0.000	0.000	0.000
(240,800)	0.000	0.000	<b>1.000</b>	<b>0.993</b>	0.007	0.000	0.000	0.000

TABLA 5.3

*Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.6$ .*

$(p, n)$	Distribución de $\hat{d}$ penalizado y de K–N							
	3		4		5		6	
	HD–MDL	K–N	HD–MDL	K–N	HD–MDL	K–N	HD–MDL	K–N
(60,100)	0.003	0.002	<b>0.978</b>	<b>0.974</b>	0.019	0.023	0.000	0.001
(120,200)	0.000	0.000	<b>0.995</b>	<b>0.989</b>	0.005	0.011	0.000	0.000
(240,400)	0.000	0.000	<b>1.000</b>	<b>0.991</b>	0.000	0.009	0.000	0.000
(480,800)	0.000	0.000	<b>1.000</b>	<b>0.997</b>	0.000	0.003	0.000	0.000

Tal como lo hicimos en el Capítulo 4, resumimos los resultados sobre la proporción de veces que el método estima correctamente la verdadera dimensión en las Figuras 5.2a y 5.2b. Como podemos observar, nuestro método resulta muy superior al método de Passemier and Yao [36] excepto en el caso en que  $p/n = 1/0.3 = 3.33$  y  $p/n = 1/0.1 = 10$ .

También replicamos los escenarios que se muestran en las Figuras 7 y 8 en [25]. Los resultados obtenidos los presentamos en la Figura 5.3. En ambos casos graficamos la probabilidad

TABLA 5.4

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 1 con  $p/m = 0.9$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado y de K-N							
	3		4		5		6	
	HD-MDL	K-N	HD-MDL	K-N	HD-MDL	K-N	HD-MDL	K-N
(90,100)	0.008	0.013	<b>0.922</b>	<b>0.969</b>	0.067	0.018	0.003	0.000
(180,200)	0.000	0.000	<b>0.978</b>	<b>0.989</b>	0.021	0.011	0.001	0.000
(360,400)	0.000	0.000	<b>0.996</b>	<b>0.994</b>	0.004	0.006	0.000	0.000
(720,800)	0.000	0.000	<b>1.000</b>	<b>0.997</b>	0.000	0.003	0.000	0.000

TABLA 5.5

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.9 = 1.11$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado y de K-N							
	3		4		5		6	
	HD-MDL	K-N	HD-MDL	K-N	HD-MDL	K-N	HD-MDL	K-N
(100,90)	0.039	0.051	<b>0.930</b>	<b>0.936</b>	0.031	0.013	0.000	0.000
(200,180)	0.002	0.002	<b>0.989</b>	<b>0.982</b>	0.009	0.016	0.000	0.000
(400,360)	0.000	0.000	<b>0.998</b>	<b>0.989</b>	0.002	0.011	0.000	0.000
(800,720)	0.000	0.000	<b>1.000</b>	<b>0.998</b>	0.000	0.002	0.000	0.000

TABLA 5.6

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.6 = 1.67$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado y de K-N							
	3		4		5		6	
	HD-MDL	K-N	HD-MDL	K-N	HD-MDL	K-N	HD-MDL	K-N
(100,60)	0.384	0.352	<b>0.583</b>	<b>0.616</b>	0.008	0.011	0.000	0.000
(200,120)	0.145	0.100	<b>0.849</b>	<b>0.884</b>	0.005	0.016	0.000	0.000
(400,240)	0.017	0.010	<b>0.983</b>	<b>0.977</b>	0.000	0.013	0.000	0.000
(800,480)	0.001	0.000	<b>0.999</b>	<b>0.993</b>	0.000	0.007	0.000	0.000

TABLA 5.7

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.3 = 3.33$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado y de K-N							
	3		4		5		6	
	HD-MDL	K-N	HD-MDL	K-N	HD-MDL	K-N	HD-MDL	K-N
(100,30)	0.502	0.351	<b>0.228</b>	<b>0.124</b>	0.012	0.007	0.000	0.000
(200,60)	0.495	0.565	<b>0.419</b>	<b>0.211</b>	0.025	0.007	0.000	0.000
(400,120)	0.312	0.500	<b>0.653</b>	<b>0.461</b>	0.029	0.011	0.000	0.000
(800,240)	0.114	0.243	<b>0.856</b>	<b>0.747</b>	0.030	0.010	0.000	0.000

de detección errónea de la dimensión cuando solamente existe un autovalor spike. El valor del autovalor spike se encuentra en el eje de abscisas. El nuevo estimador es denotado por HD-MDL (high-dimensional MDL). A partir de este gráfico, vemos cómo un análisis detallado del

TABLA 5.8

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 1 con  $p/m = 1/0.1 = 10$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado						
	1	2	3	4	5	6	7
(100,10)	0.781 (0.229)	0.191 (0.187)	0.028 (0.044)	<b>0.000 (0.005)</b>	0.000	0.000	0.000
(200,20)	0.621 (0.365)	0.331 (0.181)	0.044 (0.083)	<b>0.004 (0.006)</b>	0.000	0.000	0.000
(400,40)	0.365 (0.400)	0.506 (0.280)	0.124 (0.108)	<b>0.005 (0.017)</b>	0.000 (0.001)	0.000	0.000
(800,80)	0.170 (0.365)	0.597 (0.438)	0.219 (0.128)	<b>0.014 (0.010)</b>	0.000 (0.001)	0.000	0.000

TABLA 5.9

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.1$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado									
	1	2	3	4	5	6	7	8	9	10
(10,100)	0.000	0.000	0.000	0.000	<b>0.999</b>	0.001	0.000	0.000	0.000	0.000
(20,200)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000
(40,400)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000
(80,800)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000

TABLA 5.10

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.3$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado									
	1	2	3	4	5	6	7	8	9	10
(30,100)	0.000	0.000	0.000	0.000	<b>0.994</b>	0.006	0.000	0.000	0.000	0.000
(60,200)	0.000	0.000	0.000	0.000	<b>0.996</b>	0.004	0.000	0.000	0.000	0.000
(120,400)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000
(240,800)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000

TABLA 5.11

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.6$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado									
	1	2	3	4	5	6	7	8	9	10
(60,100)	0.000	0.000	0.000	0.000	<b>0.975</b>	0.025	0.000	0.000	0.000	0.000
(120,200)	0.000	0.000	0.000	0.000	<b>0.993</b>	0.007	0.000	0.000	0.000	0.000
(240,400)	0.000	0.000	0.000	0.000	<b>0.999</b>	0.001	0.000	0.000	0.000	0.000
(480,800)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000

crecimiento del logaritmo del estadístico de cociente de verosimilitudes nos permite mejorar notablemente la estimación de la dimensión. Sin embargo, una advertencia para nuestro enfoque es que debemos elegir un valor para  $\epsilon$ . Elegir valores de  $\epsilon$  muy grandes nos hará perder autovalores que se encuentren cerca del umbral pero minimizará la probabilidad de perder los autovalores grandes. Por el contrario, valores muy chicos para  $\epsilon$  aumentarán, ligeramente, la probabilidad de perder los autovalores más grandes (cuando  $p$  y  $n$  no sean lo suficientemente grandes).

TABLA 5.12

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 2 con  $p/m = 0.9$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado									
	1	2	3	4	5	6	7	8	9	10
(90,100)	0.000	0.000	0.000	0.000	<b>0.933</b>	0.066	0.001	0.000	0.000	0.000
(180,200)	0.000	0.000	0.000	0.000	<b>0.981</b>	0.019	0.000	0.000	0.000	0.000
(360,400)	0.000	0.000	0.000	0.000	<b>0.996</b>	0.004	0.000	0.000	0.000	0.000
(720,800)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000

TABLA 5.13

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.9 = 1.11$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado									
	1	2	3	4	5	6	7	8	9	10
(100,90)	0.000	0.000	0.000	0.003	<b>0.966</b>	0.030	0.001	0.000	0.000	0.000
(200,180)	0.000	0.000	0.000	0.000	<b>0.993</b>	0.007	0.000	0.000	0.000	0.000
(400,360)	0.000	0.000	0.000	0.000	<b>0.998</b>	0.002	0.000	0.000	0.000	0.000
(800,720)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000

TABLA 5.14

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.6 = 1.67$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado									
	1	2	3	4	5	6	7	8	9	10
(100,60)	0.000	0.000	0.000	0.084	<b>0.905</b>	0.011	0.000	0.000	0.000	0.000
(200,120)	0.000	0.000	0.000	0.006	<b>0.989</b>	0.005	0.000	0.000	0.000	0.000
(400,240)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000
(800,480)	0.000	0.000	0.000	0.000	<b>1.000</b>	0.000	0.000	0.000	0.000	0.000

TABLA 5.15

Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.3 = 3.33$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado									
	1	2	3	4	5	6	7	8	9	10
(100,30)	0.000	0.003	0.132	0.671	<b>0.193</b>	0.001	0.000	0.000	0.000	0.000
(200,60)	0.000	0.000	0.009	0.563	<b>0.428</b>	0.000	0.000	0.000	0.000	0.000
(400,120)	0.000	0.000	0.000	0.347	<b>0.653</b>	0.000	0.000	0.000	0.000	0.000
(800,240)	0.000	0.000	0.000	0.195	<b>0.805</b>	0.000	0.000	0.000	0.000	0.000

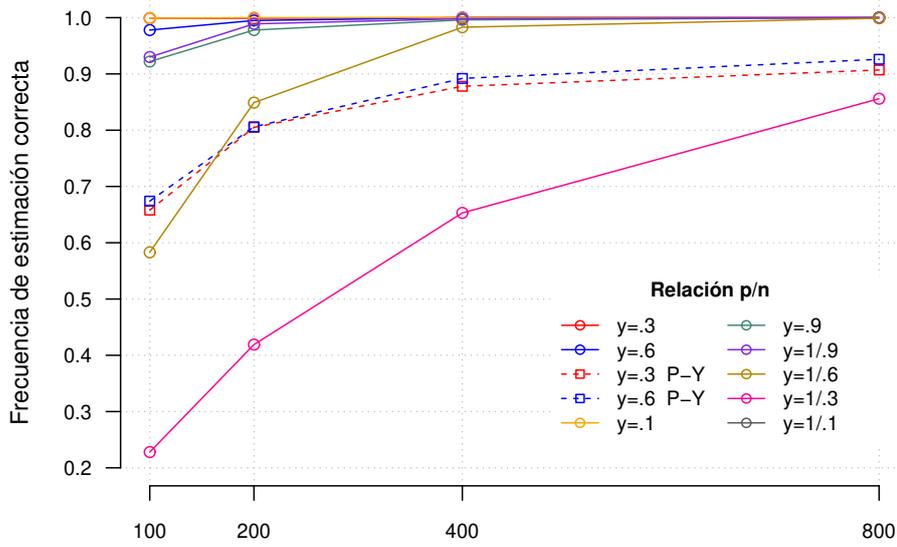
A partir de los resultados de las simulaciones, vemos que el comportamiento del estadístico de cociente de verosimilitudes penalizado es excelente, excepto en los casos en que  $p/n = 1/0.3 = 3.33$  y  $p/n = 1/0.1 = 10$ . La explicación de este fenómeno la encontramos en la Figura 5.4. Esta figura es el gráfico del  $\log(LRT_d)$  como función de la dimensión  $d$  para el **Modelo 1**, donde la verdadera dimensión es  $d_1 = 4$  y los autovalores spike iguales a  $(7, 6, 5, 4)$ , cuando  $p/n = 0.1$  y  $p/n = 1/0.1 = 10$ . Como podemos observar, para el caso en que  $p/n = 1/0.1 = 10$ , aún cuando

TABLA 5.16

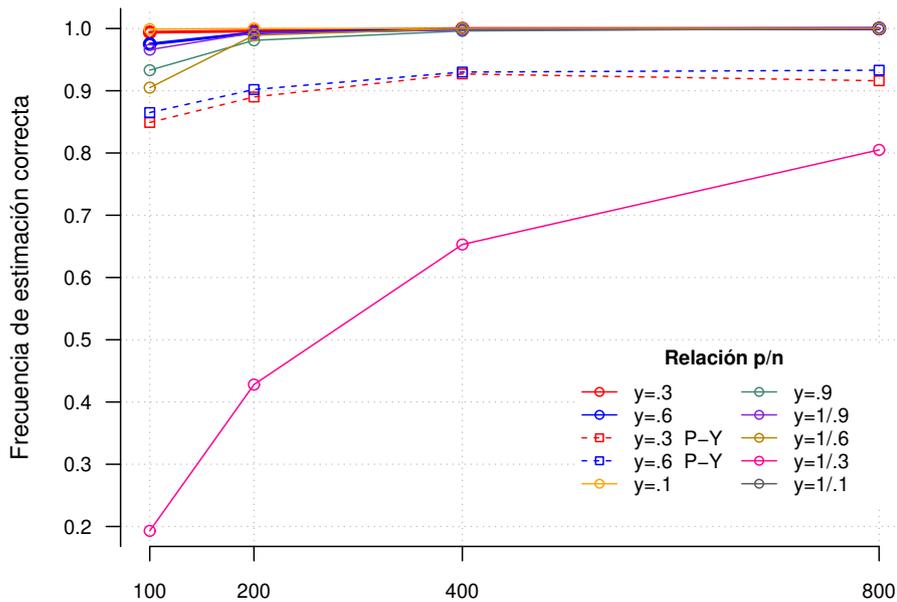
Distribución empírica de  $\hat{d}$  penalizado sobre 1000 repeticiones para el Modelo 2 con  $p/m = 1/0.1 = 10$ .

$(p, n)$	Distribución de $\hat{d}$ penalizado									
	1	2	3	4	5	6	7	8	9	10
(100,10)	0.061	0.457	0.428	0.054	<b>0.000</b>	0.000	0.000	0.000	0.000	0.000
(200,20)	0.000	0.083	0.526	0.368	<b>0.023</b>	0.000	0.000	0.000	0.000	0.000
(400,40)	0.000	0.006	0.249	0.649	<b>0.092</b>	0.004	0.000	0.000	0.000	0.000
(800,80)	0.000	0.000	0.030	0.751	<b>0.217</b>	0.002	0.000	0.000	0.000	0.000

la función presenta un cambio en la tasa de crecimiento en la verdadera dimensión  $d_1 = 4$ , dicho cambio es mínimo y por lo tanto más difícil de detectar. Además, en este caso en particular, el umbral de detección  $\sigma^2(1 + \sqrt{y})$  es igual a 4.16, con lo cual tenemos un autovalor que es indetectable (es decir, queda por debajo del umbral) y por lo tanto el método falla en encontrar la verdadera dimensión.

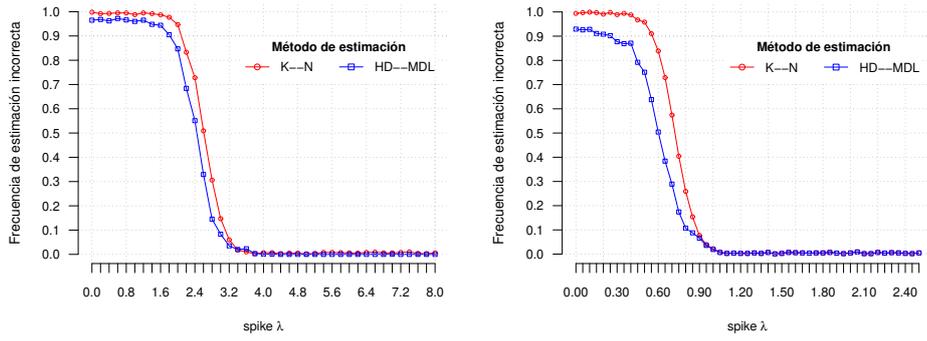


(a) Modelo 1 ( $d = 4$ )



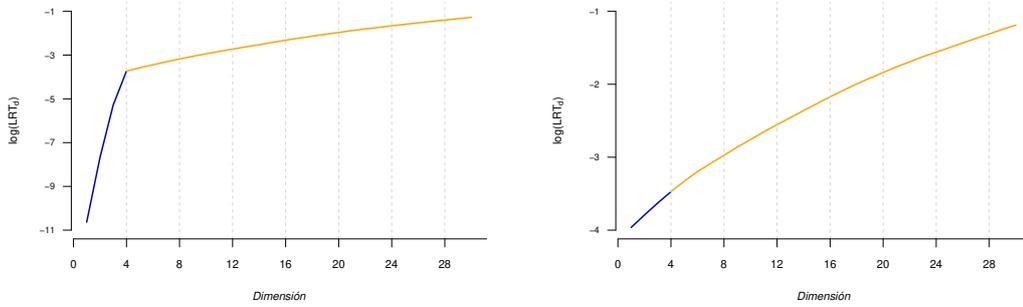
(b) Modelo 2 ( $d = 5$ )

Figura 5.2. Frecuencia de estimación correcta sobre 1000 repeticiones independientes para el método de Passemier and Yao [36] (línea de puntos) y el estimador propuesto en este capítulo.



(a)  $p = 2000, n = 500, d_1 = 1, \sigma^2 = 1, \epsilon = .01$ . (b)  $p = 200, n = 8000, d_1 = 1, \sigma^2 = 1, \epsilon = .025$ .

**Figura 5.3.** Probabilidad de estimación errónea para el método de Kritchman and Nadler [25] (KN) y el estimador penalizado (HD-MDL).



(a) Modelo 1,  $d_1 = 4, p/n = 0.1$ .

(b) Modelo 1,  $d_1 = 4, p/n = 1/0.1 = 10$ .

**Figura 5.4.** Comportamiento del  $\log(LRT_d)$  cuando  $p/n = 0.1$  y  $p/n = 1/0.1$ , siendo la verdadera dimensión  $d_1 = 4$  con autovalores spikes  $(7, 6, 5, 4)$ .



---

# Conclusiones generales

---

A lo largo de esta tesis se han introducido herramientas para identificar la dimensión (lineal) efectiva de un conjunto de datos, recurriendo a una mirada distinta sobre un estadístico clásico como es el cociente de verosimilitudes. A diferencia de resultados previos existentes en la literatura, la metodología propuesta es aplicable para cualquier relación  $p/n$  finita.

Para escenarios de alta dimensionalidad ( $p < n$  y  $p, n \rightarrow \infty$  tales que  $p/n \rightarrow y \in (0, 1]$ ) hemos estudiado la distribución asintótica del logaritmo del cociente de verosimilitudes para el test de esfericidad parcial basado en el modelo de covarianza spiked introducido por Johnstone [23]. Además, consideramos el caso de ultra-alta dimensión ( $p > n$  y  $p, n \rightarrow \infty$  tales que  $p/n \rightarrow y \geq 1$ ) y estudiamos la distribución asintótica del estadístico de cociente de verosimilitudes resultante luego de intercambiar los roles de  $p$  y de  $n$ . Conocer tales distribuciones nos permitió desarrollar un test secuencial para determinar la dimensión del subespacio de los autovalores spike, que se centra en la parte no-spike de los autovalores (los correspondientes al ruido). Una buena característica de este enfoque es que no requiere el conocimiento a priori de la varianza común  $\sigma^2$  de los autovalores de la parte del ruido, sino que se estima a partir de los datos. El estudio de la potencia del test de cociente de verosimilitudes nos permitió modificar el estadístico de prueba agregando un término de penalización. La idea de esta corrección está conectada al *método del codo* (elbow method) utilizado en el análisis de clusters para elegir el número de grupos [46] y, también, a los enfoques de teoría de la información como AIC y MDL [25, 50]. A partir del estudio del comportamiento de la media de la distribución asintótica del logaritmo del estadístico de cociente de verosimilitudes para valores por debajo y por encima de la verdadera dimensión  $d_1$  del subespacio spiked, pudimos modificar el estadístico y probar que el estimador

resultante es consistente. Más allá de estas garantías teóricas, el estimador penalizado de la dimensión mostró ser sumamente competitivo en los estudios con datos simulados.

Resulta también evidente que los escenarios de alta dimensionalidad ( $y \in (1, \infty)$ ) son intrínsecamente más dificultosos. Esto se debe a que para ser detectables, los autovalores spike tienen que tener una magnitud mayor a un valor umbral que depende de  $y$  y de  $\sigma^2$ . Así, cuando  $p \gg n$ , se requiere una relación señal-ruido muy grande para lograr buena potencia en la detección de la dimensión efectiva; caso contrario, esta dimensión resulta subestimada.

## Trabajos futuros

Los resultados encontrados en esta tesis descansan en la validez del modelo spike para la matriz de covarianza de los datos. No obstante, en escenarios de muy alta dimensionalidad, es razonable cuestionarse hasta qué punto es adecuado este modelo cuando los autovalores spike están fijos y  $p$  tiende a infinito. Expresado de otro modo, ¿cuáles son los supuestos implícitos de este modelo?

