



CIMAT

Centro de Investigación en Matemáticas, A.C.

COMPARACIÓN DE MODELOS GRÁFICOS GAUSSIANOS CON APLICACIÓN EN DATOS DE fMRI

T E S I S

Que para obtener el grado de
Maestro en Ciencias
con Especialidad en
Probabilidad y Estadística

Presenta

Natalia Monserrath Urquiza Robles

Director de Tesis:

Dr. Rogelio Ramos Quiroga

Autorización de la versión final



Centro de Investigación en Matemáticas, A.C.

Acta de Examen de Grado

Acta No.: 120
 Libro No.: 002
 Foja No.: 120

En la Ciudad de Guanajuato, Gto., siendo las 15:30 horas del día 22 de septiembre del año 2017, se reunieron los miembros del jurado integrado por los señores:

DR. ROLANDO JOSÉ BISCAY LIRIO (CIMAT)
DRA. LILIA LETICIA RAMÍREZ RAMÍREZ (CIMAT)
DR. ROGELIO RAMOS QUIROGA (CIMAT)

bajo la presidencia del primero y con carácter de secretario el segundo, para proceder a efectuar el examen que para obtener el grado de

**MAESTRO EN CIENCIAS
 CON ESPECIALIDAD EN PROBABILIDAD Y ESTADÍSTICA**

Sustenta

NATALIA MONSERRATH URQUIZA ROBLES

en cumplimiento con lo establecido en los reglamentos y lineamientos de estudios de posgrado del Centro de Investigación en Matemáticas, A.C., mediante la presentación de la tesis

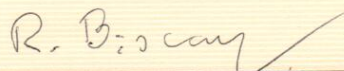
**"COMPARACIÓN DE MODELOS GRÁFICOS GAUSSIANOS
 CON APLICACIÓN EN DATOS DE fMRI"**

Los miembros del jurado examinaron alternadamente al (la) sustentante y después de deliberar entre sí resolvieron declararlo (a):

APROBADA




CIMAT
 DIRECCIÓN
 GENERAL
DR. VÍCTOR MANUEL RÍVERO MERCADO
 Director General Interino


DR. ROLANDO JOSÉ BISCAY LIRIO
 Presidente


DRA. LILIA LETICIA RAMÍREZ RAMÍREZ
 Secretario


DR. ROGELIO RAMOS QUIROGA
 Vocal

Agradecimientos

A mis padres, Herminia y Ramón, quienes me han apoyado siempre y me han alentado a seguir mis sueños. Gracias por todos los consejos de vida y por haber confiado en mí en cada etapa que he emprendido. Gracias a ustedes soy la persona de ahora.

Al Dr. Rogelio Ramos por brindarme su apoyo y compartir su conocimiento y experiencias conmigo, por el tiempo dedicado y por todos los consejos que me dio.

A mis sinodales, la Dra. Leticia Ramírez y el Dr. Rolando Biscay, por el tiempo que destinaron para la revisión de este trabajo

A mis maestros, tanto de la maestría como de la licenciatura, quienes han hecho de mí la profesionalista que ahora soy. Especialmente a la Dra. Eloísa Díaz-Francés, quien me dio muy valiosos consejos no sólo académicos sino también personales.

A mis amigos, con quienes he compartido tantas experiencias y han estado a mi lado para escucharme y aconsejarme.

A Alí, por estar siempre a mi lado, por ser mi apoyo en los momentos más difíciles, por el amor que me brinda y la confianza depositada en mí.

Al CONACYT por la beca otorgada durante estos dos años para la realización de mis estudios y al CIMAT por sus apoyos y facilidades.

Resumen

En el área de Neurociencia surge el problema de estudiar las diferencias que puedan existir en las relaciones de conectividad cerebral entre dos grupos de individuos. La conectividad cerebral de un individuo puede ser abordada mediante modelos gráficos gaussianos basándose en datos fMRI. En esta tesis se retoma esta problemática considerando que para realizar esta comparación entre grupos se debe realizar un estudio de pruebas de hipótesis múltiples. La comparación de la conectividad cerebral entre dos grupos de personas ha sido abordada en distintos estudios. En esta tesis se detalla e implementa el método introducido por Narayan et al. (2015), en el cual se consideran asuntos estadísticamente relevantes que surgen en el proceso de análisis que inicia en la preparación de los datos para estimar los correspondientes modelos gráficos hasta realizar la corrección por tratarse de una prueba de hipótesis múltiple. Además, este método se implementó para contrastar la conectividad funcional entre personas con autismo y un grupo control y se comparan los resultados obtenidos con la teoría previa de autismo.

Contenido

1	Introducción	3
2	Preliminares Teóricos	7
2.1	Modelos gráficos no-dirigidos	7
2.1.1	Modelos Gráficos Gaussianos	8
2.1.2	Estimación de la matriz de precisión sparse	10
2.1.3	Estimación del parámetro de regularización	15
2.1.4	Penalización aleatoria	17
2.2	Pruebas de hipótesis múltiples	18
2.2.1	Procedimiento Benjamini-Hochberg	19
2.2.2	Procedimiento Benjamini-Yekutieli	21
3	Contraste de hipótesis en GGMs	23
3.1	Estructura de los datos de fMRI	25
3.2	Planteamiento del contraste de hipótesis	25
3.3	Estimación de la matriz de precisión por individuo	28
3.3.1	Blanqueamiento de las series de tiempo	28
3.3.2	Remuestreo	30
3.3.3	Penalización aleatoria en el método de Narayan et al. (2015)	31
3.4	Estadístico de prueba y algoritmo resumen	32
3.4.1	Estadístico de prueba	32
3.4.2	Algoritmo Narayan et al. (2015)	34
4	Aplicación a los datos fMRI	37
4.1	Preprocesamiento de los datos	38
4.2	Regiones de interés	39
4.3	Implementación y resultados	40
5	Conclusiones	45
	Bibliografía	47

Capítulo 1

Introducción

Los modelos gráficos son una herramienta que ha sido estudiada arduamente en años recientes. El interés por estos modelos radica en que proveen una estructura en donde se pueden llevar a cabo estudios en datos en altas dimensiones. Específicamente, una gráfica consiste en un conjunto de nodos, que representan variables aleatorias, y un conjunto de aristas que unen pares de nodos. Un tipo de modelos gráficos son los modelos gráficos gaussianos (GGMs, Gaussian Graphical Models), los cuales poseen la interesante propiedad de que la ausencia de una arista entre dos nodos significa que las correspondientes variables aleatorias son condicionalmente independientes, dadas las variables restantes. Esta especial propiedad es una de las principales razones de la aplicación de este tipo de modelos gráficos, en casos donde son adecuados, a un gran número de estudios. Uno de ellos es el llevado a cabo en Wu et al. (2003), donde los modelos gráficos gaussianos son utilizados para estudiar las interacciones presentes entre los genes de levadura utilizando un microarreglo de niveles de expresión de genes. Este es sólo un ejemplo de la aplicación de estos modelos, pero hay una gran variedad de problemáticas en las que pueden ser utilizados, como lo son las redes sociales o redes de narcotráfico, por lo que es un tema de latente desarrollo.

En el área médica es recurrente el interés en comparar la efectividad de un tratamiento médico. Una forma usual de llevar a cabo este estudio es haciendo una comparación de medias, más específicamente, un contraste de hipótesis donde la hipótesis nula sea que las medias son iguales para el grupo tratado y el no tratado. Existen procedimientos muy bien establecidos para llevar a cabo estas pruebas. Por otro lado, en el área de Neurociencia es de interés estudiar las diferencias que puedan existir en las relaciones de conectividad funcional entre dos grupos de individuos. La conectividad funcional se define como la dependencia de la actividad neuronal entre distintas regiones del cerebro, Smith et al. (2013). Existen estudios que sustentan el uso de los modelos gráficos gaussianos para modelar esta conectividad. Así pues, la comparación de la conectividad funcional entre dos grupos se puede llevar a cabo comparando los modelos gráficos gaussianos que se estiman para cada grupo de estudio. Sin embargo, este problema es más complejo que una prueba de hipótesis sobre diferencia de medias, ya que un modelo gráfico gaussiano involucra más que un sólo parámetro, haciendo más difícil el planteamiento de la prueba de hipótesis adecuada.

Dada la gran cantidad de elementos a comparar si se quiere hacer un contraste de modelos gráficos gaussianos surge la necesidad de llevar a cabo un estudio de pruebas de hipótesis

múltiples, la cual es una metodología que se debe implementar cuando se quieren contrastar simultáneamente varias hipótesis. Existen una gran cantidad de investigaciones donde se abordan distintas maneras para llevar a cabo esta tarea. Así, considerar este tópico es de gran importancia en el planteamiento del contraste en las relaciones de conectividad funcional entre dos grupos de individuos.

La comparación de la conectividad funcional entre dos grupos de personas ha sido abordada en distintos estudios. Uno de los más recientes es el llevado a cabo en Narayan et al. (2015) donde se tiene como objetivo introducir una nueva metodología que responde a las problemáticas que surgen en este contexto. En esta metodología se propone hacer uso del método Lasso Gráfico para la estimación de los modelos gráficos gaussianos, lo que conlleva a considerar la estimación del parámetro de regularización, así como la problemática de la selección del verdadero modelo gráfico. Además, el enfoque de Narayan et al. (2015) considera distintos asuntos estadísticamente relevantes que surgen en el proceso que inicia en la preparación de los datos para emplear el método Lasso Gráfico hasta obtener los p -valores que deben ser corregidos por tratarse de una prueba de hipótesis múltiple.

Los datos por resonancia magnética funcional consisten en una serie de registros de las imágenes por resonancia magnética funcional (fMRI, functional Magnetic Resonance Imaging) tomadas en un determinado tiempo, con un tiempo de espaciado fijo. Dichas imágenes miden la cantidad de oxígeno en la sangre (BOLD, Blood Oxygenation Level-Dependent). Este tipo de datos son una herramienta útil en el estudio de la conectividad funcional. Se han conducido estudios utilizando este tipo de datos para analizar la actividad neurobiológica en personas con autismo, Dichter (2012). En la página web del proyecto ABIDE (“Autism Brain Imaging Data Exchange”, Craddock et al. (2013a)) se ofrece una recopilación de datos fMRI en estado de reposo de un total de 16 instituciones. La base de datos consiste en los datos fMRI para sujetos que presentan trastorno de espectro autista (ASD, Autism Spectrum Disorder) y sujetos de un grupo control. Estos datos pueden ser utilizados para efectuar un estudio con el fin de comparar las relaciones de conectividad funcional entre personas control y personas que sufren de autismo. De modo que el método propuesto en Narayan et al. (2015) puede ser utilizado para analizar este tipo de datos. Además, en este método se considera también la naturaleza de los datos fMRI.

Esta tesis busca explicar de manera introductoria la modelación mediante modelos gráficos gaussianos. Se presentan resultados relevantes en este tema, además de que se expone a detalle uno de los métodos más utilizados para efectuar la estimación de un modelo gráfico con estructura no conocida, el algoritmo Lasso Gráfico, Friedman et al. (2008). En el algoritmo Lasso Gráfico es de suma importancia el parámetro de regularización, por lo que se efectúa un análisis referente a la estimación de este parámetro. Se citan específicamente dos métodos para estimarlo y se analizan las propiedades de cada uno. Además, se exponen métodos clásicos para efectuar correcciones por pruebas de hipótesis múltiples y se mencionan resultados relevantes en esta área.

Por otro lado, esta tesis tiene como objetivo detallar el método introducido por Narayan et al. (2015), y todos los temas estadísticos que se unen para derivar el método, además de los ya discutidos en los tópicos que se mencionaron anteriormente. En este sentido, se exponen las problemáticas que surgen para comparar las relaciones de conectividad funcional

entre dos grupos utilizando datos fMRI y se discute cómo es que el método propuesto en Narayan et al. (2015) atiende a éstos, además de que se desarrollan demostraciones clave de los resultados utilizados para plantear el estadístico de prueba a utilizar en el contraste de hipótesis.

Por último, se realiza la aplicación del método de Narayan et al. (2015) a un subconjunto de la base de datos de ABIDE, los cuales son los provenientes de la Universidad de California en los Ángeles (UCLA). Se detalla la implementación del algoritmo y se discuten las conclusiones que emanan de la aplicación en estos datos así como de la implementación. Además de que se comenta sobre posibles direcciones a seguir en trabajos posteriores.

La manera en la que está estructurada esta tesis es la siguiente: el Capítulo 2 contiene una introducción a los modelos gráficos no-dirigidos y los modelos gráficos gaussianos utilizando referencias como Wasserman (2004), Hastie et al. (2008) y Hastie et al. (2015). Se deriva el método Lasso Gráfico de Friedman et al. (2008) y se presentan dos métodos para la estimación del parámetro de regularización, siendo éstos el presentado en Bien y Tibshirani (2001), que se efectúa utilizando validación cruzada, y el presentado en Liu et al. (2010). Además se comenta sobre el método de Penalización Aleatoria recientemente introducido por Meinshausen y Bühlman (2010). Por último, se presenta la teoría referente a las pruebas de hipótesis múltiples, haciendo mención de dos de los trabajos más influyentes en este tópico, el de Benjamini y Hochberg (1995) y Benjamini y Yekutieli (2001).

En el Capítulo 3 se describe la estructura de los datos fMRI y se detalla el método de Narayan et al. (2015). Específicamente se realiza el planteamiento formal de la prueba de hipótesis, se expone la teoría necesaria para realizar blanqueamiento a los datos, se detalla la manera a efectuar la estimación de la matriz de precisión concerniente al modelo gráfico, la cual involucra la idea del método de Penalización Aleatoria, y se conduce el tema de remuestreo necesario para el método. Por último, se plantea el estadístico de prueba.

En el Capítulo 4 se comentan algunas teorías sobre la conectividad funcional en personas con trastorno del espectro autista, las cuales son principalmente mencionadas en Dichter et al. (2012). Se detalla la implementación del algoritmo de Narayan et al. (2015) en el software estadístico R, R Core Team (2014), y se efectúan distintos experimentos en la aplicación a los datos de la UCLA. Se comentan las conclusiones que se derivan de dicha implementación. Resulta que las hipótesis rechazadas respaldan las observaciones que se tienen sobre la conectividad local y de larga distancia en las personas que presentan ASD. Por último, también se hace mención de recomendaciones para trabajos posteriores.

Capítulo 2

Preliminares Teóricos

En este capítulo se abordan principalmente dos temas: modelos gráficos no-dirigidos y pruebas de hipótesis múltiples. Ambos tópicos son esenciales para el desarrollo de la tesis y se procede a su escrito debido a la necesidad de que el lector esté familiarizado con ellos. Por un lado, el entendimiento de los modelos gráficos no-dirigidos, en particular los modelos gráficos gaussianos, desemboca en la comprensión del porqué se utilizan para modelar las relaciones de conectividad funcional y, por el otro, las pruebas de hipótesis múltiples resultan ser una herramienta imprescindible a utilizar en el objetivo de comparar las relaciones de conectividad entre dos grupos de individuos.

2.1. Modelos gráficos no-dirigidos

Los *modelos gráficos* son una herramienta estadística que ha tenido reciente crecimiento debido a que su implementación está en parsimonia con datos multivariados. La necesidad de utilizar estos modelos en la presente tesis radica en el hecho de que una gráfica representa las relaciones de independencia condicional que puedan existir entre las variables que los suscitan. Debido a que el interés primordial en este trabajo es sobre las relaciones de conectividad funcional, resulta natural emplear estos modelos como base fundamental para el análisis que atañe.

A continuación, para situar al lector en contexto, se da una breve introducción sobre los modelos gráficos no-dirigidos para después ahondar en los modelos gráficos gaussianos. Dicha introducción se aborda en la forma en la que está expuesta en Wasserman (2004), Hastie et al. (2008) y Hastie et al. (2015). Se recomienda al lector consultar dichas referencias para obtener mayor información acerca de los modelos gráficos no-dirigidos y sobre los modelos gráficos gaussianos.

Definición 2.1. La *independencia condicional* entre dos variables se define como: sean X, Y y Z tres variables aleatorias, se dice que X y Y son condicionalmente independientes dado Z , denotado por $X \perp\!\!\!\perp Y | Z$, si

$$f_{X,Y|Z}(x, y | z) = f_{X|Z}(x | z) f_{Y|Z}(y | z) \quad \forall x, y \text{ y } z.$$

Definición 2.2. Una *gráfica no-dirigida* $\mathcal{G} = (V, E)$ consiste en un conjunto finito V llamado el conjunto de nodos (o vértices) y E un conjunto de aristas, $E \subset V \times V$. Los nodos corresponden a variables aleatorias X, Y, Z, \dots y $(X, Y) \in E$ si existe una arista entre X y Y . Se llama gráfica no-dirigida si no existe distinción entre $(X, Y) \in E$ y la arista (Y, X) .

La definición que se enuncia a continuación desemboca en lo que ya se había mencionado antes, sobre el vínculo que se establece entre un modelo gráfico y las relaciones de independencia condicional que existan entre las variables involucradas en el modelo.

Definición 2.3. Si se considera una gráfica no-dirigida donde existen aristas que unen a todos los nodos, salvo aquellos que son independientes dado el resto de variables, es decir,

$$\text{no existe arista entre } X \text{ y } Y \Leftrightarrow X \perp\!\!\!\perp Y | V \setminus \{X, Y\}$$

entonces a esa gráfica se le conoce como *gráfica Markov a pares*.

