

CIMAT

CENTRO DE INVESTIGACIÓN EN
MATEMÁTICAS, A.C.

ESTIMACIÓN MEDIANTE NÚCLEOS POSITIVOS
DE FUNCIONES DE REGRESIÓN CON VALORES
MATRICIALES DEFINIDOS NO NEGATIVOS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
MAESTRO EN CIENCIAS

CON ESPECIALIDAD EN PROBABILIDAD Y ESTADÍSTICA

P R E S E N T A:

I L C E A N A H I M E D I N A M E N D O Z A

DIRECTOR DE TESIS:
Dr. Rolando José Biscay Lirio

2015

Hoja de Datos del Jurado

I. Datos del alumno

Medina
Mendoza
Ilce Anahi.
Centro de Investigación en Matemáticas, A. C.
Maestría en Ciencias con especialidad en Probabilidad y Estadística.

II. Datos del presidente

Dr.
Miguel
Nakamura
Savoy.

III. Datos del secretario

Dra.
Graciela Ma. de los Dolores
González
Farías.

IV. Datos del vocal (asesor)

Dr.
Rolando José
Biscay
Lirio.

V. Datos del trabajo escrito

Estimación mediante núcleos positivos de funciones de regresión con valores matriciales definidos no negativos.
18 de diciembre de 2015.

Catarsis

Agradecimientos

A mi padre, porque siempre me apoyaste aun cuando me necesitabas a tu lado. Te amaré por siempre.

A mi madre, porque eres mi pilar.

A mi asesor el Dr. Rolando Biscay quien me brindó generosamente su tiempo, apoyó y guía. Porque más que un asesor se volvió un amigo.

A mis sinodales el Dr. Miguel Nakamura y la Dra. Graciela González Farías, por el tiempo y las aportaciones que hicieron de este un mejor trabajo.

A los maravillosos amigos y a todas las personas que me acompañaron durante mi camino en CIMAT, en especial al Dr. Joaquín Ortega que me apoyó en las grandes decisiones académicas que tomé. Sin ellos nada habría sido igual.

A Emmanuel por ser un gran amigo, por la buena vibra, los consejos y por estar a mi lado en los momentos difíciles.

Finalmente al Consejo Nacional de Ciencia y Tecnología, CONACYT, por la beca que durante dos años me dio la posibilidad de estudiar esta maestría.

Índice general

Lista de símbolos y abreviaturas	IX
Resumen	XI
1. Introducción	1
2. Estimación no paramétrica de funciones de regresión con valores en una variedad convexa	5
2.1. Construcción del estimador mediante núcleos positivos	5
2.2. Propiedades asintóticas	7
2.3. Demostraciones de teoremas	9
3. Aplicación a datos simulados	17
3.1. Procedimientos de simulación y estimación	17
3.2. Modelo 1 con muestras grandes	18
3.3. Modelo 1 con muestras pequeñas	22
3.4. Modelo 2	24
3.5. Modelo 3	26
3.6. Modelo 4	29
4. Aplicación a la caracterización de la maduración del EEG multi-canal	33
4.1. Objetivo del estudio	33
4.2. Datos de EEG	34
4.3. Descriptores espectrales del EEG	36
4.4. Procedimientos de estimación de curvas de regresión de matrices de conectividad del EEG	37
4.5. Resultados del análisis	39
4.5.1. Potencias relativas	39
4.5.2. Coherencias	43
4.5.3. Coherencias parciales	45

5. Conclusiones	49
A. Variedades diferenciales	51
B. Teoría espectral de series de tiempo multivariadas	55
C. Probabilidad en variedades diferenciales	59
Bibliografía	61

Lista de símbolos y abreviaturas

E^D	espacio euclideo D -dimensional
\mathcal{M}_C	variedad convexa contenida en E^D
\mathcal{M}	variedad de las matrices d.n.n. reales de dimensión p
$\mathcal{M}^{\mathbb{C}}$	variedad de las matrices d.n.n. complejas de dimensión p
$\mu(x)$	media extrínseca de una variable aleatoria \mathcal{M}_C -valuada Y según una probabilidad condicional $P(dy x)$ con covariable $x \in \mathbb{R}^m$
$\Sigma(u)$	media extrínseca de una variable aleatoria \mathcal{M} -valuada o $\mathcal{M}^{\mathbb{C}}$ -valuada $S = S(u)$ según una probabilidad condicional $P(dS u)$ con covariable escalar $u \in \mathbb{R}$
d.n.n.	definido no negativo(a)
RP	potencia relativa ("Relative Power")
CP	coherencia parcial

Resumen

La presente tesis aborda el problema de la estimación no paramétrica de funciones de regresión con valores en una variedad convexa, haciendo énfasis en el caso de la variedad de matrices definidas no negativas (d.n.n.). Su motivación práctica principal es describir la maduración de la conectividad del electroencefalograma (EEG) a través del análisis de regresión de sus matrices de densidad espectral como funciones de la edad del sujeto. En el primer año de vida esto tiene gran importancia debido a que la detección de anomalías de conectividad a temprana edad puede tener un impacto decisivo en la rehabilitación de niños con patologías asociadas a riesgos perinatales. Actualmente, para describir la maduración de la conectividad se suelen llevar a cabo metodologías ([Bosch-Bayard et al., 2012], [Otero et al., 2011], [Bosch-Bayard et al., 2001], [Szava et al., 1994]) que no resuelven el punto central de la problemática. Es así que este dominio de aplicación nos motivó a considerar la función de regresión en el sentido de media extrínseca en una variedad según el encajamiento de inclusión en un espacio euclideo. Para su estimación se propone un método no paramétrico basado en núcleos positivos. La positividad del núcleo garantiza que el estimador propuesto toma valores solamente en la variedad de las matrices d.n.n. Se estudian teóricamente propiedades del estimador, incluyendo consistencia, orden de convergencia en media cuadrática y normalidad asintótica. También se explora computacionalmente su comportamiento en diversas condiciones a través de datos simulados. Finalmente, se aplica la metodología propuesta para obtener una caracterización de la maduración de la conectividad del EEG multi-canal a diferentes frecuencias durante el primer año de vida.

CAPÍTULO 1

Introducción

Un problema relevante dentro del área de la neurofisiología, y que es una motivación práctica del presente trabajo, es el estudio de la maduración de la conectividad del electroencefalograma (EEG) a diferentes frecuencias al aumentar la edad del sujeto. Durante el primer año de vida esto tiene gran importancia debido a que la detección de anomalías de conectividad a temprana edad puede tener un impacto decisivo en la rehabilitación de niños con patologías asociadas a riesgos perinatales ([Bosch-Bayard et al., 2012], [Otero et al., 2011]). En tal situación, para cada sujeto i con edad u_i el tipo de dato relevante para nuestro trabajo es la matriz de covarianza muestral $S(u_i) = (s_{jk}(u_i))$. El interés principal es estudiar la correspondiente matriz de covarianza poblacional $\Sigma(u) = (\sigma_{jk}(u))$ como función de regresión con respecto a la edad u del sujeto como covariable.

Para este propósito, se suele llevar a cabo un análisis de regresión lineal o cuadrática diferente para cada elemento $\sigma_{jk}(u)$ de la matriz $\Sigma(u)$ ([Bosch-Bayard et al., 2012], [Otero et al., 2011], [Bosch-Bayard et al., 2001], [Szava et al., 1994]). Pero este proceder adolece de varias desventajas: a) La estimación así obtenida $\hat{\Sigma}(u) = (\hat{\sigma}_{jk}(u))$ de la matriz de covarianza $\Sigma(u)$ para cada edad u no necesariamente es una matriz de covarianza, pues no necesariamente resulta definida no negativa (d.n.n.). Como consecuencia, el análisis simultáneo de diferentes covarianzas predichas para una edad, digamos $\hat{\sigma}_{jk}(u)$ y $\hat{\sigma}_{j'k'}(u)$, puede carecer de sentido estadístico porque no sean compatibles entre sí. b) El uso de regresiones polinomiales de grado bajo puede resultar incapaz de captar tendencias no lineales presentes en los datos.

La dificultad (a) que acabamos de indicar motiva la necesidad de considerar la estimación de la matriz de regresión $\Sigma(u)$ como un problema de estimación en la variedad de las matrices d.n.n. \mathcal{M} . Relacionada con esta problemática existe una creciente literatura estadística (motivada principalmente por el análisis de imágenes de tensores de difusión) que aborda la definición y estimación de medias y funciones de regresión con valores en esta variedad ([Bhattacharya and Bhattacharya, 2015], [Turaga and Srivastava, 2016],

[Bhattacharya and Patrangenaru, 2003], [Chikuse, 2003], [Abhishek Bhattacharya, 2008], [Yuan et al., 2012], [Shi et al., 2009], [Zhu et al., 2009], [Pérez and González-Farías, 2013], [Bhattacharya and Patrangenaru, 2005]). La mayoría de los enfoques publicados basan su análisis en algún tipo de media intrínseca (ver Apéndice C para definiciones de medias intrínseca y extrínseca). Tal media depende de la distancia geodésica asociada a la métrica Riemanniana adoptada sobre la variedad \mathcal{M} . Las dos más usadas son las siguientes.

a) Distancia geodésica asociada a la métrica Riemanniana Log-Euclidean:

$$d_L(A, B) = \sqrt{\text{tr}([\log(A) - \log(B)][\log(A) - \log(B)]^T)} \quad (1.1)$$

para A y B en \mathcal{M} .

b) Distancia geodésica asociada a la métrica Riemanniana traza:

$$d_T(A, B) = \sqrt{\text{tr}([\log^2\{B^{-1/2}AB^{-1/2}\}])} \quad (1.2)$$

para A y B en \mathcal{M} .

Es importante aclarar que la pertinencia de usar una u otra media intrínseca, o una media extrínseca, depende del modelo generador de los datos. Por ejemplo, si $S(u_i)$ fuera una matriz d.n.n. generada con acuerdo en un modelo log-normal, es decir, si $\log(S(u_i))$ tuviera una distribución $N(\log(\Sigma(u_i)), \Psi)$, entonces el parámetro $\Sigma(u_i)$ sería la media intrínseca asociada a la distancia geodésica Log-Euclidean (1.1). Pero si el modelo generador de datos fuera otro, el parámetro de localización $\Sigma(u_i)$ del dato $S(u_i)$ pudiera ser una media extrínseca o a una media intrínseca asociada a una métrica distinta.

En el caso de datos de EEG, el cual es de especial interés práctico en el presente trabajo, la matriz de conectividad $S(u_i)$ suele ser una matriz de densidad espectral muestral. En tal situación, el parámetro que se desea estimar es la correspondiente matriz de densidad espectral $\Sigma(u_i)$. Es conocido que una matriz de densidad espectral muestral tiene distribución asintótica Wishart compleja $W_p^C(l, \Sigma(u_i))$ (ver, e.g., [Brillinger, 2001], y también Apéndice B). Por tanto, se cumple (asintóticamente) que $\Sigma(u_i) = E(S(u_i))$, donde el valor esperado representa la media extrínseca asociada a la métrica euclidean del espacio ambiente, \mathbb{R}^{p^2} , de la variedad \mathcal{M} . De modo que, para nuestro propósito práctico, se impone considerar la función de regresión $\Sigma(u)$ en el sentido de una media extrínseca según el encajamiento de inclusión o identidad dado por $id : A \in \mathcal{M} \subset \mathbb{R}^{p^2} \rightarrow A \in \mathbb{R}^{p^2}$.