Como un primer abordaje a este problema, consideremos el siguiente escenario. Supongamos que estamos considerando  $p - 1$  variables aleatorias  $\mathbf{X}_{p-1} = (X_1, \dots, X_{p-1})$  y agregamos una nueva variable  $X_p$  para obtener el vector aleatorio  $\mathbf{X}_p = (X_1, \dots, X_p)$ . Si consideramos la matriz de covarianza de  $\mathbf{X}_p$ ,  $\Sigma_p$ , tenemos

$$\Sigma_p = \begin{pmatrix} \Sigma_{p-1} & \mathbf{a}_p \\ \mathbf{a}_p^\top & b_p \end{pmatrix}, \quad (5.3)$$

con  $\mathbf{a}_p \in \mathbb{R}^{p-1}$  y  $b_p$  un número real. Supongamos que la matriz de covarianza sigue un modelo spike de dimensión  $d$ ; de modo que podemos escribir<sup>1</sup>

$$\Sigma_p = \Psi_p \Lambda_p \Psi_p^\top + \sigma_p^2 \mathbf{I}_p, \quad (5.4)$$

con  $\Psi_p \in \mathbb{R}^{p \times d}$  semi-ortogonal y consideramos  $\Lambda_p > \sigma_p^2 \mathbf{I}_d$  diagonal de dimensión  $d$ . Dado que  $\Sigma_{p-1}$  es una submatriz principal de  $\Sigma_p$ , usando (1) de Johnson and Robinson [22] tenemos que  $\sigma_p^2 = \sigma_{p-1}^2$  y  $\Lambda_p \geq \Lambda_{p-1}$ , donde consideramos que los autovalores en  $\Lambda_p$  y  $\Lambda_{p-1}$  están ordenados en forma decreciente. Usando (5.3) y (5.4) y llamando  $\Psi_p = \begin{pmatrix} \Psi_{p,1} \\ \Psi_{p,p} \end{pmatrix}$ , con  $\Psi_{p,1} \in \mathbb{R}^{(p-1) \times d}$  y

<sup>1</sup>Notar que esta parametrización del modelo spiked es diferente de la utilizada en los capítulos previos. Con esta parametrización los autovalores spike resultan iguales a  $\lambda_i + \sigma^2$ .

$\Psi_{p,p}^\top \in \mathbb{R}^d$ , tenemos

$$\Psi_{p-1} \Lambda_{p-1} \Psi_{p-1}^\top + \sigma_{p-1}^2 \mathbf{I}_{p-1} = \Psi_{p,1} \Lambda_p \Psi_{p,1}^\top + \sigma_p^2 \mathbf{I}_{p-1}, \quad (5.5)$$

$$\mathbf{a}_p = \Psi_{p,1} \Lambda_p \Psi_{p,p}^\top, \quad (5.6)$$

$$b_p - \sigma_p^2 = \Psi_{p,p} \Lambda_p \Psi_{p,p}^\top. \quad (5.7)$$

Ahora, de (5.5) y dado que  $\Psi_{p-1}^\top \Psi_{p-1} = \mathbf{I}_d$  y  $\Psi_{p,1}^\top \Psi_{p,1} = \mathbf{I}_d - \Psi_{p,p}^\top \Psi_{p,p} \geq 0$  tenemos

$$\text{tr}(\Lambda_{p-1}) + (p-1)(\sigma_{p-1}^2 - \sigma_p^2) = \text{tr}(\Lambda_p(\mathbf{I}_d - \Psi_{p,p}^\top \Psi_{p,p})) \quad (5.8)$$

y por lo tanto

$$\text{tr}(\Lambda_p) - \text{tr}(\Lambda_{p-1}) = \text{tr}(\Lambda_p \Psi_{p,p}^\top \Psi_{p,p}) + (p-1)(\sigma_p^2 - \sigma_{p-1}^2). \quad (5.9)$$

Pero por ser  $\sigma_p^2 = \sigma_{p-1}^2$  obtenemos

$$\text{tr}(\Lambda_p) - \text{tr}(\Lambda_{p-1}) = \text{tr}(\Lambda_p \Psi_{p,p}^\top \Psi_{p,p}). \quad (5.10)$$

En este punto es interesante preguntarnos bajo qué condiciones la matriz  $\Lambda_p$  converge a una matriz fija  $\Lambda$  que representa el límite del modelo spiked. Consideremos que  $\Lambda_p \rightarrow \Lambda$ , con  $\Lambda$  una matriz fija con autovalores  $\lambda_i$  finitos, y que  $\sigma_p^2 \rightarrow \sigma^2$ , con  $\Lambda > \sigma^2 \mathbf{I}_d$ . Tenemos el siguiente resultado:

**Proposición 5.5.** *Si  $\Lambda_p \rightarrow \Lambda$  un matriz fija con autovalores  $\lambda_i$  finitos (el límite spiked), entonces  $\|\mathbf{a}_p\|^2 \rightarrow 0$ .*

**Demostración de la Proposición 5.5:** Usando (5.10) y la hipótesis de que  $\Lambda_p \rightarrow \Lambda$  tenemos que  $\text{tr}(\Lambda_p \Psi_{p,p}^\top \Psi_{p,p}) \rightarrow 0$ . Ahora, de (5.6) y dado que  $\Psi_{p-1}^\top \Psi_{p-1} = \mathbf{I}_d$  y  $\Psi_{p,1}^\top \Psi_{p,1} = \mathbf{I}_d - \Psi_{p,p}^\top \Psi_{p,p} \geq 0$  tenemos

$$\begin{aligned} \|\mathbf{a}_p\|^2 &= \text{tr}(\Psi_{p,1} \Lambda_p \Psi_{p,p}^\top \Psi_{p,p} \Lambda_p \Psi_{p,1}^\top) \\ &= \text{tr}[\Lambda_p \Psi_{p,p}^\top \Psi_{p,p} \Lambda_p (\mathbf{I}_d - \Psi_{p,p}^\top \Psi_{p,p})] \\ &\leq \text{tr}(\Lambda_p \Psi_{p,p}^\top \Psi_{p,p}) \text{tr}[\Lambda_p (\mathbf{I}_d - \Psi_{p,p}^\top \Psi_{p,p})] \\ &= [\text{tr}(\Lambda_p) - \text{tr}(\Lambda_{p-1})] \text{tr}(\Lambda_{p-1}), \end{aligned}$$

donde la última igualdad es consecuencia de (5.10), que tiende a cero y por lo tanto todos los componentes de  $\mathbf{a}_p$  tienden a cero.

□

**Corolario 5.6.** *Bajo el modelo poblacional spiked de dimensión  $d$  fijo, tenemos que cuando  $p$  crece, estamos agregando variables tales que la suma del cuadrado de la correlación con las variables previas se va a 0.*

**Demostración del Corolario 5.6:** De (5.7) y (5.10) tenemos que  $b_p \rightarrow \sigma^2$ . Además,

$$\sum_{i=1}^{p-1} \rho^2(X_i, X_p) = \sum_{i=1}^{p-1} \frac{a_{p,i}^2}{b_p \text{var}(X_i)} \quad (5.11)$$

$$\leq \frac{1}{b_p \min_{i=1, \dots, p} \text{var}(X_i)} \|\mathbf{a}_p\|^2. \quad (5.12)$$

Como  $\min_{i=1, \dots, p-1} \text{var}(X_i) = \min_{i=1, \dots, p-1} (\sigma_p^2 + (\Psi_{p-1} \Lambda_{p-1} \Psi_{p-1}^T)_{ii}) \geq \sigma_p^2$ ,  $b_p \rightarrow \sigma^2$  y  $\|\mathbf{a}_p\|_p \rightarrow 0$ , obtenemos el corolario.

□

Estos resultados implican que el modelo spiked supone que cuando  $p \rightarrow \infty$ , las nuevas variables no aportan información nueva. Si bien esto es una suposición comúnmente aceptada, no resulta adecuada en algunos contextos, tales como los problemas de fusión de datos. Un modelo más adecuado en estos escenarios es el de predictores *abundantes*, en el sentido que agregar una variable no agrega ruido sino que realmente incorpora información al problema bajo estudio. En este caso es necesario que al menos uno de los autovalores spike crezca con el aumento de  $p$ . Determinar la dimensión del subespacio spike bajo estas condiciones es un problema abierto.

## Resultados conocidos

A continuación enunciaremos una serie de resultados conocidos que utilizaremos en las pruebas presentadas en esta tesis. Si no se aclara nada al respecto,  $n$  y  $p$  pueden tener cualquier relación.

**Proposición A.1.** (Proposición 5.10, Eaton [12])

Sean  $\mathbf{A}$  y  $\mathbf{B}$  dos matrices inversibles de  $n \times n$  y  $k \times k$  respectivamente, y sea la función  $g : \mathbb{R}^{n \times k} \rightarrow \mathbb{R}^{n \times k}$  tal que  $g(\mathbf{X}) = \mathbf{A}\mathbf{X}\mathbf{B}$ . Entonces el jacobiano de la transformación es  $J_g(\mathbf{X}) = |\mathbf{A}|^k |\mathbf{B}|^n$ .

**Proposición A.2.** (Proposición 8.1, Eaton [12])

Si  $\mathbf{S} \sim \mathcal{W}_p(m, \boldsymbol{\Sigma})$  y  $\mathbf{A}$  es una matriz de  $r \times p$ , entonces  $\mathbf{A}\mathbf{S}\mathbf{A}^\top \sim \mathcal{W}_r(m, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$ .

**Proposición A.3.** (Teorema 3.2.7, Muirhead [33])

Si  $\mathbf{S} \sim \mathcal{W}_p(m, \boldsymbol{\Sigma})$  y particionamos  $\mathbf{S}$  y  $\boldsymbol{\Sigma}$  del siguiente modo

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

donde  $\mathbf{S}_{11}$  y  $\boldsymbol{\Sigma}_{11}$  son matrices de  $k \times k$  y  $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ , entonces  $\mathbf{S}_{11}$  y  $\mathbf{S}_{22}$  son independientes y sus distribuciones son, respectivamente,  $\mathcal{W}_k(m, \boldsymbol{\Sigma}_{11})$  y  $\mathcal{W}_{p-k}(m, \boldsymbol{\Sigma}_{22})$ .

**Proposición A.4.** (Teorema 3.2.20, Muirhead [33])

Si  $\mathbf{S} \sim \mathcal{W}_p(m, \sigma^2 \mathbf{I}_p)$  donde  $m (\geq p)$  es un entero, entonces  $u = \frac{|\mathbf{S}|}{\left[\frac{1}{p} \text{tr}(\mathbf{S})\right]^p}$  y  $\text{tr}(\mathbf{S})$  son independientes y  $\text{tr}(\mathbf{S}/\sigma^2) \sim \chi_{pm}^2$ .

**Corolario A.5.** (Corolario 8.3.6, Muirhead [33])

Sean  $\mathbf{X}_1, \dots, \mathbf{X}_n$  vectores aleatorios independientes, idénticamente distribuidos como  $\mathcal{N}_p(\mathbf{0}, \Sigma)$ .

Sea  $\mathbf{A} = n\hat{\Sigma} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top \sim \mathcal{W}_p(m, \Sigma)$ , donde  $m = n - 1$ . Si  $\Sigma = \sigma^2 \mathbf{I}_p$ , entonces los momentos del estadístico de esfericidad  $V$ , definido por

$$V = \frac{|\mathbf{A}|}{\left(\frac{1}{p} \text{tr} \mathbf{A}\right)^p}, \quad (\text{A.1})$$

están dados por

$$\mathbb{E}(V^h) = p^{ph} \frac{\Gamma(\frac{1}{2}pm)}{\Gamma(\frac{1}{2}pm + ph)} \frac{\Gamma_p(\frac{1}{2}m + h)}{\Gamma_p(\frac{1}{2}m)}.$$

**Lema A.6.** (Lema 5.1, Jiang and Yang [21])

Sea  $b := b(x)$  una función a valores reales definida sobre  $(0, \infty)$ . Entonces

$$\log \frac{\Gamma(x + b)}{\Gamma(x)} = b \log x + \frac{b^2 - b}{2x} + c(x)$$

cuando  $x \rightarrow +\infty$ , donde

$$c(x) = \begin{cases} O(x^{-1/2}), & \text{si } b(x) = \mathcal{O}(\sqrt{x}); \\ O(x^{-2}), & \text{si } b(x) = \mathcal{O}(1). \end{cases}$$

Además, para cualesquiera constantes  $d > c$ , si  $x \rightarrow +\infty$ ,

$$\sup_{c \leq t \leq d} \left| \log \frac{\Gamma(x + t)}{\Gamma(x)} - t \log(x) \right| \rightarrow 0.$$

**Lema A.7.** (Lema 5.4, Jiang and Yang [21])

Sea  $\Gamma_p(z)$  definida por

$$\Gamma_p(z) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma\left(z - \frac{1}{2}(i-1)\right). \quad (\text{A.2})$$

Sea  $n > p = p_n$  y  $r_n = \left[-\log\left(1 - \frac{p}{n}\right)\right]^{1/2}$ . Supongamos que  $\frac{p}{n} \rightarrow y \in (0, 1]$ ,  $s = s_n = \mathcal{O}(1/r_n)$  y  $t = t_n = \mathcal{O}(1/r_n)$  cuando  $n \rightarrow \infty$ . Entonces

$$\log \frac{\Gamma_p\left(\frac{n}{2} + t\right)}{\Gamma_p\left(\frac{n}{2} + s\right)} = p(t-s)(\log n - 1 - \log 2) + r_n^2 \left[ (t^2 - s^2) - \left(p - n + \frac{1}{2}\right)(t-s) \right] + o(1),$$

siempre que  $n \rightarrow \infty$ .

**Teorema A.8.** (Teorema 1, Jiang and Yang [21])

Sea  $m =: n - 1 > p$  para todo  $n \geq 3$  y  $\mathbf{V}_m$  definido como en (A.1). Supongamos que  $p/m \rightarrow y \in (0, 1]$ . Entonces, bajo la hipótesis nula  $H_0 : \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$ ,  $(\log \mathbf{V}_m - \mu_m)/\sigma_m$  converge en distribución a una  $\mathcal{N}(0, 1)$  cuando  $n \rightarrow \infty$ , donde

$$\begin{aligned}\mu_n &= -p - \left(m - p - \frac{1}{2}\right) \log \left(1 - \frac{p}{m}\right), \\ \sigma_n^2 &= -2 \left[\frac{p}{m} + \log \left(1 - \frac{p}{m}\right)\right].\end{aligned}$$

**Corolario A.9.** (Corolario 4.2, Magnus and Neudecker [29])

Sea la matriz  $\mathbf{S}$  de con distribución Wishart  $\mathcal{W}_p(m, \boldsymbol{\Sigma})$ . Entonces

$$\begin{aligned}\mathbb{E}(\mathbf{S}) &= m\boldsymbol{\Sigma}, \\ \text{var}(\text{vec}(\mathbf{S})) &= m(\mathbf{I}_{p^2} + \mathbf{K}_p)(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}),\end{aligned}$$

donde  $\mathbf{K}_p$  es la matriz de conmutación de orden  $p^2$ .

**Teorema A.10.** (von Rosen [47], ver también Cook and Forzani [11])

Sea  $\mathbf{S} \sim \mathcal{W}_p(m, \mathbf{I}_p)$ . Entonces si  $m > p + 3$ ,  $\text{rank}(\mathbf{S}) = p$  con probabilidad 1 y

$$\begin{aligned}\mathbb{E}(\mathbf{S}^{-1}) &= a_1 \mathbf{I}_p, \\ \mathbb{E}(\mathbf{S}^{-1} \mathbf{S}^{-1}) &= b_1 \mathbf{I}_p, \\ \text{var}[\text{vec}(\mathbf{S}^{-1})] &= c_1(\mathbf{I}_{p^2} + \mathbf{K}_p) + 2d_1 \text{vec}(\mathbf{I}_p) \text{vec}^\top(\mathbf{I}_p),\end{aligned}$$

donde  $a_1 = (m - p - 1)^{-1}$ ,  $b_1 = (m - 1)c_1$ ,  $c_1^{-1} = (m - p)(m - p - 1)(m - p - 3)$ ,  $d_1^{-1} = (m - p)(m - p - 1)^2(m - p - 3)$  y  $\mathbf{K}_p$  es la matriz de conmutación de orden  $p^2$ .

**Proposición A.11.** (Ghazal and Neudecker [15])

Consideremos la matriz aleatoria  $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ , donde  $\mathbf{X} \in \mathbb{R}^{n \times p}$  e  $\mathbf{Y} \in \mathbb{R}^{n \times q}$ . Sea  $\mathbb{E}(\mathbf{Z}) =$

$$(\mathbf{M}, \mathbf{N}) \text{ y } \text{var}(\text{vec}(\mathbf{Z})) = \begin{pmatrix} \boldsymbol{\Omega}_{XX} & \boldsymbol{\Omega}_{XY} \\ \boldsymbol{\Omega}_{YX} & \boldsymbol{\Omega}_{YY} \end{pmatrix}. \text{ Entonces}$$

$$\mathbb{E}(\mathbf{X} \otimes \mathbf{Y}) = \sum_{i,j} (E_{ij}^\top \otimes \mathbf{I}_n) \boldsymbol{\Omega}_{YX} (\mathbf{I}_p \otimes E_{ij}^\top) + \mathbf{M} \otimes \mathbf{N},$$

donde  $E_{ij}$  (para  $i = 1, \dots, q$ ,  $j = 1, \dots, n$ ) es una matriz de  $q \times n$  con un 1 en la  $ij$ -ésima posición y 0 en el resto de las posiciones.



## Resultados auxiliares

En este apartado enunciamos y demostramos algunos lemas que resultarán útiles para derivar los resultados principales de esta tesis.

**Lema B.1.** *Sea  $\mathbf{S} \sim \mathcal{W}_p(m, \lambda \mathbf{I}_p)$ , donde  $m(> p)$  es un entero y sea  $\mathbf{A}$  una matriz de  $p \times r$ ,  $r \leq p$ , semiortogonal (e.i.  $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_r$ ). Entonces,*

$$\text{tr} \left( \frac{\mathbf{A}^\top \mathbf{S} \mathbf{A}}{rm} \right) \rightarrow \lambda \quad (\text{B.1})$$

si  $m \rightarrow \infty$ .

**Demostración del Lema B.1:** Por Proposición A.2, tenemos que  $\mathbf{A}^\top \mathbf{S} \mathbf{A} \sim \mathcal{W}_r(m, \lambda \mathbf{I}_r)$ . Luego, aplicando Proposición A.4 tenemos que  $\frac{1}{\lambda} \text{tr} \mathbf{A}^\top \mathbf{S} \mathbf{A} \sim \chi_{rm}^2$ . Por lo tanto,

$$\begin{aligned} \mathbb{E} \left( \frac{\text{tr} \mathbf{A}^\top \mathbf{S} \mathbf{A}}{rm} \right) &= \lambda \rightarrow \lambda, \\ \text{var} \left( \frac{\text{tr} \mathbf{A}^\top \mathbf{S} \mathbf{A}}{rm} \right) &= \frac{2\lambda^2}{rm} \rightarrow 0, \end{aligned}$$

lo cual implica (B.1).

□

**Lema B.2.** *En todos los modelos spiked de dimensión  $d$ ,  $\hat{\sigma}^2 = \frac{1}{p-d} \sum_{i=d+1}^p l_i$  converge casi seguro a  $\sigma^2$ .*

**Demostración del Lema B.2** Observemos en primer lugar que

$$\frac{1}{p-d} \sum_{i=d+1}^p l_i = \frac{1}{p-d} \operatorname{tr}(\widehat{\Sigma}) - \frac{1}{p-d} \sum_{i=1}^d l_i.$$

Entonces, veamos la convergencia de cada uno de los términos en la suma anterior.

En primer lugar veamos que  $\frac{1}{p-d} \operatorname{tr}(\widehat{\Sigma}) \rightarrow \sigma^2$  si  $p, m \rightarrow \infty$ . Para ello, recordemos que  $m\widehat{\Sigma} \sim \mathcal{W}_p(m, \Sigma)$  y por lo tanto  $\mathbb{E}(m\widehat{\Sigma}) = m\Sigma$  y  $\operatorname{var}[\operatorname{vec}(m\widehat{\Sigma})] = m(\mathbf{I}_{p^2} + \mathbf{K}_p)(\Sigma \otimes \Sigma)$ . Luego,

$$\begin{aligned} \mathbb{E}\left[\frac{1}{p-d} \operatorname{tr}(\widehat{\Sigma})\right] &= \frac{1}{(p-d)m} \mathbb{E}\left[\operatorname{tr}(m\widehat{\Sigma})\right] \\ &= \frac{1}{(p-d)m} \operatorname{tr}(m\Sigma) \\ &= \frac{1}{p-d} \sum_{i=1}^h s_i \lambda_i + \frac{1}{p-d} \sum_{i=d+1}^p \sigma^2 \\ &= \frac{1}{p-d} \sum_{i=1}^h s_i \lambda_i + \sigma^2 \rightarrow \sigma^2. \end{aligned}$$

Por otro lado,

$$\begin{aligned} \operatorname{var}\left[\frac{1}{p-d} \operatorname{tr}(\widehat{\Sigma})\right] &= \operatorname{var}\left[\frac{1}{(p-d)m} \operatorname{vec}^\top(\mathbf{I}_p) \operatorname{vec}(m\widehat{\Sigma})\right] \\ &= \frac{1}{(p-d)^2 m^2} \operatorname{vec}^\top(\mathbf{I}_p) \operatorname{var}\left[\operatorname{vec}(m\widehat{\Sigma})\right] \operatorname{vec}(\mathbf{I}_p) \\ &= \frac{1}{(p-d)^2 m} \operatorname{vec}^\top(\mathbf{I}_p) (\mathbf{I}_{p^2} + \mathbf{K}_p) (\Sigma \otimes \Sigma) \operatorname{vec}(\mathbf{I}_p) \\ &= \frac{2}{(p-d)^2 m} \operatorname{vec}^\top(\mathbf{I}_p) \operatorname{vec}(\Sigma^2) \\ &= \frac{2}{(p-d)^2 m} \operatorname{tr}(\Sigma^2) \\ &= \frac{2}{(p-d)^2 m} \left[ \sum_{i=1}^h s_i \lambda_i^2 + \sum_{i=d+1}^p \sigma^4 \right] \\ &= \frac{2}{(p-d)^2 m} \sum_{i=1}^h s_i \lambda_i^2 + \frac{2\sigma^4}{(p-d)m} \rightarrow 0. \end{aligned}$$

Por lo tanto, tenemos que  $\frac{1}{p-d} \operatorname{tr}(\widehat{\Sigma}) \rightarrow \sigma^2$  si  $p, m \rightarrow \infty$ , como queríamos.