Ejemplo 2.4. Para ejemplificar la definición anterior se muestra la siguiente figura. La gráfica corresponde a los vértices $V = \{X, Y, Z\}$ con aristas $E = \{(X, Y), (Y, Z)\}$.

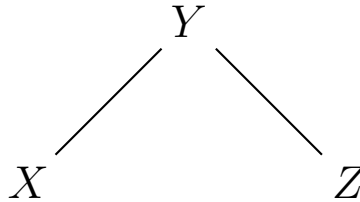


Figura 1. Gráfica con $V = \{X, Y, Z\}$ y aristas $E = \{(X, Y), (Y, Z)\}$.

La gráfica de la Figura 1 ejemplifica la relación de independencia condicional $X \perp\!\!\!\perp Z | Y$ ya que no hay arista entre X y Z . Hasta aquí se ha expuesto la definición de un modelo gráfico no-dirigido, así como los conceptos fundamentales de que se harán uso para abordar los modelos gráficos gaussianos que se estudian a continuación, los cuales tienen la propiedad de ser una gráfica Markov a pares.

2.1.1. Modelos Gráficos Gaussianos

Los *modelos gráficos gaussianos* (GGMs, Gaussian Graphical Models) son una clase de modelos gráficos populares debido a sus propiedades derivadas de la distribución Normal. Por ejemplo, si dos variables aleatorias tienen distribución conjunta Normal entonces su distribución condicional y marginal también es Normal. Suponga que se quiere estimar la gráfica de un conjunto de observaciones. Si dichas observaciones provienen de una distribución Normal

entonces los parámetros del modelo gráfico subyacente son computacionalmente sencillos de estimar.

Sea X una variable aleatoria con distribución Normal p -variada con matriz de medias, μ , igual al vector cero y matriz de covarianzas Σ , esto es, $X = (X_1, X_2, \dots, X_p)^T \sim N_p(0, \Sigma)$. En general, la media en aplicaciones de modelos gráficos no juega un papel importante, así que, sin pérdida de generalidad, se considera $\mu = 0$. Sea $\Theta = \Sigma^{-1} = \{\theta_{ij}\}_{i,j=1}^p$. A la matriz Θ se le conoce como *matriz de precisión* o matriz de concentración. En el área de modelos gráficos es conveniente usar la parametrización de la distribución Normal con respecto a la matriz Θ debido a que esta matriz contiene la información sobre la gráfica minimal subyacente, Giraud (2015). El siguiente lema vislumbra la relación entre la matriz de precisión y la gráfica no-dirigida correspondiente, es decir, las relaciones de independencia condicional.

Lema 2.5. *La ij -ésima entrada de Θ es distinta de 0 si y sólo si hay una arista entre X_i y X_j .*

La demostración del lema anterior se hará de manera similar en la que se expone en Giraud (2015). Primero es necesario recordar cuál es la inversa de una matriz particionada.

Lema 2.6. *Inversa de una matriz particionada: Sea Σ una matriz de tamaño $p \times p$ invertible y considere su partición dada por*

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix},$$

donde $\Sigma_{aa}, \Sigma_{ab}, \Sigma_{ba}, \Sigma_{bb}$ son de tamaño $p_1 \times p_1$, $p_1 \times p_2$, $p_2 \times p_1$ y $p_2 \times p_2$ respectivamente y $p = p_1 + p_2$. La inversa Θ de Σ es:

$$\Theta = \begin{pmatrix} \theta_{aa} & \theta_{ab} \\ \theta_{ba} & \theta_{bb} \end{pmatrix} = \Sigma^{-1} = \begin{pmatrix} \Sigma_{a|b}^{-1} & -\Sigma_{a|b}^{-1}\Sigma_{ab}\Sigma_{bb}^{-1} \\ -\Sigma_{bb}^{-1}\Sigma_{ba}\Sigma_{a|b}^{-1} & \Sigma_{bb}^{-1} + \Sigma_{bb}^{-1}\Sigma_{ba}\Sigma_{a|b}^{-1}\Sigma_{ab}\Sigma_{bb}^{-1} \end{pmatrix}, \quad (2.1)$$

donde $\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$.

De igual modo, para la demostración del Lema 2.5, se emplea el siguiente resultado: Si $X = \begin{pmatrix} X_a \\ X_b \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}\right)$ entonces $X_a|X_b \sim N(\Sigma_{ab}\Sigma_{bb}^{-1}X_b, \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})$. Note entonces que

$$\begin{aligned} \text{Var}(X_a|X_b) &= \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \\ &= \Sigma_{a|b} && \text{con la notación de (2.1)} \\ &= \theta_{aa}^{-1} && \text{utilizando (2.1).} \end{aligned} \quad (2.2)$$

Luego de enunciar estos resultados se procede a la demostración del Lema 2.5. Sea $X = (X_1, X_2, \dots, X_p)^T$ un vector aleatorio con distribución Normal con media cero y varianza Σ . Considere $X_a = (X_1, X_2)^T$ y $X_b = (X_3, \dots, X_p)^T$ y de esta manera la partición de Σ y $K = \Sigma^{-1}$ como

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \text{ y } K = \begin{pmatrix} k_{aa} & k_{ab} \\ k_{ba} & k_{bb} \end{pmatrix}.$$

Note que k_{aa} es de tamaño 2×2 . Utilizando la notación *resto* para hacer referencia a X_b se tiene lo siguiente

$$\begin{aligned} \text{Var}(X_1, X_2 | \text{resto}) &= k_{aa}^{-1} && \text{utilizando (2.2)} \\ &= \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}^{-1}, \end{aligned}$$

donde $k_{11}, k_{12}, k_{21}, k_{22}$ son las entradas correspondientes de K , de modo que

$$\begin{aligned} \text{Var}(X_1, X_2 | \text{resto}) &= \frac{1}{k_{11}k_{22} - k_{21}k_{12}} \begin{pmatrix} k_{22} & -k_{12} \\ -k_{21} & k_{11} \end{pmatrix} \\ \Rightarrow \text{Cov}(X_1, X_2 | \text{resto}) &= -\frac{k_{12}}{k_{11}k_{22} - k_{21}k_{12}}, \end{aligned}$$

así

$$X_1 \perp\!\!\!\perp X_2 | \text{resto} \Leftrightarrow k_{12} = 0.$$

Ahora considere la matriz de permutación P_{ij} que intercambia la fila i con la fila j . Observe entonces lo siguiente

$$P_{2j}P_{1i}X = \begin{pmatrix} X_i \\ X_j \\ \vdots \\ X_p \end{pmatrix}.$$

Sea $A = P_{2j}P_{1i}$. Note que $AX \sim N(0, A\Sigma A^T)$ y $A\Sigma A^T = P_{2j}P_{1i}\Sigma(P_{2j}P_{1i})^T = P_{2j}P_{1i}\Sigma P_{1i}P_{2j}$. Así, las entradas (1,1), (1,2), (2,1) y (2,2) de la varianza de AX son $\sigma_{ii}, \sigma_{ij}, \sigma_{ji}$ y σ_{jj} respectivamente. Por lo que utilizando este desarrollo se llega al siguiente caso general:

$$\text{Var}(X_i, X_j | \text{resto}) = \begin{pmatrix} k_{ii} & k_{ij} \\ k_{ji} & k_{jj} \end{pmatrix} = \frac{1}{k_{ii}k_{jj} - k_{ji}k_{ij}} \begin{pmatrix} k_{jj} & -k_{ij} \\ -k_{ji} & k_{ii} \end{pmatrix}$$

de modo que se ha probado que $X_i \perp\!\!\!\perp X_j | \text{resto} \Leftrightarrow k_{ij} = 0$, con lo que queda demostrado el lema.

El Lema 2.5 apunta a que, en un modelo gráfico gaussiano, conocer la matriz de precisión permite conocer la gráfica no-dirigida subyacente. Esto indica que la estimación de esta matriz dada una muestra de datos es el objetivo si se quiere conocer su gráfica.

2.1.2. Estimación de la matriz de precisión sparse

Ahora se desea estimar la gráfica subyacente para X . En la literatura se hace diferencia entre estimar el modelo gráfico gaussiano cuando este tiene una estructura conocida y cuando no se conoce la estructura. Típicamente, y en particular en esta tesis, la estructura del grafo es no conocida, por lo que se expondrá la teoría del segundo caso. La estimación del modelo gráfico se ejecuta realizando la estimación de la matriz de precisión basados en la colección $X = \{x_1, \dots, x_N\}$ de observaciones que se tenga disponible.

Un modelo estadístico sparse es aquél que tiene un número pequeño de parámetros que son distintos a cero. En la actualidad, con el auge de la llamada era del *Big Data*, surgen problemas en los que el tamaño de muestra disponible es igual o menor al número de características que se miden de cada observación. Dicho fenómeno provoca que resolver el problema estadístico *denso* no sea adecuado debido a que las estimaciones pierden propiedades importantes, por lo que el verdadero interés recae en atender el problema sparse. Este es el caso de la mayoría de las ocasiones en que se desea estimar un modelo gráfico gaussiano.

Para entender de mejor manera el obstáculo de tener muestras de tamaño similar o mayor a la dimensión de cada muestra se menciona lo siguiente. Después de realizar cálculos sencillos de álgebra, se obtiene que la función de logverosimilitud de la muestra aleatoria $X = \{x_1, \dots, x_N\}$, con X proveniente de una distribución Normal p -variada con media μ y matriz de precisión Θ , es la siguiente:

$$l(\mu, \Theta) = \text{cte} + \frac{N}{2} \log |\Theta| - \frac{N}{2} \text{tr}(\Theta S) - \frac{N}{2} \text{tr}(\Theta(\bar{x} - \mu)(\bar{x} - \mu)^T), \quad (2.3)$$

donde $S = \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T$, $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$, $|\cdot|$ es el determinante de una matriz y $\text{tr}(\cdot)$ es la traza de una matriz. Si se sustituye el estimador de máxima verosimilitud de μ en (2.3), es decir, $\hat{\mu}_{MV} = \bar{x}$, se obtiene la función de logverosimilitud perfil para Θ (omitiendo constantes):

$$l(\Theta) = \log |\Theta| - \text{tr}(\Theta S).$$

La teoría clásica asegura que el estimador de máxima verosimilitud de Θ converge a la verdadera matriz de precisión conforme N tiende a infinito. Sin embargo, cuando se tiene que el número de parámetros p es comparable a N o más grande, no existe el estimador de máxima verosimilitud, como se comenta en Hastie et al. (2015). Esto provoca la necesidad de plantear el problema sparse.

Además de la observación anterior, otro aspecto que motiva al planteamiento de estimar la matriz de precisión resolviendo un problema sparse es que la matriz de precisión adquiere una interpretación más clara. Estimar la solución sparse de la matriz de precisión se traduce en estimar dicha matriz motivando que algunas de sus entradas sean igual a 0. Como ya se había mencionado, que la entrada θ_{ij} sea igual a 0 equivale a que X_i y X_j sean independientes dado el resto. Por lo que la solución sparse adquiere una interpretación clara.

Entonces, se quiere estimar la matriz de precisión con algunas de sus entradas igual a cero, es decir, de manera que algunas aristas no estén presentes en la gráfica correspondiente. Friedman et al. (2008) propusieron el método *Lasso Gráfico* que consiste en resolver la estimación sparse de Θ regularizando la logverosimilitud a través de la norma $l_1 = \|\cdot\|_1$, la suma de los valores absolutos de los elementos de Θ , lo que produce un problema de optimización convexo. Así, la expresión que se desea minimizar con respecto a Θ , llamada negativo de la logverosimilitud regularizada, es la siguiente:

$$f(\Theta)_\lambda = -\log |\Theta| + \text{tr}(S\Theta) + \lambda \|\Theta\|_1, \quad (2.4)$$

donde el parámetro λ es llamado *parámetro de regularización* y debe ser una cantidad fija

para realizar la optimización. El método Lasso Gráfico consiste en optimizar la función en (2.4) mediante un procedimiento iterativo realizando regresión Lasso, Tibshirani (1996), en cada iteración. A continuación se derivará dicho método de manera similar en la que se expone en Hastie et al. (2008).

Se busca minimizar $f(\Theta)_\lambda$ sobre el espacio de matrices positivo-definidas por lo que se desea encontrar el argumento $\hat{\Theta}$ tal que

$$\frac{\partial}{\partial \Theta} f(\hat{\Theta})_\lambda = 0.$$

En Giraud (2015) se demuestra el lema de condición de primer orden para optimalidad, el cual indica que una condición equivalente a $\partial f(\hat{\Theta})_\lambda / \partial \Theta = 0$ es que 0 pertenezca al subgradiente de $f(\Theta)_\lambda$. Asimismo, en Fukunaga (1990) se demuestran los siguientes enunciados, de los que se hará uso,

$$\begin{aligned} \frac{\partial}{\partial \Theta} \log|\Theta| &= 2\Theta^{-1} - \text{diag}(\Theta^{-1}), \\ \frac{\partial}{\partial \Theta} \text{tr}(S\Theta) &= 2S - \text{diag}(S). \end{aligned}$$

Si se utilizan estos dos resultados entonces ya se puede derivar la primera parte de $f(\Theta)_\lambda$ y resta derivar $\|\Theta\|_1$. Sea θ_{rs} un elemento cualesquiera de Θ , entonces

$$\frac{\partial}{\partial \theta_{rs}} \|\Theta\|_1 = \frac{\partial}{\partial \theta_{rs}} \left(\sum_{i \neq j} |\theta_{ij}| + \sum_{i=1}^p |\theta_{ii}| \right). \quad (2.5)$$

De igual modo en Fukunaga (1990) se muestra que para $u \in \mathbb{R}$,

$$\frac{d}{du} |u| = \begin{cases} \text{signo}(u) & \text{si } u \neq 0 \\ \in [-1, 1] & \text{si } u = 0 \end{cases},$$

donde $\text{signo}(u)$ mapea a u al 1 si $u > 0$ y al 0 si $u < 0$. Si se define $\text{Sign}(u) = d|u|/du$ siguiendo en la expresión (2.5) se obtiene que

$$\frac{\partial}{\partial \theta_{rs}} \|\Theta\|_1 = 2\text{Sign}(\theta_{rs}) \text{ o } \text{Sign}(\theta_{rr}),$$

entonces

$$\frac{\partial}{\partial \Theta} \|\Theta\|_1 = 2\text{Sign}(\Theta) - \text{diag}(\text{Sign}(\Theta)),$$

así,

$$\begin{aligned} \frac{\partial}{\partial \Theta} f(\Theta)_\lambda &= -2\Theta^{-1} + \text{diag}(\Theta^{-1}) + 2S - \text{diag}(S) + \lambda 2\text{Sign}(\Theta) - \text{diag}(\lambda \text{Sign}(\Theta)) = 0 \\ &\Leftrightarrow 2(-\Theta^{-1} + S + \lambda \text{Sign}(\Theta)) - \text{diag}(-\Theta^{-1} + S + \lambda \text{Sign}(\Theta)) = 0 \end{aligned}$$

de modo que se desea resolver para Θ , sobre matrices positivo-definidas:

$$-\Theta^{-1} + S + \lambda \text{Sign}(\Theta) = 0.$$

Note que $\Theta = \Sigma^{-1}$ es positivo-definida por lo que $\theta_{jj} > 0$, entonces todas los elementos de la diagonal de $Sign(\Theta)$ son igual a 1. Luego, sea $W = \Theta^{-1}$, se desea encontrar

$$W = S + \lambda Sign(\Theta), \quad w_{kk} = s_{kk} + \lambda,$$

o escrito de distinta manera, sea $\Gamma(\Theta) = (\gamma_{ij})_{p \times p}$ donde $\gamma_{ij} \equiv Sign(\theta_{ij})$, se desea

$$-\Theta^{-1} + S + \lambda \Gamma(\Theta) = 0. \quad (2.6)$$

El procedimiento iterativo del método Lasso Gráfico es el siguiente:

Inicializar W como $S + \lambda I$. La diagonal de W permanecerá sin cambios en los siguientes procedimientos. Particionar $\Theta, W, \Gamma(\Theta)$ y S de la siguiente manera:

$$\Theta = \begin{pmatrix} \Theta_{11} & \theta_{12} \\ \theta_{12}^T & \theta_{22} \end{pmatrix}, \quad W = \begin{pmatrix} W_{11} & w_{12} \\ w_{12}^T & w_{22} \end{pmatrix}, \quad \Gamma(\Theta) = \begin{pmatrix} \Gamma_{11} & \gamma_{12} \\ \gamma_{12}^T & \gamma_{22} \end{pmatrix}, \quad S = \begin{pmatrix} S_{11} & s_{12} \\ s_{12}^T & s_{22} \end{pmatrix}$$

donde Θ_{11}, θ_{12} , y θ_{22} son de tamaño $(p-1) \times (p-1)$, $(p-1) \times 1$ y 1×1 respectivamente, y las demás matrices están particionadas de la misma manera.