El análisis de regresión con valores en la variedad \mathcal{M} en el sentido de media extrínseca ha sido abordado por diversos autores (ver, e.g., [Bhattacharya and Patrangenaru, 2005], [Bhattacharya and Patrangenaru, 2003], [Lin et al., 2011], [Bhattacharya and Bhattacharya, 2015]). Sin embargo, éstos adoptan una media extrínseca asociada a algún encajamiento $J : \mathcal{M} \rightarrow \mathbb{R}^{p^2}$ que trata de conservar aspectos de una media intrínseca de interés; pues sabemos que el uso de una u otra media intrínseca así como media extrínseca depende del modelo generador de datos con el cual se cuente. Con este propósito, imponen invarianzas al encajamiento con respecto a algún grupo especificado de transformaciones, lo que conduce a que el encajamiento no sea el encajamiento de inclusión o identidad. La media

extrínseca así resultante no coincide con el concepto de media extrínseca que acabamos de fundamentar como el adecuado para el caso de datos de conectividad de EEG.

Los argumentos anteriores nos motivan a considerar el problema de regresión para matrices d.n.n. desde el punto de vista del concepto de media extrínseca asociada al encajamiento de inclusión, y a proponernos construir un estimador no paramétrico de tal media. Para esto, hay que enfrentar la dificultad de que los estimadores de regresión usuales según este concepto de media extrínseca no necesariamente toman valores en la variedad \mathcal{M} , lo cual es el punto crucial del problema que abordamos debido a que es indispensable que nuestra covariable en la función de regresión sea una matriz de covarianza.

Específicamente, nos proponemos los siguientes

Objetivos

1. Elaborar una método de estimación de funciones de regresión con valores en la variedad \mathcal{M} de las matrices d.n.n. que satisfaga los siguientes requisitos: estar basado en el concepto de media extrínseca asociado al encajamiento de inclusión, y proveer estimadores que tomen valores solamente en la variedad \mathcal{M} . Planteamos estimadores no paramétricos ya que son flexibles, computacionalmente eficientes y permiten modelar patrones complejos en los modelos.
2. Utilizar el método propuesto para describir la maduración de la conectividad del EEG.

Estructura de la tesis

En el Capítulo 2 se elabora un enfoque basado en núcleos positivos, haciendo especial uso de la convexidad de la variedad \mathcal{M} en la construcción del estimador, proponiéndose un estimador no paramétrico de funciones de regresión con valores en una variedad convexa \mathcal{M}_C . Los métodos y resultados de este capítulo son aplicables al caso particular en que la variedad convexa esté constituida por las matrices d.n.n. (reales o complejas). Se estudian teóricamente algunas propiedades del estimador propuesto, incluyendo consistencia, orden de convergencia en media cuadrática y normalidad asintótica. En el Capítulo 3 se explora su comportamiento mediante datos simulados bajo diversas condiciones. El Capítulo 4 aplica esta metodología al estudio de la maduración de la conectividad del electroencefalograma (EEG) multi-canal a diferentes frecuencias durante el primer año de vida. Finalmente, el Capítulo 5 contiene las conclusiones del trabajo y propone algunas líneas de investigación futura.

Estimación no paramétrica de funciones de regresión con valores en una variedad convexa

En este capítulo se construye un método de estimación no paramétrica, basado en núcleos positivos, para funciones de regresión con valores en una variedad convexa. Ello abarca como casos particulares las variedades de las matrices definidas no negativas reales y complejas. El estimador propuesto está basado en el concepto de media extrínseca asociada al encajamiento de inclusión en el espacio ambiente, y toma valores solamente en la variedad. Se demuestran teóricamente varias propiedades deseables que posee el estimador.

2.1. Construcción del estimador mediante núcleos positivos

Sea \mathcal{M} la variedad de las matrices definidas no negativas (d.n.n.) de dimensión p , y sea u una covariable con valores en \mathbb{R}^m . Sea además $(X, Y) \in \mathbb{R}^m \times \mathcal{M}$ una variable aleatoria con distribución $P(dX, dY)$, y $P(dY|x)$ la distribución condicional de Y dado X . Denotemos por $\mu(x)$ a la media extrínseca de Y según la distribución condicional $P(dY|x)$ asociada al encajamiento de inclusión o identidad $J = \text{id} : A \in \mathcal{M} \subset E^D \rightarrow A \in E^D$, E^D un espacio euclideo D -dimensional, (ver Apéndice C para definición de media extrínseca).

Según se motivó en el Capítulo 1, nuestro propósito es estimar la función de regresión $\mu(x) \in \mathcal{M}$ de la variable respuesta Y como función de la covariable X , sobre la base una muestra de observaciones independientes (x_i, y_i) , $i = 1, \dots, n$, de (X, Y) . Queremos además que la estimación satisfaga los requisitos de ser una estimación no paramétrica y tomar valores solamente en la variedad \mathcal{M} .

Es bien conocido que la variedad \mathcal{M} es convexa. Más generalmente, en todo este capítulo consideraremos un espacio euclideo arbitrario E^D de dimensión finita D , con norma $\|\cdot\|$, y una *variedad convexa arbitraria* $\mathcal{M}_C \subset E^D$ completa (con la métrica de E^D).

El espacio euclideo $\mathbb{R}^{p \times p}$ de matrices de dimensión p dotado de la métrica de Frobenius es isomorfo al espacio euclideo \mathbb{R}^{p^2} . En efecto, la operación de vectorización de matrices (ver, e.g., [Lütkepohl, 1996]), $A \in \mathbb{R}^{p \times p} \rightsquigarrow \text{vec}(A) \in \mathbb{R}^{p^2}$, es un isomorfismo. Por tanto, haciendo la identificación $\mathcal{M} \cong \text{vec}(\mathcal{M})$, las propiedades que se demostrarán en este capítulo abarcan el caso de la *variedad \mathcal{M} de matrices d.n.n. reales* de dimensión p , tomando $\mathcal{M}_C = \mathcal{M}$ y $E^D = \mathbb{R}^{p^2}$.

Los resultados de este capítulo también son aplicables al caso de la *variedad $\mathcal{M}^{\mathbb{C}}$ de las matrices d.n.n. complejas* de dimensión p , $\mathcal{M}^{\mathbb{C}} \subset \mathbb{C}^p \times \mathbb{C}^p$. Para ello basta, sobre la isometría $A \in \mathbb{C}^p \times \mathbb{C}^p \rightsquigarrow (\text{Re}(\text{vec}(A)), \text{Im}(\text{vec}(A))) \in \mathbb{R}^{2p^2}$, tomar $\mathcal{M}_C = \mathcal{M}^{\mathbb{C}}$ y $E^D = \mathbb{R}^{2p^2}$. Esto es importante para la aplicación a datos de EEG que se hará en el Capítulo 4, pues allí trabajaremos con la variedad $\mathcal{M}^{\mathbb{C}}$.

La convexidad de \mathcal{M}_C implica propiedades específicas de la media extrínseca de una variable \mathcal{M}_C -valuada. En particular, como se enuncia en la siguiente proposición, la media extrínseca pertenece a \mathcal{M}_C .

Proposición 2.1. *Si Z es una variable aleatoria \mathcal{M}_C -valuada con $E\|Z\|^2 < \infty$, entonces se cumple que:*

- a) $\gamma \in \mathcal{M}_C$, donde $\gamma = E(Z) = (E(Z_1), \dots, E(Z_D))$, y $Z = (Z_1, \dots, Z_D)$,
- b) γ es la única media extrínseca de Z .

Demostración. Parte a). Sea (Ω, \mathcal{F}, Q) el espacio de probabilidad sobre el que está definida la variable Z . Sea una función simple arbitraria

$$G(\omega) = \sum_{j=1}^k \delta_j 1_{A_j}(\omega)$$

donde $\delta_j \in \mathcal{M}_C$ y $\{A_1, \dots, A_k\}$ es una partición de Ω por conjuntos medibles. La convexidad de \mathcal{M}_C implica que $E(G) = \sum_{j=1}^k \delta_j P(A_j) \in \mathcal{M}_C$. Aproximando la integral $\gamma = \int_{\mathcal{M}} Z dQ$ mediante valores esperados de funciones simples, y teniendo en cuenta la completitud de \mathcal{M} , se deduce que $\gamma \in \mathcal{M}_C$.

Parte b). $E\|Z - \alpha\|^2$ se minimiza con respecto a $\alpha \in E^D$ en el único punto $\alpha = \gamma$. Esto y (a) implican que γ es la única media extrínseca de Z . \square

La convexidad de \mathcal{M}_C desempeña un rol clave en la construcción que haremos de un estimador no paramétrico $\hat{\mu}(x)$ de $\mu(x)$ que tome valores solamente en \mathcal{M}_C . Específicamente, consideramos un estimador del tipo Nadaraya Watson,

$$\hat{\mu}(x) = \sum_{i=1}^n \frac{y_i K_H(x_i - x)}{\sum_{i=1}^n K_H(x_i - x)} \quad (2.1)$$

donde $K_H(x) = \frac{1}{|H|} K(H^{-1}x)$, $H = \text{Diag}(h_1, \dots, h_m)$ con $h_l > 0$ para $l \in \{1, \dots, m\}$, $|H| = h_1 \dots h_m$, y $K : \mathbb{R}^m \rightarrow \mathbb{R}$ es una función (núcleo) que cumple las siguientes condiciones:

$$\text{K1)} \int_{\mathbb{R}^m} K(x) dx = 1$$

$$\text{K2)} \int_{\mathbb{R}^m} x K(x) dx = 0$$

K3) K es función simétrica en torno al cero,

$$\text{K4)} \int_{\mathbb{R}^m} \|x\|^2 K(x) dx < \infty$$

El siguiente resultado es consecuencia directa de la convexidad de \mathcal{M}_C .

Proposición 2.2. Si K es una función no negativa, entonces $\hat{\mu}(x) \in \mathcal{M}_C$ para todo $x \in \mathbb{R}^m$.

Definición 2.3. Definimos el estimador $\hat{\mu}(x)$ de la función de regresión $\mu(x)$ mediante la expresión (2.1), agregando la condición de que el núcleo K utilizado sea no negativo.

Notemos que la Proposición (2.2) garantiza que el estimador propuesto satisface $\hat{\mu}(x) \in \mathcal{M}_C$.

Comentario. Se puede demostrar que cualquier estimador de $\mu(x)$ por polinomios locales con un núcleo no negativo también satisface la restricción de tomar valores solamente en la variedad \mathcal{M}_C . Pero en esta tesis nos limitaremos a estimadores del tipo (2.1), es decir, a estimadores por polinomios locales de grado cero.

2.2. Propiedades asintóticas

A continuación se demuestran algunas propiedades teóricas del estimador propuesto.

En adelante supondremos que el núcleo K cumple la condición de no negatividad y los supuestos (K1)-(K4). Además supondremos que la covariable x posee una densidad f diferenciable y con derivadas parciales acotadas.

El siguiente teorema garantiza la normalidad asintótica del estimador $\hat{\mu}(x)$ bajo condiciones adecuadas de regularidad.