Veamos ahora que  $\frac{1}{p-d} \sum_{i=1}^d l_i \rightarrow 0$ . Para ello, vamos a definir los conjuntos de índices

$$\begin{aligned} J_1 &= \{i \leq h : \lambda_i > \sigma^2(1 + \sqrt{y})\}, \\ J_2 &= \{i \leq h : \lambda_i \in (\sigma^2, \sigma^2(1 + \sqrt{y})]\}. \end{aligned}$$

Entonces, siguiendo a [7], tenemos que

$$\begin{aligned}\lambda_i &\xrightarrow{c.s.} \lambda_i \left(1 + \frac{y\sigma^2}{\lambda_i - \sigma^2}\right), \quad \text{si } i \in J_1, \\ \lambda_i &\xrightarrow{c.s.} \sigma^2 (1 + \sqrt{y})^2, \quad \text{si } i \in J_2.\end{aligned}$$

Luego,

$$\sum_{i=1}^d l_i \xrightarrow{c.s.} \left[ \sum_{i \in J_1} \lambda_i \left(1 + \frac{y\sigma^2}{\lambda_i - \sigma^2}\right) + \sum_{i \in J_2} \sigma^2 (1 + \sqrt{y})^2 \right] < \infty$$

y por lo tanto

$$\frac{1}{p-d} \sum_{i=1}^d l_i \xrightarrow{c.s.} 0.$$

□

**Lema B.3.** Si  $a_p \rightarrow a < \infty$  cuando  $p \rightarrow \infty$ , entonces

$$(p-d) \log \left(1 + \frac{1}{p-d} a_p\right) \rightarrow a.$$

**Demostración del Lema B.3:** Recordemos que, por definición, el valor límite de  $\left(1 + \frac{1}{x}\right)^x$  para  $x \rightarrow \infty$  es el número  $e$ . De ahí,

$$\left(1 + \frac{1}{p-d} a_p\right)^{p-d} = \left(1 + \frac{1}{\frac{p-d}{a_p}}\right)^{\frac{p-d}{a_p} a_p} \xrightarrow{p \rightarrow \infty} e^a,$$

donde hemos usado la continuidad de la función exponencial. Ahora, usando que el logaritmo también es una función continua, tenemos

$$(p-d) \log \left(1 + \frac{1}{p-d} a\right) \rightarrow \log e^a = a,$$

como queríamos demostrar.

□

**Lema B.4.** Sean  $p, q, t$  enteros tales que  $q + t = p$ . Entonces se verifica que

$$\Gamma_p(a) = \pi^{qt/2} \Gamma_t\left(a - \frac{q}{2}\right) \Gamma_q(a).$$

**Demostración del Lema B.4:** Por definición de la función Gamma multivariada (ver (A.2))

$$\begin{aligned}\Gamma_p(a) &= \pi^{p(p-1)/4} \prod_{j=1}^p \Gamma\left(a + \frac{1-j}{2}\right), \\ \Gamma_t\left(a - \frac{q}{2}\right) &= \pi^{t(t-1)/4} \prod_{j=1}^t \Gamma\left(a - \frac{q}{2} + \frac{1-j}{2}\right), \\ \Gamma_q(a) &= \pi^{q(q-1)/4} \prod_{j=1}^q \Gamma\left(a + \frac{1-j}{2}\right).\end{aligned}\tag{B.2}$$

Haciendo el cambio de variable en los índices  $h = j + q$  de la productoria en (B.2), resulta

$$\Gamma_t\left(a - \frac{q}{2}\right) = \pi^{t(t-1)/4} \prod_{h=q+1}^p \Gamma\left(a + \frac{1-h}{2}\right).$$

Por lo tanto tenemos

$$\begin{aligned}& \pi^{qt/2} \Gamma_t\left(a - \frac{q}{2}\right) \Gamma_q(a) \\ &= \pi^{qt/2 + t(t-1)/4 + q(q-1)/4} \prod_{h=q+1}^p \Gamma\left(a + \frac{1-h}{2}\right) \prod_{j=1}^q \Gamma\left(a + \frac{1-j}{2}\right) \\ &= \pi^{p(p-1)/4} \prod_{j=1}^p \Gamma\left(a + \frac{1-j}{2}\right),\end{aligned}$$

que es exactamente  $\Gamma_p(a)$ .

□

**Lema B.5.** Sea  $\lambda_i$ ,  $1 \leq i \leq h$ , un autovalor spike de la matriz de covarianza  $\Sigma$  y sea  $\tilde{\lambda}_i$  un estimador para  $\lambda_i$  definido por

$$\tilde{\lambda}_i = \frac{1}{2} \left\{ l_i + \hat{\sigma}^2 - \hat{\sigma}^2 \frac{p}{m} + \sqrt{-4l_i \hat{\sigma}^2 + \left( l_i + \hat{\sigma}^2 - \hat{\sigma}^2 \frac{p}{m} \right)^2} \right\}.\tag{B.3}$$

Consideremos los conjuntos de índices  $J_1 = \{i \leq h : \lambda_i > \sigma^2(1 + \sqrt{y})\}$  y  $J_2 = \{i \leq h : \lambda_i \in (\sigma^2, \sigma^2(1 + \sqrt{y})]\}$ . Entonces cuando  $p, m \rightarrow \infty$  con  $p/m \rightarrow y$ ,  $\tilde{\lambda}_i \rightarrow \lambda_i$  para  $i \in J_1$  y  $\tilde{\lambda}_i \rightarrow \sigma^2(1 + \sqrt{y})$  para  $i \in J_2$ .

**Demostración del Lema B.5:**

Consideremos primero el caso en que  $i \in J_1$ . Queremos ver que el límite cuando  $p, m \rightarrow \infty$  con  $p/m \rightarrow y$  de  $\tilde{\lambda}_i$  es igual a  $\lambda_i$ . Comencemos notando que, dado que  $i \in J_1$ , el correspondiente autovalor muestral  $l_i$  converge a  $\lambda_i \left(1 + \frac{y\sigma^2}{\lambda_i - \sigma^2}\right)$ . Además, por el Lema B.2,  $\hat{\sigma}^2 \rightarrow \sigma^2$ . Luego,

reemplazando en (B.3)  $l_i$  por  $\lambda_i \left(1 + \frac{y\sigma^2}{\lambda_i - \sigma^2}\right)$ ,  $\hat{\sigma}^2$  por  $\sigma^2$  y  $p/m$  por  $y$ , obtendremos el valor del límite buscado.

Por un lado tenemos

$$\begin{aligned} \lambda_i \left(1 + \frac{y\sigma^2}{\lambda_i - \sigma^2}\right) + \sigma^2 - \sigma^2 y &= \frac{\lambda_i^2 - \sigma^4 + \sigma^4 y}{\lambda_i - \sigma^2} \\ &= \lambda_i + \sigma^2 + \frac{\sigma^4 y}{\lambda_i - \sigma^2}. \end{aligned} \quad (\text{B.4})$$

Por otro lado, el discriminante resulta

$$\begin{aligned} &-4\lambda_i\sigma^2 - 4\frac{\lambda_i\sigma^4 y}{\lambda_i - \sigma^2} + \left[\lambda_i \left(1 + \frac{y\sigma^2}{\lambda_i - \sigma^2}\right) + \sigma^2 - \sigma^2 y\right]^2 \\ &= -4\lambda_i\sigma^2 - 4\frac{\lambda_i\sigma^4 y}{\lambda_i - \sigma^2} + \left[\lambda_i + \sigma^2 + \frac{\sigma^4 y}{\lambda_i - \sigma^2}\right]^2 \\ &= -4\lambda_i\sigma^2 - 4\frac{\lambda_i\sigma^4 y}{\lambda_i - \sigma^2} + (\lambda_i + \sigma^2)^2 + \frac{\sigma^8 y^2}{(\lambda_i - \sigma^2)^2} + 2(\lambda_i + \sigma^2)\frac{\sigma^4 y}{\lambda_i - \sigma^2} \\ &= -4\lambda_i\sigma^2 - 4\frac{\lambda_i\sigma^4 y}{\lambda_i - \sigma^2} + \lambda_i^2 + \sigma^4 + 2\lambda_i\sigma^2 + \frac{\sigma^8 y^2}{(\lambda_i - \sigma^2)^2} + 2\frac{\lambda_i\sigma^4 y}{\lambda_i - \sigma^2} + 2\frac{\sigma^6 y}{\lambda_i - \sigma^2} \\ &= \frac{\lambda_i^4 - 4\lambda_i^3\sigma^2 - 4\lambda_i\sigma^6 + 6\lambda_i^2\sigma^4 - 2\lambda_i^2\sigma^4 y + 4\lambda_i\sigma^6 y + \sigma^8(1 - y)^2}{(\lambda_i - \sigma^2)^2} \\ &= \left[\frac{\lambda_i^2 - 2\lambda_i\sigma^2 + \sigma^4(1 - y)}{\lambda_i - \sigma^2}\right]^2 \end{aligned} \quad (\text{B.5})$$

Por lo tanto, sumando (B.4) más la raíz cuadrada de (B.5) resulta que

$$\begin{aligned} \tilde{\lambda} &\rightarrow \frac{1}{2} \left[ \lambda_i + \sigma^2 + \frac{\sigma^4 y}{\lambda_i - \sigma^2} + \frac{\lambda_i^2 - 2\lambda_i\sigma^2 + \sigma^4(1 - y)}{\lambda_i - \sigma^2} \right] \\ &= \frac{1}{2} \left[ \frac{2\lambda_i(\lambda_i - \sigma^2)}{\lambda_i - \sigma^2} \right] \\ &= \lambda_i, \end{aligned}$$

lo que demuestra el Lema en el caso en que  $i \in J_1$ .

Veamos ahora el caso en que  $i \in J_2$ . Nuevamente, siguiendo a [7], sabemos que  $l_i \rightarrow \sigma^2(1 + \sqrt{y})^2$ . Luego, en el límite, B.3 es igual a

$$\frac{1}{2} \left\{ \sigma^2(1 + \sqrt{y})^2 + \sigma^2 - \sigma^2 y + \sqrt{-4\sigma^2(1 + \sqrt{y})^2\sigma^2 + [\sigma^2(1 + \sqrt{y})^2 + \sigma^2 - \sigma^2 y]^2} \right\}.$$

Ahora bien, notando que  $\sigma^2(1 + \sqrt{y})^2 + \sigma^2 - \sigma^2 y = 2\sigma^2(1 + \sqrt{y})$ , tenemos

$$\begin{aligned}\tilde{\lambda}_i &\rightarrow \frac{1}{2} \left\{ 2\sigma^2(1 + \sqrt{y}) + \sqrt{-4\sigma^4(1 + \sqrt{y})^2 + 4\sigma^4(1 + \sqrt{y})^2} \right\} \\ &= \frac{1}{2} [2\sigma^2(1 + \sqrt{y})] \\ &= \sigma^2(1 + \sqrt{y}),\end{aligned}$$

lo que completa la demostración.

□

## Demostraciones del Capítulo 2

### C.1 Demostración de la Proposición 2.4

Para la demostración de este resultado seguiremos la técnica desarrollada por [33] y las aproximaciones dadas en [20]. En primer lugar notemos que la hipótesis nula  $\mathcal{H}_d$  es equivalente a tener

$$\begin{aligned}\Sigma &= (\Psi, \Psi_0)\Lambda(\Psi, \Psi_0)^\top \\ &= \Psi\Lambda_d\Psi^\top + \sigma^2\Psi_0\Psi_0^\top,\end{aligned}\tag{C.1}$$

donde  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_h, \sigma^2, \dots, \sigma^2)$ ,  $\Lambda_d$  es la matriz truncada obtenida luego de eliminar las últimas  $p - d$  filas y columnas de  $\Lambda$ ,  $d = \sum_{i=1}^h s_i$ ,  $\lambda_1 > \dots > \lambda_h > \sigma^2$ ,  $\Psi = (\Psi_1, \dots, \Psi_h) \in \mathbb{R}^{p \times d}$  semiortogonal con  $\Psi_i \in \mathbb{R}^{p \times s_i}$  y  $\Psi_0$  el complemento semiortogonal de  $\Psi$  de dimensión  $p \times (p - d)$ .

Denotemos con  $\hat{\Psi}_i$ ,  $i = 0, \dots, h$  a la versión muestral de  $\Psi_i \in \mathbb{R}^{p \times s_i}$ ,  $i = 0, \dots, h$  y  $s_0 = p - d$ , esto es, para  $i = 1, \dots, h$ ,  $\hat{\Psi}_i$  es la matriz cuyas columnas son los autovectores correspondientes a los autovalores  $l_{\sum_{j=1}^{i-1} s_j + 1}, \dots, l_{\sum_{j=1}^i s_j}$  de  $\hat{\Sigma}$  (bajo la hipótesis nula  $\mathcal{H}_d$ , las matrices  $\Psi_i$  para  $i = 1, \dots, h$ , corresponden a los  $d$  autovalores spike de la matriz de covarianza  $\Sigma$ ), y  $\hat{\Psi}_0$  es la matriz cuyas columnas son los autovectores correspondientes a los  $p - d$  autovalores más chicos (en la población, bajo  $\mathcal{H}_d$ , la matriz  $\Psi_0$  corresponde a los autovalores de la parte de ruido, que son todos iguales a  $\sigma^2$ ). Utilizando estas matrices, podemos descomponer a la matriz de

covarianza muestral del siguiente modo

$$\widehat{\Sigma} = (\widehat{\Psi}, \widehat{\Psi}_0) \widehat{\Lambda} \begin{pmatrix} \widehat{\Psi}^\top \\ \widehat{\Psi}_0^\top \end{pmatrix}$$

con  $\widehat{\Lambda} = \text{diag}(l_1, \dots, l_p)$ . Es preciso notar que, dado que  $m > p$ , los autovalores de la matriz de covarianzas muestral, con probabilidad 1, son todos diferentes. Ahora bien, a partir de la definición del estadístico  $LRT_d$ , podemos escribir

$$\begin{aligned} LRT_d &= \frac{l_{d+1} \cdots l_p}{\left( \frac{1}{p-d} \sum_{i=d+1}^p l_i \right)^{p-d}}, \\ &= LRT_* A_{\Psi_1} \cdots A_{\Psi_h} A_{\Psi_0}, \end{aligned} \quad (\text{C.2})$$

donde

$$\begin{aligned} LRT_* &= \frac{|\widehat{\Sigma}|}{\left\{ \frac{\text{tr}(\widehat{\Psi}_1^\top \widehat{\Sigma} \widehat{\Psi}_1)}{s_1} \right\}^{s_1} \cdots \left\{ \frac{\text{tr}(\widehat{\Psi}_h^\top \widehat{\Sigma} \widehat{\Psi}_h)}{s_h} \right\}^{s_h} \left\{ \frac{\text{tr}(\widehat{\Psi}_0^\top \widehat{\Sigma} \widehat{\Psi}_0)}{p-d} \right\}^{p-d}}, \\ A_{\Psi_i} &= \frac{\left\{ \frac{\text{tr}(\widehat{\Psi}_i^\top \widehat{\Sigma} \widehat{\Psi}_i)}{s_i} \right\}^{s_i}}{|\widehat{\Psi}_i^\top \widehat{\Sigma} \widehat{\Psi}_i|}, \quad i = 1, \dots, h, \text{ y} \\ A_{\Psi_0} &= \left\{ \frac{\text{tr}(\widehat{\Psi}_0^\top \widehat{\Sigma} \widehat{\Psi}_0)}{\text{tr}(\widehat{\Psi}_0^\top \widehat{\Sigma} \widehat{\Psi}_0)} \right\}^{p-d}. \end{aligned}$$

Dado que  $\sigma_{m,p,d} \not\rightarrow 0$ , el resultado se sigue si probamos que

$$\frac{\log LRT_* - \widetilde{\mu}_{m,p}}{\sigma_{m,p,d}} \rightsquigarrow \mathcal{N}(0, 1), \quad (\text{C.3})$$

$$\log \prod_{i=1}^h A_{\Psi_i} - \log A_{m,p,d} \rightarrow 0, \quad (\text{C.4})$$

$$\log A_{\Psi_0} - \log B_{m,p,d} \rightarrow 0. \quad (\text{C.5})$$

DEMOSTRACIÓN DE (C.3): Para imitar el procedimiento utilizado en la prueba de Muirhead, necesitamos calcular  $\mathbb{E}(LRT_*^t)$ , es decir, necesitamos la función generadora de momentos de  $\log LRT_*$ , bajo la hipótesis nula, para  $t$  en un entorno del 0. Recordemos que hemos denotado con  $s_0 = p - d$  y, para facilitar la notación, llamemos  $\lambda_0 = \sigma^2$ . Entonces, utilizando propiedades

de determinantes, podemos escribir

$$\begin{aligned}
LRT_* &= \frac{|\hat{\Sigma}|}{\prod_{i=0}^h \left[ \text{tr} \left( \frac{\Psi_i^\top \hat{\Sigma} \Psi_i}{s_i} \right) \right]^{s_i}} \\
&= \frac{|\hat{\Sigma}| |\Sigma^{-1}|}{\prod_{i=0}^h \left[ \frac{1}{\lambda_i} \text{tr} \left( \frac{\Psi_i^\top \hat{\Sigma} \Psi_i}{s_i} \right) \right]^{s_i}} \\
&= \frac{|\Sigma^{-1/2} \hat{\Sigma} \Sigma^{-1/2}|}{\prod_{i=0}^h \left[ \text{tr} \left( \frac{\Psi_i^\top \Sigma^{-1/2} \hat{\Sigma} \Sigma^{-1/2} \Psi_i}{s_i} \right) \right]^{s_i}}.
\end{aligned}$$

La última igualdad se deduce del hecho que  $\Sigma^{-1/2} \Psi_i = \lambda_i^{-1/2} \Psi_i$ . Por lo tanto, tenemos que

$$\begin{aligned}
\mathbb{E}(LRT_*^t) &= \mathbb{E} \left[ \frac{|\Sigma^{-1/2} \hat{\Sigma} \Sigma^{-1/2}|^t}{\prod_{i=0}^h \text{tr} \left( \frac{\Psi_i^\top \Sigma^{-1/2} \hat{\Sigma} \Sigma^{-1/2} \Psi_i}{s_i} \right)^{s_i t}} \right] \\
&= \mathbb{E} \left[ \frac{|\mathbf{A}|^t}{\prod_{i=0}^h \text{tr} \left( \frac{\mathbf{A}_{ii}}{s_i} \right)^{s_i t}} \right],
\end{aligned}$$

donde hemos denotado  $\mathbf{A} := m \Sigma^{-1/2} \hat{\Sigma} \Sigma^{-1/2}$  y  $\mathbf{A}_{ii} := m \Psi_i^\top \Sigma^{-1/2} \hat{\Sigma} \Sigma^{-1/2} \Psi_i$ . En esta última expresión, la esperanza se calcula respecto a la distribución de la matriz  $\mathbf{A} \sim \mathcal{W}_p(m, \mathbf{I}_p)$ .

Usando la Proposición A.2,  $\mathbf{A}_{ii} \sim \mathcal{W}_{s_i}(m, \mathbf{I}_{s_i})$  y además son independientes. Para ver esta última afirmación, llamando  $\mathbf{A}_{ij} \doteq \Psi_i^\top \mathbf{A} \Psi_j$ , tenemos

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{hh} & \mathbf{A}_{00} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{A}_{h1} & \mathbf{A}_{h2} & \dots & \mathbf{A}_{hh} & \mathbf{A}_{h0} \\ \mathbf{A}_{01} & \mathbf{A}_{02} & \dots & \mathbf{A}_{0h} & \mathbf{A}_{00} \end{pmatrix} = \begin{pmatrix} \Psi_1^\top \\ \Psi_2^\top \\ \vdots \\ \Psi_h^\top \\ \Psi_0^\top \end{pmatrix} \mathbf{A} \begin{pmatrix} \Psi_1 & \Psi_2 & \dots & \Psi_h & \Psi_0 \end{pmatrix}.$$

Usando ahora que  $\mathbf{A} \sim \mathcal{W}_p(m, \mathbf{I}_p)$ , por Proposición A.2,

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{hh} & \mathbf{A}_{00} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{A}_{h1} & \mathbf{A}_{h2} & \dots & \mathbf{A}_{hh} & \mathbf{A}_{h0} \\ \mathbf{A}_{01} & \mathbf{A}_{02} & \dots & \mathbf{A}_{0h} & \mathbf{A}_{00} \end{pmatrix} \sim \mathcal{W}_p(m, (\Psi, \Psi_0)^\top (\Psi, \Psi_0)),$$

donde hemos usado que  $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_h)$ . Ahora bien,  $(\Psi, \Psi_0)^\top (\Psi, \Psi_0) = \mathbf{I}_p$ ; de aquí, utilizando la Proposición A.3, resultan  $\mathbf{A}_{ii}$  independiente de  $\mathbf{A}_{jj}$  para  $i \neq j$ .