Ahora usando el Lema 2.6 sobre la inversa de una matriz particionada, es decir las expresiones en (2.1), se tiene lo siguiente

$$W = \begin{pmatrix} W_{11} & w_{12} \\ w_{12}^T & w_{22} \end{pmatrix} = \Theta^{-1} = \begin{pmatrix} (\Theta_{11} - \theta_{12}\theta_{22}^{-1}\theta_{12}^T)^{-1} & -W_{11}\theta_{12}\theta_{22}^{-1} \\ (-W_{11}\theta_{12}\theta_{22}^{-1})^T & \theta_{22}^{-1} + \theta_{22}^{-1}\theta_{12}^T W_{11}\theta_{12}\theta_{22}^{-1} \end{pmatrix}, \quad (2.7)$$

y utilizando (2.7) junto con la expresión (2.6) se sigue que

$$\begin{aligned} W &= S + \lambda \Gamma(\Theta) \\ \Rightarrow \quad w_{12} &= s_{12} + \lambda \gamma_{12} = -W_{11}\theta_{12}\theta_{22}^{-1} \\ \Leftrightarrow \quad W_{11}\theta_{12}\theta_{22}^{-1} &+ s_{12} + \lambda \gamma_{12} = 0. \end{aligned}$$

Si se define $\beta = -\theta_{12}\theta_{22}^{-1}$ y se incluye en la expresión anterior se obtiene

$$\begin{aligned} -W_{11}\beta + s_{12} + \lambda \gamma_{12} &= 0 \\ \Rightarrow \quad W_{11}\beta - s_{12} + \lambda(-\gamma_{12}) &= 0. \end{aligned}$$

Por otra parte, se sabe que $\gamma_{12} = Sign(\theta_{12})$ y $Sign(\theta_{12}) = Sign(\theta_{12}\theta_{22}^{-1})$ ya que $\theta_{22} > 0$, entonces

$$\begin{aligned} \gamma_{12} &= Sign(-\beta) \\ \Rightarrow \quad -\gamma_{12} &= Sign(\beta), \end{aligned}$$

de modo que se tiene que

$$W_{11}\beta - s_{12} + \lambda Sign(\beta) = 0. \quad (2.8)$$

Ahora observe el siguiente problema de optimización:

$$\min_{\beta \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2} \beta^T W_{11} \beta - \beta^T s_{12} + \lambda \|\beta\|_1 \right\}, \quad (2.9)$$

note que la derivada respecto a β de la expresión a minimizar en (2.9) es precisamente la expresión (2.8), de modo que para estimar β se debe resolver dicho problema de optimización.

Por otro lado, considere la matriz X de tamaño $N \times p$ particionada como

$$X = (X_1 \ Y),$$

donde X_1 y Y son de tamaño $N \times (p-1)$ y $N \times 1$, respectivamente. El problema de regresión Lasso, expuesto en Tibshirani (1996), de Y sobre X_1 es el siguiente:

$$\min_{\beta \in \mathbb{R}^{p-1}} \|Y - X_1\beta\|^2 \quad \text{s.a.} \quad \|\beta\|_1 \leq c,$$

que es equivalente a

$$\min_{\beta \in \mathbb{R}^{p-1}} \{ \|Y - X_1\beta\|^2 + 2\lambda(\|\beta\|_1 - c) \}. \quad (2.10)$$

Si se desarrolla la función a minimizar en la expresión (2.10) se obtiene lo siguiente

$$(Y - X_1\beta)^T(Y - X_1\beta) + 2\lambda(\|\beta\|_1 - c) = Y^TY + \beta^T X_1^T X_1\beta - 2\beta^T X_1^T Y + 2\lambda(\|\beta\|_1 - c).$$

Sustituyendo este último desarrollo en el problema de optimización dado en (2.10) se obtiene

$$\min_{\beta \in \mathbb{R}^{p-1}} Y^TY - 2\lambda c + 2 \underbrace{\left\{ \frac{1}{2} \beta^T X_1^T X_1\beta - \beta^T X_1^T Y + \lambda \|\beta\|_1 \right\}}. \quad (2.11)$$

Note que la expresión que se encuentra sobre la llave en (2.11) es similar al problema de optimización mencionado en (2.9), ya que (2.9) puede ser escrita como

$$\min_{\beta \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2} \beta^T W_{11}^{1/2} W_{11}^{1/2} \beta - \beta^T W_{11}^{1/2} W_{11}^{-1/2} s_{12} + \lambda \|\beta\|_1 \right\},$$

así, si se define $X_1 = W_{11}^{1/2}$ y $Y = W_{11}^{-1/2} s_{12}$ se obtiene $\hat{\beta}$ planteando regresión Lasso de Y sobre X_1 .

Además, de (2.7) se deriva que $w_{22} = \theta_{22}^{-1} + \theta_{22}^{-1} \theta_{12}^T W_{11} \theta_{12} \theta_{22}^{-1}$ y se había definido $\beta = -\theta_{12} \theta_{22}^{-1}$, de modo que

$$\begin{aligned} \theta_{22}^{-1} &= w_{22} - \beta^T W_{11} \beta \\ \Rightarrow \hat{\theta}_{22} &= (w_{22} - \hat{\beta}^T W_{11} \hat{\beta})^{-1} \end{aligned}$$

y

$$\begin{aligned} \hat{\theta}_{12} &= -\hat{\beta} \hat{\theta}_{22}, \\ \hat{w}_{12} &= W_{11} \hat{\beta}, \end{aligned}$$

con lo que se actualizó θ_{12} y θ_{22} , es decir, la última columna de Θ , así como w_{12} .

Resumiendo los procedimientos anteriores, y extendiéndolos de manera que se modifiquen todas las entradas de la matriz Θ , se obtiene el algoritmo Lasso Gráfico:

Algoritmo GLASSO:

1. Inicializar W como $S + \lambda I$.
2. Ciclar sobre las columnas repetidamente hasta convergencia:
 - (a) Rearreglar columnas y renglones de modo que la columna a actualizar sea la última.
 - (b) Resolver el problema de regresión Lasso de $Y = W_{11}^{-1/2} s_{12}$ sobre $X_1 = W_{11}^{1/2}$.
 - (c) Actualizar renglón y columna correspondiente de W usando w_{12} (w_{22} no se actualiza).
 - (c) Actualizar $\hat{\theta}_{12}$ y $\hat{\theta}_{22}$.

Una generalización del algoritmo Lasso Gráfico es sustituir al parámetro λ por una matriz simétrica, de modo que la regularización para cada arista sea distinta. El procedimiento se sigue de manera inmediata en el algoritmo anterior pero variando el valor de λ por los correspondientes elementos de la matriz. El método Lasso Gráfico es uno de los más populares en la literatura, pero existen otros métodos como el propuesto por Meinshausen y Bühlman (2006). Sin embargo, a lo largo de este trabajo de tesis se empleará el método Lasso Gráfico.

2.1.3. Estimación del parámetro de regularización

Para resolver el problema Lasso Gráfico el parámetro de regularización λ debe ser una cantidad fija. El resultado de la estimación de la matriz de precisión sparse depende de la elección que se haga de λ , por lo que establecer este parámetro es de suma relevancia. Si se utiliza $\lambda = 0$ entonces la expresión (2.4) se reduce a sólo considerar el negativo de la logverosimilitud, de modo que no se inducirá esparsidad en la matriz de precisión. Conforme el valor de λ sea mayor que 0, como resultado inherente del problema de Regresión Lasso, sucede que la matriz de precisión estimada tendrá mayor número de entradas igual a cero, es decir, la gráfica correspondiente aumentará su esparsidad. Esto demuestra que establecer el valor al que será igual λ es una tarea delicada.

En la literatura se pueden encontrar distintos métodos para establecer el parámetro de regularización, cada uno desde un enfoque diferente y con distintos objetivos. Una manera de definir un valor para λ es utilizar validación cruzada mediante el siguiente algoritmo presentado en Bien y Tibshirani (2011):

Fijar λ ,

- 1) Supongamos que se tienen los datos A y se particionan en K conjuntos, $A = \{A_1, A_2, \dots, A_K\}$.
- 2) Para cada $k = 1, 2, \dots, K$ resolver la optimización de (2.4) para el conjunto $A_k^c = A \setminus A_k$ obteniendo $\hat{\Theta}_{A_k^c}^\lambda$.
- 3) Calcular $\hat{\alpha}_{CV}(\lambda) = \frac{1}{K} \sum_{i=1}^K l(\hat{\theta}_{A_i^c}^\lambda; S_{A_i})$ donde $S_{A_i} = \frac{1}{|A_i|} \sum_{j \in A_i} (x_j - \bar{x}_{A_i})(x_j - \bar{x}_{A_i})^T$ y $l(\hat{\theta}_{A_i^c}^\lambda; S_{A_i}) = \log |\hat{\Theta}_{A_i^c}^\lambda| - \text{tr}(\hat{\Theta}_{A_i^c}^\lambda S_{A_i})$.

Realizar el procedimiento anterior para distintos valores de λ y escoger $\hat{\lambda}_{CV} = \arg \max_{\lambda} \hat{\alpha}_{CV}(\lambda)$.

Esta manera de seleccionar el valor de λ busca encontrar el valor del parámetro de regularización que arroje el valor más alto en la función de logverosimilitud sin regularización. Sin embargo, no hay ninguna relación clara entre el resultado que se obtenga y la calidad de la gráfica que se produce con el valor seleccionado. Otro método para establecer el parámetro de regularización es el procedimiento StARS (Stability Approach to Regularization Selection) introducido por Liu et al. (2010). Este método tiene como propósito minimizar la inestabilidad sobre si una arista está presente o no en la gráfica estimada. Además, resulta que utilizando este método, con probabilidad alta, la gráfica real está contenida en la gráfica estimada, al mismo tiempo que se estima la matriz de precisión tan sparse como sea posible. La razón por la que esta última propiedad es útil es debido a que en distintas áreas de aplicación de los modelos gráficos es aceptable obtener un falso positivo, a diferencia de los falsos negativos que no se toleran.

El procedimiento StARS es el siguiente: extraer M submuestras sin remplazo S_1, S_2, \dots, S_M de la muestra original x_1, x_2, \dots, x_N , cada una de tamaño $b = b(N) < N$. Basta con que M sea un número grande. Sea $\Lambda = 1/\lambda$. Para cada valor de Λ de entre posibles valores $\mathbf{G}_k = \{\Lambda_1, \Lambda_2, \dots, \Lambda_k\}$ estimar la matriz de precisión correspondiente a cada submuestra por medio del algoritmo Lasso Gráfico, obteniendo las matrices $\hat{E}_1^b(\Lambda), \hat{E}_2^b(\Lambda), \dots, \hat{E}_M^b(\Lambda)$. Luego, sea $\psi^\Lambda(\cdot)$ que denota el algoritmo Lasso Gráfico con parámetro Λ , sea

$$\psi_{st}^\Lambda(S_j) = \begin{cases} 1 & \text{si el algoritmo une a los nodos } s \text{ y } t \\ 0 & \text{si el algoritmo no une a los nodos } s \text{ y } t \end{cases}.$$

Se define $\theta_{st}^b(\Lambda) = P(\psi_{st}^\Lambda(S_j) = 1)$, que es la probabilidad de que la arista (s, t) esté presente en la gráfica estimada mediante el algoritmo Lasso Gráfico. Para estimar esta probabilidad se utiliza

$$\hat{\theta}_{st}^b(\Lambda) = \frac{1}{M} \sum_{j=1}^M \psi_{st}^\Lambda(S_j).$$

Sea $\xi_{st}^b(\Lambda) = 2\hat{\theta}_{st}^b(\Lambda)(1 - \hat{\theta}_{st}^b(\Lambda))$ y sea

$$\hat{\xi}_{st}^b(\Lambda) = 2\hat{\theta}_{st}^b(\Lambda)(1 - \hat{\theta}_{st}^b(\Lambda)) = 2\frac{1}{M} \sum_{j=1}^M \psi_{st}^\Lambda(S_j) \left(1 - \frac{1}{M} \sum_{j=1}^M \psi_{st}^\Lambda(S_j)\right)$$

su estimador. Note que, además de que $\xi_{st}^b(\Lambda)$ es dos veces la varianza de la variable aleatoria Bernoulli $\psi_{st}^\Lambda(S_j)$, también es una medida de la presencia de la arista (s, t) en cada submuestra en el siguiente sentido. Si la gráfica obtenida para cada submuestra es la misma entonces $\hat{\theta}_{st}^b(\Lambda)$ es igual a 1 (si la arista (s, t) estuvo siempre presente) o 0 (si la arista (s, t) nunca estuvo presente), y en estos dos casos $\hat{\xi}_{st}^b(\Lambda)$ es igual a 0. De este modo $\xi_{st}^b(\Lambda)$ es una medida de cuánto varió la presencia o ausencia de la arista (s, t) en la gráfica de cada submuestra y $0 \leq \xi_{st}^b(\Lambda) \leq 1/2$. A esta medida se le llama *inestabilidad*. La inestabilidad total se define como:

$$\hat{D}_b(\Lambda) = \frac{\sum_{s < t} \hat{\xi}_{st}^b}{\binom{p}{2}}.$$

Como ya se había mencionado, si el valor del parámetro de regularización aumenta, el algoritmo Lasso Gráfico estima una matriz de precisión con mayor número de entradas igual a cero. En el límite, si λ tiende a infinito entonces se tiene una matriz igual a la matriz cero. Así, si $\Lambda = \frac{1}{\lambda} = 0$ entonces $\hat{D}_b(0) = 0$ ya que para todas las submuestras su gráfica correspondiente nunca tendrá la presencia de alguna arista. Conforme Λ aumente, $\hat{D}_b(\Lambda)$ tenderá a ser mayor, pero si Λ es muy grande, es decir, λ tiende a 0, entonces la gráfica correspondiente tendrá la presencia de casi todas las aristas, por lo que $\hat{D}_b(\Lambda)$ tenderá a disminuir.

Al tener interés en la estabilidad de gráficas sparse se define $\bar{D}_b(\Lambda) = \sup_{\{t \in \mathbf{G}_k, t \leq \Lambda\}} \hat{D}_b(t)$. Finalmente, la estimación de Λ utilizando el procedimiento StARS es $\hat{\Lambda}_s = \sup\{\lambda : \bar{D}_b(\Lambda) \leq \beta\}$. Los autores de este procedimiento sugieren establecer $\beta = 0.05$ y $b = \lfloor 10\sqrt{N} \rfloor$.

2.1.4. Penalización aleatoria

Como se explicó en la sección 2.1.2 del presente escrito, en la estimación de la matriz de precisión sparse mediante el algoritmo Lasso Gráfico se realiza de manera iterativa regresión Lasso. Regresión Lasso, al ser un método que permite realizar la selección del modelo de los datos concernientes, requiere ser estudiado con el fin de saber en qué casos el modelo seleccionado es el verdadero. Esta cuestionante fue analizada en Zhao y Yu (2006). Ellos encontraron dos condiciones suficientes y necesarias, cada una en un escenario, para que la regresión Lasso seleccionara al verdadero modelo, a las cuales llamaron *Condiciones Irrepresentables*.

La regresión Lasso de Y_N de tamaño $N \times 1$ sobre X_N de tamaño $N \times p$, encuentra el vector $\hat{\beta}^N$, de tamaño $p \times 1$ de la siguiente manera

$$\hat{\beta}^N(\lambda) = \arg \min_{\beta} \|Y_N - X_N \beta\|_2^2 + \lambda \|\beta\|_1. \quad (2.12)$$

En Zhao y Yu (2006) se dan dos definiciones de consistencia de Lasso. La primera de ellas dicta que Lasso es *Strongly Sign Consistent* si existe un parámetro λ_N que es función de N e independiente de Y_N o X_N tal que

$$\lim_{N \rightarrow \infty} P(\hat{\beta}^N(\lambda_N) =_s \beta^N) = 1,$$

donde $=_s$ significa igualdad en signo y $signo(\cdot)$ mapea las entradas positivas al 1, las negativas al -1 y los ceros al cero; y β^N denota al verdadero modelo. La segunda definición de consistencia de Lasso describe que éste es *General Sign Consistent* si

$$\lim_{N \rightarrow \infty} P(\exists \lambda \geq 0, \hat{\beta}^N(\lambda) =_s \beta^N) = 1.$$

Así, estas definiciones describen la consistencia de Lasso en el sentido de que, en el límite, este método elige con probabilidad 1 al verdadero modelo.

Las Condiciones Irrepresentables son las siguientes. Considere a β^N ordenado de manera que las primeras q entradas sean distintas de 0, $\beta_{(1)}^N$, y las restantes $p - q$ entradas sean iguales

a cero, $\beta_{(2)}^N$. Con X_N particionada de la misma manera, es decir, en las primeras p y últimas $p - q$ columnas, se deriva la partición de $C_N = (1/N)X_N^T X_N$ como

$$C^N = \begin{pmatrix} C_{11}^N & C_{12}^N \\ C_{21}^N & C_{22}^N \end{pmatrix},$$

Si se supone que C_{11}^N es invertible entonces la *Condición Fuerte Irrepresentable* es que existe un vector de constantes positivo η tal que

$$|C_{21}^N (C_{11}^N)^{-1} \text{signo}(\beta_{(1)}^N)| < 1 - \eta,$$

mientras que la *Condición Débil Irrepresentable* es

$$|C_{21} (C_{11})^{-1} \text{signo}(\beta_{(1)}^N)| < 1,$$

donde las desigualdades se satisfacen entrada a entrada.

Bajo dos escenarios, para p y q pequeños y para grandes p y q , y algunas condiciones de regularidad extras, se encontró que estas dos condiciones son equivalentes a las definiciones de consistencia. Para mayor detalle consultar Zhao y Yu (2006). Así, este estudio permitió saber cuándo el método Lasso tiene el desempeño deseado.