Teorema 2.4. Supongamos que $\mu(x) \in C^2$ y $n|H| \rightarrow \infty$. Denotemos $x = (x^1, \dots, x^m)$. Sea $\tilde{\mu}(x) = \mu(x) + \frac{Z(x)}{f(x)}$, donde el k -ésimo componente $Z_k(x)$ ($k = 1, \dots, D$) de $Z(x)$ está dado por

$$Z_k(x) = h_1^2 \left(\frac{\partial f}{\partial x^1} \frac{\partial \mu_k}{\partial x^1} + \frac{1}{2} f(x) \frac{\partial^2 \mu_k}{\partial (x^1)^2} \right) \int v_1^2 K_1(v_1) dv_1 + \dots \quad (2.2)$$

$$+ h_m^2 \left(\frac{\partial f}{\partial x^m} \frac{\partial \mu_k}{\partial x^m} + \frac{1}{2} f(x) \frac{\partial^2 \mu_k}{\partial (x^m)^2} \right) \int v_m^2 K_m(v_m) dv_m. \quad (2.3)$$

Si además $P(dy|x)$ tiene segundo momento finito, entonces

$$\sqrt{n|H|}(\hat{\mu}(x) - \tilde{\mu}(x)) \xrightarrow{d} N(0, \bar{\Sigma}(x)), \quad (2.4)$$

cuando $n \rightarrow \infty$, donde

$$\bar{\Sigma} = (\bar{\Sigma}_{tk}),$$

$$\bar{\Sigma}_{tk} = \sigma(y^t, y^k) \frac{\int_{\mathbb{R}^m} K(v)^2 dv}{f(x)}$$

y $\sigma(y^t, y^k) = \text{cov}(y^t, y^k|x)$, $y = (y_1, \dots, y_D)$.

. Denotaremos convergencia en distribución como d .

Sea $\mathcal{X} \subset \mathbb{R}^m$ el rango de la covariable x . Se define el *error cuadrático medio integrado* (MISE) del estimador $\hat{\mu}(x)$ como

$$\text{MISE}(\hat{\mu}) = E \int_{\mathcal{X}} \|\mu(x) - \hat{\mu}(x)\|^2 dx \quad (2.5)$$

El teorema siguiente caracteriza el componente asintótico del $\text{MISE}(\hat{\mu})$ cuando el tamaño de muestra n crece.

Teorema 2.5. *Bajo las mismas condiciones del teorema (2.4), y suponiendo además que \mathcal{X} es acotado, para cualesquiera $h_1, \dots, h_m > 0$ fijos se cumple que*

$$\text{MISE}(\hat{\mu}) = \frac{C}{n|H|} + \sum_{i=1}^m O(h_i^4).$$

Donde C es una constante.

La parte (a) del siguiente corolario establece la razón de convergencia óptima $\text{MISE}(\hat{\mu})$ con una elección adecuada de los anchos h_j del núcleo. Su parte (b) ofrece una condición suficiente para que la media asintótica del estimador $\hat{\mu}(x)$ según el teorema (2.4) coincida con la verdadera media $\mu(x)$.

Corolario 2.6. *Bajo las mismas condiciones que los teoremas anteriores, se cumple que:*

a) Si $h_j = O(n^{-1/(m+4)})$ para $j = 1, \dots, m$, entonces

$$\text{MISE}(\hat{\mu}) = O(n^{-4/(m+4)})$$

b) Si $h_j = O(n^{-\beta})$ con $\frac{1}{m} < \beta < \frac{1}{m+4}$, para $j = 1, \dots, m$, entonces

$$\sqrt{n|H|}(\hat{\mu}(x) - \mu(x)) \xrightarrow{d} N(0, \bar{\Sigma}(x)), \quad (2.6)$$

y

$$\text{MISE}(\hat{\mu}) \rightarrow 0 \quad (2.7)$$

cuando $n \rightarrow \infty$.

2.3. Demostraciones de teoremas

En [Lin et al., rint] se aborda la estimación de la función de regresión en el sentido de media extrínseca en la siguiente situación. La variable respuesta Y toma valores en una variedad \mathbf{M} no necesariamente convexa; se especifica un encajamiento $J : \mathbf{M} \rightarrow E^D$, donde E^D es un espacio euclideo (sin suponer $\mathbf{M} \subset E^D$); y la función de regresión $\mu(x)$ se considera la media extrínseca de la distribución condicional $P(dY|x)$ según el encajamiento J (ver Apéndice C para definición de media extrínseca).

Para esta situación, dicho artículo construye un estimador $\hat{F}(x)$ de $\mu(x)$ según el procedimiento de tres pasos siguiente: 1) calcula $\hat{\mu}(x)$ como en (2.1) con un núcleo no necesariamente no negativo; 2) calcula la proyección $P\hat{\mu}(x)$ de $\hat{\mu}(x)$ sobre $J(\mathbf{M})$; 3) y finalmente, calcula $\hat{F}(x) = J^{-1}(P(\hat{\mu}(x)))$. Además, estudia propiedades asintóticas de este estimador.

Seguiremos técnicas de demostración similares a las utilizadas en [Lin et al., rint]. Pero dichas técnicas se simplifican notablemente debido a los supuestos de convexidad de la variedad \mathcal{M}_C y de no negatividad del núcleo K propios del contexto de regresión que nos interesa. Más específicamente, la Proposición (2.2) garantiza que $\hat{\mu}(x) \in \mathcal{M}_C$. Por tanto, los pasos 2) y 3) que acabamos de describir no son necesarios en nuestro método de estimación (equivalentemente, podemos considerar los operadores J y P como identidades).

Por simplicidad, en demostraciones que siguen supondremos que el núcleo K es un producto de núcleos univariados, $K(x) = \prod_{i=1}^m K_i(x^i)$, donde $x = (x^1, \dots, x^m)$.

Demostración del Teorema 2.4. Un estimador por núcleo de la densidad f está dado por

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_H(x_i - x) = \frac{1}{n|H|} \sum_{i=1}^n K(H^{-1}(x_i - x)). \quad (2.8)$$

Es estándar demostrar que para toda x fija

$$\hat{f}(x) \xrightarrow{P} f_X(x). \quad (2.9)$$

Por otra parte, notemos que

$$\begin{aligned}
E\left(\frac{1}{n}\sum_{i=1}^n y_i K_H(x_i - x)\right) &= \frac{1}{n}\sum_{i=1}^n E(y_i K_H(x_i - x)) \\
&= \frac{1}{n}\sum_{i=1}^n E[E(y_i K_H(x_i - x)) | x_i] \\
&= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^m} E(y_i K_H(x_i - x) | x_i) f(x_i) dx_i \\
&= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^m} E(y_i | x_i) K_H(x_i - x) f(x_i) dx_i \\
&= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^m} \mu(x_i) K_H(x_i - x) f(x_i) dx_i \\
&= \int_{\mathbb{R}^m} \mu(z) K_H(z - x) f(z) dz.
\end{aligned}$$

Haciendo el cambio de variable $v = H^{-1}(z - x)$ y notando que $dz = |H|dv$, se obtiene

$$E\left(\frac{1}{n}\sum_{i=1}^n J(y_i) K_H(x_i - x)\right) = \int_{\mathbb{R}^m} \mu(Hv + x) K(v) f(Hv + x) dv. \quad (2.10)$$

Dado que f es diferenciable, podemos hacer un desarrollo de Taylor multivariado en \mathbb{R}^m hasta primer orden:

$$f(x + Hv) = f(x) + (Hv) \cdot (\nabla f(x)) + R(Hv, x), \quad (2.11)$$

donde $a \cdot b$ denota producto escalar de vectores $a, b \in \mathbb{R}^m$.

Dado que $\mu : \mathbb{R}^m \rightarrow \mathbb{R}^D$ y $\mu \in C^2$, es posible aplicar el desarrollo de Taylor hasta segundo orden en cada una de las componentes $\mu_k(x)$, $k = 1, \dots, D$. De esta manera

$$\begin{aligned}
\mu_k(x + Hv) &= \mu_k(x) + \sum_{j=1}^m (Hv)_j \frac{\partial \mu_k}{\partial x^j}(x) + \frac{1}{2} \sum_{j,l=1}^m (Hv)_j (Hv)_l \frac{\partial^2 \mu_k}{\partial x^j \partial x^l}(x) + R_k(Hv, x) \\
&= \mu_k(x) + \sum_{j=1}^m h_j v_j \frac{\partial \mu_k}{\partial x^j}(x) + \frac{1}{2} \sum_{j,l=1}^m h_j v_j h_l v_l \frac{\partial^2 \mu_k}{\partial x^j \partial x^l}(x) + R_k(Hv, x).
\end{aligned}$$

Para escribir de manera matricial los desarrollos anteriores, introduzcamos la matriz A de $D \times m$ con renglones $A_k = (\nabla \mu_k(x))^T$ y el vector V de $D \times 1$ con componentes $V_k = \left(\frac{1}{2}(Hv)^T (\nabla^2 \mu_k(x)) (Hv)\right)$. De este modo

$$\mu(x + Hv) = \mu(x) + A(Hv) + V + R(Hv, x). \quad (2.12)$$

Notemos que los órdenes de los residuos en las ecuaciones (2.11) y (2.12) son $o(\max\{h_1, \dots, h_m\})$ y $o(\max\{h_1^2, \dots, h_m^2\})$, respectivamente.

De (2.11), (2.12) y (2.10) obtenemos que

$$E \left(\frac{1}{n} \sum_{i=1}^n J(y_i) K_H(x_i - x) \right) \quad (2.13)$$

$$\begin{aligned} &\approx \int_{\mathbb{R}^m} (\mu(x) + A(Hv) + V) K(v) (f(x) + (Hv) \cdot (\nabla f(x))) dv \\ &= \int_{\mathbb{R}^m} f(x) \mu(x) + f(x) K(v) A(Hv) + f_X(x) K(v) V \\ &+ \mu(x) (\nabla f) \cdot (Hv) K(v) + (\nabla f) \cdot (Hv) K(v) A(Hv) + (\nabla f) \cdot (Hv) K(v) V dv \\ &= f(x) \mu(x) + f(x) \int_{\mathbb{R}^m} K(v) A(Hv) dv + f(x) \int_{\mathbb{R}^m} K(v) V dv \end{aligned} \quad (2.14)$$

$$+ \mu(x) \int_{\mathbb{R}^m} (\nabla f) \cdot (Hv) K(v) dv + \int_{\mathbb{R}^m} (\nabla f) \cdot (Hv) K(v) A(Hv) dv + \int_{\mathbb{R}^m} (\nabla f) \cdot (Hv) K(v) V dv. \quad (2.15)$$

A continuación demostraremos que el segundo término de la ecuación (2.14) es igual a cero. En efecto,

$$f(x) \int_{\mathbb{R}^m} K(v) A(Hv) dv = f(x) \int_{\mathbb{R}^m} \prod_{s=1}^m K_s(v_s) A(Hv) dv.$$

Notemos que $A(Hv)$ es de dimensión $D \times 1$. Su componente k -ésima es la siguiente:

$$\begin{aligned} f(x) \int_{\mathbb{R}^m} K(v) (A(Hv))_k dv &= \int_{\mathbb{R}^m} \prod_{s=1}^m K_s(v_s) (A(Hv))_k dv \\ &= \int_{\mathbb{R}^m} \prod_{s=1}^m K_s(v_s) \sum_{j=1}^m \frac{\partial \mu_k}{\partial x^j}(x) h_j v_j dv \\ &= \sum_{j=1}^m \frac{\partial \mu_k}{\partial x^j}(x) h_j \int_{\mathbb{R}^m} v_j \prod_{s=1}^m K_s(v_s) dv. \end{aligned}$$

Sea $v_{-j} = (v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m)^T$. Notemos que por hipótesis (K2), cada K_j satisface $\int_{\mathbb{R}} v_j K_j(v_j) = 0$. Aplicando el teorema de Vitali-Fubini obtenemos que

$$\begin{aligned} &= \sum_{j=1}^m \frac{\partial \mu_k}{\partial x^j}(x) h_j \left(\int_{\mathbb{R}} v_j K_j(v_j) \right) \int_{\mathbb{R}^{m-1}} \prod_{s=1, s \neq j}^m K_s(v_s) dv_{-j} \\ &= \sum_{j=1}^m \frac{\partial \mu_k}{\partial x^j}(x) h_j \left(\int_{\mathbb{R}} v_j K_j(v_j) \right) \prod_{s=1, s \neq j}^m \int_{\mathbb{R}} K_s(v_s) dv_s \\ &= 0. \end{aligned}$$

De lo anterior concluimos que

$$f(x) \int_{\mathbb{R}^m} K(v) A(Hv) dv = \bar{0}, \quad (2.16)$$

donde $\bar{0} = (0, \dots, 0) \in \mathbb{R}^D$.