Como consecuencia,

$$\begin{aligned}
\mathbb{E}(LRT_*^t) &= \frac{1}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \int \frac{|\mathbf{A}|^{t+\frac{m-p-1}{2}} \exp\left\{-\frac{1}{2} \operatorname{tr}(\mathbf{A})\right\}}{\prod_{i=0}^h \left(\frac{\operatorname{tr}\mathbf{A}_{ii}}{s_i}\right)^{s_i t}} d\mathbf{A} \\
&= \frac{2^{\frac{(m+2t)p}{2}} \Gamma_p(\frac{m+2t}{2})}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \mathbb{E} \left\{ \frac{1}{\prod_{i=0}^h \left(\frac{\operatorname{tr}\mathbf{A}_{ii}}{s_i}\right)^{s_i t}} \right\} \\
&= \frac{2^{\frac{(m+2t)p}{2}} \Gamma_p(\frac{m+2t}{2})}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \prod_{i=0}^h \mathbb{E} \left\{ \left(\frac{\operatorname{tr}\mathbf{A}_{ii}}{s_i}\right)^{-s_i t}\right\}, \tag{C.6}
\end{aligned}$$

donde la esperanza la calculamos con respecto a la matriz  $\mathbf{A} \sim \mathcal{W}_p(m+2t, \mathbf{I}_p)$  y utilizamos la independencia de los  $\mathbf{A}_{ii}$  en la última expresión.

Para calcular esta esperanza, vamos a recurrir a un cambio de variables. Pero antes notemos que, si llamamos con  $\Psi_{i0}$  al complemento ortogonal de  $\Psi_i$ , entonces  $\mathbf{A}$  puede escribirse de la forma  $\mathbf{A} = (\Psi_i \ \Psi_{i0}) \begin{pmatrix} \mathbf{A}_{ii} & \mathbf{A}_{i0} \\ \mathbf{A}_{i0}^\top & \mathbf{A}_{00} \end{pmatrix} \begin{pmatrix} \Psi_i^\top \\ \Psi_{i0}^\top \end{pmatrix}$ , donde  $\mathbf{A}_{i0} = \Psi_{i0}^\top \mathbf{A} \Psi_i$ . Además, se verifica que

$$|\mathbf{A}| = \begin{vmatrix} \mathbf{A}_{ii} & \mathbf{A}_{i0} \\ \mathbf{A}_{i0}^\top & \mathbf{A}_{00} \end{vmatrix} \quad \text{y} \quad \operatorname{tr}(\mathbf{A}) = \operatorname{tr} \begin{pmatrix} \mathbf{A}_{ii} & \mathbf{A}_{i0} \\ \mathbf{A}_{i0}^\top & \mathbf{A}_{00} \end{pmatrix},$$

y por Proposición A.2,  $\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_{ii} & \mathbf{A}_{i0} \\ \mathbf{A}_{i0}^\top & \mathbf{A}_{00} \end{pmatrix} = \begin{pmatrix} \Psi_i^\top \\ \Psi_{i0}^\top \end{pmatrix} \mathbf{A} (\Psi_i \ \Psi_{i0}) \sim \mathcal{W}_p(m+2t, \mathbf{I}_p)$ . Además, por ser  $(\Psi_i \ \Psi_{i0})$  ortogonal, el jacobiano de la transformación  $\tilde{\mathbf{A}} = h(\mathbf{A})$  es igual a 1. Luego, podemos cambiar  $\mathbf{A}$  por  $\tilde{\mathbf{A}}$  en el cálculo de la esperanza.

Proponemos entonces el cambio de variables:

$$\begin{aligned}
\tilde{\mathbf{A}} &= g : \mathcal{S}_{s_i}^+ \times \mathbb{R}^{s_i \times (p-s_i)} \times \mathcal{S}_{p-s_i}^+ = g(\mathbf{V}_{ii}, \mathbf{V}_{i0}, \mathbf{V}_{00}), \\
\begin{pmatrix} \mathbf{A}_{ii} & \mathbf{A}_{i0} \\ \mathbf{A}_{i0}^\top & \mathbf{A}_{00} \end{pmatrix} &= \begin{pmatrix} \mathbf{V}_{ii} & \mathbf{V}_{ii} \mathbf{V}_{i0} \\ \mathbf{V}_{i0}^\top \mathbf{V}_{ii} & \mathbf{V}_{00} + \mathbf{V}_{i0}^\top \mathbf{V}_{ii} \mathbf{V}_{i0} \end{pmatrix}
\end{aligned}$$

Tenemos entonces que

$$\begin{aligned}
\mathbf{V}_{ii} &= \mathbf{A}_{ii}, \\
\mathbf{V}_{i0} &= \mathbf{A}_{ii}^{-1} \mathbf{A}_{i0}, \\
\mathbf{V}_{00} &= \mathbf{A}_{00} - \mathbf{A}_{i0}^\top \mathbf{A}_{ii}^{-1} \mathbf{A}_{i0}.
\end{aligned}$$

También resulta

$$\begin{aligned} |\tilde{\mathbf{A}}| &= |\mathbf{V}_{ii}||\mathbf{V}_{00}|, \\ \text{tr}(\tilde{\mathbf{A}}) &= \text{tr}(\mathbf{V}_{ii}) + \text{tr}(\mathbf{V}_{00} + \mathbf{V}_{i0}^\top \mathbf{V}_{ii} \mathbf{V}_{i0}). \end{aligned}$$

De acuerdo a la Proposición A.1, el jacobiano de esta transformación es  $|\mathbf{V}_{ii}|^{p-s_i}$ . Por lo tanto, tenemos que

$$\begin{aligned} \mathbb{E} \left[ (\text{tr} \mathbf{A}_{ii}/s_i)^{-s_i t} \right] &= \frac{1}{2^{(m+2t)p/2} \Gamma_p \left( \frac{m+2t}{2} \right)} \int_{\mathcal{S}_p^+} \frac{|\tilde{\mathbf{A}}|^{(m+2t-p-1)/2} e^{-\frac{1}{2} \text{tr}(\tilde{\mathbf{A}})}}{[\text{tr}(\mathbf{A}_{ii}/s_i)]^{s_i t}} d\tilde{\mathbf{A}} \\ &= \frac{1}{2^{(m+2t)p/2} \Gamma_p \left( \frac{m+2t}{2} \right)} \\ &\quad \int_B \frac{|\mathbf{V}_{ii}|^{(m+2t+p-2s_i-1)/2}}{[\text{tr}(\mathbf{A}_{ii}/s_i)]^{s_i t}} |\mathbf{V}_{00}|^{(m+2t-p-1)/2} e^{-\frac{1}{2} \text{tr}(\mathbf{V}_{ii}) - \frac{1}{2} \text{tr}(\mathbf{V}_{00} + \mathbf{V}_{i0}^\top \mathbf{V}_{ii} \mathbf{V}_{i0})} d\mathbf{V}_{ii} d\mathbf{V}_{00} d\mathbf{V}_{i0}, \end{aligned}$$

donde hemos llamado  $B = \mathcal{S}_{s_i}^+ \times \mathbb{R}^{s_i \times (p-s_i)} \times \mathcal{S}_{p-s_i}^+$ . Integrando respecto de  $\mathbf{V}_{i0}$ , resulta

$$\begin{aligned} \mathbb{E} \left[ (\text{tr} \mathbf{A}_{ii}/s_i)^{-s_i t} \right] &= \frac{(2\pi)^{s_i(p-s_i)/2}}{2^{(m+2t)p/2} \Gamma_p \left( \frac{m+2t}{2} \right)} \times \\ &\quad \times \int_{\mathcal{S}_{p-s_i}^+ \times \mathcal{S}_{s_i}^+} \frac{|\mathbf{V}_{ii}|^{(m+2t-s_i-1)/2} e^{-\frac{1}{2} \text{tr}(\mathbf{V}_{ii})} e^{-\frac{1}{2} \text{tr}(\mathbf{V}_{00})} |\mathbf{V}_{00}|^{(m+2t-p-1)/2}}{[\text{tr}(\mathbf{V}_{ii}/s_i)]^{s_i t}} d\mathbf{V}_{00} d\mathbf{V}_{ii} \end{aligned}$$

Integrando ahora respecto de  $\mathbf{V}_{00}$  nos queda

$$\begin{aligned} \mathbb{E} \left[ (\text{tr} \mathbf{A}_{ii}/s_i)^{-s_i t} \right] &= \frac{(2\pi)^{s_i(p-s_i)/2} 2^{(m+2t-s_i)(p-s_i)/2} \Gamma_{p-s_i} \left( \frac{m+2t-s_i}{2} \right)}{2^{(m+2t)p/2} \Gamma_p \left( \frac{m+2t}{2} \right)} \int_{\mathcal{S}_{s_i}^+} \frac{|\mathbf{V}_{ii}|^{(m+2t-s_i-1)/2} e^{-\frac{1}{2} \text{tr}(\mathbf{V}_{ii})} d\mathbf{V}_{ii}}{(\text{tr} \mathbf{V}_{ii}/s_i)^{s_i t}} \\ &= \frac{\pi^{s_i(p-s_i)/2} \Gamma_{p-s_i} \left( \frac{m+2t-s_i}{2} \right)}{2^{(m+2t)s_i/2} \Gamma_p \left( \frac{m+2t}{2} \right)} \int_{\mathcal{S}_{s_i}^+} \frac{|\mathbf{V}_{ii}|^t |\mathbf{V}_{ii}|^{(m-s_i-1)/2} e^{-\frac{1}{2} \text{tr} \mathbf{V}_{ii}} d\mathbf{V}_{ii}}{(\text{tr} \mathbf{V}_{ii}/s_i)^{s_i t}} \\ &= \frac{\pi^{s_i(p-s_i)/2} \Gamma_{p-s_i} \left( \frac{m+2t-s_i}{2} \right) \Gamma_{s_i} \left( \frac{m}{2} \right)}{2^{ts_i} \Gamma_p \left( \frac{m+2t}{2} \right)} \mathbb{E} \left[ \left( \frac{|\mathbf{V}_{ii}|}{\left( \frac{1}{s_i} \text{tr} \mathbf{V}_{ii} \right)^{s_i}} \right)^t \right], \end{aligned}$$

en donde la última esperanza se calcula respecto de  $\mathbf{V}_{ii} \sim \mathcal{W}_{s_i}(m, \mathbf{I}_{s_i})$ . Así, reemplazando en C.6, tenemos

$$\mathbb{E}(LRT_*^t) = \frac{2^{pt} \Gamma_p \left( \frac{m+2t}{2} \right)}{\Gamma_p \left( \frac{m}{2} \right)} \prod_{i=0}^h \frac{\pi^{s_i(p-s_i)/2} \Gamma_{p-s_i} \left( \frac{m+2t-s_i}{2} \right) \Gamma_{s_i} \left( \frac{m}{2} \right)}{2^{ts_i} \Gamma_p \left( \frac{m+2t}{2} \right)} \mathbb{E} \left\{ \left[ \frac{|\mathbf{V}_{ii}|}{\left( \frac{1}{s_i} \text{tr} \mathbf{V}_{ii} \right)^{s_i}} \right]^t \right\}. \quad (\text{C.7})$$

Utilizando el Corolario A.5,

$$\mathbb{E} \left[ \left( \frac{|\mathbf{V}_{ii}|}{\left( \frac{1}{s_i} \text{tr} \mathbf{V}_{ii} \right)^{s_i}} \right)^t \right] = \frac{s_i^{s_i t} \Gamma \left( \frac{s_i m}{2} \right) \Gamma_{s_i} \left( \frac{m}{2} + t \right)}{\Gamma \left( \frac{s_i m}{2} + s_i t \right) \Gamma_{s_i} \left( \frac{m}{2} \right)}. \quad (\text{C.8})$$

Así, reemplazando en (C.7), resulta

$$\mathbb{E}(LRT_*^t) = \frac{\Gamma_p\left(\frac{m+2t}{2}\right)}{\Gamma_p\left(\frac{m}{2}\right)} \prod_{i=0}^h \frac{\pi^{s_i(p-s_i)/2} \Gamma_{p-s_i}\left(\frac{m+2t-s_i}{2}\right) s_i^{s_i t} \Gamma\left(\frac{s_i m}{2}\right) \Gamma_{s_i}\left(\frac{m}{2} + t\right)}{\Gamma\left(\frac{s_i m}{2} + s_i t\right) \Gamma_p\left(\frac{m+2t}{2}\right)}.$$

En esta última expresión, usando el Lema B.4, resulta

$$\mathbb{E}(LRT_*^t) = \frac{\Gamma_p\left(\frac{m+2t}{2}\right)}{\Gamma_p\left(\frac{m}{2}\right)} \prod_{i=0}^h \frac{s_i^{s_i t} \Gamma\left(\frac{1}{2} s_i m\right)}{\Gamma\left(\frac{1}{2} s_i m + s_i t\right)}. \quad (\text{C.9})$$

Usemos ahora el Lema A.6, el cual es consecuencia de la expansión de Stirling para la función Gamma. Para ello, tomemos  $b(x) = 2tx/m$  con  $x = m(p-d)/2$ . Dado que  $b(x) = \mathcal{O}(\sqrt{x}) = \mathcal{O}(m)$  para  $t$  finito,  $-\log\left[\frac{\Gamma\{m(p-d)/2 + t(p-d)\}}{\Gamma\{m(p-d)/2\}}\right]$  es igual a

$$\begin{aligned} & -t(p-d) \log\left[\frac{m}{2}(p-d)\right] - \frac{t^2(p-d)^2 - t(p-d)}{m(p-d)} + \mathcal{O}\left(\frac{1}{m}\right) \\ &= -t(p-d) \log\left[\frac{m}{2}(p-d)\right] - \frac{t^2(p-d)}{m} + \mathcal{O}\left(\frac{1}{m}\right) \end{aligned}$$

y tomando  $b(x) = s_i t$ , para  $t$  finito y  $x = m s_i/2$ ,  $b(x) = \mathcal{O}(1)$ ,

$$\begin{aligned} -\log\frac{\Gamma\left(\frac{m}{2} s_i + s_i t\right)}{\Gamma\left(\frac{m}{2} s_i\right)} &= -s_i t \log\left(\frac{m}{2} s_i\right) - \frac{s_i^2 t^2 - s_i t}{m s_i} + \mathcal{O}\left(\frac{1}{m^2}\right) \\ &= -s_i t \log\left(\frac{m}{2} s_i\right) + \mathcal{O}\left(\frac{1}{m}\right). \end{aligned}$$

Por otro lado, aplicando el Lema A.7 para  $r_m^2 = -\log(1-p/m)$  y  $t = \mathcal{O}(1/r_m)$ , tenemos

$$\log\frac{\Gamma_p\left(\frac{m}{2} + t\right)}{\Gamma_p\left(\frac{m}{2}\right)} = pt(\log m - 1 - \log 2) + r_m^2 \left[ t^2 - \left(p - m + \frac{1}{2}\right) t \right] + o(1).$$

Tomando ahora logaritmo en (C.9) y utilizando las aproximaciones antes expuestas, resulta

$$\begin{aligned} \log \mathbb{E}(LRT_*^t) &= \log\frac{\Gamma_p\left(\frac{m}{2} + t\right)}{\Gamma_p\left(\frac{m}{2}\right)} + (p-d)t \log(p-d) \\ &\quad - \log\left(\frac{\Gamma\left[\frac{m(p-d)}{2} + t(p-d)\right]}{\Gamma\left[\frac{m(p-d)}{2}\right]}\right) + \sum_{i=1}^h s_i t \log s_i - \sum_{i=1}^h \log\frac{\Gamma\left(\frac{m}{2} s_i + s_i t\right)}{\Gamma\left(\frac{m}{2} s_i\right)} \\ &= pt[\log(m) - 1 - \log(2)] + r_m^2 \left[ t^2 - \left(p - m + \frac{1}{2}\right) t \right] + (p-d)t \log(p-d) \\ &\quad - t(p-d) \log\left[\frac{m}{2}(p-d)\right] - \frac{t^2(p-d)}{m} + \sum_{i=1}^h s_i t \log(s_i) - \sum_{i=1}^h s_i t \log\left(\frac{m}{2} s_i\right) + o(1) \\ &= 2 \left[ -\frac{(p-d)}{m} + r_m^2 \right] \frac{t^2}{2} + \left[ -p + r_m^2 \left(m - p - \frac{1}{2}\right) \right] t + o(1). \quad (\text{C.10}) \end{aligned}$$

De aquí, recordando que hemos llamado  $\tilde{\mu}_{m,p} = -p - \left(m - p - \frac{1}{2}\right) \log\left(1 - \frac{p}{m}\right)$  y  $\sigma_{m,p,d}^2 = -2 \left[ \frac{p-d}{m} + \log\left(1 - \frac{p}{m}\right) \right]$ , en (C.10) tenemos

$$\log \mathbb{E}(LRT_*^t) = \sigma_{m,p,d}^2 \frac{t^2}{2} + \tilde{\mu}_{m,p} t + o(1),$$

o, equivalentemente

$$\mathbb{E}(LRT_*^t) = e^{\frac{t^2}{2} \sigma_{m,p,d}^2 + \tilde{\mu}_{m,p} t + o(1)},$$

de donde obtenemos (C.3).

DEMOSTRACIÓN DE (C.4): Como  $m \Psi_i^\top \hat{\Sigma} \Psi_i \sim \mathcal{W}_{s_i}(m, \lambda_i \mathbf{I}_{s_i})$ , tenemos que si  $m \rightarrow \infty$   $\text{tr}(\Psi_i^\top \hat{\Sigma} \Psi_i)/s_i \rightarrow \lambda_i$  (ver Lema B.1). Por lo tanto

$$\sum_{i=1}^h \log A_{\Psi_i} - \log A_{m,p,d} = \sum_{i=1}^h s_i \left( \log \frac{\text{tr}(\Psi_i^\top \hat{\Sigma} \Psi_i)}{s_i} - \log \lambda_i \right) \rightarrow 0.$$

DEMOSTRACIÓN DE (C.5): Por definición de  $A_{\Psi_0}$ , necesitamos calcular

$$\frac{\text{tr}(\Psi_0^\top \hat{\Sigma} \Psi_0)}{\text{tr}(\hat{\Psi}_0^\top \hat{\Sigma} \hat{\Psi}_0)} = \frac{\sum_{i=1}^p l_i - \text{tr}(\Psi^\top \hat{\Sigma} \Psi)}{\text{tr}(\hat{\Psi}_0^\top \hat{\Sigma} \hat{\Psi}_0)}.$$

Por lo tanto,

$$\log(A_{\Psi_0}/B_{m,p,d}) = (p-d) \log \left\{ 1 + \frac{1}{p-d} \frac{\sum_{i=1}^h s_i \lambda_i - \text{tr}(\Psi^\top \hat{\Sigma} \Psi)}{\frac{1}{p-d} (\sum_{i=1}^p l_i - \sum_{i=1}^h s_i \lambda_i)} \right\},$$

y, por Lema B.3, la prueba de (C.5) se sigue si

$$a = \frac{\sum_{i=1}^h s_i \lambda_i - \text{tr}(\Psi^\top \hat{\Sigma} \Psi)}{\frac{1}{p-d} (\sum_{i=1}^p l_i - \sum_{i=1}^h s_i \lambda_i)} \rightarrow 0. \quad (\text{C.11})$$

Por un lado, que el numerador tiende a 0 es consecuencia del Lema B.1 y del hecho de que  $\text{tr}(\Psi^\top \hat{\Sigma} \Psi) = \sum_{i=1}^h \text{tr}(\Psi_i^\top \hat{\Sigma} \Psi_i)$ . Por otro lado, para el denominador tenemos

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{p-d} \left( \sum_{i=1}^p l_i - \sum_{i=1}^h s_i \lambda_i \right) \right] &= \sigma^2, \\ \text{var} \left[ \frac{1}{p-d} \left( \sum_{i=1}^p l_i - \sum_{i=1}^h s_i \lambda_i \right) \right] &= \frac{2}{m(p-d)^2} \left[ \sum_{i=1}^h s_i \lambda_i^2 + \sigma^4(p-d) \right] \rightarrow 0. \end{aligned} \quad (\text{C.12})$$

En efecto, notemos primero que

$$\sum_{i=1}^p l_i - \sum_{i=1}^h s_i \lambda_i = \operatorname{tr} \widehat{\Sigma} - \sum_{i=1}^h s_i \lambda_i.$$

Luego, utilizando el Corolario A.9, tenemos que, por ser  $m\widehat{\Sigma} \sim \mathcal{W}_p(m, \Sigma)$ ,

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{p-d} \left( \sum_{i=1}^p l_i - \sum_{i=1}^h s_i \lambda_i \right) \right] &= \mathbb{E} \left( \frac{1}{p-d} \operatorname{tr} \widehat{\Sigma} - \frac{1}{p-d} \sum_{i=1}^h s_i \lambda_i \right), \\ &= \frac{1}{m(p-d)} \mathbb{E} \left[ \operatorname{tr} (m\widehat{\Sigma}) \right] - \frac{1}{p-d} \sum_{i=1}^h s_i \lambda_i, \\ &= \frac{1}{m(p-d)} \operatorname{tr} \left[ \mathbb{E} (m\widehat{\Sigma}) \right] - \frac{1}{p-d} \sum_{i=1}^h s_i \lambda_i, \\ &= \frac{1}{p-d} \operatorname{tr} \Sigma - \frac{1}{p-d} \sum_{i=1}^h s_i \lambda_i, \\ &= \frac{1}{p-d} \sum_{i=d+1}^p \lambda_i = \frac{1}{p-d} \sum_{i=d+1}^p \sigma^2 = \sigma^2. \end{aligned}$$

Por otro lado,

$$\begin{aligned} \operatorname{var} \left[ \frac{1}{p-d} \left( \sum_{i=1}^p l_i - \sum_{i=1}^h s_i \lambda_i \right) \right] &= \operatorname{var} \left( \frac{1}{p-d} \operatorname{tr} \widehat{\Sigma} \right) \\ &= \frac{1}{(p-d)^2 m^2} \operatorname{var} \left[ \operatorname{tr} (m\widehat{\Sigma}) \right] \\ &= \frac{1}{(p-d)^2 m^2} \operatorname{var} \left[ \operatorname{vec}^\top (\mathbf{I}_p) \operatorname{vec} (m\widehat{\Sigma}) \right] \\ &= \frac{1}{(p-d)^2 m^2} \operatorname{vec}^\top (\mathbf{I}_p) \operatorname{var} \left[ \operatorname{vec} (m\widehat{\Sigma}) \right] \operatorname{vec} (\mathbf{I}_p) \\ &= \frac{1}{(p-d)^2 m} \operatorname{vec}^\top (\mathbf{I}_p) \left[ (\mathbf{I}_{p^2} + \mathbf{K}_p) (\Sigma \otimes \Sigma) \right] \operatorname{vec} (\mathbf{I}_p) \\ &= \frac{2}{(p-d)^2 m} \operatorname{vec}^\top (\mathbf{I}_p) (\Sigma \otimes \Sigma) \operatorname{vec} (\mathbf{I}_p) \\ &= \frac{2}{(p-d)^2 m} \operatorname{tr} \Sigma^2 \\ &= \frac{2}{(p-d)^2 m} \left( \sum_{i=1}^h s_i \lambda_i^2 + \sigma^4 (p-d) \right), \end{aligned}$$

de lo cual se sigue (C.12) que prueba (C.11).