Más tarde, Meinshausen y Bühlman (2010) propusieron una manera de asegurar que el método Lasso seleccionara el modelo real aún cuando estas condiciones no se satisfacen. Este método recibe el nombre de *Penalización Aleatoria*, que consiste en penalizar cada entrada de β en (2.12) de manera distinta y aleatoria. En el artículo concerniente se estudian resultados teóricos de esta propuesta, así como los resultados en estudios de simulación. Estas ideas fueron implementadas en el algoritmo de Narayan et al. que se detalla en el siguiente capítulo.

Así, se completa la teoría introductoria sobre los modelos gráficos no-dirigidos necesaria para entender que la naturaleza del problema abordado en esta tesis sugiere utilizar estos modelos para llevar a cabo el objetivo de interés.

2.2. Pruebas de hipótesis múltiples

En el área de Estadística las pruebas de hipótesis desempeñan un rol de particular importancia. Son una de las herramientas más populares y más utilizadas en casi todas las ciencias, por mencionar algunas se encuentran la medicina, las ciencias sociales, la biología o la química. El control del error tipo I es un tópico clave y de gran sensibilidad. Pero este criterio es más común en los análisis en los que se está interesado en probar una sola hipótesis. Sin embargo, existen otros tipos de problemas en los que surgen las pruebas de hipótesis múltiples, para las cuales es necesario plantear un criterio distinto. En esta sección se abordará el tema de las pruebas de hipótesis múltiples, el cual es de gran importancia para el estudio de los datos fMRI que tiene como principal objetivo esta tesis. La sección se desarrollará desde el punto de vista de Efron (2010) e Iturriaga (2011) quienes ofrecen una revisión del tema, así como haciendo referencia a Benjamini y Hochberg (1995) y Benjamini y Yekutieli (2001).

Supongamos que se quieren probar simultáneamente m hipótesis nulas $H_1^0, H_2^0, \dots, H_m^0$ basados en un conjunto de datos X . Para extender el error tipo I que se utiliza en el contraste de una sola hipótesis nula, el primer concepto introducido fue el de *FWER* (family-wise error rate) que es la probabilidad de realizar uno o más *falsos descubrimientos*, es decir, la probabilidad de cometer al menos un error de tipo I en el conjunto de contrastes. Más formalmente, sea V la variable aleatoria que denota el número total de falsos descubrimientos, o mejor dicho, el número total de hipótesis nulas verdaderas que fueron rechazadas, entonces la FWER se define como

$$FWER = P(V > 0).$$

El método más popular para realizar la corrección por pruebas de hipótesis múltiples sujetos a $FWER \leq \alpha$ es el de Bonferroni. Si p_1, p_2, \dots, p_m son los p -valores correspondientes a las m hipótesis, entonces este método consiste en rechazar las hipótesis nulas tales que $p_i \leq \alpha/m$. Notemos que si el número de contrastes a realizar, m , es muy grande entonces la cantidad α/m será muy pequeña. En el siguiente ejemplo, tomado de Iturriaga (2011), se expone el porqué controlar la FWER al realizar pruebas de hipótesis múltiples puede resultar no ser la decisión adecuada.

Ejemplo 2.7. Sea $V = \sum_{i=1}^m V_i$ donde V_i es la función indicadora que toma el valor de 1 si se realizó un error de tipo I para la hipótesis H_i^0 y 0 en otro caso, $i = 1, 2, \dots, m$. Suponga que las V_i 's son variables aleatorias independientes y que el nivel de significancia utilizado para cada uno de los m contrastes es α , o sea, $P(V_i = 1) = \alpha$. Entonces

$$\begin{aligned} FWER &= P(V > 0) = 1 - P(V = 0) \\ &= 1 - P\left(\sum_{i=1}^m V_i = 0\right) \\ &= 1 - P(V_1 = 0, V_2 = 0, \dots, V_m = 0) \\ &= 1 - (P(V_1 = 0))^m \\ &= 1 - (1 - P(V_1 = 1))^m \\ &= 1 - (1 - \alpha)^m. \end{aligned}$$

Si se tiene un número grande de hipótesis a contrastar, es decir m es grande, note del Ejemplo 2.7 que la probabilidad de cometer al menos un falso descubrimiento tiende a 1. Por lo que, si se quiere contrastar un gran número de hipótesis, un procedimiento de corrección por pruebas de hipótesis múltiples basado en el control de la FWER no sería conveniente.

2.2.1. Procedimiento Benjamini-Hochberg

En la mayoría de los casos, no sólo se está interesado en controlar la FWER, sino que se quisiera controlar el número total de falsos descubrimientos. En 1995 Benjamini y Hochberg propusieron el concepto de *tasa de falsos descubrimientos* (FDR, False Discovery Rate) que es la proporción esperada de errores dentro de todos los rechazos. Para definir de manera formal la FDR se utilizará la siguiente tabla, con la notación de Benjamini y Hochberg (1995).

	Decisión		
	No rechazo	Rechazo	Total
Hipótesis nulas verdaderas	U	V	m_0
Hipótesis nulas falsas	T	S	$m - m_0$
	$m - R$	R	m

Tabla 1. Número de errores cometidos cuando se contrastan m hipótesis.

Bajo la notación de la Tabla 1, la FWER tiene la misma expresión que antes. Note también que la variable aleatoria R es observable, mientras que U, V, S y T son variables aleatorias no observables, ya que no se conoce m_0 , es decir, el número real de hipótesis nulas verdaderas. La proporción de falsos descubrimientos cometidos, es decir, la proporción de errores que se cometieron al rechazar una hipótesis es

$$Q = \frac{V}{V + S} = \frac{V}{R}.$$

La variable aleatoria Q tampoco es observable y se define como 0 si $R = 0$. La FDR se define como la esperanza de Q :

$$FDR = E(Q) = E\left(\frac{V}{R}\right).$$

La tasa FDR tiene propiedades interesantes que son probadas en Benjamini y Hochberg (1995). A continuación se enuncian dos de ellas:

- Si todas las hipótesis nulas son verdaderas, es decir $S = 0$, entonces controlar la FWER es equivalente a controlar la FDR ya que si $S = 0$, y por lo tanto $V = R$, entonces $V = 0$ implica $Q = 0$, mientras que $V > 0$ implica $Q = 1$. De modo que $E(Q) = P(V > 0)$.
- Si $m_0 < m$ entonces $FDR \leq FWER$.

La segunda propiedad quiere decir que, en el caso en el que no todas las hipótesis nulas son ciertas, si se controla la FWER entonces también se controla la FDR , pero no necesariamente al revés.

En Benjamini y Hochberg (1995) se propone el siguiente algoritmo para realizar la corrección por pruebas de hipótesis múltiples basadas en el control de la FDR cuando los estadísticos empleados para el contraste son independientes, la demostración se omite en este escrito pero el lector puede encontrarla en el artículo de Benjamini y Hochberg (1995). Sean $H_1^0, H_2^0, \dots, H_m^0$ las hipótesis nulas a contrastar y p_1, p_2, \dots, p_m sus correspondientes p -valores. Sean $p_{(1)}, p_{(2)}, \dots, p_{(m)}$ los p -valores ordenados y sea $H_{(i)}^0$ la hipótesis nula correspondiente a $p_{(i)}$. Sea

$$k = \max\{i : p_{(i)} \leq \frac{i}{m}q\},$$

entonces rechazar $H_{(1)}^0, \dots, H_{(k)}^0$. Si tal i no existe entonces no realizar ningún rechazo. Así, para cualquier configuración de hipótesis nulas falsas, es decir, sin importar si $m_0 = m$ o si $m_0 < m$, este procedimiento controla la FDR al nivel q , o lo que es lo mismo,

$$FDR = E(Q) \leq \frac{m_0}{m}q \leq q.$$

2.2.2. Procedimiento Benjamini-Yekutieli

Una limitante del algoritmo Benjamini-Hochberg es que el resultado se enuncia como verdadero únicamente cuando los estadísticos con los que se llevan a cabo las pruebas son independientes. Años más tarde, en Benjamini y Yekutieli (2001) se probó que si los estadísticos de prueba son continuos y conjuntamente independientes entonces la FDR es estrictamente igual a $\frac{m_0}{m}q$ y, más aún, demostraron que si se sustituye la hipótesis de independencia entre los estadísticos de prueba por que sólo cumplan la propiedad de *dependencia de regresión positiva en cada uno de los estadísticos de prueba del subconjunto I_0* , donde I_0 es el subconjunto de estadísticos que apoyan a las verdaderas hipótesis nulas (PRDS, positive regression dependency on each one from a subset I_0) se cumple que $FDR \leq \frac{m_0}{m}q$.

A continuación se explica la propiedad PRDS. Un subconjunto D se dice ser creciente si para cualquier $x \in D$ se cumple que $y \in D$ si $y \geq x$. Sea $Y = \{Y_1, Y_2, \dots, Y_m\}$ el vector de estadísticos de prueba, entonces Y satisface ser PRDS en I_0 si para cualquier conjunto creciente D y para cualquier $i \in I_0$ se satisface que $P(Y \in D | Y_i = x)$ es no decreciente en x . Para interpretar el significado de esta propiedad recuerde que en el procedimiento Benjamini-Hochberg se ordenan los p -valores de menor a mayor y se rechazan las hipótesis que son menores a cierto índice; la propiedad PRDS sugiere que la probabilidad de que un estadístico de prueba sea parte del conjunto de estadísticos que apoyan la hipótesis nula es menor para el estadístico con menor p -valor y aumenta de ahí en adelante.

Pero el resultado más fuerte demostrado en el artículo de Benjamini y Yekutieli (2001) no es el mencionado anteriormente sino que se probó que, para el caso general de dependencia entre los estadísticos de prueba, si en el procedimiento Benjamini-Hochberg se sustituye la definición de k por

$$k' = \max\{i : p_{(i)} \leq \frac{i}{m} \frac{q}{\sum_{j=1}^m 1/j}\},$$

entonces $FDR \leq \frac{m_0}{m}q$. Así, el procedimiento modificado de Benjamini-Hochberg, al cual se le conoce como Benjamini-Yekutieli, resulta tener robusticidad en el sentido de que la FDR se controla para cualquier estructura de dependencia entre los estadísticos de prueba.

A partir de la introducción de la definición de la FDR han surgido una amplia gama de procedimientos para controlar distintos criterios diferentes a esta tasa, incluso el problema de las pruebas de hipótesis múltiples se ha abordado con un enfoque bayesiano. Cabe destacar que en la literatura de pruebas de hipótesis múltiple existe una amplia discusión sobre cuál es la mejor tasa a controlar para llevar a cabo una prueba de hipótesis múltiples. Una recomendación es entender específicamente el problema que se está abordando.

En el siguiente capítulo se motiva y plantea de manera formal el contraste de hipótesis para contestar al objetivo principal en esta tesis, el cual busca comparar las relaciones de conectividad funcional de dos grupos de individuos. Además, se detalla el procedimiento a seguir para poder realizar dicho contraste.

Capítulo 3

Contraste de hipótesis en GGMs

En el capítulo anterior se estudiaron los modelos Gráficos Gaussianos debido a que estos modelos permiten directamente estudiar las relaciones de correlación parcial entre variables. Específicamente en esta tesis se tiene el interés de conocer las diferencias que puedan existir en las relaciones de conectividad funcional para dos grupos de individuos a través de datos fMRI. Los grupos de estudio son, por un lado, individuos que presentan trastorno del espectro austista (ASD, Autism Spectrum Disorder) y, por el otro, un grupo control. Los datos fMRI consisten en una serie de imágenes del cerebro tomadas en un período de tiempo. En este capítulo se detallarán los datos con los que se trabajó, de modo que su estructura sea la motivación del método empleado para el problema de interés.

El problema básico a considerar consiste en abordar el tema de la conectividad entre las distintas regiones del cerebro, es decir, la asociación o relación de estas, que de manera natural se hace a través de la matriz de precisión de un modelo gráfico. Así, se está interesado en, planteándolo de manera general, realizar la prueba de hipótesis

$$\mathbb{H}_0 : \Theta_A = \Theta_B \quad vs \quad \mathbb{H}_1 : \Theta_A \neq \Theta_B, \quad (3.1)$$

donde Θ_A y Θ_B son las matrices de precisión asociadas a la gráfica del grupo A (grupo con ASD) y del grupo B (grupo control), respectivamente. El contraste de hipótesis planteado en (3.1) tiene que ser reformulado dada la estructura de los datos fMRI y los desafíos estadísticos que tienen que ser atendidos. En el Capítulo 2 se detalló el algoritmo Lasso Gráfico, el cual estima la gráfica subyacente de un conjunto de datos dado el parámetro de regularización λ . Sin embargo, se necesita que las observaciones sean independientes. Como se detallará en la siguiente sección, las observaciones en los datos fMRI son dependientes en el tiempo, por lo que debe utilizarse una técnica de blanqueamiento para poder usar el algoritmo Lasso Gráfico. El parámetro de regularización es crucial para la estimación de la gráfica, por lo que merece ser atendido a detalle. Meinshausen y Bühlman (2010) recientemente introdujeron la *Penalización Aleatoria*, que es una alternativa para fijar el parámetro de regularización con propiedades muy interesantes. Además, en el desafío de plantear una adecuada prueba de hipótesis se debe atender el hecho de que se tienen diferencias no sólo entre las gráficas de cada grupo, sino que también hay diferencias entre las gráficas de los individuos dentro de un grupo. Observe por ejemplo la siguiente figura.

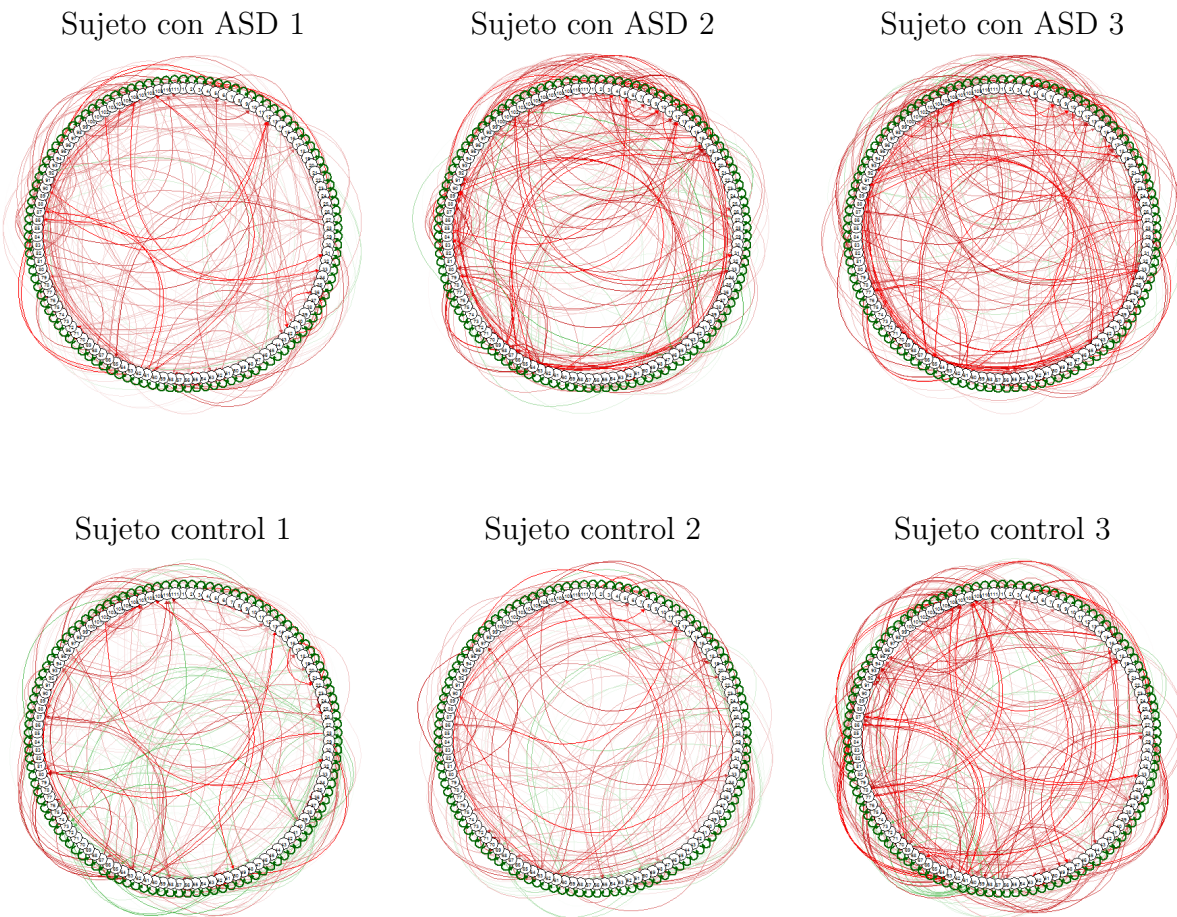


Figura 2. Gráfica estimada de 6 distintos sujetos.

En la Figura 2 se ilustran las gráficas de 6 distintos individuos, 3 del grupo con ASD y 3 del grupo control. En el perímetro del círculo se encuentran los nodos de la gráfica mientras que las líneas rojizas representan las aristas. Las gráficas fueron estimadas utilizando el método Lasso Gráfico y el algoritmo StARS. Note de las figuras que no sólo son diferentes las gráficas entre el grupo que presenta ASD y el grupo control, sino que para cada individuo la gráfica varía. Así, esta observación debe ser contemplada en el planteamiento de la prueba de hipótesis que interesa. Además, también se debe considerar la varianza que se deriva del hecho de que las gráficas son estimadas para cada individuo y no solamente observadas, lo que puede ser atendido mediante remuestreo.