Consideraremos a continuación el tercer término de la ecuación (2.14). Comencemos notando que $V \in \mathbb{R}^D$, por lo cual el k -ésimo elemento de este término está dado por:

$$\begin{aligned} f(x) \int_{\mathbb{R}^m} K(v) V_k dv &= f(x) \int_{\mathbb{R}^m} \prod_{s=1}^m K_s(v_s) V_k dv \\ &= f(x) \int_{\mathbb{R}^m} \prod_{s=1}^m K_s(v_s) \sum_{j,l=1}^m h_j h_l v_j v_l \frac{\partial^2 \mu_k}{\partial x^j \partial x^l}(x) dv \\ &= f(x) \sum_{j=1}^m h_j^2 \frac{\partial^2 \mu_k}{\partial (x^j)^2}(x) \int_{\mathbb{R}^m} \prod_{s=1}^m K_s(v_s) v_j^2 dv \\ &\quad + f(x) \sum_{j \neq l}^m h_j h_l \frac{\partial^2 \mu_k}{\partial x^j \partial x^l}(x) \int_{\mathbb{R}^m} \prod_{s=1}^m K_s(v_s) v_j v_l dv. \end{aligned}$$

Aplicando el teorema de Vitali-Fubini obtenemos que

$$\begin{aligned} f(x) \int_{\mathbb{R}^m} K(v) V_k dv &= f(x) \sum_{j=1}^m h_j^2 \frac{\partial^2 \mu_k}{\partial (x^j)^2}(x) \left(\int_{\mathbb{R}} v_j^2 K_j(v_j) dv_j \right) \prod_{s=1, s \neq j}^m \int_{\mathbb{R}} K_s(v_s) dv_s \\ &\quad + f(x) \sum_{j \neq l}^m h_j h_l \frac{\partial^2 \mu_k}{\partial x^j \partial x^l}(x) \left(\int_{\mathbb{R}} v_j K_j(v_j) dv_j \right) \left(\int_{\mathbb{R}} v_l K_l(v_l) dv_l \right) \prod_{s \neq j, s \neq l}^m \int_{\mathbb{R}} K_s(v_s) dv_s \\ &= f(x) \sum_{j=1}^m h_j^2 \frac{\partial^2 \mu_k}{\partial (x^j)^2}(x) \int_{\mathbb{R}} v_j^2 K_j(v_j) dv_j. \end{aligned}$$

De aquí concluimos que

$$f(x) \int_{\mathbb{R}^m} K(v) V_k dv = f(x) \sum_{j=1}^m h_j^2 \frac{\partial^2 \mu_k}{\partial (x^j)^2}(x) \int_{\mathbb{R}} v_j^2 K_j(v_j) dv_j. \quad (2.17)$$

El primer término de la ecuación (2.15) es $\bar{0}$, pues

$$\begin{aligned} \mu(x) \int_{\mathbb{R}^m} (\nabla f) \cdot (Hv) K(v) dv &= \mu(x) \int_{\mathbb{R}^m} \sum_{j=1}^m \frac{\partial f}{\partial x^j}(x) h_j v_j \prod_{s=1}^m K_s(v_s) dv \\ &= \mu(x) \sum_{j=1}^m \frac{\partial f}{\partial x^j}(x) h_j \int_{\mathbb{R}} v_j K_j(v_j) dv_j \\ &= \bar{0}. \end{aligned}$$

De lo anterior concluimos que

$$\mu(x) \int_{\mathbb{R}^m} (\nabla f) \cdot (Hv) K(v) dv = \bar{0}. \quad (2.18)$$

El segundo término de la ecuación (2.15) puede reescribirse

$$\int_{\mathbb{R}^m} (\nabla f) \cdot (Hv) K(v) A(Hv) dv = \int_{\mathbb{R}^m} \sum_{j=1}^m \frac{\partial f}{\partial x^j}(x) h_j v_j \prod_{s=1}^m K_s(v_s) A(Hv) dv.$$

Notemos que $A(Hv)$ es de dimensión $D \times 1$. Su componente k -ésima está dada por:

$$\begin{aligned} \int_{\mathbb{R}^m} (\nabla f) \cdot (Hv)K(v)(A(Hv))_k dv &= \int_{\mathbb{R}^m} \sum_{j=1}^m \frac{\partial f}{\partial x^j}(x) h_j v_j \prod_{s=1}^m K_s(v_s) (A(Hv))_k dv \\ &= \int_{\mathbb{R}^m} \left(\sum_{j=1}^m \frac{\partial f}{\partial x^j}(x) h_j v_j \right) \left(\prod_{s=1}^m K_s(v_s) \right) \sum_{l=1}^m \frac{\partial \mu_k}{\partial x^l}(x) h_l v_l dv \\ &= \sum_{j,l=1}^m \frac{\partial f}{\partial x^j}(x) \frac{\partial \mu_k}{\partial x^l}(x) \int_{\mathbb{R}^m} \left(\prod_{s=1}^m K_s(v_s) \right) h_j h_l v_j v_l dv. \end{aligned}$$

Teniendo en cuenta el supuesto (K4) y aplicando el teorema de Vitali-Fubini obtenemos que

$$\begin{aligned} \int_{\mathbb{R}^m} (\nabla f) \cdot (Hv)K(v)(A(Hv))_k dv &= \sum_{j=1}^m \frac{\partial f}{\partial x^j}(x) \frac{\partial \mu_k}{\partial x^j}(x) h_j^2 \left(\int_{\mathbb{R}} v_j^2 K_j(v_j) dv_j \right) \int_{\mathbb{R}^{m-1}} \prod_{s=1, s \neq j}^m K_s(v_s) dv_{-j} \\ &+ \sum_{j,l=1; j \neq l}^m \frac{\partial f}{\partial x^j}(x) \frac{\partial \mu_k}{\partial x^l}(x) h_j h_l \left(\int_{\mathbb{R}} v_j K_j(v_j) dv_j \right) \int_{\mathbb{R}^{m-1}} v_l \prod_{s=1, s \neq j}^m K_s(v_s) dv_{-j} \\ &= \sum_{j=1}^m \frac{\partial f}{\partial x^j}(x) \frac{\partial \mu_k}{\partial x^j}(x) h_j^2 \int_{\mathbb{R}} v_j^2 K_j(v_j) dv_j. \end{aligned}$$

De lo anterior concluimos que

$$\int_{\mathbb{R}^m} (\nabla f) \cdot (Hv)K(v)(A(Hv))_k dv = \sum_{j=1}^m \frac{\partial f}{\partial x^j}(x) \frac{\partial \mu_k}{\partial x^j}(x) h_j^2 \int_{\mathbb{R}} v_j^2 K_j(v_j) dv_j. \quad (2.19)$$

A continuación comprobaremos que el tercer término de la ecuación (2.15) es nulo. Su k -ésima componente es

$$\begin{aligned} \int_{\mathbb{R}^m} (\nabla f) \cdot (Hv)K(v)V_k dv &= \int_{\mathbb{R}^m} \left(\sum_{j=1}^m \frac{\partial f}{\partial x^j}(x) h_j v_j \right) \prod_{s=1}^m K_s(v_s) \left(\sum_{l,r=1}^m h_r h_l v_r v_l \frac{\partial^2 \mu_k}{\partial x^r \partial x^l}(x) \right) dv \\ &= \sum_{l,r,j=1}^m \frac{\partial f}{\partial x^j}(x) \frac{\partial^2 \mu_k}{\partial x^r \partial x^l}(x) h_j h_r h_l \int_{\mathbb{R}^m} v_j v_r v_l \prod_{s=1}^m K_s(v_s) dv. \end{aligned}$$

Teniendo en cuenta los supuestos (K2) y (K4), y aplicando el teorema de Vitali Fubini, obtenemos que la expresión anterior es igual a:

$$\begin{aligned}
& \sum_{l=r=j}^m \frac{\partial f}{\partial x^j}(x) \frac{\partial^2 \mu_k}{\partial (x^j)^2}(x) h_j^3 \int_{\mathbb{R}} v_j^3 K_j(v_j) dv_j \\
& + \sum_{l=r, r \neq j}^m \frac{\partial f}{\partial x^j}(x) \frac{\partial^2 \mu_k}{\partial (x^r)^2}(x) h_j h_r^2 \left(\int_{\mathbb{R}} v_r^2 K_r(v_r) dv_r \right) \left(\int_{\mathbb{R}} v_j K_j(v_j) v_j \right) \\
& + \sum_{l \neq r, r \neq j, j \neq l}^m \frac{\partial f}{\partial x^j}(x) \frac{\partial^2 \mu_k}{\partial x^r \partial x^l}(x) h_j h_r h_l \left(\int_{\mathbb{R}} v_j K_j(v_j) v_j \right) \left(\int_{\mathbb{R}} v_r K_r(v_r) v_r \right) \left(\int_{\mathbb{R}} v_l K_l(v_l) v_l \right) \\
& = \sum_{l=r=j}^m \frac{\partial f}{\partial x^j}(x) \frac{\partial^2 \mu_k}{\partial (x^j)^2}(x) h_j^3 \int_{\mathbb{R}} v_j^3 K_j(v_j) dv_j.
\end{aligned}$$

Esta última expresión es nula, dado que el núcleo K es simétrico (supuesto (K3)) y todos los momentos impares de un kernel simétrico son iguales a 0.

De lo anterior concluimos que

$$\int_{\mathbb{R}^m} (\nabla f) \cdot (Hv) K(v) V_k dv = \bar{0}. \quad (2.20)$$

Definamos Z_k mediante

$$Z_k = \sum_{j=1}^m h_j^2 \left(f_X(x) \frac{\partial^2 \mu_k}{\partial (x^j)^2}(x) + \frac{\partial f}{\partial x^j}(x) \frac{\partial \mu_k}{\partial x^j}(x) \right) \int_{\mathbb{R}} v_j^2 K_j(v_j) dv_j. \quad (2.21)$$

Las ecuaciones (2.20), (2.19), (2.18), (2.17) y (2.16) implican que

$$E \left(\frac{1}{n} \sum_{i=1}^n J(y_i) K_H(x_i - x) \right) \approx f(x) \mu(x) + Z, \quad (2.22)$$

donde $Z = (Z_1, \dots, Z_D)^T$.