□

## C.2 Demostración de la Proposición 2.5

Primero notemos que por ser  $m\widehat{\Sigma} \sim \mathcal{W}_p(m, \Sigma)$ , podemos escribir  $m\widehat{\Sigma} = \mathbf{Z}^\top \mathbf{Z}$  con  $\mathbf{Z} \in \mathbb{R}^{m \times p} \sim \mathcal{N}(0, \mathbf{I}_m \otimes \Sigma)$ . Definamos ahora  $m\widetilde{\Sigma} := \mathbf{Z}\mathbf{Z}^\top = \mathbf{Z}(\Psi, \Psi_0)(\Psi, \Psi_0)^\top \mathbf{Z}^\top$ , donde  $\Psi$  y  $\Psi_0$  son las definidas en (C.1). En consecuencia,

$$m\widetilde{\Sigma} = \mathbf{Z}\Psi\Psi^\top \mathbf{Z}^\top + \mathbf{Z}\Psi_0\Psi_0^\top \mathbf{Z}^\top.$$

Ahora bien,  $\mathbf{Z}\Psi \in \mathbb{R}^{m \times d} \sim \mathcal{N}(0, \mathbf{I}_m \otimes \Lambda_d)$  y por lo tanto  $\mathbf{Z}\Psi = \mathbf{Z}_d \Lambda_d^{1/2}$  para algún  $\mathbf{Z}_d \sim \mathcal{N}(0, \mathbf{I}_m \otimes \mathbf{I}_d)$ . Análogamente,  $\mathbf{Z}\Psi_0 \in \mathbb{R}^{m \times (p-d)} \sim \sigma^2 \mathcal{N}(0, \mathbf{I}_m \otimes \mathbf{I}_{p-d})$  y por lo tanto  $\mathbf{Z}\Psi_0 = \sigma \widetilde{\mathbf{Z}}$  para algún  $\widetilde{\mathbf{Z}} \in \mathbb{R}^{m \times (p-d)} \sim \mathcal{N}(0, \mathbf{I}_m \otimes \mathbf{I}_{p-d})$ . Más aún,  $\mathbf{Z}_d$  y  $\widetilde{\mathbf{Z}}$  son independientes pues  $\mathbf{Z}\Psi$  y  $\mathbf{Z}\Psi_0$  lo son. Así,

$$\widetilde{\Sigma} = \frac{1}{m} \mathbf{Z}_d \Lambda_d \mathbf{Z}_d^\top + \frac{\sigma^2}{m} \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top, \quad (\text{C.13})$$

donde  $\widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \sim \mathcal{W}_m(p-d, \mathbf{I}_m)$  independiente de  $\mathbf{Z}_d \sim \mathcal{N}(0, \mathbf{I}_m \otimes \mathbf{I}_d)$ . Además,

$$\begin{aligned} |\widetilde{\Sigma}| &= \left| \frac{1}{m} \mathbf{Z}_d \Lambda_d \mathbf{Z}_d^\top + \frac{\sigma^2}{m} \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \right| \\ &= \left| \frac{1}{\sigma^2} \mathbf{Z}_d \Lambda_d \mathbf{Z}_d^\top \left( \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \right)^{-1} + \mathbf{I}_m \right| \left| \frac{\sigma^2}{m} \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \right| \\ &= \left| \frac{\sigma^2}{m} \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \right| \left| \mathbf{I}_d + \frac{1}{\sigma^2} \Lambda_d^{1/2} \mathbf{Z}_d^\top \left( \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \right)^{-1} \mathbf{Z}_d \Lambda_d^{1/2} \right|. \end{aligned}$$

Notemos también que los autovalores de  $\widetilde{\Sigma}$  son iguales a los autovalores no-nulos de  $\widehat{\Sigma}$  ( $l_1, \dots, l_m$ ) y por lo tanto  $\text{tr}(\widehat{\Sigma}) = \text{tr}(\widetilde{\Sigma})$ . Luego resulta

$$\begin{aligned} LRT_d &= \frac{l_{d+1} \cdots l_m}{\left( \frac{1}{m-d} \sum_{i=d+1}^m l_i \right)^{m-d}} \\ &= \frac{|\widetilde{\Sigma}|}{\left( \frac{1}{m-d} \sum_{i=d+1}^m l_i \right)^{m-d} \prod_{i=1}^d l_i} \\ &= \frac{\left| \frac{\sigma^2}{m} \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \right| \left| \mathbf{I}_d + \frac{1}{\sigma^2} \Lambda_d^{1/2} \mathbf{Z}_d^\top \left( \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \right)^{-1} \mathbf{Z}_d \Lambda_d^{1/2} \right|}{\left( \frac{1}{m-d} \sum_{i=d+1}^m l_i \right)^{m-d} \prod_{i=1}^d l_i} \\ &= \frac{\left| \sigma^2 \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \right|}{\left[ \frac{1}{m} \text{tr}(\sigma^2 \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top) \right]^m} \left[ \frac{\frac{1}{m} \text{tr}(\sigma^2 \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top)}{\frac{1}{m-d} \sum_{i=d+1}^m l_i} \right]^{m-d} \left[ \frac{1}{m} \text{tr} \left( \sigma^2 \frac{\widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top}{m} \right) \right]^d \\ &= \frac{\left| \mathbf{I}_d + \frac{1}{\sigma^2} \Lambda_d^{1/2} \mathbf{Z}_d^\top \left( \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \right)^{-1} \mathbf{Z}_d \Lambda_d^{1/2} \right|}{\prod_{i=1}^d l_i} \\ &= LRT^*BCD, \end{aligned}$$

donde  $B$ ,  $C$  y  $D$  son los factores en exactamente el mismo orden que en la cuarta línea de la ecuación anterior.

Nuevamente, dado que  $\sigma_{m,p,d} \not\rightarrow 0$ , para probar el resultado, bastará con probar que

$$\frac{\log LRT^* - \tilde{\mu}_{m,p,d}}{\sigma_{m,p,d}} \rightarrow \mathcal{N}(0, 1), \quad (\text{C.14})$$

$$\log B - \log B_{m,p,d} \rightarrow 0, \quad (\text{C.15})$$

$$\log C - \log C_{m,p,d} \rightarrow 0, \quad (\text{C.16})$$

$$\log D - \log D_{m,p,d} \rightarrow 0. \quad (\text{C.17})$$

DEMOSTRACIÓN DE (C.14): Por definición,

$$LRT^* = \frac{|\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top|}{\left[\frac{1}{m} \text{tr}(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top)\right]^m}.$$

Ahora bien, dado que  $\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top \sim \mathcal{W}_m(p-d, \mathbf{I}_m)$ , con  $p-d > m$ , el resultado se sigue del Teorema A.8.

DEMOSTRACIÓN DE (C.15): Primero notemos que, usando (C.13) y algunas manipulaciones algebraicas, podemos reescribir a  $B$  del siguiente modo

$$\begin{aligned} B &= \left[ \frac{\frac{1}{m} \text{tr}(\sigma^2 \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top)}{\frac{1}{m-d} \sum_{i=d+1}^m l_i} \right]^{m-d} \\ &= \left[ \frac{\text{tr}\left(\tilde{\Sigma} - \frac{1}{m} \mathbf{Z}_d \Lambda_d \mathbf{Z}_d^\top\right)}{\sum_{i=d+1}^m l_i} \right]^{m-d} \left(\frac{m-d}{m}\right)^{m-d}. \end{aligned}$$

Por otro lado, recordemos que

$$\begin{aligned} B_{m,p,d} &= \left(1 + \frac{\sum_{i=1}^d l_i - \sum_{i=1}^h s_i \lambda_i}{\sum_{i=d+1}^m l_i}\right)^{m-d} \left(\frac{m-d}{m}\right)^{m-d} \\ &= \left(\frac{\text{tr}\tilde{\Sigma} - \sum_{i=1}^h s_i \lambda_i}{\sum_{i=d+1}^m l_i}\right)^{m-d} \left(\frac{m-d}{m}\right)^{m-d}. \end{aligned}$$

Luego

$$\begin{aligned}
\log(B/B_{m,p,d}) &= (m-d) \log \frac{\text{tr} \left( \tilde{\Sigma} - \frac{1}{m} \mathbf{Z}_d \mathbf{\Lambda}_d \mathbf{Z}_d^\top \right)}{\text{tr}(\tilde{\Sigma}) - \sum_{i=1}^h s_i \lambda_i} \\
&= (m-d) \log \left( 1 + \frac{\sum_{i=1}^h s_i \lambda_i - \frac{1}{m} \text{tr}(\mathbf{Z}_d \mathbf{\Lambda}_d \mathbf{Z}_d^\top)}{\text{tr}(\hat{\Sigma}) - \sum_{i=1}^h s_i \lambda_i} \right) \\
&= (m-d) \log \left[ 1 + \frac{1}{m-d} \frac{\sum_{i=1}^h s_i \lambda_i - \frac{1}{m} \text{tr}(\mathbf{Z}_d \mathbf{\Lambda}_d \mathbf{Z}_d^\top)}{\frac{1}{m-d} \left( \text{tr}(\hat{\Sigma}) - \sum_{i=1}^h s_i \lambda_i \right)} \right],
\end{aligned}$$

y la prueba de (C.15) se sigue si probamos que

$$a = \frac{\sum_{i=1}^h s_i \lambda_i - \frac{1}{m} \text{tr}(\mathbf{Z}_d \mathbf{\Lambda}_d \mathbf{Z}_d^\top)}{\frac{1}{m-d} \left[ \text{tr}(\hat{\Sigma}) - \sum_{i=1}^h s_i \lambda_i \right]} \rightarrow 0. \quad (\text{C.18})$$

En efecto, por un lado tenemos que

$$\begin{aligned}
\mathbb{E} \left[ \text{tr}(\mathbf{Z}_d \mathbf{\Lambda}_d \mathbf{Z}_d^\top / m) \right] &= \sum_{i=1}^h s_i \lambda_i, \\
\text{var} \left[ \text{tr}(\mathbf{Z}_d \mathbf{\Lambda}_d \mathbf{Z}_d^\top / m) \right] &= \mathcal{O}(m^{-1}).
\end{aligned}$$

Para ver esto, en primer lugar notemos que, por ser  $\mathbf{Z}_d \mathbf{\Lambda}_d^{1/2} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m \otimes \mathbf{\Lambda}_d)$ , tenemos que  $\mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2} \sim \mathcal{W}_d(m, \mathbf{\Lambda}_d)$ . Así, usando la propiedad cíclica de la traza y el Corolario A.9, tenemos que

$$\begin{aligned}
\mathbb{E} \left[ \text{tr}(\mathbf{Z}_d \mathbf{\Lambda}_d \mathbf{Z}_d^\top / m) \right] &= \mathbb{E} \left[ \text{tr}(\mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2} / m) \right] \\
&= \frac{1}{m} \text{tr} \mathbb{E}(\mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2}) \\
&= \frac{1}{m} \text{tr}(m \mathbf{\Lambda}_d) \\
&= \sum_{i=1}^h s_i \lambda_i.
\end{aligned}$$

Por otro lado,

$$\begin{aligned}
\text{var} [\text{tr} (\mathbf{Z}_d \mathbf{\Lambda}_d \mathbf{Z}_d^\top / m)] &= \frac{1}{m^2} \text{var} [\text{tr} (\mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2})] \\
&= \frac{1}{m^2} \text{var} [\text{vec}^\top (\mathbf{I}_d) \text{vec} (\mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2})] \\
&= \frac{1}{m^2} \text{vec}^\top (\mathbf{I}_d) \text{var} (\mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2}) \text{vec} (\mathbf{I}_d) \\
&= \frac{1}{m} \text{vec}^\top (\mathbf{I}_d) (\mathbf{I}_{d^2} + \mathbf{K}_d) (\mathbf{\Lambda}_d \otimes \mathbf{\Lambda}_d) \text{vec} (\mathbf{I}_d) \\
&= \frac{2}{m} \text{tr} (\mathbf{\Lambda}_d^2) = \frac{2}{m} \sum_{i=1}^h s_i \lambda_i^2.
\end{aligned}$$

Por lo tanto, el numerador tiende a 0 cuando  $p, m \rightarrow \infty$ . Veamos ahora que el denominador tiende a  $y\sigma^2 \neq 0$  si  $p, m \rightarrow \infty$ . En efecto, usando que  $p/m \rightarrow y$  y el Lema B.2, tenemos

$$\begin{aligned}
\frac{1}{m-d} \text{tr} \widehat{\Sigma} &= \frac{1}{m-d} \sum_{i=1}^d l_i + \frac{p}{m-d} \frac{p-d}{p} \frac{1}{p-d} \sum_{i=d+1}^p l_i, \\
&\rightarrow 0 + y \cdot 1 \cdot \sigma^2 \neq 0,
\end{aligned}$$

de donde se sigue (C.18) y en consecuencia tenemos (C.15).

DEMOSTRACIÓN DE (C.16): Por definición de  $C$  y  $C_{m,p,d}$  tenemos

$$\begin{aligned}
\log C - \log C_{m,p,d} &= d \left\{ \log \left[ \frac{1}{m} \text{tr} \left( \frac{\sigma^2}{m} \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \right) \right] - \log \left[ \frac{\sigma^2(p-d)}{m} \right] \right\} \\
&= d \log \frac{\text{tr} (\widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top)}{m(p-d)}.
\end{aligned}$$

Dado que  $\widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top \sim \mathcal{W}_m(p-d, \mathbf{I}_m)$ , por el Corolario A.9

$$\begin{aligned}
\mathbb{E} \frac{\text{tr} (\widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top)}{m(p-d)} &= \frac{\text{tr} \mathbb{E} (\widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top)}{m(p-d)} \\
&= \frac{(p-d) \text{tr} (\mathbf{I}_m)}{m(p-d)} = 1,
\end{aligned}$$

y

$$\begin{aligned}
\text{var} \left[ \frac{\text{tr} (\widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top)}{m(p-d)} \right] &= \frac{1}{m^2(p-d)^2} \text{vec}^\top (\mathbf{I}_m) \text{var} [\text{vec} (\widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^\top)] \text{vec} (\mathbf{I}_m) \\
&= \frac{1}{m^2(p-d)} \text{vec}^\top (\mathbf{I}_m) (\mathbf{I}_{m^2} + \mathbf{K}_m) \text{vec} (\mathbf{I}_m) \\
&= \frac{2}{m^2(p-d)} \text{tr} (\mathbf{I}_m) \\
&= \frac{2}{m(p-d)} \rightarrow 0.
\end{aligned}$$

Con esto se sigue (C.16).

DEMOSTRACIÓN DE (C.17):

Comencemos recordando las definiciones de  $D$  and  $D_{m,p,d}$ :

$$D = \frac{|\mathbf{I}_d + \frac{1}{\sigma^2} \mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2}|}{\prod_{i=1}^d l_i},$$

$$D_{m,p,d} = \frac{\prod_{i=1}^h (1 + \frac{\lambda_i}{\sigma^2} \frac{m}{p-d-m-1})^{s_i}}{\prod_{i=1}^d l_i}.$$

Por lo tanto, será suficiente probar que

$$\log \left| \mathbf{I}_d + \frac{1}{\sigma^2} \mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2} \right| - \sum_{i=1}^h s_i \log \left( 1 + \frac{\lambda_i}{\sigma^2} \frac{m}{p-d-m-1} \right) \rightarrow 0. \quad (\text{C.19})$$

Para ver esto, veamos que

$$\mathbb{E} \left( \mathbf{I}_d + \frac{1}{\sigma^2} \mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2} \right) = \mathbf{I}_d + \frac{1}{\sigma^2} \frac{m}{p-d-m-1} \mathbf{\Lambda}_d > 0, \quad (\text{C.20})$$

$$\text{var} \left[ \text{vec} \left( \mathbf{I}_d + \frac{1}{\sigma^2} \mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2} \right) \right] \rightarrow 0. \quad (\text{C.21})$$

En efecto, utilizando la Ley de las esperanzas iteradas, resulta

$$\begin{aligned} \mathbb{E} \left( \mathbf{I}_d + \frac{1}{\sigma^2} \mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2} \right) &= \mathbf{I}_d + \frac{1}{\sigma^2} \mathbf{\Lambda}_d^{1/2} \mathbb{E}_{\mathbf{Z}_d} \left\{ \mathbb{E}_{\tilde{\mathbf{Z}}|\mathbf{Z}_d} \left[ \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \mid \mathbf{Z}_d \right] \mathbf{\Lambda}_d^{1/2} \right\} \\ &= \mathbf{I}_d + \frac{1}{\sigma^2} \mathbf{\Lambda}_d^{1/2} \mathbb{E}_{\mathbf{Z}_d} \left\{ \mathbf{Z}_d^\top \mathbb{E}_{\tilde{\mathbf{Z}}} \left[ (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \right] \mathbf{Z}_d \right\} \mathbf{\Lambda}_d^{1/2}, \end{aligned} \quad (\text{C.22})$$

donde la última igualdad es consecuencia de la independencia entre  $\mathbf{Z}_d$  y  $\tilde{\mathbf{Z}}$ . Ahora, dado que  $\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top \sim \mathcal{W}_m(p-d, \mathbf{I}_m)$ , por el Teorema A.10 ,

$$\mathbb{E}_{\tilde{\mathbf{Z}}} \left[ (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \right] = \frac{1}{p-d-m-1} \mathbf{I}_m.$$

Reemplazando en (C.22), resulta

$$\mathbb{E} \left( \mathbf{I}_d + \frac{1}{\sigma^2} \mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2} \right) = \mathbf{I}_d + \frac{1}{\sigma^2(p-d-m-1)} \mathbf{\Lambda}_d^{1/2} \mathbb{E}_{\mathbf{Z}_d} (\mathbf{Z}_d^\top \mathbf{Z}_d) \mathbf{\Lambda}_d^{1/2}.$$

Finalmente, ya que  $\mathbf{Z}_d^\top \mathbf{Z}_d \sim \mathcal{W}_d(m, \mathbf{I}_d)$ , tenemos que  $\mathbb{E}_{\mathbf{Z}_d} (\mathbf{Z}_d^\top \mathbf{Z}_d) = m \mathbf{I}_d$ . Luego,

$$\mathbb{E} \left( \mathbf{I}_d + \frac{1}{\sigma^2} \mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2} \right) = \mathbf{I}_d + \frac{1}{\sigma^2} \frac{m}{p-d-m-1} \mathbf{\Lambda}_d.$$

Veamos ahora (C.21). Primero observemos que

$$\begin{aligned} \text{var} \left\{ \text{vec} \left[ \mathbf{I}_d + \frac{1}{\sigma^2} \boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \boldsymbol{\Lambda}_d^{1/2} \right] \right\} &= \frac{1}{\sigma^4} \text{var} \left\{ \text{vec} \left[ \boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \boldsymbol{\Lambda}_d^{1/2} \right] \right\} \\ &= \frac{1}{\sigma^4} \text{var} \left[ (\boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \otimes \boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top) \text{vec}(\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \right]. \end{aligned}$$

Ahora utilizaremos la fórmula de Descomposición de la Varianza o Ley de Varianzas Iteradas, según la cual, si  $\mathbf{X}$  e  $\mathbf{Y}$  son vectores aleatorios en el mismo espacio de probabilidad,

$$\text{var}(\mathbf{Y}) = \mathbb{E}_{\mathbf{X}}[\text{var}(\mathbf{Y}|\mathbf{X})] + \text{var}_{\mathbf{X}}[\mathbb{E}(\mathbf{Y}|\mathbf{X})].$$