Una manera de replantear el contraste de hipótesis en la expresión (3.1) es la que se plantea en Narayan et al. (2015). Así, en este capítulo se detalla el método introducido en dicho estudio. Se explica primero la estructura de los datos fMRI, para después dar paso al planteamiento del contraste de hipótesis. Luego se detalla el procedimiento a seguir para realizar la estimación de la matriz de precisión por individuo, donde se incluye el blanqueamiento de los datos, remuestreo y Penalización aleatoria. Por último, se presenta el estadístico de prueba y el algoritmo resumen.

3.1. Estructura de los datos de fMRI

Los datos fMRI consisten en una serie de imágenes por Resonancia Magnética funcional tomadas al cerebro. Las imágenes son tomadas una cantidad determinada de veces con un tiempo de espaciamiento fijo. Una manera de trabajar con los datos es particionando el cerebro, bajo cierto criterio, en nodos o regiones de interés (ROIs, Regions of Interest), para después promediar las series de tiempo que están involucradas en un nodo y de esta forma tener la información para las regiones de la partición. Las particiones hechas al cerebro reciben el nombre de Atlas y existen estudios en los que se proponen distintas maneras de configuración de las regiones según distintos intereses. Por mencionar algunos se encuentran el Atlas de Harvard-Oxford presentado en Fischl et al. (2004) que particiona al cerebro de forma anatómica o el Atlas de Power et al. (2011) que hace la partición por regiones funcionales.

Los datos de fMRI que se utilizan en esta tesis fueron extraídos del proyecto “Preprocessed Connectomes Project (PCP)” el cual ofrece de manera pública una recopilación de un total de 16 instituciones que comparten sus imágenes por Resonancia Magnética funcional de individuos que presentan trastorno de espectro autista (ASD) e individuos control. Dichos datos recopilados llevan por nombre “Autism Brain Imaging Data Exchange (ABIDE)” (Craddock et al. (2013a)).

En particular, en esta tesis se usan los datos de fMRI en estado de reposo de la Universidad de California en Los Ángeles (UCLA) con el preprocesamiento, ofrecido en la misma página de ABIDE, C-PAC (Craddock et al. (2013b)) y con filtro de paso de banda y regresión de señal global. Este tratamiento de preproceso es explicado con mayor detalle en el Capítulo 4, en donde se hablará de la aplicación del método descrito en este capítulo. Se eligió el Atlas de Harvard-Oxford que desemboca en 111 regiones de interés (ROIs) por individuo. Las imágenes fueron tomadas en 116 tiempos cada una con 3 segundos de espaciamiento.

Los datos fMRI en estado de reposo que ofrece la UCLA son de un total de 63 individuos, de los cuales 37 sujetos pertenecen al grupo con ASD y 26 sujetos pertenecen al grupo control. Cada sujeto tiene la información de 111 regiones y 116 puntos en el tiempo. Así, para cada individuo se tienen 116 observaciones, cada una de dimensión 111 (número de ROIs). Note en particular que las observaciones son dependientes en el tiempo, como se había mencionado anteriormente, por lo que en las siguientes secciones esta observación debe considerarse.

3.2. Planteamiento del contraste de hipótesis

Los datos de fMRI tienen una estructura muy específica. Como se mencionó anteriormente, por individuo se tienen observaciones que son dependientes en el tiempo. Sin embargo, en el estudio de interés, se quiere indagar sobre la dependencia espacial entre las distintas ROIs en las que se encuentra particionado el cerebro. De esta forma, la estructura natural de los datos fMRI obliga a que el planteamiento del problema básico en (3.1) se adecue y sea replanteado.

Un primer fundamento para plantear un método adecuado, es el hecho de que se tie-

ne evidencia de que los modelos Gráficos Gaussianos (GGMs) son adecuados para estimar las relaciones de conectividad funcional dadas por las señales de las imágenes fMRI, como se argumenta en Varoquaux et al. (2010). Además, hay distintas cuestiones que deben ser atendidas si se quiere realizar la prueba de hipótesis general descrita en (3.1). Dichas cuestiones tienen respuesta en el método presentado en Narayan et al. (2015), el cual se explica en lo siguiente de este capítulo.

Considere primero la estructura general de los datos: sea n_A y n_B el número de individuos pertenecientes al grupo con ASD (grupo A) y al grupo control (grupo B) respectivamente, sea $n = n_A + n_B$ y $X_{T \times p}^1, \dots, X_{T \times p}^n$ los datos de fMRI de todos los individuos, donde T es el número de puntos en el tiempo en los que fueron tomadas las imágenes y p el número de ROIs. Se asume que cada sujeto sigue un modelo Gráfico Gaussiano, es decir,

$$x_j^i \sim^{iid} N(0, (\Theta^i)^{-1}) \quad \forall j = 1, \dots, T,$$

para todo $i = 1, \dots, n$. Además, se asume que el *soporte* de cada arista en la estructura de la gráfica sigue una probabilidad que depende del grupo al que pertenece el individuo, y no al individuo en sí, es decir

$$Y_{k,l}^i = I(\theta_{k,l}^i \neq 0) \sim \text{Bern}(\pi_{k,l}^g) \quad \forall i \in \mathbb{G}_g \ \& \ \forall 1 \leq k < l \leq p,$$

donde $g \in \{A, B\}$ y \mathbb{G}_g denota a los individuos del grupo g . Entonces, note que $Y_{k,l}^i$ modela el que la arista (k, l) esté presente o no en la gráfica del individuo i con una probabilidad que depende, además de la arista, del grupo al que pertenece tal individuo. Así $\pi_{k,l}^g$ es el soporte de la arista (k, l) .

Con este modelo se da paso al planteamiento de las hipótesis que se desean comparar, que consisten en realizar una prueba sobre la diferencia en el soporte de cada arista (k, l) de los dos grupos mediante el contraste de las siguientes hipótesis:

$$\mathbb{H}_0 : \pi_{k,l}^A = \pi_{k,l}^B \quad vs \quad \mathbb{H}_1 : \pi_{k,l}^A \neq \pi_{k,l}^B, \quad \forall 1 \leq k < l \leq p. \quad (3.2)$$

De esta manera, en lugar de plantear las hipótesis como la comparación de dos matrices de precisión, cada una concerniente a la gráfica de los dos grupos de sujetos, se compara el soporte de cada arista. La razón por la que se recurre a este cambio es debido a que asumir que por grupo se tiene una sola gráfica resulta sumamente restrictivo, ya que aún dentro de los grupos los individuos presentan variabilidad entre ellos, como se había mencionado en las observaciones de la Figura 2.

Otro factor por el que se replantea la prueba de hipótesis de esta forma, es el hecho de que de esta manera se puede atender un segundo tipo de variabilidad, el cual consiste en tener en cuenta que la gráfica de un individuo es *estimada* y no solamente observada. La gráfica es seleccionada utilizando técnicas de esparsidad ya que se está interesado en contrastar el soporte sparse de las matrices de precisión, además de que los datos con que se cuenta tienen una dimensión del mismo orden que el número de observaciones, por lo que se requiere utilizar estimación por máxima verosimilitud regularizada. Incorporar en la inferencia la variabilidad inducida por la selección de un modelo basado en los datos es un tema de reciente estudio

llamado *Inferencia después de la selección del modelo*. Este enfoque tiene como objetivo realizar estimaciones con mayor exactitud. Para obtener un mayor panorama sobre este tema se puede consultar Efron y Hastie (2016).

Específicamente para esta tesis, la inferencia después de la selección del modelo consiste en que se desea realizar un problema de dos niveles. El primero de ellos es un procedimiento de selección de variables para la estimación de parámetros de cada individuo, es decir, se trata de la selección de la gráfica para estimar la conexión funcional en cada individuo. El segundo nivel corresponde a la inferencia que se lleva a cabo cuando se contrasta el soporte de una arista entre los dos grupos de sujetos. Así pues, se tiene el interés de incorporar la variabilidad que se produce cuando se hace la selección de la gráfica por individuo. Además, otra motivación para incorporar esta variabilidad proviene del hecho de que en estudios en los que se conduce la extracción de imágenes en un mismo sujeto en repetidas ocasiones se tiene el resultado de que las gráficas estimadas en cada sesión tienen variabilidad entre ellas, como se muestra en Wang et al. (2011)

Para modelar las dos fuentes de variabilidad, la dada por la diferencia en la gráfica de un individuo a otro y la que surge en la selección de la gráfica de un individuo, se da paso a modelar a la variable $Y_{k,l}^i$ con un modelo jerárquico. Un modelo natural para describir datos binarios, en donde la probabilidad de éxito es también aleatoria, es el modelo Beta-Binomial. En este caso se puede entender como éxito el que la arista (k, l) esté presente en la gráfica. El modelo Beta-Binomial se plantea, en el problema que se tiene de interés, de la siguiente manera.

Denote por $\mu_{k,l}^i$ la probabilidad de que la arista (k, l) esté presente o no en la gráfica del individuo i , es decir, $\mu_{k,l}^i = P(\theta_{k,l}^i \neq 0)$. Entonces se modela $Y_{k,l}^i \sim \text{Bern}(\mu_{k,l}^i)$. Pero, como ya se había mencionado, $\mu_{k,l}^i$ es una variable aleatoria, que depende del grupo al cual pertenece el individuo i , por lo que se asume $\mu_{k,l}^i \sim \text{Beta}(\alpha_{k,l}^g, \beta_{k,l}^g)$.

Una reparametrización común en el modelo Beta-Binomial es definir $\pi_{k,l}^g = \frac{\alpha_{k,l}^g}{\alpha_{k,l}^g + \beta_{k,l}^g}$, de manera que $\pi_{k,l}^g = \mathbb{E}(\mu_{k,l}^i)$, y definir $\rho_{k,l}^g = \frac{1}{\alpha_{k,l}^g + \beta_{k,l}^g + 1}$ el cual es un parámetro que tiene relación con la varianza de $\mu_{k,l}^i$, ya que

$$\text{Var}(\mu_{k,l}^i) = \frac{\alpha_{k,l}^g \beta_{k,l}^g}{(\alpha_{k,l}^g + \beta_{k,l}^g + 1)(\alpha_{k,l}^g + \beta_{k,l}^g)^2} = \rho_{k,l}^g \pi_{k,l}^g (1 - \pi_{k,l}^g).$$

Suponiendo que se tienen m observaciones de este modelo y tomando $Z_{k,l}^i$ igual a la suma de estas observaciones, es decir, $Z_{k,l}^i = \sum_{j=1}^m Y_{k,l}^{j,i}$ se tiene el modelo Beta-Binomial:

$$Z_{k,l}^i | \mu_{k,l}^i \sim \text{Bin}(\mu_{k,l}^i, m), \quad \mu_{k,l}^i \sim \text{Beta}(\pi_{k,l}^g, \rho_{k,l}^g). \quad (3.3)$$

El modelo Beta-Binomial, al modelar datos binarios cuya probabilidad de éxito es una variable aleatoria, provoca un resultado que es conveniente y necesario para el problema que se está tratando. Este resultado, demostrado en Kleinman (1973), es el hecho de que la varianza no condicional de $Z_{k,l}^i$ tiene una cantidad extra si se compara con la que se tendría asumiendo a $Z_{k,l}^i$ como solamente la suma de variables binarias sin considerar la probabilidad de éxito como variable:

$$\text{Var}(Z_{k,l}^i) = m\pi_{k,l}^g(1 - \pi_{k,l}^g) + \underbrace{m(m-1)\rho_{k,l}^g\pi_{k,l}^g(1 - \pi_{k,l}^g)}_{}, \quad (3.4)$$

note que el término que se encuentra sobre la llave en la expresión (3.4) es la varianza extra por considerar la variabilidad en la estimación de la gráfica de un individuo, por lo que si no se tuviera en cuenta este hecho entonces se estaría subestimando la varianza de $Z_{k,l}^i$.

3.3. Estimación de la matriz de precisión por individuo

Inicialmente se está interesado en estimar la matriz de precisión de cada individuo, es decir, su gráfica de conexión funcional. Para un individuo cualquiera i se tienen T observaciones cada una de dimensión p . Una manera para estimar la matriz de precisión fue explicada en la sección 2.1.2 de este escrito, en donde se derivó el algoritmo GLASSO. Note que el algoritmo es aplicable cuando las observaciones que se tienen son independientes ya que de este modo es como se llega a la expresión (2.3). Sin embargo, para cada individuo, en el estudio de interés, se tienen observaciones que son dependientes en el tiempo. Debido a la observación anterior se debe realizar primero un blanqueamiento a las series de tiempo para así obtener observaciones independientes.

A continuación se expondrá el método utilizado para el blanqueamiento de las series de tiempo. También se tratará el tema de remuestreo, el cual es una solución que se ofrece para poder considerar la variabilidad derivada de la selección de la gráfica basada en los datos. Además, se detallará la manera en la que se establece el parámetro de regularización.

3.3.1. Blanqueamiento de las series de tiempo

El *blanqueamiento* es una transformación que se realiza a un conjunto de variables aleatorias con el fin de producir nuevas variables ortogonales. Escrito de otra manera, con notación de Kessy et al. (2016), el blanqueamiento consiste en transformar un vector aleatorio $Y = (Y_1, Y_2, \dots, Y_d)$ con matriz de covarianzas $\text{Var}(Y) = \Sigma$ en un nuevo vector aleatorio $Z = (Z_1, Z_2, \dots, Z_d) = WY$ tal que la varianza de Z sea la identidad, $\text{Var}(Z) = I$. A la matriz W se le llama matriz de blanqueo. Así, las variables blanqueadas Z son no correlacionadas y además tienen varianza unitaria.

Note entonces que se debe cumplir

$$\begin{aligned} \text{Var}(Z) &= \text{Var}(WY) \\ &= W\Sigma W^T \\ &= I \\ \Rightarrow (W\Sigma W^T)W &= W, \end{aligned} \quad (3.5)$$

la última igualdad en (3.5) se cumple si W es tal que

$$W^T W = \Sigma^{-1}. \quad (3.6)$$

Sin embargo, esta condición no determina únicamente a W , como se menciona en Kessy et al. (2015). Así, dada Σ , se tienen infinitas W que satisfacen (3.6), es decir, infinitos vectores aleatorios Z . Es por esta razón que existen distintos métodos para realizar el blanqueamiento de un vector aleatorio.

Uno de dichos métodos es el blanqueamiento ZCA (zero-phase components analysis) o también conocido como blanqueamiento de Mahalanobis. Este blanqueamiento consiste en definir W como

$$W = \Sigma^{-1/2},$$

donde la raíz cuadrada de Σ se calcula usando su descomposición espectral, $\Sigma = VUV^T$, donde V es la matriz de vectores propios ortogonales de Σ y $U = \text{diag}(u_1, \dots, u_d)$ con u_1, \dots, u_d sus valores propios, de modo que $\Sigma^{1/2} = VU^{1/2}V$ donde $U^{1/2} = \text{diag}(u_1^{1/2}, \dots, u_d^{1/2})$.

Este procedimiento es empleado en Worsley et al. (2002) para el tratamiento de datos fMRI. En este artículo se menciona que el blanqueamiento se hace para cada nodo por separado de la siguiente manera. Suponga que la respuesta fMRI en un vóxel en particular al tiempo t está dado por $x(t)$, de modo que se tiene $x_i = x(t_i)$ para cada tiempo t_1, t_2, \dots, t_T . Así, el dato observado fMRI al tiempo i es

$$y_i = x_i\beta + e_i,$$

donde β es un parámetro que varía de vóxel a vóxel y e_i es un error aleatorio.

Los errores e_i , $i = 1, \dots, T$, no son independientes en el tiempo. Uno de los modelos más simples para modelar dicha dependencia es un modelo autorregresivo de primer orden AR(1):

$$e_i = \rho e_{i-1} + \xi_i,$$

donde $|\rho| < 1$ y $\{\xi_i\}_{i=1}^T$ son variables aleatorias independientes e idénticamente distribuidas, $\xi_i \sim N(0, \sigma^2)$. Este modelo puede ser extendido a un modelo autorregresivo de orden $p > 1$.

Mediante las ecuaciones de Yule-Walker, ver Brockwell y Davis (1991), se estima

$$\text{Cov}(e_t, e_{t+h}) = \Sigma(h),$$

de modo que se obtiene la estimación de la matriz de covarianzas $\Sigma = (\Sigma_{ij})_{i,j=1}^T$, donde $\Sigma_{ij} = \Sigma(|i-j|)$. Así, sea $Y = (y_1, y_2, \dots, y_T)'$ y $e = (e_1, e_2, \dots, e_T)'$, haciendo uso del blanqueamiento ZCA se obtiene lo siguiente:

$$\begin{aligned} \text{Var}(\Sigma^{-1/2}Y) &= \Sigma^{-1/2}\text{Var}(Y)\Sigma^{-1/2} \\ &= \Sigma^{-1/2}\text{Var}(e)\Sigma^{-1/2} \\ &= \Sigma^{-1/2}\Sigma\Sigma^{-1/2} \\ &= I. \end{aligned}$$

Por lo que $Z = (z_1, z_2, \dots, z_T)' = \Sigma^{-1/2}Y$ son los datos fMRI blanqueados. Así, se tiene la colección de datos blanqueados en cada nodo $\{Z^r\}_{r=1}^p$. Parte de los objetivos de esta tesis

fue la implementación del método de Narayan et al. (2015), en la cual se usó esta técnica de blanqueamiento. Esto se explicará con mayor detalle en el siguiente capítulo.