A continuación calcularemos la matriz de covarianza de $\frac{1}{n} \sum_{i=1}^n J(y_i) K_H(x_i - x)$, que será denotada por $\Sigma(x) = (\Sigma_{jk})$. Para $y_i \in \mathbb{R}^D$, usaremos la notación $y_i = (y_i^1, \dots, y_i^D)$. Entonces

$$\begin{aligned}
\Sigma_{tk} &= E\left[\left(\frac{1}{n}\sum_{i=1}^n y_i^t K_H(x_i - x) - E\left(\frac{1}{n}\sum_{i=1}^n y_i^t K_H(x_i - x)\right)\right)\right. \\
&\quad \left.\left(\frac{1}{n}\sum_{i=1}^n y_i^k K_H(x_i - x) - E\left(\frac{1}{n}\sum_{i=1}^n y_i^k K_H(x_i - x)\right)\right)\right] \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^n \left(y_i^t K_H(x_i - x) - \int \mu_t(x_i) K_H(x_i - x) f(x_i) d(x_i)\right)\right)\right. \\
&\quad \left.\left(\frac{1}{n}\sum_{i=1}^n \left(y_i^k K_H(x_i - x) - \int \mu_k(x_i) K_H(x_i - x) f(x_i) d(x_i)\right)\right)\right) \\
&= \frac{1}{n^2} E\left[\int_{\mathbb{R}} \left(\sum_{i=1}^n \left(y_i^t K_H(x_i - x) - \int \mu_t(x_i) K_H(x_i - x) f(x_i) d(x_i)\right)\right)\right. \\
&\quad \left.\left(\sum_{i=1}^n \left(y_i^k K_H(x_i - x) - \int \mu_k(x_i) K_H(x_i - x) f(x_i) d(x_i)\right)\right) \Big| x_1 \right] f(x_1) dx_1 \\
&= \frac{1}{n} E\left[\int_{\mathbb{R}} \left(y_1^t K_H(x_1 - x) - \int \mu_t(x_1) K_H(x_1 - x) f(x_1) d(x_1)\right)\right. \\
&\quad \left.\left(y_1^k K_H(x_1 - x) - \int \mu_k(x_1) K_H(x_1 - x) f(x_1) d(x_1)\right) \Big| x_1 \right] f(x_1) dx_1.
\end{aligned}$$

Denotando por $\sigma_{tk}(y_1^t, y_1^k)$ a la covarianza condicional con respecto a x_1 de la entrada t -ésima y j -ésima del vector y , tenemos

$$\begin{aligned}
&= \frac{1}{n} \int_{\mathbb{R}} \sigma(y_1^t K_H(x_1 - x), y_1^k K_H(x_1 - x)) f_X(x_1) dx_1 \\
&= \frac{1}{n} \int_{\mathbb{R}} K_H(x_1 - x)^2 \sigma(y_1^t, y_1^k) f_X(x_1) dx_1.
\end{aligned}$$

Por el cambio de variable $v = H^{-1}(x_1 - x)$, obtenemos

$$\begin{aligned}
\Sigma_{tk} &= \frac{1}{n|H|} \int_{\mathbb{R}} K(v)^2 \sigma(y_{x+Hv}^t, y_{x+Hv}^k | x + Hv) f(Hv + x) dv \\
&= \frac{1}{n|H|} \int_{\mathbb{R}} K(v)^2 \sigma(y_{x+Hv}^t, y_{x+Hv}^k | x + Hv) (f(x) + \nabla f \cdot (Hv) + o(\max\{h_1, \dots, h_m\})) dv \\
&= \frac{1}{n|H|} \int_{\mathbb{R}} K(v)^2 \sigma(y_{x+Hv}^t, y_{x+Hv}^k | x + Hv) (f(x) dv + o\left(\frac{1}{n|H|}\right) dv.
\end{aligned}$$

De lo anterior concluimos que

$$\Sigma_{tk} = \frac{1}{n|H|} \int_{\mathbb{R}} K(v)^2 \sigma(y_{x+Hv}^t, y_{x+Hv}^k | x + Hv) f(x) dv + o\left(\frac{1}{n|H|}\right).$$

Por razonamientos similares, basados en desarrollos de Taylor, se puede comprobar que

$$\Sigma_{tk} = \frac{1}{n|H|} \int_{\mathbb{R}} K(v)^2 \sigma(y^t, y^k) f(x) dv + o\left(\frac{1}{n|H|}\right), \quad (2.23)$$

donde $\sigma(y^t, y^k) = \text{cov}(y^t, y^k | x)$.

De aquí, aplicando el teorema del límite central y el teorema de Slutsky, concluimos la tesis dada por la ecuación 2.4. □

Demostración el Teorema 2.5. Usando razonamientos análogos a los utilizados para demostrar el corolario 4.2 en [Lin et al., rint] se obtiene la tesis del teorema. Omitimos los detalles. □

Demostración del Corolario 2.6. La parte a) del corolario es consecuencia del teorema (2.5), minimizando el MISE con respecto a h .

La parte b) es consecuencia de los teoremas (2.4) y (2.5), notando que con la elección $h_j = O(n^{-\beta})$ ($j = 1, \dots, n$) de los anchos de banda se cumple que $n|H| \rightarrow \infty$ y $\sqrt{n|H|h_j^2} \rightarrow 0$ cuando $n \rightarrow \infty$. □

Aplicación a datos simulados

En este capítulo se explora, a través de datos simulados, el comportamiento del estimador $\hat{\Sigma}(u)$ que fue definido y estudiado teóricamente en el capítulo anterior.

3.1. Procedimientos de simulación y estimación

El procedimiento que seguiremos para obtener las muestras de los datos a ser usados en este estudio por simulación es el siguiente:

- 1) Se especifica la dimensión p de las matrices d.n.n. que constituyen la variedad \mathcal{M} .
- 2) Se elige un modelo de función de regresión $\Sigma(u)$ con valores en la variedad \mathcal{M} , dependiente de una covariable u con rango $[0, 1]$.
- 3) Se toma un tamaño de muestra n .
- 4) Se especifica valores equiespaciados $u_i \in [0, 1]$, ($i = 1, \dots, n$) de la covariable u .
- 5) Se elige un número $m \in \mathbb{N}$ de grados de libertad.
- 6) Se simulan observaciones $M(u_i)$ independientes con distribución Wishart $W_p(m, \Sigma(u_i))$, $i = 1, \dots, n$.
- 7) Se calcula

$$S(u_i) = \frac{M(u_i)}{m} \tag{3.1}$$

para $i = 1, \dots, n$.

Notar que $E(S(u_i)) = \Sigma(u_i)$, donde el valor esperado es en el sentido de la media ex-rínseca según métrica euclídeana. Por tanto, la muestra de datos así obtenida $(u_i, S(u_i))$,

$i = 1, \dots, n$, constituyen una muestra de valores u_i de una covariable y los correspondientes valores $S(u_i)$ de una variable respuesta que tiene media extrínseca $\Sigma(u_i)$. De esta manera $\Sigma(u)$ es la función de regresión de la variable respuesta $S(u)$ con covariable u .

El procedimiento de estimación de la función de regresión es el siguiente.

Dado un ancho de banda $h > 0$, sobre la base de dichos datos, se calcula el estimador $\hat{\Sigma}(u; h)$ de la función de regresión $\Sigma(u)$ propuesto en el capítulo anterior, usando un núcleo Gaussiano K :

$$\hat{\Sigma}(u) = \hat{\Sigma}(u; \hat{h}) = \frac{\sum_{i=1}^n S(u_i) K\left(\frac{u_i - u}{\hat{h}}\right)}{\sum_{i=1}^n K\left(\frac{u_i - u}{\hat{h}}\right)}. \quad (3.2)$$

La elección \hat{h} del ancho de banda h se llevo acabo mediante el siguiente procedimiento de validación cruzada:

$$\hat{h} = \arg \min_{h>0} \text{CV}(h), \quad (3.3)$$

donde

$$\text{CV}(h) = \sum_{i=1}^n \sum_{j \leq k} (S_{jk}(u_i) - \hat{\sigma}_{kj}^{(i)}(u_i; h))^2 \quad (3.4)$$

y $\hat{\Sigma}^{(i)}(u_i; h) = (\hat{\sigma}_{jk}^{(i)}(u_i))_{1 \leq j, k \leq p}$ es la estimación obtenida al quitar de la muestra el dato $(u_i, S(u_i))$ en la expresión 3.2.

Se exploró el comportamiento del estimador $\hat{\Sigma}(u)$ en una amplia variedad de escenarios, variando el modelo de función de regresión y los parámetros p, m y n . A continuación se presentan algunos de los resultados que ilustran el desempeño del estimador. Cabe mencionar que parte de la complejidad de este capítulo viene de encontrar gráficos para visualizar las estimaciones de la función de regresión que nos permitan hacer análisis comparativos.

3.2. Modelo 1 con muestras grandes

Este modelo de función de regresión permite evaluar el comportamiento del estimador $\hat{\Sigma}(u)$ en una situación de dimensión $p = 2$ donde los autovectores de $\Sigma(u)$ rotan al variar la covariable u , mientras que los autovalores permanecen constantes. Específicamente, se considera la función de regresión:

$$\Sigma(u) = \begin{pmatrix} \sin(2\pi u) & \cos(2\pi u) \\ \cos(2\pi u) & -\sin(2\pi u) \end{pmatrix} \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{pmatrix} \begin{pmatrix} \sin(2\pi u) & \cos(2\pi u) \\ \cos(2\pi u) & -\sin(2\pi u) \end{pmatrix}^T, \quad (3.5)$$

donde $\lambda_1 = 3$ y $\lambda_2 = 1$. En las gráficas que se presentan a continuación se muestran resultados obtenidos con tamaño de muestra $n = 100$ y grados de libertad $m = 30$.

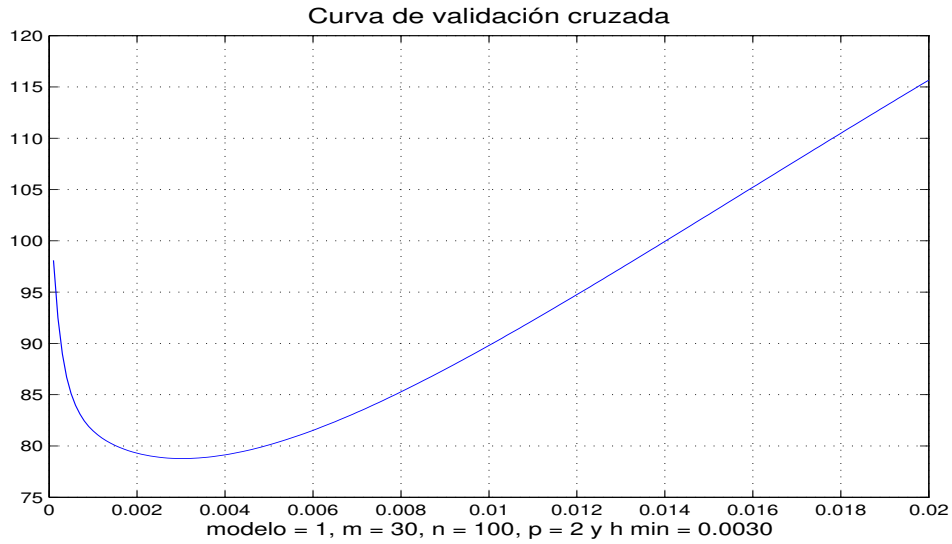


Figura 3.1

En la Figura 3.1 se puede observar que la función de validación cruzada $CV(\cdot)$ es suave y alcanza un mínimo global, el cual es un escenario óptimo para la selección del ancho de banda. Cabe mencionar que en todos los escenarios de simulación explorados se observó un comportamiento de la función CV similar.