Tenemos pues,

$$\begin{aligned} \text{var} \left\{ \text{vec} \left[ \mathbf{I}_d + \frac{1}{\sigma^2} \boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top (\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \mathbf{Z}_d \boldsymbol{\Lambda}_d^{1/2} \right] \right\} &= \frac{1}{\sigma^4} \mathbb{E}_{\mathbf{Z}_d} \left\{ \text{var} \left[ \mathbf{M} \text{vec}(\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} | \mathbf{Z}_d \right] \right\} \\ &\quad + \frac{1}{\sigma^4} \text{var}_{\mathbf{Z}_d} \left\{ \mathbb{E} \left[ \mathbf{M} \text{vec}(\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} | \mathbf{Z}_d \right] \right\} \\ &= (I) + (II), \end{aligned} \tag{C.23}$$

donde hemos denotado  $\mathbf{M} = (\boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \otimes \boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top)$ . Veamos (I). Para ello utilizaremos el Teorema A.10 con  $a_1 = (p - d - m - 1)^{-1}$ ,  $c_1 = [(p - d - m)(p - d - m - 1)(p - d - m - 3)]^{-1}$  y  $d_1 = [(p - d - m)(p - d - m - 1)^2(p - d - m - 3)]^{-1}$ . Resulta

$$\begin{aligned} (I) &= \frac{1}{\sigma^4} \mathbb{E}_{\mathbf{Z}_d} \left\{ \text{var} \left[ \mathbf{M} \text{vec}(\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} | \mathbf{Z}_d \right] \right\} \\ &= \frac{1}{\sigma^4} \mathbb{E}_{\mathbf{Z}_d} \left\{ \mathbf{M} \text{var} \left[ \text{vec}(\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^\top)^{-1} \right] \mathbf{M}^\top \right\} \\ &= \frac{1}{\sigma^4} \mathbb{E}_{\mathbf{Z}_d} \left\{ \mathbf{M} [c_1 (\mathbf{I}_{m^2} + \mathbf{K}_m) + 2d_1 (\text{vec}(\mathbf{I}_m) \text{vec}^\top(\mathbf{I}_m))] \mathbf{M}^\top \right\} \\ &= \frac{1}{\sigma^4} \mathbb{E}_{\mathbf{Z}_d} \left\{ c_1 (\boldsymbol{\Lambda}_d^{1/2} \otimes \boldsymbol{\Lambda}_d^{1/2}) (\mathbf{Z}_d^\top \mathbf{Z}_d \otimes \mathbf{Z}_d^\top \mathbf{Z}_d) (\boldsymbol{\Lambda}_d^{1/2} \otimes \boldsymbol{\Lambda}_d^{1/2}) (\mathbf{I}_{d^2} + \mathbf{K}_d) \right. \\ &\quad \left. + 2d_1 \text{vec}(\boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \boldsymbol{\Lambda}_d^{1/2}) \text{vec}^\top(\boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \boldsymbol{\Lambda}_d^{1/2}) \right\} = (III) + (IV). \end{aligned}$$

Para estudiar (III), necesitamos calcular  $\mathbb{E}_{\mathbf{Z}_d}(\mathbf{Z}_d^\top \mathbf{Z}_d \otimes \mathbf{Z}_d^\top \mathbf{Z}_d)$ . Llamemos  $\mathbf{S} = \mathbf{Z}_d^\top \mathbf{Z}_d$ . Recordemos que  $\mathbf{S} \sim \mathcal{W}_d(m, \mathbf{I}_d)$ . Por lo tanto  $\mathbb{E}(\mathbf{S}) = m\mathbf{I}_d$  y  $\text{var}(\text{vec}(\mathbf{S})) = m(\mathbf{I}_{d^2} + \mathbf{K}_d)$ . Siguiendo a [15] (ver Proposición A.11), tenemos

$$\begin{aligned} \mathbb{E}(\mathbf{S} \otimes \mathbf{S}) &= m \sum_{i,j} (E_{ij}^\top \otimes \mathbf{I}_d) (\mathbf{I}_{d^2} + \mathbf{K}_d) (\mathbf{I}_d \otimes E_{ij}^\top) + m^2 \mathbf{I}_{d^2} \\ &= m \sum_{i,j} (E_{ij}^\top \otimes E_{ij}^\top) + m \sum_{i,j} (E_{ij}^\top E_{ij}^\top \otimes \mathbf{I}_d) \mathbf{K}_d + m^2 \mathbf{I}_{d^2}, \end{aligned}$$

donde  $E_{ij} = e_i e_j^\top$  es la base canónica para  $\mathbb{R}^{d \times d}$ . Claramente,  $\sum_i E_{ii} = \sum_i e_i e_i^\top = \mathbf{I}_d$ . Además, se verifica que  $\sum_{i,j} (E_{ij} \otimes E_{ij}) = \text{vec}(\mathbf{I}_d) \text{vec}^\top(\mathbf{I}_d)$  y  $E_{ij}^\top E_{ij} = e_j e_i^\top e_j e_i^\top = \begin{cases} \mathbf{0} & \text{si } i \neq j \\ e_i e_i^\top & \text{si } i = j \end{cases}$ .

Luego, tenemos

$$\begin{aligned} \sum_{i,j} (E_{ij}^\top E_{ij} \otimes \mathbf{I}_d) \mathbf{K}_d &= \sum_i (e_i e_i^\top \otimes \mathbf{I}_d) \mathbf{K}_d \\ &= \left( \sum_i e_i e_i^\top \otimes \mathbf{I}_d \right) \mathbf{K}_d \\ &= (\mathbf{I}_d \otimes \mathbf{I}_d) \mathbf{K}_d \\ &= \mathbf{K}_d. \end{aligned}$$

De este modo resulta

$$\mathbb{E}(\mathbf{S} \otimes \mathbf{S}) = m \text{vec}(\mathbf{I}_d) \text{vec}^\top(\mathbf{I}_d) + m(m\mathbf{I}_{d^2} + \mathbf{K}_d).$$

Por lo tanto,

$$\begin{aligned} (III) &= \frac{c_1}{\sigma^4} (\boldsymbol{\Lambda}_d^{1/2} \otimes \boldsymbol{\Lambda}_d^{1/2}) \mathbb{E}_{\mathbf{Z}_d} (\mathbf{Z}_d^\top \mathbf{Z}_d \otimes \mathbf{Z}_d^\top \mathbf{Z}_d) (\boldsymbol{\Lambda}_d^{1/2} \otimes \boldsymbol{\Lambda}_d^{1/2}) (\mathbf{I}_{d^2} + \mathbf{K}_d) \\ &= \frac{c_1 m}{\sigma^4} (\boldsymbol{\Lambda}_d^{1/2} \otimes \boldsymbol{\Lambda}_d^{1/2}) [\text{vec}(\mathbf{I}_d) \text{vec}^\top(\mathbf{I}_d) + (m\mathbf{I}_{d^2} + \mathbf{K}_d)] (\boldsymbol{\Lambda}_d^{1/2} \otimes \boldsymbol{\Lambda}_d^{1/2}) (\mathbf{I}_{d^2} + \mathbf{K}_d) \\ &= \mathcal{O}\left(\frac{1}{m}\right), \end{aligned} \tag{C.24}$$

donde usamos que  $c_1 \approx \frac{1}{m^3}$  (ya que  $y > 1$ ).

Veamos ahora (IV). Necesitamos calcular  $\frac{1}{\sigma^4} \mathbb{E}_{\mathbf{Z}_d} (2d_1 \text{vec}(\boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \boldsymbol{\Lambda}_d^{1/2}) \text{vec}^\top(\boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \boldsymbol{\Lambda}_d^{1/2}))$ . Llamemos  $\mathbf{A} = (\boldsymbol{\Lambda}_d^{1/2} \otimes \boldsymbol{\Lambda}_d^{1/2})$  y  $\mathbf{S} = \mathbf{Z}_d^\top \mathbf{Z}_d \sim \mathcal{W}_d(m, \mathbf{I}_d)$ . Entonces, podemos escribir  $\text{vec}(\boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \mathbf{Z}_d \boldsymbol{\Lambda}_d^{1/2}) = \mathbf{A} \text{vec}(\mathbf{S})$ . Así, tenemos

$$\begin{aligned} \frac{2d_1}{\sigma^4} \mathbb{E}_{\mathbf{Z}_d} [\mathbf{A} \text{vec}(\mathbf{S}) \text{vec}^\top(\mathbf{S}) \mathbf{A}] &= \frac{2d_1}{\sigma^4} \mathbf{A} \mathbb{E}_{\mathbf{Z}_d} [\text{vec}(\mathbf{S}) \text{vec}^\top(\mathbf{S})] \mathbf{A} \\ &= \frac{2d_1}{\sigma^4} \mathbf{A} \{ \text{var}_{\mathbf{Z}_d}[\text{vec}(\mathbf{S})] + \mathbb{E}_{\mathbf{Z}_d}[\text{vec}(\mathbf{S})] \mathbb{E}_{\mathbf{Z}_d}^\top[\text{vec}(\mathbf{S})] \} \mathbf{A} \\ &= \frac{2d_1}{\sigma^4} \mathbf{A} \{ m(\mathbf{I}_{d^2} + \mathbf{K}_d) + m^2 \text{vec}(\mathbf{I}_d) \text{vec}^\top(\mathbf{I}_d) \} \mathbf{A} \\ &= \mathcal{O}\left(\frac{1}{m^2}\right), \end{aligned} \tag{C.25}$$

pues  $d_1 \approx \frac{1}{m^4}$  (ya que  $y > 1$ ). De (C.24) y (C.25) tenemos que (I) =  $\mathcal{O}\left(\frac{1}{m}\right)$ .

Calculemos ahora (II) de (C.23).

$$\begin{aligned}
\frac{1}{\sigma^4} \text{var}_{\mathbf{Z}_d} \left\{ \mathbb{E} \left[ \mathbf{Mvec}(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top)^{-1} | \mathbf{Z}_d \right] \right\} &= \frac{1}{\sigma^4} \text{var}_{\mathbf{Z}_d} \left\{ \mathbf{Mvec}[\mathbb{E}(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top)^{-1}] \right\} \\
&= \frac{1}{\sigma^4} \text{var}_{\mathbf{Z}_d} [\mathbf{Mvec}(a_1 \mathbf{I}_m)] \\
&= \frac{a_1^2}{\sigma^4} \text{var}_{\mathbf{Z}_d} [(\boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top \otimes \boldsymbol{\Lambda}_d^{1/2} \mathbf{Z}_d^\top) \text{vec}(\mathbf{I}_m)] \\
&= \frac{a_1^2}{\sigma^4} \text{var}_{\mathbf{Z}_d} [\mathbf{Avec}(\mathbf{Z}_d^\top \mathbf{Z}_d)] \\
&= \frac{a_1^2}{\sigma^4} \mathbf{A} \text{var}_{\mathbf{Z}_d} [\text{vec}(\mathbf{Z}_d^\top \mathbf{Z}_d)] \mathbf{A}^\top \\
&= \frac{a_1^2 m}{\sigma^4} \mathbf{A} (\mathbf{I}_{d^2} + \mathbf{K}_d) \mathbf{A}^\top = \mathcal{O} \left( \frac{1}{m} \right),
\end{aligned}$$

dado que  $a_1 \approx \frac{1}{m}$  (ya que  $y > 1$ ). De esta manera, tenemos (C.21).

Finalmente, observemos que

$$\left| \mathbf{I}_d + \frac{1}{\sigma^2} \frac{m}{p-d-m-1} \boldsymbol{\Lambda}_d \right| = \prod_{i=1}^h \left( 1 + \frac{\lambda_i}{\sigma^2} \frac{m}{p-d-m-1} \right)^{s_i}.$$

Así, (C.19) se sigue del hecho de que el determinante y el logaritmo son funciones continuas.

□

## Demostraciones del Capítulo 3

### D.1 Demostración de la Proposición 3.1

En primer lugar, notemos que

$$\begin{aligned} \mathbb{P}\left(\frac{\log LRT_d - \hat{\mu}_{m,p,d}}{\sigma_{m,p,d}} < z_\alpha\right) &= \mathbb{P}\left(\frac{\log LRT_d - \mu_{m,p,d}}{\sigma_{m,p,d}} + \frac{\mu_{m,p,d} - \hat{\mu}_{m,p,d}}{\sigma_{m,p,d}} < z_\alpha\right) \\ &= \mathbb{P}\left(\frac{\log LRT_d - \mu_{m,p,d}}{\sigma_{m,p,d}} < z_\alpha + \frac{\hat{\mu}_{m,p,d} - \mu_{m,p,d}}{\sigma_{m,p,d}}\right) \end{aligned}$$

y por lo tanto, la Proporsición 3.1 será consecuencia de las Proposiciones 2.4 y 2.5 si probamos que

$$\frac{\hat{\mu}_{m,p,d} - \mu_{m,p,d}}{\sigma_{m,p,d}} \rightarrow L \quad \text{cuando } p, m \rightarrow \infty. \quad (\text{D.1})$$

En lo que sigue, utilizaremos repetidas veces que  $\tilde{\lambda}_i \rightarrow \lambda_i$  cuando  $i \in J_1$  y  $\tilde{\lambda}_i \rightarrow \sigma^2(1 + \sqrt{y})$  para  $i \in J_2$  (Recordemos que hemos definido  $J_1 = \{i \leq h : \lambda_i > \sigma^2(1 + \sqrt{y})\}$ ,  $J_2 = \{i \leq h : \lambda_i \in (\sigma^2, \sigma^2(1 + \sqrt{y})]\}$ ) (Ver Lema B.5).

Consideremos por separado los casos  $p < m$ ,  $p > m + d + 3$  e  $y = 1$ .

#### D.1.1 Caso $p < m$

Como mencionamos arriba, necesitamos calcular el límite cuando  $p$  y  $m$  tienden a infinito de  $\frac{\hat{\mu}_{m,p,d} - \mu_{m,p,d}}{\sigma_{m,p,d}}$ . Comencemos pues, analizando la diferencia  $\hat{\mu}_{m,p,d} - \mu_{m,p,d}$ . Por definición,

tenemos

$$\begin{aligned}\hat{\mu}_{m,p,d} - \mu_{m,p,d} &= \tilde{\mu}_{m,p} + \log \hat{A}_{m,p,d} + \log \hat{B}_{m,p,d} - \tilde{\mu}_{m,p} - \log A_{m,p,d} - \log B_{m,p,d} \\ &= \log \hat{A}_{m,p,d} - \log A_{m,p,d} + \log \hat{B}_{m,p,d} - \log B_{m,p,d}.\end{aligned}$$

Por otra parte, la diferencia entre los términos  $\log(A_{m,p,d})$  y  $\log(\hat{A}_{m,p,d})$  converge a

$$\sum_{i \in J_2} s_i \log [\sigma^2(1 + \sqrt{y})] - \sum_{i \in J_2} s_i \log \lambda_i. \quad (\text{D.2})$$

En efecto,

$$\begin{aligned}\log(\hat{A}_{m,p,d}) - \log(A_{m,p,d}) &= \sum_{i=1}^d \log \tilde{\lambda}_i - \sum_{i=1}^d \log l_i - \sum_{i=1}^h s_i \log \lambda_i + \sum_{i=1}^d \log l_i \\ &= \sum_{i=1}^d \log \tilde{\lambda}_i - \sum_{i=1}^h s_i \log \lambda_i \\ &\rightarrow \sum_{i \in J_1} s_i \log \lambda_i + \sum_{i \in J_2} s_i \log [\sigma^2(1 + \sqrt{y})] \\ &\quad - \left( \sum_{i \in J_1} s_i \log \lambda_i + \sum_{i \in J_2} s_i \log \lambda_i \right) \\ &= \sum_{i \in J_2} s_i \log [\sigma^2(1 + \sqrt{y})] - \sum_{i \in J_2} s_i \log \lambda_i.\end{aligned}$$

Finalmente, para comparar los términos en  $\log(B_{m,p,d})$  para  $\hat{\mu}_{m,p,d}$  y  $\mu_{m,p,d}$ , notemos primero que  $\log(B_{m,p,d})$  de la Proposición 2.4 es asintóticamente equivalente a

$$\tilde{B}_{m,p,d} = y \sum_{i \in J_1} s_i \frac{\lambda_i}{\lambda_i - \sigma^2} + \sum_{i \in J_2} s_i (1 + \sqrt{y})^2 - \sum_{i \in J_2} s_i \frac{\lambda_i}{\sigma^2}.$$

En efecto, dado que  $(p-d) \log\{1 + a_{p,m}/(p-d)\} = a + o(1)$  cuando  $p-d \rightarrow \infty$  si  $a = \lim a_{p,m}$  es finito (ver Lema B.3), es suficiente probar que

$$\log(B_{m,p,d}) - (p-d) \log \left( 1 + \frac{1}{p-d} \tilde{B}_{m,p,d} \right) \rightarrow 0$$

cuando  $p$  y  $m$  tienen a infinito. Pero por definición de  $\log(B_{m,p,d})$  y  $\tilde{B}_{m,p,d}$  tenemos

$$\begin{aligned} & \log(B_{m,p,d}) - (p-d) \log\left(1 + \frac{1}{p-d} \tilde{B}_{m,p,d}\right) = \\ &= (p-d) \log\left(1 + \frac{\sum_{i=1}^d l_i - \sum_{i=1}^h s_i \lambda_i}{\sum_{i=d+1}^p l_i}\right) \\ & \quad - (p-d) \log\left[1 + \frac{1}{p-d} \left\{y \sum_{i \in J_1} s_i \frac{\lambda_i}{\lambda_i - \sigma^2} + \sum_{i \in J_2} s_i (1 + \sqrt{y})^2 - \sum_{i \in J_2} s_i \frac{\lambda_i}{\sigma^2}\right\}\right] \\ &= (p-d) \log\left(1 + \frac{1}{p-d} H\right), \end{aligned}$$

donde

$$H = \frac{-\left\{y \sum_{i \in J_1} s_i \frac{\lambda_i}{\lambda_i - \sigma^2} + \sum_{i \in J_2} s_i (1 + \sqrt{y})^2 - \sum_{i \in J_2} s_i \frac{\lambda_i}{\sigma^2}\right\} + \frac{\sum_{i=1}^d l_i - \sum_{i=1}^h s_i \lambda_i}{\frac{1}{p-d} \sum_{i=d+1}^p l_i}}{1 + \frac{1}{p-d} \left\{y \sum_{i \in J_1} s_i \frac{\lambda_i}{\lambda_i - \sigma^2} + \sum_{i \in J_2} s_i (1 + \sqrt{y})^2 - \sum_{i \in J_2} s_i \frac{\lambda_i}{\sigma^2}\right\}}.$$

Así, la demostración de la afirmación se sigue si probamos que  $H \rightarrow 0$ . En efecto, usando [7], sabemos que

$$\sum_{i=1}^d l_i = \sum_{i \in J_1} l_i + \sum_{i \in J_2} l_i \rightarrow \sum_{i \in J_1} s_i \lambda_i \left(1 + \frac{y\sigma^2}{\lambda_i - \sigma^2}\right) + \sum_{i \in J_2} s_i \sigma^2 (1 + \sqrt{y})^2.$$

Por otro lado, por Lema B.2,

$$\frac{1}{p-d} \sum_{i=d+1}^p l_i \rightarrow \sigma^2.$$

Luego,

$$\begin{aligned} \frac{\sum_{i=1}^d l_i - \sum_{i=1}^h s_i \lambda_i}{\frac{1}{p-d} \sum_{i=d+1}^p l_i} &\rightarrow \sum_{i \in J_1} s_i \frac{\lambda_i}{\sigma^2} \left(1 + \frac{y\sigma^2}{\lambda_i - \sigma^2}\right) + \sum_{i \in J_2} s_i (1 + \sqrt{y})^2 - \sum_{i \in J_1} s_i \frac{\lambda_i}{\sigma^2} - \sum_{i \in J_2} s_i \frac{\lambda_i}{\sigma^2} \\ &= -\sum_{i \in J_2} s_i \frac{\lambda_i}{\sigma^2} + y \sum_{i \in J_1} s_i \frac{\lambda_i}{\lambda_i - \sigma^2} + \sum_{i \in J_2} s_i (1 + \sqrt{y})^2. \end{aligned}$$

Por lo tanto, el numerador tiende a 0 y claramente el denominador tiende a 1.

Ahora bien, la diferencia entre los términos de  $\tilde{B}_{m,p,d}$  correspondientes a  $\hat{\mu}_{m,p,d}$  y  $\mu_{m,p,d}$  están en los términos que involucran  $J_2$ , los cuales convergen a

$$-\sum_{i \in J_2} s_i (1 + \sqrt{y}) + \sum_{i \in J_2} s_i \frac{\lambda_i}{\sigma^2}. \quad (\text{D.3})$$

Finalmente, dado que  $\sigma_{m,p,d} \rightarrow \sqrt{-2\{y + \log(1-y)\}}$ , de las ecuaciones (D.2) y (D.3) obtenemos (D.1).