El método de Narayan et al. (2015) se desarrolla a partir de los datos blanqueados obteniendo resultados correspondientes a la dependencia entre estos. Note lo siguiente, para $i = 1, \dots, T$:

$$\begin{aligned}
Cov(z_i^r, z_i^l) &= \mathbb{E}(z_i^r z_i^l) \\
&= \mathbb{E} \left(\left(\sum_t \alpha_t^r y_t^r \right) \left(\sum_s \alpha_s^l y_s^l \right) \right) \\
&= \mathbb{E} \left(\sum_{t,s} \alpha_t^r \alpha_s^l y_t^r y_s^l \right) \\
&= \sum_{t,s} \alpha_t^r \alpha_s^l \mathbb{E}(y_t^r y_s^l) \\
&= \sum_{t,s} \alpha_t^r \alpha_s^l Cov(y_t^r, y_s^l).
\end{aligned}$$

Así, se tiene que la dependencia de los datos blanqueados está relacionada con la dependencia de los datos originales, en donde la dependencia entre las variables z^r y z^l describe la dependencia entre los nodos y^r y y^l a través del tiempo.

3.3.2. Remuestreo

El método de Narayan et al. (2015) atiende a dos tipos de variabilidad en las gráficas. El primero de ellos es el que se tiene en las redes de conexión funcional de un individuo a otro, mientras que el segundo es el que se produce en la selección de la gráfica de un individuo. Ya se había explicado que la manera de abordar el segundo tipo de variabilidad es con el uso del modelo Beta-Binomial. Sin embargo, observe que la expresión (3.3) se obtiene cuando se tienen m observaciones de $Y_{k,l}^i$, lo que querría decir que para cada individuo se cuenta con los datos de m sesiones de extracción de imágenes, de modo que se pudieran estimar sus correspondientes m gráficas. No obstante, únicamente se cuenta con los datos de una sola sesión por individuo. Así, es necesario resolver esta cuestión, la cual se aborda con remuestreo, específicamente bootstrap, Efron (1979).

Para cada individuo i , $i = 1, \dots, n$, se extraen T observaciones con reemplazo de $X_{T \times p}^i$ para obtener $X_{T \times p}^{b,i}$ en un total de \mathbb{B} veces, de modo que se obtiene la colección $\{X_{T \times p}^{b,i}\}_{b=1}^{\mathbb{B}}$. Suponiendo que se conoce el parámetro de regularización que se empleará, de lo cual habla la siguiente subsección, se obtiene entonces, ejecutando el método Lasso Gráfico, la colección $\{Y_{k,l}^{b,i}\}_{b=1}^{\mathbb{B}}$ para $k < l$, cada $Y_{k,l}^{b,i}$ basado en la matriz de precisión estimada para los datos $X_{T \times p}^{b,i}$. Después se define el estimador de $\mu_{k,l}^i$ utilizando *Bootstrap aggregation (Bagging)*, introducido por Breiman (1996), es decir

$$\hat{\mu}_{k,l}^i = \frac{1}{\mathbb{B}} \sum_{b=1}^{\mathbb{B}} Y_{k,l}^{b,i}.$$

En el artículo de Breiman (1996) se expone la definición de inestabilidad de un procedimiento para derivar un estimador, la cual hace referencia a que si un conjunto \mathcal{L} de datos produce un conjunto de estimadores, entonces el procedimiento para derivar dichos estimadores es inestable si un ligero cambio en \mathcal{L} produce grandes cambios en el conjunto de estimadores, es decir, si pequeñas perturbaciones a los datos producen distintas conclusiones. Además, se exponen las consecuencias de utilizar procedimientos inestables, las cuales son no deseadas en los estimadores. Se motiva así la necesidad de crear un procedimiento estable a partir de uno inestable. La manera de realizarlo es, a grandes rasgos, promediar los estimadores producidos con ciertos subconjuntos \mathcal{L} . De modo que $\hat{\mu}_{k,l}^i$ así definido cumple con ser un procedimiento estable debido a que se realiza a partir de muestras bootstrap, las cuales son un tipo de perturbación a los datos.

Adicionalmente Meinshausen y Bühlmann (2010) mostraron las ventajas de seleccionar las aristas presentes en un Modelo Gráfico Gaussiano empleando perturbaciones en los datos. Desarrollaron un procedimiento que selecciona las aristas presentes en la gráfica subyacente de un conjunto de datos, de tal forma que se maximice la probabilidad de que se encuentren a lo largo de las gráficas estimadas para distintas submuestras. Lo que refuerza las bondades de utilizar un estimador basado en estabilidad.

Por otro lado, Efron y Hastie (2016) muestran, en su Capítulo 20 sobre Inferencia después de la selección del modelo, la mejora en la estimación de una curva cuando se usa la técnica de Bagging. El ejemplo consiste en ajustar una curva a datos sobre el decrecimiento de los niveles de colesterol según la obediencia de los pacientes para tomar las dosis de medicamentos indicadas. Proceden a realizar una selección del modelo basados en un criterio para determinar el grado de una curva polinomial. Luego de utilizar la técnica de Bagging para ajustar la curva por mínimos cuadrados ordinarios se observa que la precisión en la estimación de ésta es mejor que si no se usara la técnica introducida por Breiman (1996). Así, Efron y Hastie (2016) ilustran la ventaja de utilizar ésta técnica en problemas donde se seleccionan modelos basados en los datos.

3.3.3. Penalización aleatoria en el método de Narayan et al. (2015)

Como se comentó en la sección 2.1.4 del capítulo anterior, la Penalización aleatoria consiste en penalizar el parámetro β en el método Lasso de manera distinta y aleatoria. Narayan et al. (2015) incorpora esta idea en su método de la siguiente manera. Para cada muestra bootstrap $b = 1, \dots, \mathbb{B}$ se genera una matriz simétrica $\Lambda^{b,i}$ de la siguiente manera:

$$\Lambda_{k,l}^{b,i} = \lambda^i + c\lambda_{max}^i W_{k,l}^b \quad \forall k < l,$$

donde λ^i es el parámetro de regularización fijado mediante el método StARS (Liu et al. (2010)), que fue introducido en el Capítulo 2, para el sujeto i con la muestra original, es decir, con $X_{T \times p}^i$; λ_{max}^i es el menor parámetro de regularización que produce una gráfica totalmente sparse para los datos $X_{T \times p}^i$; $P(W_{k,l}^b = 1) = P(W_{k,l}^b = -1) = 1/2$ y c es una constante. Luego se modifica la función objetivo del método Lasso Gráfico (2.4) a

$$\hat{\Theta}_{\Lambda^{b,i}} = \arg \min_{\Theta > 0} \{-\log |\Theta| + tr(S\Theta) + \|\Lambda^{b,i} \circ \Theta\|_1\},$$

donde $\Theta \succ 0$ se utiliza como notación de Θ sobre el espacio de matrices positivo-definidas y \circ es el producto matricial Hadamard.

Así, se usa el algoritmo Lasso Gráfico modificado de tal manera que cada entrada de la matriz Θ es penalizado de manera distinta. Los autores delimitan $c \in (0, 0.5)$. Además, en sus resultados de simulación se estudia el desempeño del método incluyendo o no penalización aleatoria, resultando que la consistencia al elegir la gráfica verdadera es mucho mejor cuando se utiliza esta técnica.

3.4. Estadístico de prueba y algoritmo resumen

En la sección anterior se describió la manera puntual en la que se realiza la estimación de la matriz de precisión por individuo. Luego de esto, al tener como interés un contraste múltiple de hipótesis, es necesario definir el estadístico de prueba que se utilice para realizar dicho contraste. A continuación se deriva el estadístico de prueba para después resumir los procedimientos hasta el momento detallados en el Algoritmo completo de Narayan et al. (2015).

3.4.1. Estadístico de prueba

Luego de realizar el remuestreo de los datos y estimar la matriz de precisión para cada muestra se obtiene para cada sujeto i , con $i = 1, \dots, n$, $\{Y_{k,l}^{b,i}\}_{b=1}^{\mathbb{B}}$. Esta colección permite solucionar el hecho que se había mencionado sobre la necesidad de incorporar variabilidad en la selección de la gráfica de un individuo, en lugar de sólo estimarla directamente con los datos con que se cuenta. De este modo se permite establecer el estadístico de prueba basados en el modelo Beta-Binomial que se planteó en la sección 3.2.

Para una arista (k, l) , $k < l$, se define $Z_{k,l}^i = \sum_{b=1}^{\mathbb{B}} Y_{k,l}^{b,i}$ que es el estadístico asociado con la probabilidad de que la arista (k, l) esté presente en la gráfica del individuo i , $\mu_{k,l}^i$. De este modo se puede reescribir el modelo Beta-Binomial en (3.3) como:

$$Z_{k,l}^i | \mu_{k,l}^i \sim \text{Bin}(\mu_{k,l}^i, \mathbb{B}), \quad \mu_{k,l}^i \sim \text{Beta}(\pi_{k,l}^g, \rho_{k,l}^g).$$

Kleinman (1973) deriva los estimadores por el método de momentos de los parámetros π y ρ en el modelo Beta-Binomial. Los estimadores derivados tienen una expresión que resulta fácil de calcular computacionalmente y además, en el mismo artículo, se muestra que los estimadores son asintóticamente normales. Por lo que se procede a utilizar tales estimadores para después desarrollar el estadístico de prueba que se utilizará en el contraste de hipótesis dado en la expresión (3.2). Sea $Z_{k,l}^{*,i} = (1/\mathbb{B})Z_{k,l}^i$, los estimadores son los siguientes

$$\hat{\pi}_{k,l}^g = \frac{1}{n_g} \sum_{i \in G_g} Z_{k,l}^{*,i}, \quad \hat{\rho}_{k,l}^g = \frac{\mathbb{B}}{\mathbb{B} - 1} \frac{\sum_{i \in G_g} (\hat{\pi}_{k,l}^g - Z_{k,l}^{*,i})^2}{\hat{\pi}_{k,l}^g (1 - \hat{\pi}_{k,l}^g) (n_g - 1)} - \frac{1}{\mathbb{B} - 1}.$$

El estadístico de prueba que se utiliza es el estadístico de Wald, de modo que se necesita conocer la varianza de $\hat{\pi}_{k,l}^g$, la cual se puede encontrar utilizando (3.4) de la siguiente manera

$$\begin{aligned}
Var(\hat{\pi}_{k,l}^g) &= Var\left(\frac{1}{n_g} \sum_{i \in G_g} Z_{k,l}^{*,i}\right) \\
&= \frac{1}{n_g^2} Var\left(\sum_{i \in G_g} Z_{k,l}^{*,i}\right) \\
&= \frac{1}{n_g^2} Var\left(\sum_{i \in G_g} \frac{1}{\mathbb{B}} \sum_{b=1}^{\mathbb{B}} Y_{k,l}^{b,i}\right) \\
&= \frac{1}{n_g^2 \mathbb{B}^2} \sum_{i \in G_g} Var\left(\sum_{b=1}^{\mathbb{B}} Y_{k,l}^{b,i}\right) \\
&= \frac{1}{n_g^2 \mathbb{B}^2} \sum_{i \in G_g} (\mathbb{B} \pi_{k,l}^g (1 - \pi_{k,l}^g) + \mathbb{B}(\mathbb{B} - 1) \rho_{k,l}^g \pi_{k,l}^g (1 - \pi_{k,l}^g)) \\
&= \frac{1}{n_g^2 \mathbb{B}^2} n_g (\mathbb{B} \pi_{k,l}^g (1 - \pi_{k,l}^g) + \mathbb{B}(\mathbb{B} - 1) \rho_{k,l}^g \pi_{k,l}^g (1 - \pi_{k,l}^g)) \\
&= \frac{\pi_{k,l}^g (1 - \pi_{k,l}^g)}{\mathbb{B} \cdot n_g} (1 + (\mathbb{B} - 1) \rho_{k,l}^g),
\end{aligned}$$

de modo que,

$$\hat{Var}(\hat{\pi}_{k,l}^g) = \frac{\hat{\pi}_{k,l}^g (1 - \hat{\pi}_{k,l}^g)}{\mathbb{B} \cdot n_g} (1 + (\mathbb{B} - 1) \hat{\rho}_{k,l}^g).$$

Así, se llega al Estadístico Wald para realizar la prueba de hipótesis dada en (3.2)

$$T_{k,l} = \frac{\hat{\pi}_{k,l}^A - \hat{\pi}_{k,l}^B}{\hat{sd}(\hat{\pi}_{k,l}^A - \hat{\pi}_{k,l}^B)}, \quad (3.7)$$

donde $\hat{sd}(\hat{\pi}_{k,l}^A - \hat{\pi}_{k,l}^B) = \sqrt{\hat{Var}(\hat{\pi}_{k,l}^A - \hat{\pi}_{k,l}^B)} = \sqrt{\hat{Var}(\hat{\pi}_{k,l}^A) + \hat{Var}(\hat{\pi}_{k,l}^B)}$.

En Kleinman (1973) se muestra que el estadístico $T_{k,l}$ es asintóticamente normal bajo la hipótesis nula. Sin embargo, en los datos con los que se cuenta, n_g es una cantidad pequeña, $n_A = 37$ y $n_B = 26$. Por lo que no resultaría adecuado utilizar los p -valores de la distribución normal. Una alternativa es calcular los p -valores comparando el estadístico en una prueba por permutación. A continuación se introducirán las pruebas por permutación como se detalla en Wasserman (2004).

Una prueba por permutación es un método no paramétrico que comúnmente es utilizado para saber si dos muestras provienen de una misma distribución. Estas pruebas evitan la necesidad de plantear un modelo paramétrico pues se utiliza una estimación empírica del estadístico de prueba bajo la hipótesis nula. Además, las pruebas por permutación son “exactas” en el sentido de que no se basan en aproximaciones de teoría asintótica. Suponga que se

tienen dos muestras independientes $X_1, \dots, X_M \sim F_X$ y $Y_1, \dots, Y_N \sim F_Y$ y se quiere probar la hipótesis

$$\mathbb{H}_0 : F_X = F_Y \quad vs \quad \mathbb{H}_1 : F_X \neq F_Y.$$

Se define un estadístico de prueba $T(X_1, \dots, X_M, Y_1, \dots, Y_N)$ que sea adecuado según sea el caso, por ejemplo

$$T(X_1, \dots, X_M, Y_1, \dots, Y_N) = |\bar{X}_M - \bar{Y}_N|.$$

Sea $R = N + M$ y considere las $R!$ posibles permutaciones de los datos $X_1, \dots, X_M, Y_1, \dots, Y_N$. Para cada permutación j , $j = 1, \dots, R!$, calcule el estadístico de prueba T y llámelo T_j . Bajo la hipótesis nula cada uno de estos valores tiene la misma probabilidad de ocurrir. Sea t_{obs} el valor del estadístico de prueba evaluado en la muestra original. Suponiendo que la hipótesis nula se rechaza para valores grandes de T , entonces el p -valor es

$$p\text{-valor} = P_0(T > t_{obs}) = \frac{1}{R!} \sum_{j=1}^{R!} I(T_j > t_{obs}).$$

En la práctica, resulta prácticamente imposible calcular las $R!$ permutaciones posibles, note tan solo que $15! \approx 1.31 \times 10^{12}$. Por lo que se recurre a aproximar los p -valores utilizando una muestra aleatoria del conjunto de permutaciones. La fracción de veces que ocurre $T_j > t_{obs}$ sobre el total de permutaciones seleccionadas es entonces el p -valor aproximado.

Así, se pueden calcular los p -valores aproximados para el estadístico $T_{k,l}$ definido en (3.7) comparando el estadístico en una prueba por permutación. Se calcula el valor del estadístico $T_{k,l}$ para la muestra original $X_{T \times p}^1, \dots, X_{T \times p}^{n_A}, X_{T \times p}^{n_A+1}, \dots, X_{T \times p}^n$ y se obtiene su valor absoluto. Dicho cálculo recibe el nombre de t_{obs} . Luego se selecciona aleatoriamente un conjunto de permutaciones de los datos de tamaño C del total posible y se calculan los correspondientes valores del estadístico de prueba en valor absoluto $|T_{k,l}^j|$, $j = 1, \dots, C$. Los p -valores aproximados son entonces la fracción de veces que ocurre $|T_{k,l}^j| > t_{obs}$ sobre el total de permutaciones seleccionadas.

3.4.2. Algoritmo Narayan et al. (2015)

Las secciones anteriores se resumen en el siguiente algoritmo:

0. Realizar el blanqueamiento de los datos.
1. Para cada sujeto, $i = 1, \dots, n$, obtener λ^i utilizando el método StARS (Liu et al. (2010)).
2. *Remuestreo y penalización aleatoria:*
Para $b = 1, \dots, \mathbb{B}$:
 - (a) Obtener la muestra bootstrap $X_{T \times p}^{b,i}$
 - (b) Estimar la matriz de precisión mediante el problema Lasso Gráfico con Penalización Aleatoria.
 - (c) Calcular $Y_{k,l}^{b,i} = I(\hat{\theta}_{k,l}^{b,i} \neq 0)$.

3. Eliminar aristas ausentes para ambos grupos, obteniendo el conjunto ε_F para realizar la prueba.

4. *Inferencia:*

Para $(k, l) \in \varepsilon_F$:

- i. Calcular el estadístico $T_{k,l}$.
- ii. Calcular los p -valores usando una prueba por permutación de $T_{k,l}$

Fin

Hacer corrección por pruebas de hipótesis múltiples mediante el procedimiento Benjamini-Yekutieli.