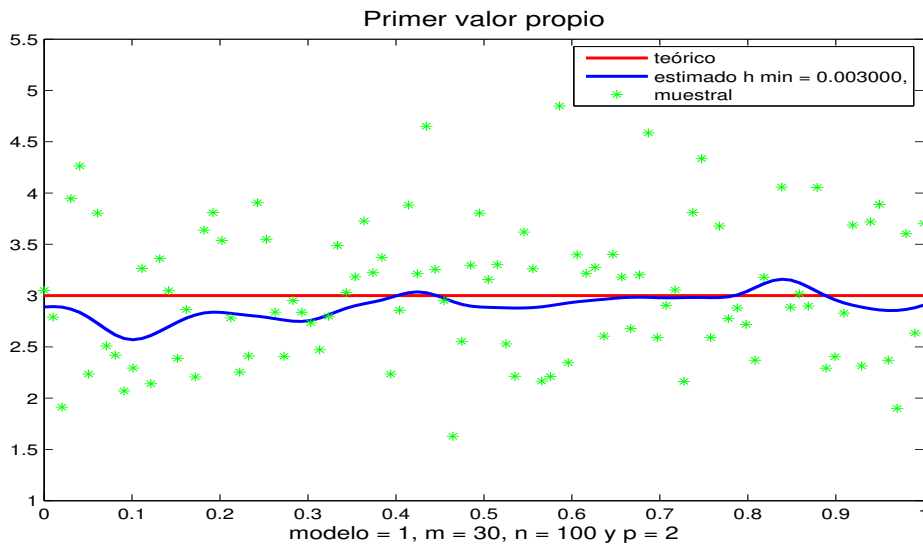


Figura 3.2

En la Gráfica 3.2 se muestran las curva del primer autovalor $\hat{\lambda}_1(u)$ de la regresión

estimada $\hat{\Sigma}(u)$ (en azul) y del primer autovalor $\lambda_1(u) \equiv \lambda(u) = 3$ de la matriz de regresión teórica $\Sigma(u)$ (en rojo), ambas como función de la covariable u . Se puede observar que ambas curvas son muy cercanas entre sí, y por lo tanto empíricamente el estimador parece ser bueno.

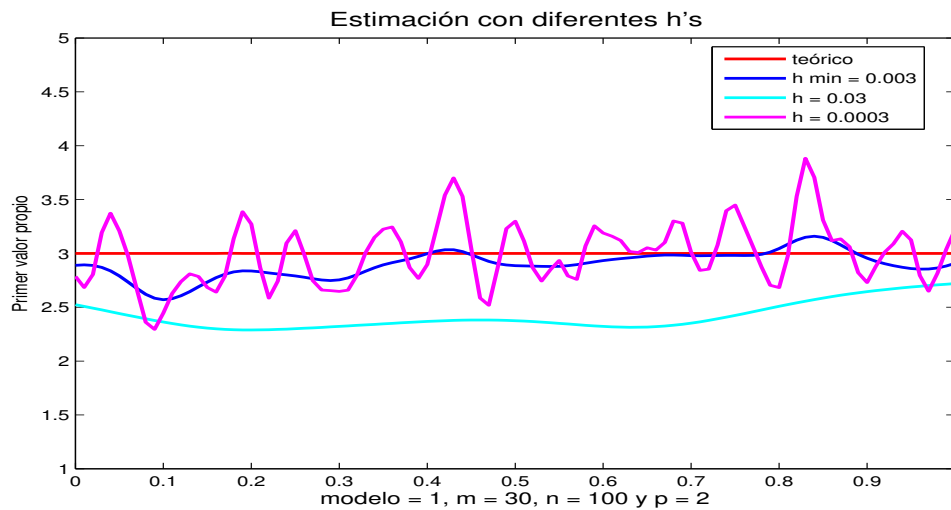


Figura 3.3

La Figura 3.3 ilustra que el ancho de banda $\hat{h} = h_{min}$ seleccionado por validación cruzada conduce a un buen comportamiento del estimador del primer autovalor. En efecto, se observa que con un ancho ligeramente menor la estimación es muy ruidosa, y con un ancho ligeramente mayor resulta ser muy sesgada.

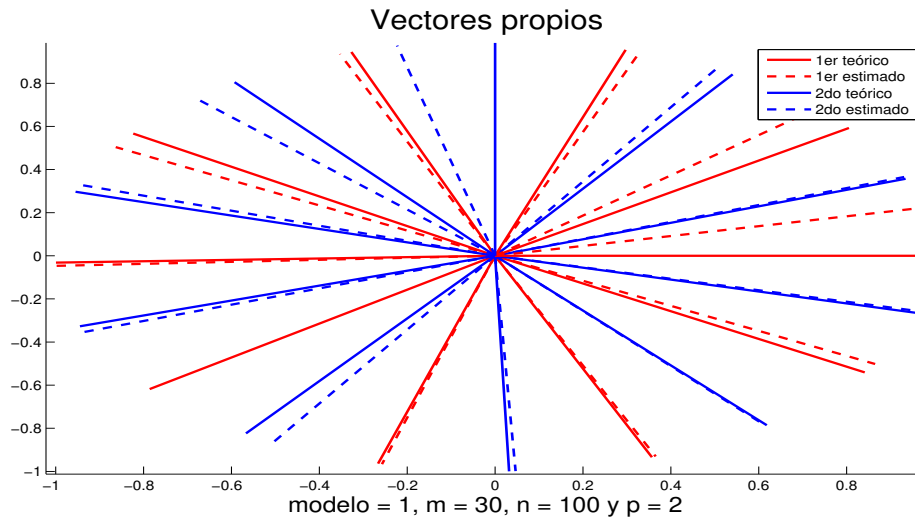


Figura 3.4

En la Gráfica 3.4 se observa la gran cercanía entre los vectores propios de la matriz estimada $\hat{\Sigma}(u)$ (las líneas punteadas) y los correspondientes de la matriz teórica $\Sigma(u)$ (líneas continuas), para un conjunto de valores de la covariable u . Es decir, que la dirección de los valores propios teóricos y los estimados, es en casi todos los casos, casi indistinguible. Notemos que en este caso nos encontramos viendo dos diferentes vectores propios debido a que la dimensión de las matrices es justo $p=2$.

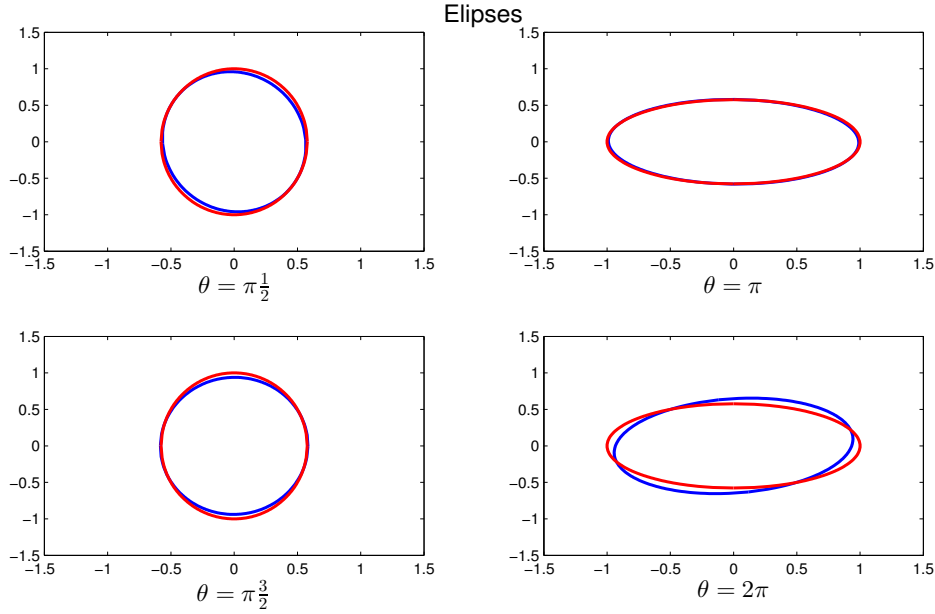


Figura 3.5

El gráfico 3.5 muestra, para varios valores del ángulo $\theta = \theta(u) = 2\pi u$, las elipses de radio 1 correspondientes a la matriz teórica $\Sigma(u)$ (en rojo) y a su estimación $\hat{\Sigma}(u)$ (en azul):

$$\{x \in \mathbb{R}^p | x^T \Sigma(u) x = 1\} \quad (3.6)$$

y

$$\{x \in \mathbb{R}^p | x^T \hat{\Sigma}(u) x = 1\} \quad (3.7)$$

donde $p = 2$. Notar que, por construcción, estas elipses dependen tanto de los valores propios como de los vectores propios de su correspondiente matriz, $\Sigma(u)$ ó $\hat{\Sigma}(u)$. Como se observa en este gráfico las elipses estimadas en general aproximan aceptablemente a las correspondientes poblacionales. Es destacable el buen comportamiento mostrado a pesar de las no linealidades involucradas en este modelo de regresión.

3.3. Modelo 1 con muestras pequeñas

Al disminuir los grados de libertad m en el paso (5) del procedimiento de simulación, aumenta la varianza de la variable respuesta $S(u_i)$. Este aumento de varianza de las observaciones es de esperar que degrade el comportamiento del estimador $\hat{\Sigma}(u)$, al igual que una disminución del tamaño de muestra n . A continuación ilustraremos esta degradación para valores $m = 25$ y $n = 40$, los cuales son menores que los utilizados en las gráficas de la sección anterior.

Las Gráficas 3.6, 3.7 y 3.8 muestran, para este nuevo escenario de simulación, los mismos aspectos que las Gráficas 3.2, 3.4 y 3.5, respectivamente.