**D.1.2 Caso  $p > m + d + 3$** 

En este caso, en la Proposición 2.5  $\log(B_{m,p,d})$  puede reemplazarse por

$$\sum_{i \in J_1} s_i \frac{\lambda_i}{\lambda_i - \sigma^2} + \sum_{i \in J_2} s_i \frac{(1 + \sqrt{y})^2}{y} - \sum_{i \in J_2} s_i \frac{\lambda_i}{y\sigma^2} + (m - d) \log \left( \frac{m - d}{m} \right) \quad (\text{D.4})$$

y  $\log(D_{m,p,d})$  por

$$\sum_{i=1}^h s_i \log \left\{ 1 + \frac{\lambda_i}{\sigma^2(y-1)} \right\} - \sum_{i \in J_1} s_i \log \left\{ \lambda_i \left( 1 + \frac{y\sigma^2}{\lambda_i - \sigma^2} \right) \right\} - \sum_{i \in J_2} s_i \log [\sigma^2(1 + \sqrt{y})^2] \quad (\text{D.5})$$

con  $J_1$  y  $J_2$  como antes. Las pruebas de estas observaciones son completamente análogas a las del caso previo. De esto, y ya que los otros términos no involucran a  $\lambda_i$ ,

$$\hat{\mu}_{m,p,d} - \mu_{m,p,d} \rightarrow \sum_{i \in J_2} s_i \left\{ \frac{\lambda_i}{y\sigma^2} - \frac{1 + \sqrt{y}}{y} - \log \frac{\sigma^2(y-1) + \lambda_i}{\sqrt{y}\sigma^2(\sqrt{y}+1)} \right\}$$

y  $\sigma_{m,p,d}$  tiende a  $\sqrt{-2 \{1/y + \log(1 - 1/y)\}} < \infty$  cuando ambos  $p$  y  $m$  tienden a infinito.

**D.1.3 Caso  $y = 1$** 

Si  $y = 1$  tenemos que  $L = 0$ . En efecto, si  $p/m \rightarrow 1$ ,  $\sigma_{m,p,d} \rightarrow \infty$  mientras que los términos en las sumatorias permanecen sin cambios. Por lo tanto  $L \rightarrow 0$  si  $p/m \rightarrow 1$ . En consecuencia, cuando  $y = 1$  tenemos que el nivel asintótico del test 3.1 es igual a  $\alpha$ .

□

## Demostraciones del Capítulo 4

### E.1 Demostración de la Proposición 4.1

A los efectos de facilitar la lectura, comencemos recordando que hemos definido

$$\begin{aligned} J_{1,k} &= \{i \leq h_k : \lambda_i > \sigma^2(1 + \sqrt{y})\} \quad \text{para } k = 0, 1, \\ J_{2,k} &= \{i \leq h_k : \lambda_i \in (\sigma^2, \sigma^2(1 + \sqrt{y})]\} \quad \text{para } k = 0, 1, \\ k_i &= \begin{cases} \frac{\lambda_i}{\sigma^2} \left(1 + \frac{y\sigma^2}{\lambda_i - \sigma^2}\right) & \text{si } i \in J_{1,k}, \\ (1 + \sqrt{y})^2 & \text{si } i \in J_{2,k} \text{ o cuando } \lambda_i = \sigma^2, \end{cases} \end{aligned}$$

y llamemos  $r = \min(m, p)$ . En lo que sigue, cada vez que escribamos  $d > d_1$ , recordar que estamos suponiendo  $d_1 < d \leq q_0$  para  $q_0 \ll \min(p, m)$ , independiente de  $p$  y  $m$ .

Recordando la definición de  $LRT_d$  y considerando los casos  $d < d_1$  y  $d > d_1$ , podemos escribir

$$\begin{aligned} LRT_d &= \frac{\prod_{i=d+1}^r l_i}{\left(\frac{1}{r-d} \sum_{i=d+1}^r l_i\right)} \\ &= LRT_{d_1} \prod_{i=d+1}^{d_1} l_i \frac{\left(\frac{1}{r-d_1} \sum_{i=d_1+1}^r l_i\right)^{r-d_1}}{\left(\frac{1}{r-d} \sum_{i=d+1}^r l_i\right)^{r-d}} \quad \text{si } d < d_1, \end{aligned} \tag{E.1}$$

$$= LRT_{d_1} \frac{1}{\prod_{i=d_1+1}^d l_i} \frac{\left(\frac{1}{r-d_1} \sum_{i=d_1+1}^r l_i\right)^{r-d_1}}{\left(\frac{1}{r-d} \sum_{i=d+1}^r l_i\right)^{r-d}} \quad \text{si } d > d_1. \tag{E.2}$$

Usando la Proposición 2.4, sabemos que, si la verdadera dimensión es  $d_1$ ,

$$\frac{\log(LRT_{d_1}) - \mu_{m,p,d_1}}{\sigma_{m,p,d_1}} \rightsquigarrow \mathcal{N}(0, 1). \quad (\text{E.3})$$

Ahora, dado que para  $i \in J_{1,k}$ ,  $l_i \rightarrow \lambda_i \{1 + y\sigma^2/(\lambda_i - \sigma^2)\}$  y para  $i \in J_{2,k}$ ,  $l_i \rightarrow \sigma^2(1 + \sqrt{y})^2$  cuando  $p, m \rightarrow \infty$  y  $p/m \rightarrow y$ , resulta

$$\log(l_{d+1} \cdots l_{d_1}) - \sum_{i=h_0+1}^{h_1} s_i \log(\sigma^2 k_i) \rightarrow 0 \quad \text{si } d < d_1. \quad (\text{E.4})$$

Por otro lado, notemos que en el caso en que la verdadera dimensión es  $d_1$  y  $d > d_1$ , todos los autovalores poblacionales  $\lambda_i$  para  $i = d_1 + 1, \dots, d$  son iguales a  $\sigma^2$ . En consecuencia, y ya que estamos asumiendo la Condición  $Q_0$ , esto es,  $d \leq q_0$  para  $q_0$  fijo e independiente de  $p$  y  $m$ , los correspondientes autovalores muestrales convergen a  $\sigma^2(1 + \sqrt{y})^2$  cuando  $p, m \rightarrow \infty$  y  $p/m \rightarrow y$  ([7]). Por lo tanto,

$$\log(l_{d+1} \cdots l_{d_1}) - (d - d_1) \log[\sigma^2(1 + \sqrt{y})^2] \rightarrow 0 \quad \text{cuando } d > d_1. \quad (\text{E.5})$$

Por último,

$$\frac{\left(\frac{1}{r-d_1} \sum_{i=d_1+1}^r l_i\right)^{r-d_1}}{\left(\frac{1}{r-d} \sum_{i=d+1}^r l_i\right)^{r-d}} = ABC, \quad (\text{E.6})$$

con

$$A = \left(\frac{\sum_{i=d_1+1}^r l_i}{\sum_{i=d+1}^r l_i}\right)^{r-d} \rightarrow \begin{cases} \exp\left\{-\frac{1}{\tilde{y}} \sum_{i=h_0+1}^{h_1} s_i k_i\right\} & \text{si } d < d_1, \\ \exp\left\{\frac{1}{\tilde{y}}(d - d_1)(1 + \sqrt{y})^2\right\} & \text{si } d > d_1, \end{cases} \quad (\text{E.7})$$

$$B = \left(\frac{1}{r-d_1} \sum_{i=d_1+1}^r l_i\right)^{d-d_1} \rightarrow (\tilde{y} \sigma^2)^{d-d_1}, \quad (\text{E.8})$$

$$C = \left(\frac{r-d}{r-d_1}\right)^{r-d} \rightarrow \exp\{d_1 - d\}, \quad (\text{E.9})$$

donde  $\tilde{y} = \max\{1, y\}$  (es decir que  $\tilde{y} = 1$  si  $p < m$  mientras que  $\tilde{y} = y$  si  $p > m + d_1 + 3$ ). En efecto, por un lado tenemos

$$A = \left(\frac{\sum_{i=d_1+1}^r l_i}{\sum_{i=d+1}^r l_i}\right)^{r-d} = \begin{cases} \left(1 - \frac{1}{r-d} \frac{\sum_{i=d_1+1}^{d_1} l_i}{\frac{1}{r-d} \sum_{i=d+1}^r l_i}\right)^{r-d} & \text{si } d < d_1, \\ \left(1 + \frac{1}{r-d} \frac{\sum_{i=d_1+1}^d l_i}{\frac{1}{r-d} \sum_{i=d+1}^r l_i}\right)^{r-d} & \text{si } d > d_1. \end{cases}$$

Entonces, si llamamos

$$a_r = \begin{cases} \frac{-\sum_{i=d+1}^{d_1} l_i}{\frac{1}{r-d} \sum_{i=d+1}^r l_i} \rightarrow -\frac{1}{\tilde{y}} \sum_{i=h_0+1}^{h_1} s_i k_i & \text{si } d < d_1, \\ \frac{\sum_{i=d+1}^d l_i}{\frac{1}{r-d} \sum_{i=d+1}^r l_i} \rightarrow \frac{1}{\tilde{y}} (d-d_1)(1+\sqrt{y})^2 & \text{si } d > d_1, \end{cases}$$

por el Lema B.3 obtenemos (E.7).

Para ver (E.8) vamos a utilizar el Lema B.2.

$$\begin{aligned} B &= \left( \frac{1}{r-d_1} \sum_{i=d_1+1}^r l_i \right)^{d-d_1} \\ &= \left( \frac{p-d_1}{r-d_1} \frac{1}{p-d_1} \sum_{i=d_1+1}^r l_i \right)^{d-d_1} \\ &\rightarrow (\tilde{y} \sigma^2)^{d-d_1}. \end{aligned}$$

Finalmente, (E.9) se obtiene aplicando nuevamente el Lema B.3. En efecto,

$$\begin{aligned} C &= \left( \frac{r-d}{r-d_1} \right)^{r-d} \\ &= \left( 1 + \frac{d_1-d}{r-d_1} \right)^{(r-d_1) \frac{r-d}{r-d_1}} \\ &\rightarrow \exp\{d_1-d\}. \end{aligned}$$

Por lo tanto, dado que  $\lim_{p,m \rightarrow \infty} \sigma_{m,p,d_1} \neq 0$ , aplicando (E.3)–(E.9) a (E.1) para el caso  $d < d_1$ , tenemos

$$\begin{aligned} \frac{\log(LRT_d) - \mu_{m,p,d,d_1}}{\sigma_{m,p,d_1}} &= \frac{\log(LRT_d) - \mu_{m,p,d_1} - \sum_{i=h_0+1}^{h_1} s_i (\log \frac{k_i}{\tilde{y}} - \frac{k_i}{\tilde{y}} + 1)}{\sigma_{m,p,d_1}} \\ &= \frac{\log(LRT_{d_1}) + \sum_{i=d+1}^{d_1} \log(l_i) + \log(A) + \log(B) + \log(C)}{\sigma_{m,p,d_1}} \\ &\quad - \frac{\mu_{m,p,d_1} - \sum_{i=h_0+1}^{h_1} s_i (\log \frac{k_i}{\tilde{y}} - \frac{k_i}{\tilde{y}} + 1)}{\sigma_{m,p,d_1}} \\ &= \frac{\log(LRT_{d_1}) - \mu_{m,p,d_1}}{\sigma_{m,p,d_1}} + \frac{\sum_{i=d+1}^{d_1} \log(l_i) + \log(B) - \sum_{i=h_0+1}^{h_1} s_i \log \frac{k_i}{\tilde{y}}}{\sigma_{m,p,d_1}} \\ &\quad + \frac{\log(A) + \sum_{i=h_0+1}^{h_1} s_i \frac{k_i}{\tilde{y}}}{\sigma_{m,p,d_1}} + \frac{\log(C) - \sum_{i=h_0+1}^{h_1} s_i}{\sigma_{m,p,d_1}} \\ &\rightsquigarrow \mathcal{N}(0, 1), \end{aligned}$$

que es lo que queríamos demostrar.

Análogamente, aplicando (E.3)–(E.9) a (E.2) obtenemos el caso  $d > d_1$ .

□

## E.2 Demostración de la Proposición 4.2

Recordemos que el test que hemos definido es secuencial; es decir, testeamos secuencialmente las hipótesis  $\mathcal{H}_d$ : la dimensión del subespacio spiked es  $d$  versus  $\mathcal{H}_a$ : la dimensión del subespacio spiked es mayor que  $d$  para  $d = 0, \dots, p-2$ , donde  $\lambda_1, \dots, \lambda_p$  son los autovalores de  $\Sigma$ . Por tal motivo, y de acuerdo con la región de rechazo definida en 3.1, para calcular la potencia del test necesitamos calcular

$$\mathbb{P}(\log LRT_d < z_\alpha \sigma_{m,p,d} + \hat{\mu}_{m,p,d} | d < d_1)$$

donde la notación  $\mathbb{P}(\cdot | d < d_1)$  significa la probabilidad condicionada a que *la verdadera dimensión es  $d_1 > d$* .

Ahora bien, dado que todos los autovalores spike son mayores que el umbral, es equivalente (asintóticamente) utilizar  $\hat{\mu}_{m,p,d}$  o  $\mu_{m,p,d}$ . Usando la Proposición 4.1, tenemos:

a) Caso  $p < m$ :

$$\begin{aligned} \Psi(d_1) &= \mathbb{P}\{\log(LRT_d) < \mu_{m,p,d} + z_\alpha \sigma_{m,p,d} | d < d_1\} \\ &= \mathbb{P}\left(Z < \frac{\mu_{m,p,d} - \mu_{m,p,d,d_1} + z_\alpha \sigma_{m,p,d}}{\sigma_{m,p,d_1}}\right) \\ &= \mathbb{P}\left\{Z < \frac{\mu_{m,p,d} - \mu_{m,p,d_1} + \sum_{i=h_0+1}^{h_1} s_i(k_i - \log k_i - 1) + z_\alpha \sigma_{m,p,d}}{\sigma_{m,p,d_1}}\right\}. \end{aligned} \quad (\text{E.10})$$

b) Caso  $p > m + d_1 + 3$ :

$$\begin{aligned} \Psi(d_1) &= \mathbb{P}\{\log(LRT_d) < \mu_{m,p,d} + z_\alpha \sigma_{m,p,d} | d < d_1\} \\ &= \mathbb{P}\left(Z < \frac{\mu_{m,p,d} - \mu_{m,p,d,d_1} + z_\alpha \sigma_{m,p,d}}{\sigma_{m,p,d_1}}\right) \\ &= \mathbb{P}\left\{Z < \frac{\mu_{m,p,d} - \mu_{m,p,d_1} + \sum_{i=h_0+1}^{h_1} s_i\left(\frac{k_i}{y} - \log \frac{k_i}{y} - 1\right) + z_\alpha \sigma_{m,p,d}}{\sigma_{m,p,d_1}}\right\}, \end{aligned} \quad (\text{E.11})$$

donde  $Z \sim \mathcal{N}(0, 1)$ .

Veamos primero el caso  $p < m$ . Por la definición de  $\mu_{m,p,d}$  resulta

$$\mu_{m,p,d} - \mu_{m,p,d_1} = \log A_{m,p,d} + \log B_{m,p,d} - \log A_{m,p,d_1} - \log B_{m,p,d_1}.$$

Por un lado tenemos

$$\begin{aligned}
\log A_{m,p,d} - \log A_{m,p,d_1} &= \sum_{i=d+1}^{d_1} \log l_i - \sum_{i=h_0+1}^{h_1} s_i \log \lambda_i \\
&\rightarrow \sum_{i=h_0+1}^{h_1} s_i \log(\sigma^2 k_i) - \sum_{i=h_0+1}^{h_1} s_i \log \lambda_i \\
&= \sum_{i=h_0+1}^{h_1} s_i \left( \log k_i - \log \frac{\lambda_i}{\sigma^2} \right).
\end{aligned}$$

Por otra parte, desde la demostración de la Proposición 3.1 sabemos que  $\log B_{m,p,d}$  es asintóticamente equivalente a

$$\begin{aligned}
\tilde{B}_{m,p,d} &= y \sum_{i \in J_1} s_i \frac{\lambda_i}{\lambda_i - \sigma^2} + \sum_{i \in J_2} s_i (1 + \sqrt{y})^2 - \sum_{i \in J_2} s_i \frac{\lambda_i}{\sigma^2} \\
&= \sum_{i=1}^{h_0} s_i \left( k_i - \frac{\lambda_i}{\sigma^2} \right).
\end{aligned}$$

Por lo tanto, aplicando la equivalencia anterior a  $\log B_{m,p,d}$  y  $\log B_{m,p,d_1}$  respectivamente, resulta

$$\log B_{m,p,d} - \log B_{m,p,d_1} \sim - \sum_{i=h_0+1}^{h_1} s_i \left( k_i - \frac{\lambda_i}{\sigma^2} \right),$$

donde  $\sim$  significa *asintóticamente equivalente*.

En consecuencia,  $\mu_{m,p,d} - \mu_{m,p,d_1}$  converge a

$$- \sum_{i=h_0+1}^{h_1} s_i \left[ \log \left( \frac{\lambda_i}{\sigma^2} \right) - \frac{\lambda_i}{\sigma^2} \right] - \sum_{i=h_0+1}^{h_1} s_i (k_i - \log k_i). \quad (\text{E.12})$$

Reemplazando (E.12) en (E.10) (asintóticamente) se obtiene (4.4).

Veamos ahora el caso  $p > m + d_1 + 3$ . De acuerdo con la Proposición 2.5,

$$\begin{aligned}
\mu_{m,p,d} - \mu_{m,p,d_1} &= \tilde{\mu}_{m,p,d} - \tilde{\mu}_{m,p,d_1} + \log B_{m,p,d} + \log C_{m,p,d} \\
&\quad + \log D_{m,p,d} - \log B_{m,p,d_1} - \log C_{m,p,d_1} - \log D_{m,p,d_1}.
\end{aligned} \quad (\text{E.13})$$

Por un lado, tenemos

$$\begin{aligned}
\tilde{\mu}_{m,p,d} - \tilde{\mu}_{m,p,d_1} &= (p-m) \left[ \log \left( 1 - \frac{m}{p-d_1} \right) - \log \left( 1 - \frac{m}{p-d} \right) \right] \\
&\quad - (d_1 + \frac{1}{2}) \log \left( 1 - \frac{m}{p-d_1} \right) + (d + \frac{1}{2}) \log \left( 1 - \frac{m}{p-d} \right) \\
&= (p-d-m) \log \left[ 1 + \frac{1}{p-d-m} \frac{m(d-d_1)}{p-d_1} \right] + d \log \left[ 1 + \frac{1}{p-d-m} \frac{m(d-d_1)}{p-d_1} \right] \\
&\quad - (d_1 + \frac{1}{2}) \log \left( 1 - \frac{m}{p-d_1} \right) + (d + \frac{1}{2}) \log \left( 1 - \frac{m}{p-d} \right) \\
&\sim \frac{d-d_1}{y} + (d-d_1) \log \left( 1 - \frac{1}{y} \right). \tag{E.14}
\end{aligned}$$

Por otro lado, usando la equivalencia asintótica que vimos en la demostración de la Proposición 3.1, podemos reemplazar  $\log B_{m,p,d}$  por

$$\begin{aligned}
&\sum_{i \in J_1} s_i \frac{\lambda_i}{\lambda_i - \sigma^2} + \sum_{i \in J_2} s_i \frac{(1 + \sqrt{y})^2}{y} - \sum_{i \in J_2} s_i \frac{\lambda_i}{y\sigma^2} + (m-d) \log \left( \frac{m-d}{m} \right) \\
&= \sum_{i=1}^h s_i \left( \frac{k_i}{y} - \frac{\lambda_i}{y\sigma^2} \right) + (m-d) \log \left( \frac{m-d}{m} \right),
\end{aligned}$$

y  $\log(D_{m,p,d})$  por

$$\sum_{i=1}^h s_i \log \left\{ 1 + \frac{\lambda_i}{\sigma^2(y-1)} \right\} - \sum_{i=1}^h s_i \log(\sigma^2 k_i).$$

Por lo tanto,

$$\begin{aligned}
\log B_{m,p,d} - \log B_{m,p,d_1} &\sim \sum_{i=h_0+1}^{h_1} s_i \left( \frac{\lambda_i}{y\sigma^2} - \frac{k_i}{y} \right) + (m-d_1) \log \left( \frac{m-d}{m-d_1} \right) \\
&\quad + d_1 \log \left( \frac{m-d}{m-d_1} \right) - d \log \left( \frac{m-d}{m} \right) + d_1 \log \left( \frac{m-d_1}{m} \right) \\
&\sim \sum_{i=h_0+1}^{h_1} s_i \left( \frac{\lambda_i}{y\sigma^2} - \frac{k_i}{y} \right) + d_1 - d, \tag{E.15}
\end{aligned}$$

mientras que

$$\log D_{m,p,d} - \log D_{m,p,d_1} \sim - \sum_{i=h_0+1}^{h_1} s_i \left\{ \log \left[ 1 + \frac{\lambda_i}{\sigma^2(y-1)} \right] - \log(\sigma^2 k_i) \right\}.$$

Por otra parte,

$$\begin{aligned}
\log C_{m,p,d} - \log C_{m,p,d_1} &= (d-d_1) \log \frac{\sigma^2(p-d)}{m} - d_1 \log \frac{p-d_1}{p-d} \\
&\sim (d-d_1) \log(y\sigma^2). \tag{E.16}
\end{aligned}$$

Por lo tanto, aplicando las equivalencias (E.14)–(E.16) a (E.13) tenemos

$$\begin{aligned}
\mu_{m,p,d} - \mu_{m,p,d_1} &\sim \frac{d - d_1}{y} + (d - d_1) \log \left( 1 - \frac{1}{y} \right) + \sum_{i=h_0+1}^{h_1} s_i \left( \frac{\lambda_i}{y\sigma^2} - \frac{k_i}{y} \right) + d_1 - d \\
&\quad + (d - d_1) \log (y\sigma^2) - \sum_{i=h_0+1}^{h_1} s_i \left\{ \log \left[ 1 + \frac{\lambda_i}{\sigma^2(y-1)} \right] - \log(\sigma^2 k_i) \right\} \\
&= \sum_{i=h_0+1}^{h_1} s_i \left\{ -\frac{1}{y} - \log \left( 1 - \frac{1}{y} \right) + \frac{\lambda_i}{y\sigma^2} - \frac{k_i}{y} + 1 - \log (y\sigma^2) \right. \\
&\quad \left. - \log \left[ 1 + \frac{\lambda_i}{\sigma^2(y-1)} \right] + \log (\sigma^2 k_i) \right\}.
\end{aligned} \tag{E.17}$$

Finalmente, reemplazando con (E.17) en (E.11) obtenemos (4.5).