El paso 3 se refiere a eliminar del conjunto total de hipótesis a contrastar, $\binom{p}{2}$, la comparación de aristas que no estén presentes en ambos grupos de individuos. Más formalmente:

$$\varepsilon_F^C = \{(k, l) : Z_{k,l}^{*,i} \leq \tau, \forall i = 1, \dots, n\}$$

Los autores sugieren tomar $\tau \in (0.2, 0.5)$. Realizar este filtro de aristas a comparar provoca que el número de comparaciones se reduzca drásticamente. Para el estudio conducido en esta tesis las comparaciones realizadas se redujeron de $\binom{111}{2} = 12,210$ a cantidades cercanas al 1,000 o 2,000. De manera que encontrar el conjunto ε_F disminuye en gran medida el costo computacional de la implementación del método. Por otro lado, la corrección por pruebas de hipótesis múltiples se hace utilizando el procedimiento Benjamini-Yekutieli ya que los estadísticos para cada arista son dependientes entre sí debido a que dependen de la estructura de las gráficas y, dicho procedimiento permite controlar la FDR bajo cualquier tipo de dependencia entre los estadísticos de prueba.

Hasta aquí se ha expuesto y motivado el método introducido en Narayan et al. (2015) para la comparación en el soporte de las aristas de las gráficas estimadas para dos grupos distintos de individuos. Resulta que los estudios de simulación presentados en el artículo muestran que el desempeño del método propuesto es superior a métodos como el propuesto en Zalesky et al. (2010), en donde no se contempla la variabilidad que existe cuando se estima la gráfica de un individuo pero sí se considera la dependencia temporal en los datos. La manera en la que se conducen los estudios de simulación es simulando observaciones que tengan una estructura de *mundo pequeño*, ya que esta propiedad es característica de la conectividad funcional, como se menciona en Achard et al. (2006). Luego se indaga sobre el desempeño del algoritmo utilizando o no la penalización aleatoria. Mediante los estudios de simulación se llega a la conclusión de que el algoritmo completo tiene un mayor poder estadístico, menor proporción de falsos descubrimientos y mayor tasa de verdaderos positivos, a lo largo de 50 réplicas de las simulaciones, tanto si se compara con el algoritmo sin considerar la penalización aleatoria, como si se compara con el método de Zalesky et al. (2010).

En el siguiente capítulo se describe, entre otras cosas, la implementación del método introducido por Narayan et al. (2015) y los resultados obtenidos derivados de su aplicación al conjunto de datos con que se cuenta.

Capítulo 4

Aplicación a los datos fMRI

En la sección anterior se describió a detalle el método de Narayan et al. (2015), exponiendo su motivación y su desglose. Como se mencionó, la naturaleza de los datos por resonancia magnética marca una pauta para establecer el procedimiento a seguir en el objetivo de comparar las relaciones de conectividad funcional entre los dos grupos de individuos en los que se está interesado. En este capítulo se explica en mayor detalle el preprocesamiento de los datos con los que se trabaja, así como el Atlas de Harvard-Oxford que se utilizó para particionar el cerebro. Además, se detalla la manera en la que se realizó la implementación computacional así como los resultados obtenidos. A continuación se motiva el interés por el cual se estudia la conectividad funcional de individuos con autismo.

Dichter (2012) define al autismo como un trastorno del desarrollo con síntomas que incluyen deficiencias en la comunicación y en la interacción social caracterizadas por patrones de comportamiento restringidos y repetitivos. Uno de los primeros estudios que se hizo sobre el autismo fue el hecho por Asperger (1944). Dicho trabajo se trata de un caso de estudio psiquiátrico donde se identifican una serie de condiciones en niños, tales como apariencia física y funciones expresivas que desembocan en una alteración del comportamiento. A este conjunto de condiciones ahora se les conoce como trastorno del espectro autista (ASD). Los resultados reportados en el estudio llevado a cabo por Asperger (1944) sugirieron un posible origen biológico del autismo. En la década de los 90s, con el desarrollo de la Neurociencia, se llevaron a cabo estudios científicos sobre el ASD que indagaron sobre este supuesto origen y develaron resultados interesantes.

En palabras de Dichter (2012) el ASD incluye una gran cantidad de síntomas, lo que conlleva a que existan distintos tipos de autismo. Debido a esta heterogeneidad en las caracterizaciones del ASD, los estudios de imágenes cerebrales que se han llevado a cabo para describir este desorden han identificado una amplia cantidad de posibles conexiones cerebrales alteradas. Varios de estos estudios han utilizado los datos fMRI, lo que prueba que este tipo de datos son útiles para llevar a cabo estudios de esta índole.

En Dichter (2012) y Just et al. (2012) se comenta sobre las distintas alteraciones, reportadas en la literatura sobre ASD, que caracterizan este trastorno. Algunas de ellas son: disminución de la conectividad entre los sistemas *corticales frontal y posterior-temporal*, que juegan un rol en el procesamiento de la información socio-afectiva; sobreconectividad local

y baja conectividad entre regiones de larga distancia; disminución de la conectividad entre las regiones *insulares anterior y posterior*; hipoactivación en “regiones sociales del cerebro” como el *medial prefrontal cortex*, *inferior frontal gyrus*, *anterior insula*, *posterior superior temporal sulcus*, *interparietal sulcus*, *amygdala* y *fusiform gyrus*. Dentro de los resultados de esta tesis, de los que se habla en la sección 4.3, se encuentran algunos relacionados con estas observaciones.

Varios han sido los estudios dirigidos a comprender las conexiones funcionales alteradas en las personas con trastornos del espectro autista. Estos se han llevado a cabo desde perspectivas basadas en tareas o en estado de reposo. Las conexiones antes mencionadas son apenas un breve resumen sobre lo que se ha encontrado en estudios de esta índole. Para una lectura más profunda Dichter (2012) ofrece una amplia reseña sobre los estudios llevados a cabo utilizando imágenes por resonancia magnética funcional. Para los objetivos de esta tesis, se abordó la comparación de la conectividad funcional entre individuos con ASD y un grupo control, utilizando datos fMRI. El contraste de hipótesis planteado, descrito en la sección 3.2, consiste en establecer como hipótesis nula que la probabilidad de que la conexión esté presente entre dos nodos del cerebro es igual para ambos grupos mientras que la hipótesis alternativa es que esta probabilidad es distinta. A continuación se explica con mayor detalle el estudio realizado.

4.1. Preprocesamiento de los datos

El preprocesamiento de los datos por resonancia magnética funcional es sumamente importante debido a que existen muy diversos factores que pueden alterar los resultados en un análisis estadístico. Estos factores pueden ser derivados del ambiente en el que se condujo la extracción de las unidades o de la calidad de la medición de las intensidades. Cuando los datos aún no cuentan con un preprocesamiento no son comparables entre sesiones de extracción o entre sujetos. De este modo, se necesitan aplicar distintas técnicas a los datos fMRI para que se pueda proceder a su estudio específico.

Como ya se había mencionado, los datos con los que se trabajó son un subconjunto de la base de datos ABIDE (Craddock et al. (2013a)). Dicho subconjunto es el que corresponde a los datos provenientes de la UCLA. En la página web de ABIDE, <http://preprocessed-connectomes-project.org/abide/>, se ofrecen los datos con 4 distintas fuentes de preprocesamiento, las cuales son *Connectome Computation System (CCS)*, *Configurable Pipeline for the Analysis of Connectomes (C-PAC)*, *Data Processing Assistant for Resting-State fMRI (DPARSF)* y *NeuroImaging Analysis Kit*. Específicamente los datos de estudio en esta tesis tiene un preprocesamiento vía C-PAC.

El Configurable Pipeline for the Analysis of Connectomes, C-PAC, desarrollado por Craddock et al. (2013b), es un software libre que permite aplicar de manera automatizada distintas opciones de preprocesamiento a datos fMRI con la finalidad de explorarlos y alistarlos para poder estudiarlos de una forma estandarizada. Este software está implementado en Python y puede ser utilizado para el tratamiento de los datos de cientos de individuos. Los procedimientos que fueron aplicados específicamente a los datos con que se trabajó en esta tesis

son corrección del tiempo de corte, corrección por movimiento, normalización de intensidad y regresión en variables de ruido (movimiento, señales del tejido, desvíos de baja frecuencia) para eliminar la variación debida a procesos fisiológicos como latidos del corazón y respiración, movimientos de cabeza y desvíos del escáner. Por último, a los datos también se les aplicó un filtro de paso de banda (0.01-0.1 Hz) y una regresión de señal global. Para mayor detalle se puede consultar la página web de ABIDE.

Con este preprocesamiento las series de tiempo de los vóxeles se encuentran listos para que se les pueda particionar y así emprender su estudio estadístico. La cantidad de vóxeles en los que se mide la cantidad de oxígeno presente es muy grande, por lo que construir la información de las ROIs es necesario para realizar el análisis de manera factible, no sólo por el hecho de disminuir el número de los posibles nodos, sino por el mismo hecho de identificar regiones de interés según distintos objetivos. A continuación se habla sobre las regiones de interés que se utilizaron.

4.2. Regiones de interés

Las regiones de interés (ROIs) representan los nodos de las gráficas no-dirigidas que se tiene como interés estimar, por lo que para definir el modelo gráfico es necesario definir tales regiones. En los estudios en los que se utilizan datos fMRI un nodo es un conjunto contiguo de vóxeles del cerebro. En la literatura se encuentran estudios en los que se ha parcelado el cerebro bajo distintos intereses, como se había mencionado en la sección 3.1 del presente escrito. Cada una de las distintas particiones o atlas desemboca en un número diferente de ROIs, por lo que los resultados que se obtengan de los estudios que procedan después de definir estas regiones depende ampliamente de la elección del atlas que se haga.

Para este estudio se utilizaron las ROIs determinadas por el Atlas Harvard-Oxford (Fischl et al., 2004) desarrollado por el Centro de Análisis Morfométrico (CMA por sus siglas en inglés) y distribuido por la librería FSL del software FMRIB. En palabras textuales el Atlas Harvard-Oxford *es una técnica que usa datos manualmente etiquetados como base para un procedimiento de partición automatizado*, Fischl et al. (2004). Este Atlas usa técnicas neuroanatómicas para realizar las particiones y divide al cerebro en regiones corticales y subcorticales.

Cuando el Atlas con el que se realizará el estudio es elegido, la manera en la que se obtiene la serie de tiempo de una región (o nodo) es promediando las series de tiempo de los vóxeles que se encuentran dentro de dicha región. Los datos concernientes a cada ROI son ofrecidos por la página ABIDE donde se detalla que la partición se realizó de la siguiente manera: luego de realizar la partición en regiones corticales y subcorticales se bisectó en regiones derecha e izquierda y se removió del atlas subcortical las regiones izquierda/derecha WM, izquierda/derecha GM, izquierda/derecha CSF y tronco encefálico. En total se tienen un total de 111 regiones de interés.

Con los datos preprocesados y parcelados bajo el Atlas Harvard-Oxford es ya apropiado proceder a su análisis con la metodología estudiada en esta tesis. En las siguientes secciones

se describe la manera en la que el método de Narayan et al. (2015) fue implementado y los resultados obtenidos.

4.3. Implementación y resultados

Como ya se había mencionado, se cuenta con los datos de fMRI de 63 individuos, 37 del grupo con ASD (grupo A) y 26 del grupo control (grupo B), donde para cada individuo se tienen 111 series de tiempo de una longitud de 116. El método de Narayan et al. (2015) se implementó en el software estadístico R, R Core Team (2014), para realizar la presente tesis y se ejecutó en estos datos con la finalidad de indagar sobre las diferencias de conectividad funcional entre ambos grupos. A continuación se explicará primero la manera en la que el método fue implementado y luego se expondrán los resultados obtenidos con la finalidad de ofrecer un análisis de ellos.

La primera tarea a realizar es el blanqueamiento de las series de tiempo. La técnica de blanqueamiento que se utilizó fue el blanqueamiento ZCA que se describió en la sección 3.3.1. Dicha técnica se encuentra implementada en R a través del paquete *ForeCA* mediante la función *whiten*. Luego de que las series de tiempo fueron blanqueadas se procedió a establecer el parámetro de regularización λ para cada uno de los 63 individuos mediante el método StARS. Este método se encuentra implementado en el paquete *huge* a través de la función *huge.select* que selecciona el valor óptimo de λ , según el método StARS escribiendo *criterion* = “stars” en sus parámetros. El parámetro β de este método se fijó como 0.3, obteniendo valores de λ entre 0.47 y 0.62.

La manera en la que se encontró λ_{max} , para cada individuo, fue ejecutando un algoritmo de búsqueda binaria, Cormen et al. (2009), entre el 0 y el 10 a fin de encontrar el mínimo valor de λ que produjera una matriz de precisión totalmente rala mediante el método Lasso Gráfico. El método Lasso Gráfico se ejecutó a través de la librería *glasso* mediante la función *glasso*. Luego, por individuo, para cada muestra bootstrap se registraron los respectivos estadísticos $Y_{k,l}^{b,i}$ utilizando las matrices de precisión que se derivaban de ejecutar el método Lasso Gráfico con matriz de penalización $\Lambda^{b,i}$ y con valor del parámetro c igual a 0.25.

Luego de lo anterior se obtuvo el conjunto ε_F estableciendo $\tau = 0.5$. La inferencia se realizó para las aristas (k, l) del conjunto ε_F . Los p-valores para los estadísticos $T_{k,l}$ se aproximaron mediante 1000 permutaciones. Por último, se corrigió por multiplicidad mediante el procedimiento Benjamini-Yekutieli controlando la FDR a un nivel 0.1, es decir, se utilizó $q = 0.1$. Este valor es un umbral típico utilizado en el control de hipótesis múltiples.

En la siguiente tabla se enlistan las aristas (k, l) de las correspondientes hipótesis nulas rechazadas en 5 distintas ejecuciones del método de Narayan et al. (2015) a los datos, variando el valor de bootstrap \mathbb{B} . También se muestra la respectiva estimación de $\pi_{k,l}^g$.

\mathbb{B}	Arista (k, l)	$\hat{\pi}_{k,l}^A$	$\hat{\pi}_{k,l}^B$
100	·Left Putamen - Left Frontal Operculum Cortex ·Left Lateral Occipital Cortex; superior division - Left Precuneous Cortex ·Right Parietal Operculum Cortex - Left Planum Temporale	0.402432 0.34 0.170541	0.131154 0.194231 0.347308
100	·Left Putamen - Left Frontal Operculum Cortex ·Right Precentral Gyrus-Left Occipital Pole ·Right Inferior Temporal Gyrus; posterior division - Right Inferior Temporal Gyrus; temporooccipital part ·Left Postcentral Gyrus - Right Superior Parietal Lobule ·Right Cingulate Gyrus; posterior division - Left Temporal Fusiform Cortex; posterior division	0.294324 0.094595 0.299460 0.216216 0.063784	0.125 0.013077 0.122692 0.369615 0.009231
500	·Right Hippocampus - Right Superior Temporal Gyrus; posterior division ·Right Parietal Operculum Cortex - Left Planum Temporale	0.082811 0.170216	0.015 0.344770
500	·Left Putamen - Left Frontal Operculum Cortex ·Right Inferior Temporal Gyrus; posterior division - Right Inferior Temporal Gyrus; temporooccipital part ·Left Central Opercular Cortex - Left Heschl's Gyrus (includes H1 and H2) ·Right Parietal Operculum Cortex - Left Planum Temporale	0.289784 0.288162 0.431405 0.171838	0.125231 0.126077 0.320231 0.345308
600	·Left Lateral Occipital Cortex - Left Precuneous Cortex ·Right Parietal Operculum Cortex - Left Planum Temporale	0.338514 0.169054	0.196026 0.345256

Tabla 2. Aristas correspondientes a las hipótesis nulas rechazadas en 5 distintas ejecuciones.

Note en la Tabla 2 que son 4 las aristas cuya hipótesis nula correspondiente se rechaza en dos o más ejecuciones del método, aún cuando el tamaño de \mathbb{B} es distinto. Por ejemplo, la hipótesis nula de la arista formada entre los nodos Left Putamen y Left Frontal Operculum Cortex se repite tanto en las dos ejecuciones con tamaño de bootstrap 100 como en la segunda ejecución hecha con tamaño de bootstrap 500. Específicamente, las 4 aristas cuya hipótesis nula se rechazó dos veces o más son las siguientes:

1. Left Putamen- Left Frontal Operculum Cortex.
2. Left Lateral Occipital Cortex; superior division - Left Precuneous Cortex.
3. Right Parietal Operculum Cortex - Left Planum Temporale.
4. Right Inferior Temporal Gyrus; posterior division - Right Inferior Temporal Gyrus; temporooccipital part.

Las aristas 3 y 4 están altamente relacionadas con las observaciones que se hicieron al principio de este capítulo. A continuación se ilustran las regiones de la arista 3, *Right Parietal Operculum Cortex* y *Left Planum Temporale*.