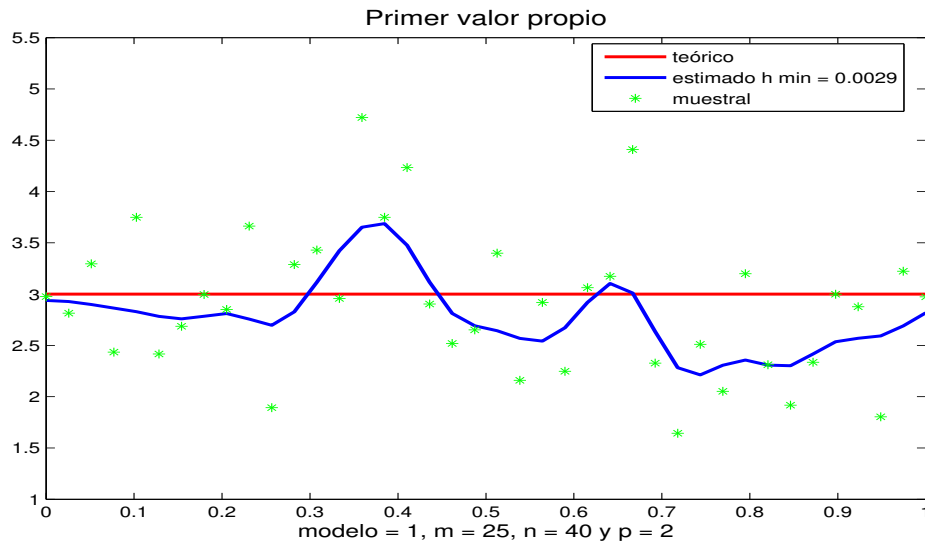


Figura 3.6

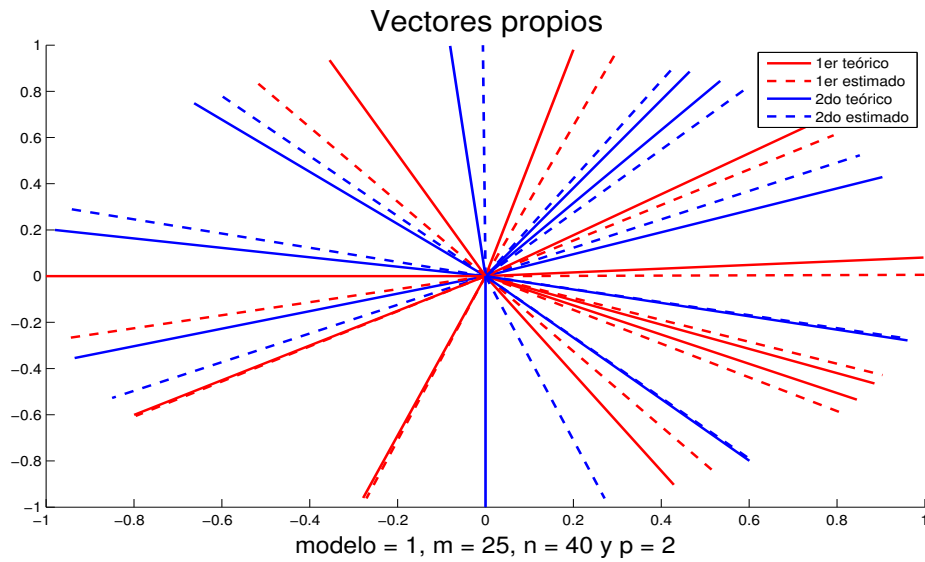


Figura 3.7

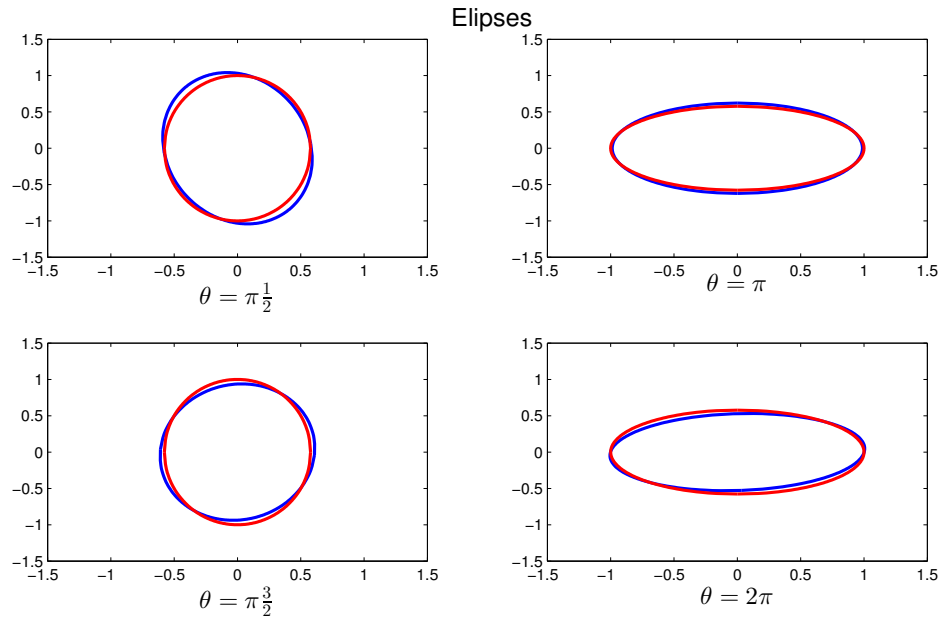


Figura 3.8

En las nuevas gráficas se observa, como era de esperar, un decaimiento en la precisión y ajuste de las estimaciones. Sin embargo, es destacable que este empeoramiento es moderado, más aun teniendo en cuenta el carácter no paramétrico del estimador y las no linealidades involucradas en la regresión.

3.4. Modelo 2

En este modelo de funciones de regresión $\Sigma(u)$ se considera la misma dependencia de los autovalores de $\Sigma(u)$ con respecto a la covariable especificada en el modelo 1; pero contiene además una dependencia lineal de los valores con propios respecto a la covariable. El modelo se define por:

$$\Sigma(u) = \begin{pmatrix} \sin(2\pi u) & \cos(2\pi u) \\ \cos(2\pi u) & -\sin(2\pi u) \end{pmatrix} \begin{pmatrix} \lambda_2(u) & 0 \\ 0 & \lambda_1(u) \end{pmatrix} \begin{pmatrix} \sin(2\pi u) & \cos(2\pi u) \\ \cos(2\pi u) & -\sin(2\pi u) \end{pmatrix}^T, \quad (3.8)$$

donde

$$\lambda_i(u) = \lambda_i u + k_i, \quad (3.9)$$

con $k = 1$, $k = 2$, $\lambda_1 = 1$ y $\lambda_2 = 2$. Los siguientes gráficos (similares a los gráficos 3.2, 3.4 y 3.5) muestran resultados obtenidos con tamaño de muestra $n = 100$ y dimensión $p = 2$.

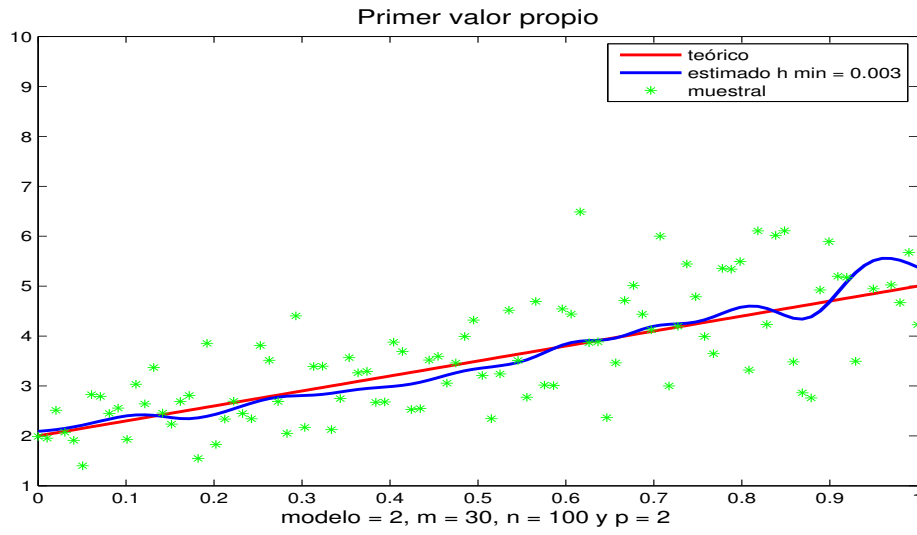


Figura 3.9

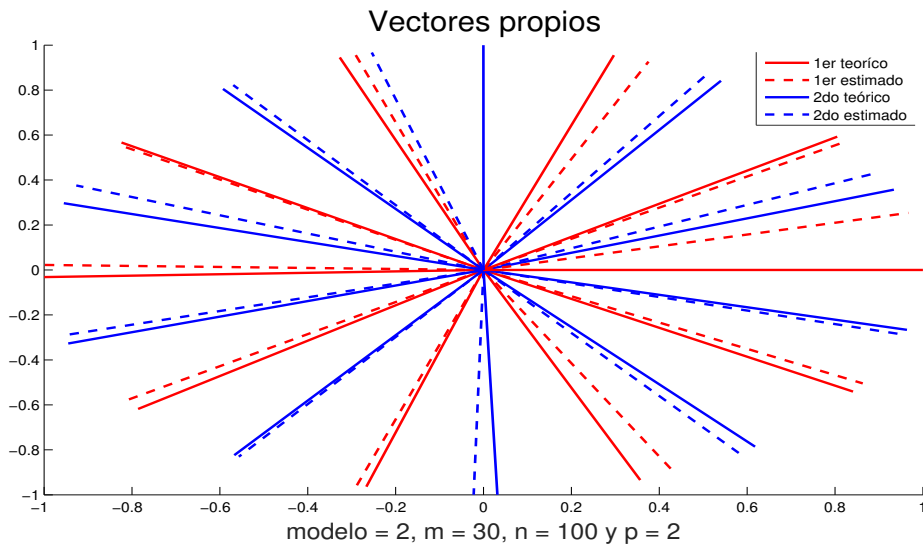


Figura 3.10

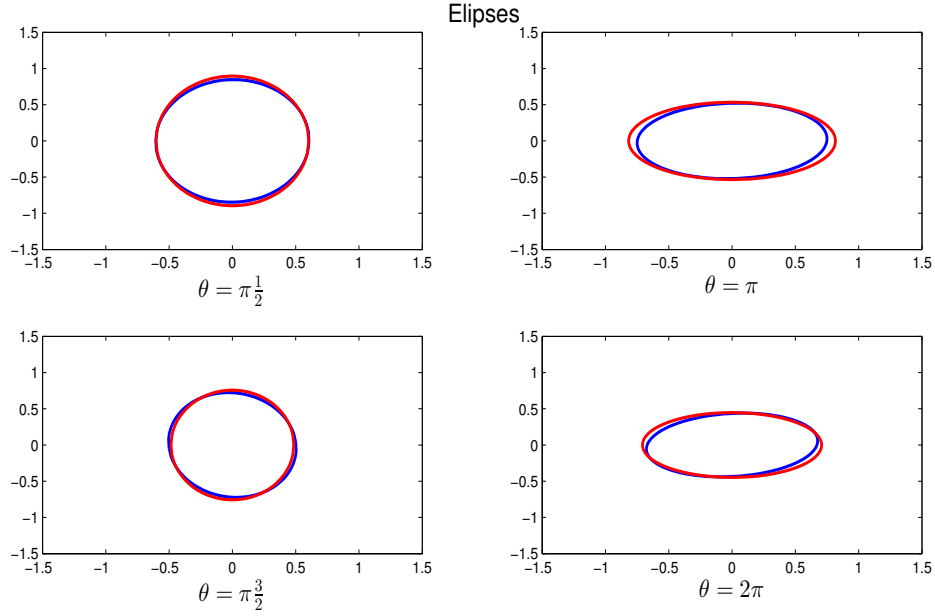


Figura 3.11

En ellas puede observarse que se obtienen buenos ajustes y estimaciones, tanto de las tendencias de los autovalores y autovectores como de las elipses correspondientes.

3.5. Modelo 3

Este modelo trata de preservar las características de dependencia del modelo anterior, pero en una dimensión p mayor. Para esto consideremos

$$\Sigma(u) = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{\mathbf{u}} \end{pmatrix}. \quad (3.10)$$

En esta expresión, \mathbf{I} es la matriz identidad de tamaño $(p-2) \times (p-2)$ y $\mathbf{M}_{\mathbf{u}}$ es una matriz de tamaño 2×2 que tiene la forma (3.8) dada en el modelo 2. En este caso se eligió un tamaño de muestra $n=100$, grados de libertad $m=30$, dimensión $p=19$ y valores de los parámetros $\lambda_1=1$, $\lambda_2=3$, $k_1=1$ y $k_2=2$ en la expresión (3.9).

Se puede comprobar que los dos mayores autovalores de $\Sigma(u)$ son los autovalores $\lambda_1(u)$ y $\lambda_2(u)$ de $\mathbf{M}_{\mathbf{u}}$ dados en (3.9), y sus correspondientes autovectores son $(0, \dots, 0, \gamma_i(u)^T)^T$ para $i=1,2$, donde los $\gamma_i(u)$ son los autovectores de $\mathbf{M}_{\mathbf{u}}$.