□



## Demostraciones del Capítulo 5

### F.1 Demostración de la Proposición 5.1

Antes de comenzar con la prueba de la proposición, tengamos presente que, por hipótesis, para todo  $i \in \{1, \dots, h_1\}$ ,  $\lambda_i > \lambda^* > \sigma^2(1 + \sqrt{y})$ . Por lo tanto, los conjuntos  $J_{2,1} = J_{2,0} = \emptyset$ . Recordemos también que hemos llamado  $\tilde{y} = \max(1, y)$ , de modo tal que  $\tilde{y} = 1$  si  $p < m$  mientras que  $\tilde{y} = y$  en el caso  $p > m + d_1 + 3$ . La función  $h(\lambda)$ , definida en (5.1), verifica:

1) Para  $\lambda > \sigma^2(1 + \sqrt{y})$ ,  $h(\lambda)$  es estrictamente creciente. (F.1)

2)  $h[\sigma^2(1 + \sqrt{y})] = \frac{(1 + \sqrt{y})^2}{\tilde{y}} - \log \frac{(1 + \sqrt{y})^2}{\tilde{y}} - 1$ . (F.2)

Para ver la primera afirmación, en primer lugar notemos que podemos escribir  $h(\lambda) = (m \circ b)(\lambda) = m(b(\lambda))$ , donde

$$\begin{aligned} m(\lambda) &= \lambda - \log(\lambda) - 1 \quad \text{y} \\ b(\lambda) &= \frac{\lambda}{\tilde{y}\sigma^2} \left( 1 + \frac{y\sigma^2}{\lambda - \sigma^2} \right). \end{aligned}$$

Tenemos,

- a)  $m$  es creciente si  $\lambda > 1$ .
- b)  $b$  es creciente si  $\lambda > \sigma^2(1 + \sqrt{y})$  o si  $\lambda < \sigma^2(1 - \sqrt{y})$ .

Luego, bajo la hipótesis de la Proposición 5.1 de que  $\lambda_i > \lambda^* > \sigma^2(1 + \sqrt{y})$ , tenemos que  $h(\lambda)$  es una función creciente (estricta).

Veamos entonces la prueba de la proposición. Usando (4.1), (4.2) y (F.2), tenemos que para  $d \geq d_1$ ,

$$g(d) = \mu_{m,p,d_1} - (d - d_1)\epsilon.$$

Por lo tanto  $g$  es una función decreciente para  $d \geq d_1$ . Por otro lado, para  $d < d_1$ , la función  $g(d)$  está dada por

$$\begin{aligned} g(d) &= \mu_{m,p,d_1} + \sum_{i=h_0+1}^{h_1} s_i \left( \log \frac{k_i}{\tilde{y}} - \frac{k_i}{\tilde{y}} + 1 \right) \\ &\quad - (d - d_1) \left[ \frac{(1 + \sqrt{y})^2}{\tilde{y}} - \log \frac{(1 + \sqrt{y})^2}{\tilde{y}} - 1 + \epsilon \right] \\ &= \mu_{m,p,d_1} - \sum_{i=h_0+1}^{h_1} s_i [h(\lambda_i) - h\{\sigma^2(1 + \sqrt{y})\} - \epsilon]. \end{aligned}$$

Usando ahora (F.1) y la definición de  $\epsilon$ ,  $h(\lambda_i) - h\{\sigma^2(1 + \sqrt{y})\} - \epsilon = h(\lambda_i) - h(\lambda^*) > 0$ . Por lo tanto, para  $d < d_1$ , cuando  $d$  crece estamos agregando menos términos negativos y en consecuencia  $g$  crece para  $d < d_1$ . De esto se deduce el resultado. □

## F.2 Demostración de la Proposición 5.3

La proposición se sigue si probamos que para todo  $\delta > 0$  existen  $m_0, p_{m_0}$  tales que para  $m > m_0, p > p_{m_0}$  y  $p/m$  cercano a  $y$

$$\mathbb{P} \left( \bigcap_{d=1}^{q_0} A_d \right) \geq 1 - \delta, \quad (\text{F.3})$$

donde

$$A_d = A_d(p, m, \epsilon) = \begin{cases} \{F(d, p, m, \epsilon) - F(d-1, p, m, \epsilon) > 0\} & \text{si } 0 \leq d \leq d_1, \\ \{F(d, p, m, \epsilon) - F(d-1, p, m, \epsilon) < 0\} & \text{para } d_1 < d \leq q_0, \end{cases} \quad (\text{F.4})$$

con  $F(d, p, m, \epsilon) = \log(LTR_d) - (d - d_1) \{h[\sigma^2(1 + \sqrt{y})] + \epsilon\}$ . En efecto, notemos que para  $\hat{g}(\lambda) = \log(LRT_d) - (d - d_1)[h(\lambda) + \epsilon]$ ,

$$\bigcap_{d=1}^{q_0} A_d \subseteq \{\hat{g}(\lambda) \text{ es creciente para } d \leq d_1 \text{ y decreciente para } d \geq d_1\}. \quad (\text{F.5})$$

Ciertamente, si calculamos la derivada de la función  $\hat{g}(\lambda)$ , tenemos

$$\hat{g}'(\lambda) = -(d - d_1)h'(\lambda).$$

Por lo tanto, recordando que  $h$  es creciente si  $\lambda > \sigma^2(1 + \sqrt{y})$  (F.2), resulta

$$\begin{cases} \hat{g}'(\lambda) > 0, & \text{si } d \leq d_1, \\ \hat{g}'(\lambda) < 0, & \text{si } d \geq d_1, \end{cases}$$

de donde se deduce (F.5).

Veamos entonces la prueba de (F.3). En lo que sigue, vamos a llamar nuevamente  $r = \min(m, p)$ . Consideremos primero  $d \leq d_1$ . Desde la definición de  $F(d, p, m, \epsilon)$ ,

$$F(d, p, m, \epsilon) - F(d - 1, p, m, \epsilon) = \log \left( \frac{LRT_d}{LRT_{d-1}} \right) - h[\sigma^2(1 + \sqrt{y})] - \epsilon.$$

Veamos ahora como podemos reescribir  $\log \left( \frac{LRT_d}{LRT_{d-1}} \right)$ .

$$\begin{aligned} \frac{LRT_d}{LRT_{d-1}} &= \frac{\prod_{i=d+1}^r l_i \left( \frac{1}{r-d+1} \sum_{i=d}^r l_i \right)^{r-d+1}}{\prod_{i=d}^r l_i \left( \frac{1}{r-d} \sum_{i=d+1}^r l_i \right)^{r-d}} \\ &= \frac{1}{l_d} \left( \frac{1}{r-d+1} \sum_{i=d+1}^r l_i \right) \left( \frac{r-d}{r-d+1} \right)^{r-d} \left( \frac{\sum_{i=d}^r l_i}{\sum_{i=d+1}^r l_i} \right)^{r-d} \\ &= \frac{1}{l_d} \left( \frac{1}{r-d+1} \sum_{i=d+1}^r l_i \right) \frac{1}{\left(1 + \frac{1}{r-d}\right)^{r-d}} \left( 1 + \frac{l_d}{\sum_{i=d+1}^r l_i} \right)^{r-d}. \end{aligned}$$

Usando nuevamente  $\tilde{y} = \min(1, y)$ , podemos deducir que

$$\frac{LRT_d}{LRT_{d-1}} \rightarrow \frac{\tilde{y}\sigma^2}{\lambda_d \left(1 + \frac{y\sigma^2}{\lambda_d - \sigma^2}\right)} \exp(-1) \exp \left\{ \frac{\lambda_d}{\tilde{y}\sigma^2} \left( 1 + \frac{y\sigma^2}{\lambda_d - \sigma^2} \right) \right\}. \quad (\text{F.6})$$

En efecto,

1. Por hipótesis,  $\lambda_d \geq \lambda_{d_1} > \lambda^* > \sigma^2(1 + \sqrt{y})$  y, por lo tanto,  $l_d \rightarrow \lambda_d \left( 1 + \frac{y\sigma^2}{\lambda_d - \sigma^2} \right)$ .
2.  $\left( 1 + \frac{1}{r-d} \right)^{r-d} \rightarrow e$ .
3. Utilizando el Lema B.3,  $\left( 1 + \frac{l_d}{\sum_{i=d+1}^r l_i} \right)^{r-d} \rightarrow \exp \left\{ \frac{\lambda_d}{\tilde{y}\sigma^2} \left( 1 + \frac{y\sigma^2}{\lambda_d - \sigma^2} \right) \right\}$ .
4. Por Lema B.2,  $\left( \frac{1}{r-d+1} \sum_{i=d+1}^r l_i \right) \rightarrow \tilde{y}\sigma^2$ .

Por lo tanto,  $\log(LRT_d/LRT_{d-1}) \rightarrow h(\lambda_d)$ . En consecuencia, recordando que  $\epsilon = h(\lambda^*) - h[\sigma^2(1 + \sqrt{y})]$ , tenemos

$$\begin{aligned} F(d, p, m, \epsilon) - F(d-1, p, m, \epsilon) &\rightarrow h(\lambda_d) - h[\sigma^2(1 + \sqrt{y})] - \epsilon \\ &= h(\lambda_d) - h(\lambda^*), \end{aligned}$$

y, dado que  $\lambda_d > \lambda^* > \sigma^2(1 + \sqrt{y})$ , la monotonía de la función  $h$  implica que para  $p, m$  suficientemente grandes,

$$\mathbb{P}\{A_d\} = \mathbb{P}\{F(d, p, m, \epsilon) - F(d-1, p, m, \epsilon) > 0\} > 1 - \frac{\delta}{q_0}.$$

El caso  $d > d_1$  es muy similar. Solamente debemos notar que aquí  $\log(LRT_d/LRT_{d-1}) \rightarrow h[\sigma^2(1 + \sqrt{y})]$ . Luego, para  $p, m$  suficientemente grandes, tenemos

$$\mathbb{P}\{F(d, p, m) - F(d-1, p, m) < 0\} > 1 - \frac{\delta}{q_0}.$$

Como consecuencia,

$$\begin{aligned} \mathbb{P}\left(\bigcap_{d=1}^{q_0} A_d\right) &= 1 - \mathbb{P}\left(\bigcup_{d=1}^{q_0} A_d^c\right) \\ &\geq 1 - \sum_{d=1}^{q_0} \mathbb{P}(A_d^c) \geq 1 - \delta, \end{aligned}$$

□

---

# Bibliografía

---

- [1] Akaike, H. *Information Theory and an Extension of the Maximum Likelihood Principle*, pages 199–213. Springer New York, New York, NY, 1973.
- [2] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974. ISSN 0018-9286. doi: 10.1109/TAC.1974.1100705.
- [3] Bai, Z. and Yao, J.-F. Central limit theorems for eigenvalues in a spiked population model. *Annales de L’Institut Henri Poincare Section (B) Probability and Statistics*, 44:447–474, June 2008. doi: 10.1214/07-AIHP118.
- [4] Bai, Z., Silverstein, J. W., and Yin, Y. A note on the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis*, 26(2):166 – 168, 1988. ISSN 0047-259X.
- [5] Bai, Z., Jiang, D., Yao, J., and Zheng, S. Corrections to lrt on large-dimensional covariance matrix by rmt. *Ann. Statist.*, 37(6B):3822–3840, 12 2009. doi: 10.1214/09-AOS694. URL <http://dx.doi.org/10.1214/09-AOS694>.
- [6] Bai, Z. D. and Yin, Y. Q. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.*, 21(3):1275–1294, 07 1993. doi: 10.1214/aop/1176989118. URL <https://doi.org/10.1214/aop/1176989118>.
- [7] Baik, J. and Silverstein, J. W. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97:1382–1408, 2006.
- [8] Baik, J., Arous, G. B., and Pécché, S. Phase transition for the largest eigenvalue of large sample covariance matrices. *Annals of Probability*, 33(5):1643–1697, 2005.
- [9] Bartlett, M. S. A note on the multiplying factors for various  $\chi^2$  approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 16(2):296–298, 1954. ISSN

00359246. URL <http://www.jstor.org/stable/2984057>.
- [10] Billingsley, P. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 1995. ISBN 9780471007104.
- [11] Cook, R. D. and Forzani, L. On the mean and variance of the generalized inverse of a singular wishart matrix. *Electron. J. Statist.*, 5:146–158, 2011. doi: 10.1214/11-EJS602. URL <https://doi.org/10.1214/11-EJS602>.
- [12] Eaton, J. *Multivariate statistics: a vector space approach*. Institute of Mathematical Statistics, 1983.
- [13] Forzani, L., Gieco, A., and Tolmasky, C. Likelihood ratio test for partial sphericity in high and ultra-high dimensions. *Journal of Multivariate Analysis*, 159:18 – 38, 2017.
- [14] Geman, S. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 04 1980. doi: 10.1214/aop/1176994775. URL <https://doi.org/10.1214/aop/1176994775>.
- [15] Ghazal, G. A. and Neudecker, H. On second-order and fourth-order moments of jointly distributed random matrices: a survey. *Linear Algebra and its Applications*, 321(1):61 – 93, 2000. ISSN 0024-3795. doi: [https://doi.org/10.1016/S0024-3795\(00\)00181-6](https://doi.org/10.1016/S0024-3795(00)00181-6). URL <http://www.sciencedirect.com/science/article/pii/S0024379500001816>. Eighth Special Issue on Linear Algebra and Statistics.
- [16] Harding, M. C., Jorgenson, D., King, G., Linton, O., Lorenzoni, G., Mazur, B., Panageas, S., and Patel, K. Structural estimation of high-dimensional factor models. Technical report, Stanford University, 2007.
- [17] Izenman, A. J. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Springer Publishing Company, Incorporated, 1 edition, 2008. ISBN 0387781889, 9780387781884.
- [18] James, A. T. Test of equality of the latent roots of the covariance matrix. In Krishnaiah, P. R., editor, *Multivariate Analysis*, volume 2, pages 205–218. Academic Press, New York, 1969.
- [19] Jiang, D., Jiang, T., and Yang, F. Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *Journal of Statistical Planning and Inference*, 142(8): 2241 – 2256, 2012. ISSN 0378-3758. doi: <http://dx.doi.org/10.1016/j.jspi.2012.02.057>.
- [20] Jiang, T. and Yang, F. Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.*, 41(4):2029–2074, 08 2013. doi: 10.1214/13-AOS1134. URL <http://dx.doi.org/10.1214/13-AOS1134>.
- [21] Jiang, T. and Yang, F. Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.*, 41(4):2029–2074, 08 2013. doi: 10.1214/13-AOS1134.

- [22] Johnson, C. R. and Robinson, H. A. Eigenvalue inequalities for principal submatrices. *Linear Algebra and its Applications*, 37:11 – 22, 1981. ISSN 0024-3795. doi: [https://doi.org/10.1016/0024-3795\(81\)90163-4](https://doi.org/10.1016/0024-3795(81)90163-4). URL <http://www.sciencedirect.com/science/article/pii/0024379581901634>.
- [23] Johnstone, I. M. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29:295–327, 2001.
- [24] Jolliffe, I. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2002. ISBN 9780387954424.
- [25] Kritchman, S. and Nadler, B. Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *IEEE Transactions on Signal Processing*, 57(10): 3930–3941, 2009.
- [26] Laloux, L., Cizeau, P., Potters, M., and Bouchaud, J. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 03(03):391–397, 2000. URL <https://doi.org/10.1142/S0219024900000255>.
- [27] Lawley, D. N. Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, 43(1-2):128–136, 1956. doi: 10.1093/biomet/43.1-2.128. URL <http://biomet.oxfordjournals.org/content/43/1-2/128.short>.
- [28] Li, Z. and Yao, J. Testing the sphericity of a covariance matrix when the dimension is much larger than the sample size. *Electron. J. Statist.*, 10(2):2973–3010, 2016. doi: 10.1214/16-EJS1199. URL <https://doi.org/10.1214/16-EJS1199>.
- [29] Magnus, J. R. and Neudecker, H. The commutation matrix: Some properties and applications. *Ann. Statist.*, 7(2):381–394, 03 1979. doi: 10.1214/aos/1176344621. URL <https://doi.org/10.1214/aos/1176344621>.
- [30] Marčenko, V. A. and Pastur, L. A. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967. URL <http://stacks.iop.org/0025-5734/1/i=4/a=A01>.
- [31] Mauchly, J. W. Significance test for sphericity of a normal  $n$ -variate distribution. *Ann. Math. Statist.*, 11(2):204–209, 06 1940. doi: 10.1214/aoms/1177731915. URL <http://dx.doi.org/10.1214/aoms/1177731915>.
- [32] McLaren, N. and Shanbhogue, R. Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, 51(2):134–140, 2011. URL <https://EconPapers.repec.org/RePEc:boe:qbullt:0052>.
- [33] Muirhead, R. J. *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Statistics, 1982, 2005.

- [34] Onatski, A. Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479, 2009. ISSN 1468-0262. doi: 10.3982/ECTA6964. URL <http://dx.doi.org/10.3982/ECTA6964>.
- [35] Passemier, D. and Yao, J. Estimation of the number of spikes, possibly equal, in the high-dimensional case. *Journal of Multivariate Analysis*, 127:173–183, 2014.
- [36] Passemier, D. and Yao, J.-F. On determining the number of spikes in a high-dimensional spiked population model. *Random Matrices: Theory and Applications*, 01(01):1150002, 2012. doi: 10.1142/S201032631150002X. URL <http://www.worldscientific.com/doi/abs/10.1142/S201032631150002X>.
- [37] Paul, D. Asymptotics of the leading sample eigenvalues for a spiked covariance model. Technical report, 2004. URL <http://anson.ucdavis.edu/~debashis/techrep/eigenlimit.pdf>.
- [38] Paul, D. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17:1617–1642, 2007.
- [39] Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., and Stanley, H. E. Random matrix approach to cross correlations in financial data. *Phys. Rev. E*, 65:066126, Jun 2002. doi: 10.1103/PhysRevE.65.066126. URL <http://link.aps.org/doi/10.1103/PhysRevE.65.066126>.
- [40] Preis, T., Moat, H. S., and Stanley, H. E. Quantifying trading behavior in financial markets using google trends. *Sci. Rep.*, 3, Apr. 2013. doi: 10.1038/srep01684.
- [41] Rissanen, J. Modeling by the shortest data description. *Automatica*, 14(5):465–471, 09 1978.
- [42] Schwarz, G. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 03 1978. doi: 10.1214/aos/1176344136. URL <https://doi.org/10.1214/aos/1176344136>.
- [43] Silverstein, J. W. The smallest eigenvalue of a large dimensional wishart matrix. *Ann. Probab.*, 13(4):1364–1368, 11 1985. doi: 10.1214/aop/1176992819. URL <https://doi.org/10.1214/aop/1176992819>.
- [44] Srivastava, M. S. Some tests concerning the covariance matrix in high dimensional data. *JOURNAL OF THE JAPAN STATISTICAL SOCIETY*, 35(2):251–272, 2005. doi: 10.14490/jjss.35.251.
- [45] Srivastava, M. S. Some tests criteria for the covariance matrix with fewer observations than the dimension. *Acta Comment. Univ. Tartu. Math.*, 10:77–93, 2006.
- [46] Tibshirani, R., Walther, G., and Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001. ISSN 1467-9868. doi: 10.1111/1467-9868.00293. URL

- <http://dx.doi.org/10.1111/1467-9868.00293>.
- [47] von Rosen, D. Moments for the inverted wishart distribution. *Scandinavian Journal of Statistics*, 15(2):97–109, 1988. ISSN 03036898, 14679469. URL <http://www.jstor.org/stable/4616090>.
- [48] Wainwright, M. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. draft, 2018.
- [49] Wang, Q. and Yao, J. On the sphericity test with large-dimensional observations. *Electron. J. Statist.*, 7:2164–2192, 2013. doi: 10.1214/13-EJS842. URL <https://doi.org/10.1214/13-EJS842>.
- [50] Wax, M. and Kailath, T. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392, 1985.
- [51] Yao, J., Zheng, S., and Bai, Z. *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2015.
- [52] Yin, Y., Bai, Z., and Krishnaiah, P. Limiting behavior of the eigenvalues of a multivariate f matrix. *Journal of Multivariate Analysis*, 13(4):508 – 516, 1983. ISSN 0047-259X. doi: [https://doi.org/10.1016/0047-259X\(83\)90036-2](https://doi.org/10.1016/0047-259X(83)90036-2). URL <http://www.sciencedirect.com/science/article/pii/0047259X83900362>.
- [53] Yin, Y. Q., Bai, Z. D., and Krishnaiah, P. R. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields*, 78(4): 509–521, Aug 1988. ISSN 1432-2064. doi: 10.1007/BF00353874. URL <https://doi.org/10.1007/BF00353874>.