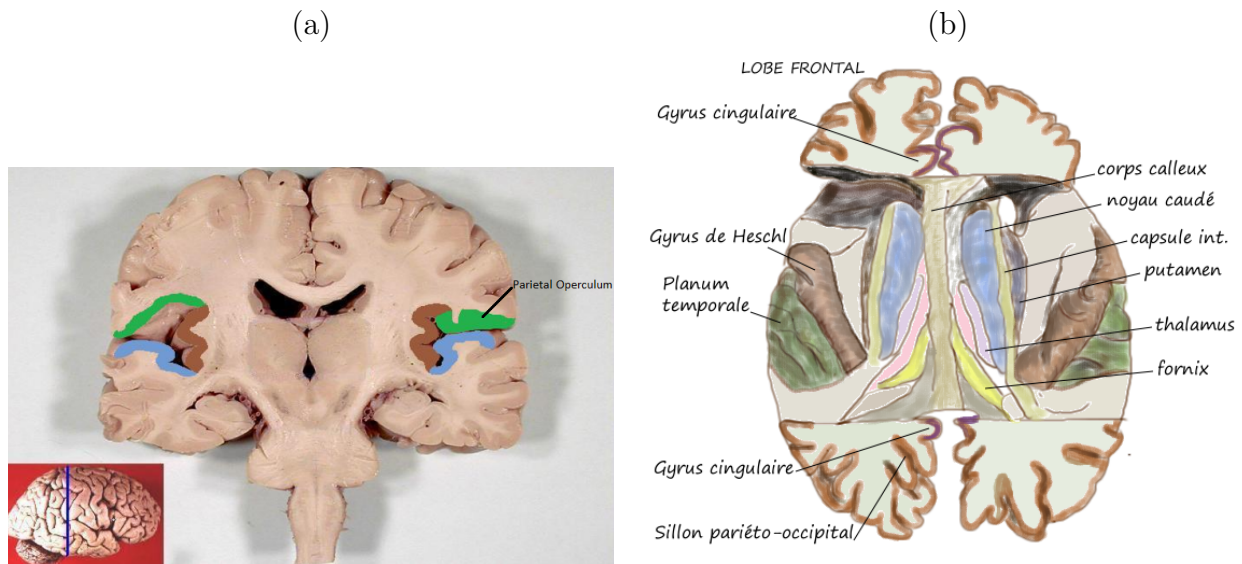


Figura 3. De izquierda a derecha se señalan las regiones Parietal Operculum y Planum Temporale.

Observe en la Figura 3(a) que en color verde se encuentra la región Right Parietal Operculum y en la Figura 3(b) se señala Left Planum Temporale. Estas dos regiones se encuentran con una amplia distancia entre sí. Además, la probabilidad de que la arista entre estos dos nodos esté presente en el grupo autista es mayor a la del grupo control, como se reporta en la Tabla 2. Lo que apunta a apoyar la observación mencionada en Dichter (2012) sobre la baja conectividad en regiones de larga distancia.

Para la arista 4 a continuación se muestra la región Inferior Temporal Gyrus.

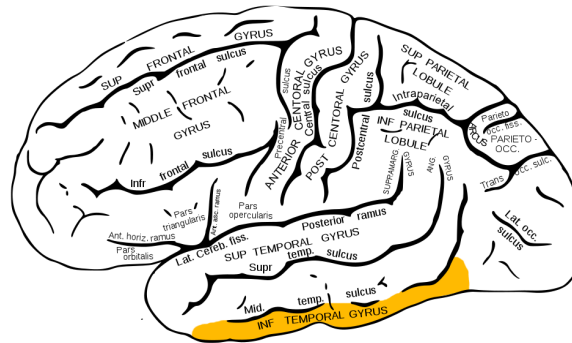


Figura 4. Inferior Temporal Gyrus en color naranja.

En el Atlas Harvard-Oxford la región Inferior Temporal Gyrus se particiona en tres nodos, anterior, posterior y temporooccipital part. Pero estas regiones son cercanas entre sí. Así pues, el que se rechace la hipótesis correspondiente a la arista entre los nodos Right Inferior Temporal Gyrus; posterior division y Right Inferior Temporal Gyrus; temporooccipital part se relaciona con la observación mencionada al principio del capítulo sobre la alta conectividad local en personas con ASD. Note, en la Tabla 2, que la probabilidad de que esta arista esté presente en el grupo con ASD es mayor que para el grupo control. De este modo, las aristas cuya hipótesis nula fueron rechazada se alinean con las tendencias observadas en los distintos estudios llevados a cabo para estudiar las particularidades de las relaciones de

conectividad funcional en sujetos con trastorno del espectro autista.

Ahora bien, es necesario mencionar las limitaciones computacionales que se tuvieron respecto al tamaño de bootstrap \mathbb{B} empleado en este estudio. En la sección de remuestreo y penalización aleatoria del Algoritmo Narayan et al. (2015), sección 3.4.2, se requiere almacenar los estadísticos $Y_{k,l}^{b,i} = I(\hat{\theta}_{k,l}^{b,i} \neq 0)$ para $b = 1, \dots, \mathbb{B}$ e $i = 1, \dots, 63$. Las matrices de precisión estimadas, $\hat{\Theta}^{b,i}$, son de tamaño 111×111 , lo que se deriva en un total de 12,321 elementos que tienen que ser almacenados en memoria por cada matriz de precisión estimada. Pero se requiere almacenar en memoria $\mathbb{B} \times 63$ matrices de precisión estimadas. Con un tamaño de \mathbb{B} de 600 el método implementado se ejecutó aún en la computadora con la que se trabajó (que cuenta con una memoria RAM de 4 GB), sin embargo, al querer ejecutar el programa estableciendo $\mathbb{B} = 700$ la memoria RAM era superada. Así, las ejecuciones estuvieron limitadas por este tamaño.

Capítulo 5

Conclusiones

En los últimos años se han presentado desafíos sin precedentes respecto a la cantidad de datos que se generan día con día. El análisis de estos datos somete a las ciencias a nuevos retos y nuevas perspectivas de mira en las cuestionantes que surgen. Los modelos gráficos gaussianos, mediante la incorporación de esparsidad, permiten reconocer patrones en la nube de datos que hagan posible elegir un modelo donde sólo un número relativamente pequeño de predictores juegan un papel importante. De modo que la esparsidad permite hacer frente a volúmenes de datos con un alto número de parámetros, respecto a las observaciones con que se cuenta. Además, estos modelos tienen la cualidad de que sus resultados pueden conducir una interpretabilidad muy clara en el área en que se estén aplicando. Así pues, el estudio de los modelos gráficos es un tema que se está desarrollando en la literatura actual pues existen distintas direcciones de investigación que se pueden seguir.

Para hacer frente a nuevos retos en la ciencia se requiere del trabajo en conjunto de expertos no sólo del área en sí, sino de áreas diferentes. El ejemplo claro está presente en esta tesis. Comparar la conectividad funcional entre dos grupos de individuos es un tema que compete no sólo a expertos en neuroimagen sino también a expertos en estadística y computación. Al querer hacer frente a preguntas aplicadas surge la necesidad de establecer nuevos métodos que brinden una alternativa y que también muestren cuáles son las ganancias respecto a métodos ya existentes. En este sentido, el algoritmo introducido en Narayan et al. (2015) tiene la cualidad, frente a otros métodos, de permitir la estimación de la gráfica de cada individuo. Asumir que cada individuo tiene potencialmente distintas relaciones de conectividad funcional permite también estudiar e incorporar la variabilidad resultante de la estimación de la gráfica por individuo. Además, la incorporación de la variabilidad derivada de la observación sobre el hecho de que la gráfica de un individuo es estimada y no simplemente observada provoca que el método responda a observaciones hechas por los especialistas.

Del trabajo realizado en la aplicación del método de Narayan et al. (2015) a los datos de la UCLA fue posible obtener varias conclusiones. Por un lado, las aristas cuyas hipótesis fueron rechazadas respaldan teorías que se han derivado del estudio de la personas que presentan ASD. Específicamente las hechas respecto a la conectividad funcional local y de larga distancia. Las aristas que resultaron tener más probabilidad de estar presentes en personas con ASD son aquellas que unen nodos cercanos, mientras que las aristas con menor probabilidad de existir en personas de este grupo unen nodos lejanos. Por otro lado, el estudio efectuado

permite conocer cuáles son las conexiones funcionales que son estadísticamente diferentes entre ambos grupos, sin embargo, los resultados no pueden extrapolarse a decir en qué grupos se encuentran las aristas cuya hipótesis nula se rechaza. Es decir, tome el ejemplo de la arista entre los nodos Right Parietal Operculum Cortex y Left Planum Temporale. Aunque en la Tabla 2 se puede ver que la probabilidad de que esta arista esté presente en individuos del grupo control es mayor a la del grupo con ASD, no se puede decir entonces que tal arista está presente en el grupo control y ausente en el grupo con ASD.

Por otro lado, el poder de cómputo fue una limitante en la ejecución del método expuesto. Como se explicó en las últimas líneas del Capítulo 4, incrementar el número de muestras bootstrap implica la necesidad de contar con una memoria RAM suficiente. Sin embargo, pudieran ser utilizados distintos paradigmas de computación como cómputo en paralelo. De aquí la observación de que es necesario el trabajo en conjunto con expertos en distintas áreas.

Existen distintas líneas de investigación a seguir como trabajo posterior a esta tesis. Por un lado, el método requiere de parámetros que deben ser calibrados. Una vez que los parámetros están fijos, el algoritmo requiere de 20 horas de ejecución, con un tamaño de muestra bootstrap B de 600, en una computadora con 4 GB de memoria RAM y un procesador Intel i3-4005U 1.70 GHz. Es de interés explorar los resultados que se obtendrían con un valor de B más grande, pero tanto el tiempo de ejecución como la limitante de la memoria RAM restringieron el análisis. Soluciones a esto sería eficientizar el código, por ejemplo utilizando cómputo en paralelo o distribuido, o hacer uso de una infraestructura con mayor poder de cómputo.

Por otro lado, para el estudio específico del autismo pueden utilizarse de la base de ABIDE los datos provenientes de otras instituciones, no únicamente de la UCLA. Además, el método de Narayan et al. (2015) puede ser aplicado no sólo a datos fMRI sino que puede ser utilizado para el estudio de distintas poblaciones para responder a cuestionantes que puedan ser modeladas a través de los modelos gráficos gaussianos, por lo que esto puede conducir estudios distintos. Por último, existen distintos métodos para estimar la matriz de precisión incorporando la dependencia temporal entre las observaciones, es decir, sin la necesidad de utilizar blanqueamiento. Tal es el caso del método introducido por Ni et al. (2017), donde se hace uso de un nuevo modelo llamado *Matrix-Variate Gaussian Graphical Model*. De modo que se abre la oportunidad para usar el método de Narayan et al. (2015) pero utilizando otras herramientas para la estimación de las gráficas.

Bibliografía

- [1] Achard, S., Salvador, R., Whitcher, B., Suckling, J. y Bullmore, E. (2006), A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs, *The Journal of Neuroscience*, **26**, 63-72.
- [2] Asperger, H. (1944), “Autistic psychopathy” in childhood, translated in autism and asperger’s syndrome, Frith, U., 1991, *Cambridge University Press*, 37-92.
- [3] Benjamini, Y. y Hochberg, Y. (1995), Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B (Methodological)*, **57**, 289-300.
- [4] Benjamini, Y. y Yekutieli, D. (2001), The control of the false discovery rate in multiple testing under dependency, *The Annals of Statistics*, **29**, 1165-1188.
- [5] Bien, J. y Tibshirani, R. J. (2011), Sparse estimation of a covariance matrix, *Biometrika*, **98**, 807-820.
- [6] Breiman, L. (1996), Heuristics of instability and stabilization in model selection, *The Annals of Statistics*, **24**, 2350-2383.
- [7] Brockwell, P. J., Davis, R. A. (1991), *Time Series: Theory and Methods* (Second Edition), Springer.
- [8] Cormen, T. H., Leiserson, C. E., Rivest, R. L. y Stein, C. (2009), *Introduction to Algorithms* (Third Edition), The MIT Press.
- [9] Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., Khundrakpam, B. S., Lewis, J. D., Li, Q., Milham, M., Yan, C. y Bellec, P. (2013a), The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives, *In Neuroinformatics 2013, Stockholm, Sweden*.
- [10] Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S. S., Yan, C., Li, Q., Lurie, D., Vogelstein, J., Burns, R., Colcombe, S., Mennes, M., Kelly, C., Di Martino, A., Castellanos, F. X. y Milham, M. (2013b), Towards automated analysis of connectomes: the configurable pipeline for the analysis of connectomes (C-PAC), *Front. Neuroinform. Conference Abstract: Neuroinformatics 2013*. doi: 10.3389/conf.fninf.2013.09.00042.
- [11] Dichter, G. S. (2012), Functional magnetic resonance imaging of autism spectrum disorders, *Dialogues in clinical neuroscience*, **14**, 319-330.
- [12] Efron, B. (1979), Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, **7**, 1-26.

- [13] Efron, B. (2010), *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing and Prediction*, Stanford University.
- [14] Efron, B., Hastie, T. (2016), *Computer Age Statistical Inference, Algorithms, Evidence, and Data Science*, Cambridge University Press.
- [15] Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A. M. (2004), Automatically parcellating the human cerebral cortex, *Cerebral Cortex*, **14**, 11-22.
- [16] Friedman, J., Hastie, T. y Tibshirani, R. (2008), Sparse inverse covariance estimation with the graphical lasso, *Biostatistics*, **98**, 1-15.
- [17] Fukunaga, K. (1990), *Introduction to Statistical Pattern Recognition* (Second Edition), Academic Press.
- [18] Giraud, C. (2015), *Introduction to High-Dimensional Statistics*, CRC Press.
- [19] Hastie, T., Tibshirani, R. y Friedman, J. (2008), *The Elements of Statistical Learning, Data Mining, Inference, and Prediction* (Second Edition), Springer.
- [20] Hastie, T., Tibshirani, R. y Wainwright, M. (2015), *Statistical Learning with Sparsity, The Lasso and Generalizations*, CRC Press.
- [21] Iturriaga-Jofré, A. M. (2011), *Teoría de Pruebas de Múltiples Hipótesis Estadísticas y Aplicaciones*, Tesis de Licenciatura, Universidad de Chile.
- [22] Just, M. A., Keller, T. A., Malave, V. L., Kana, R. K. y Varma S. (2012), Autism as a neural systems disorder: A theory of frontal-posterior underconnectivity, *Neuroscience and Biobehavioral Reviews*, **36**, 1292-1313.
- [23] Kessy, A, Lewin, A. y Strimmer, K (2016), Optimal whitening and decorrelation, *arXiv:1512.00809v4*.
- [24] Kleinman, J. C. (1973), Proportions with extraneous variance: single and independent sample, *Journal of the American Statistical Association*, **68**, 46-54.
- [25] Liu, H., Roeder, K. y Wasserman, L. (2010), Stability approach to regularization selection (StARS) for high dimensional graphical models, *Advances in Neural Information Processing Systems 23*, *arXiv:1006.3316v1*.
- [26] Meinshausen, N. y Bühlmann, P. (2006), High-dimensional graphs and variable selection with the lasso, *The Annals of Statistics*, **34**, 1436-1462.
- [27] Meinshausen, N. y Bühlmann, P. (2010), Stability selection, *Journal of the Royal Statistical Society, Series B*, **72**, 417-473.
- [28] Narayan, M., Allen, G. I. y Tomson, S. N. (2015), Two sample inference for populations of graphical models with applications to functional connectivity, *arXiv:1502.03853v1*.

- [29] Ni, Y., Stingo, F., C. y Baladandayuthapani, V. (2017), Sparse multi-dimensional graphical models: a unified bayesian framework, *Journal of the American Statistical Association*, **112**, 779-793.
- [30] Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L. y Peterson, S. E. (2011), Functional network organization of the human brain, *Neuron*, **72**, 665-678.
- [31] R Core Team (2014), R: A language and environment for statistical computing, *R Foundation for Statistical Computing*, Vienna, Austria, URL <http://www.R-project.org/>.
- [32] Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Jenkinson, K. L., Nichols, T. E., Robinson, E. C., Salimi-Khorshidi, G., Woolrich, M. W., Barch, D. M., Ugurbil, K. y Van Essen, D. C (2013), Functional connectomics from resting-state fMRI, *Trends in Cognitive Sciences*, **17**, 666-682.
- [33] Tibshirani, R. (1996), Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B (Methodological)*, **58**, 267-288.
- [34] Varoquaux, G., Gramfort, A., Poline, J.-B. y Thirion, B. (2010), Brain covariance selection: better individual functional connectivity models using population prior, *arXiv:1008.5071v4*.
- [35] Wang, J.-H., Zuo, X.-N., Gohel, S., Milham, M. P., Biswal, B. B. y He, Y. (2011), Graph theoretical analysis of functional brain networks: test-retest evaluation on short- and long-term resting state functional MRI data, *PLoS ONE*, **6**, e21976.
- [36] Wasserman, L. (2004), *All of Statistics, A Concise Course in Statistical Inference*, Springer.
- [37] Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F. y Evans, A. C. (2002), A general statistical analysis for fMRI data, *NeuroImage*, **15**, 1-15.
- [38] Wu, X., Ye, Y. y Subramanian, K. R. (2003), Interactive analysis of gene interactions using graphical gaussian model, *ACM SIGKDD Workshop on Data Mining in Bioinformatics*, **3**, 63-69.
- [39] Zalesky, A., Fornito, A. y Bullmore, E. T. (2010), Network-based statistic: identifying differences in brain networks, *NeuroImage*, **53**, 1197-1207.
- [40] Zhao, P. y Yu, B. (2006), On model selection consistency of lasso, *Journal of Machine Learning Research*, **7**, 2541-2563.