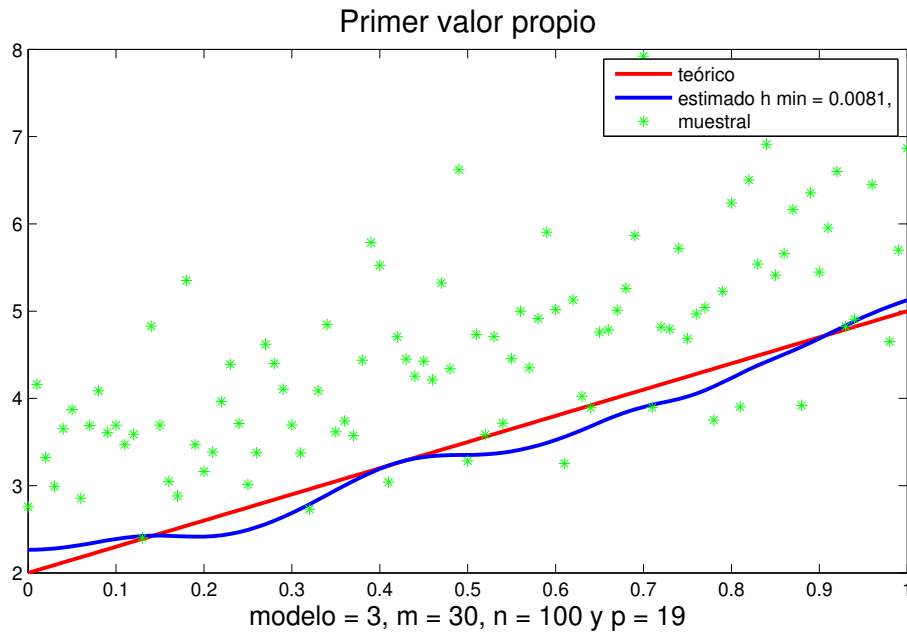


Figura 3.12

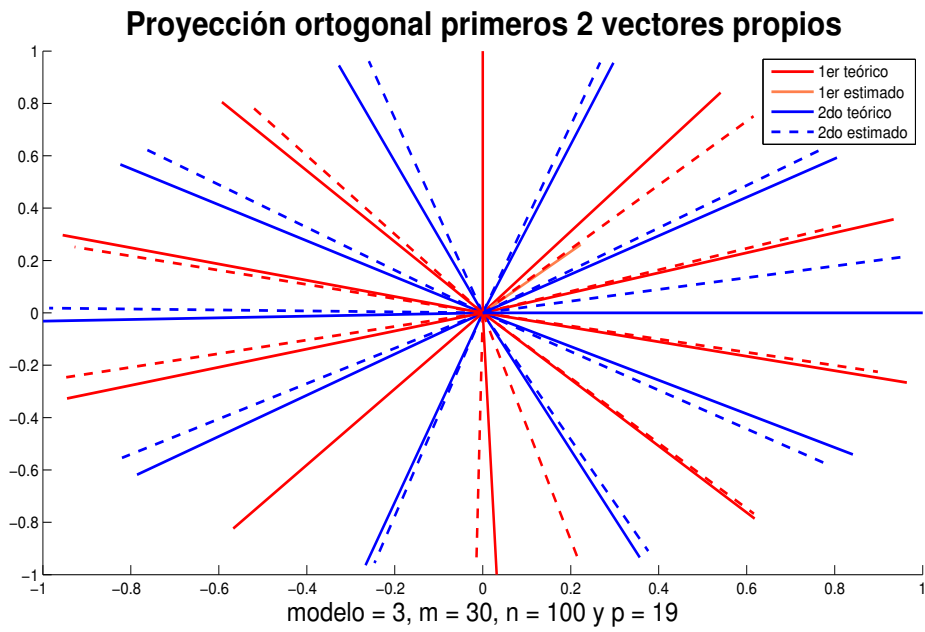


Figura 3.13

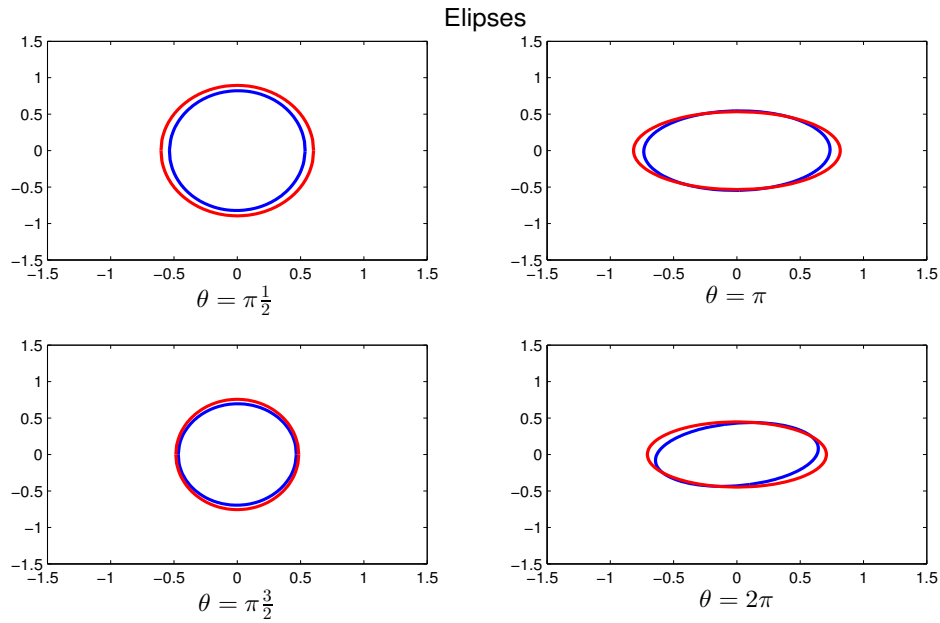


Figura 3.14

En la Figura 3.12 se observa una buena estimación empírica del primer valor propio, con un ligero efecto borde. Esto indica que el aumento de dimensión p no afectó significativamente la estimación del mayor valor propio.

Las Gráficas 3.14 muestra proyecciones sobre las dos primeras coordenadas de los vectores propios de la matriz estimada $\hat{\Sigma}(u)$ y la matriz teórica $\Sigma(u)$ para un conjunto de variables de u .

El gráfico 3.13 muestra, para varios valores del ángulo $\theta = \theta(u) = 2\pi u$, las proyecciones en las dos primeras coordenadas de las elipses de radio 1 correspondientes a la matriz teórica $\Sigma(u)$ (en rojo) y a su estimación $\hat{\Sigma}(u)$ (en azul). Como se observa en este gráfico las elipses estimadas en general aproximan aceptablemente a las correspondientes poblacionales.

En conclusión estas gráficas muestran (en caso de dimensión moderadamente alta y dependencia con respecto a la covariable u tanto de los valores propios como de los vectores propios) una aceptable precisión tanto de los dos primeros valores propios, y sus correspondientes vectores propios. El comportamiento del estimador parece indicar que el moderado aumento de la dimensión y de la variabilidad de las observaciones y valores propios no afecta en gran medida la calidad de la estimación.

3.6. Modelo 4

En este modelo presenta otro escenario de simulación de moderadamente alta dimensión ($p = 20$) y de dependencia con respecto a la covariable u tanto de los valores propios como de los vectores propios de la función de regresión $\Sigma(u)$. El modelo es el siguiente:

$$\Sigma(u) = \exp\{b_0 + b_1 u\} g g^T + c^2 \mathbf{M}. \quad (3.11)$$

En la anterior expresión,

$$\mathbf{M} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_u \end{pmatrix},$$

$$\mathbf{M}_u = \begin{pmatrix} \lambda_1 u + k_1 & 0 \\ 0 & \lambda_2 u + k_2 \end{pmatrix}.$$

Aquí $g = q/\|q\|$, y la entrada j -ésima del vector es $q(j) = k(1-u)j^2 + 20$ para $j = 1, \dots, p/2$, y $q(j) = k(1-u)(j-p/2)^2 + 20$ para $j = p/2, \dots, p$. Es decir, el primer vector propio g tiene forma de parábola, y controlando la covariable u se controla el ancho de esta. Además, $\lambda_1 = 1$, $\lambda_2 = 3$, $k_1 = 1$, $k_2 = 2$, $c = 0,5$, $b_0 = 1$ y $b_1 = 1$.

Mostraremos resultados de simulación con tamaño de muestra $n = 100$ y $m = 30$.

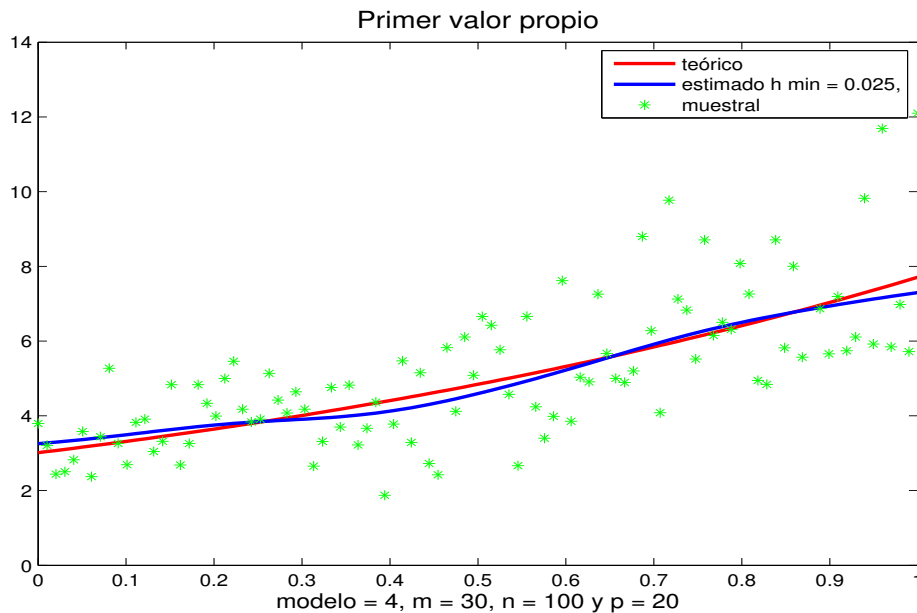


Figura 3.15

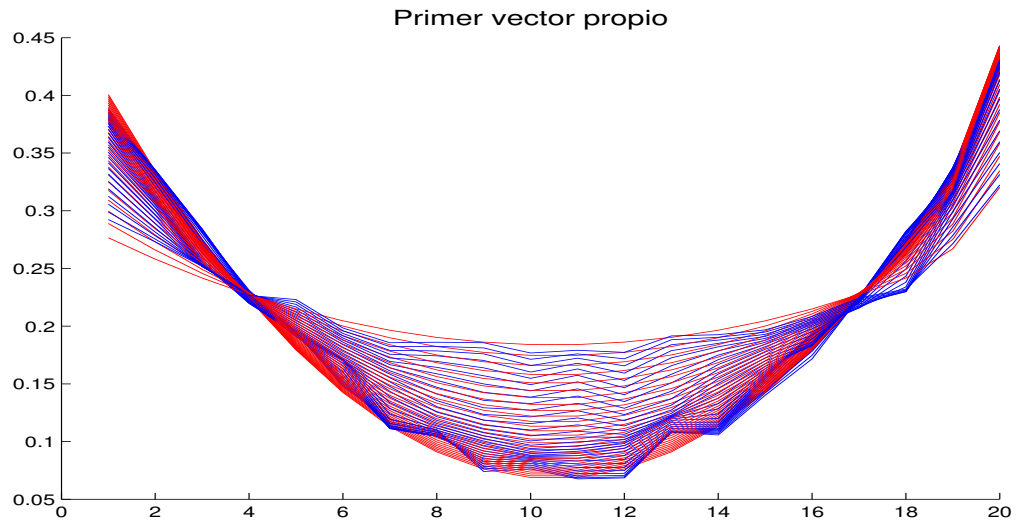


Figura 3.16

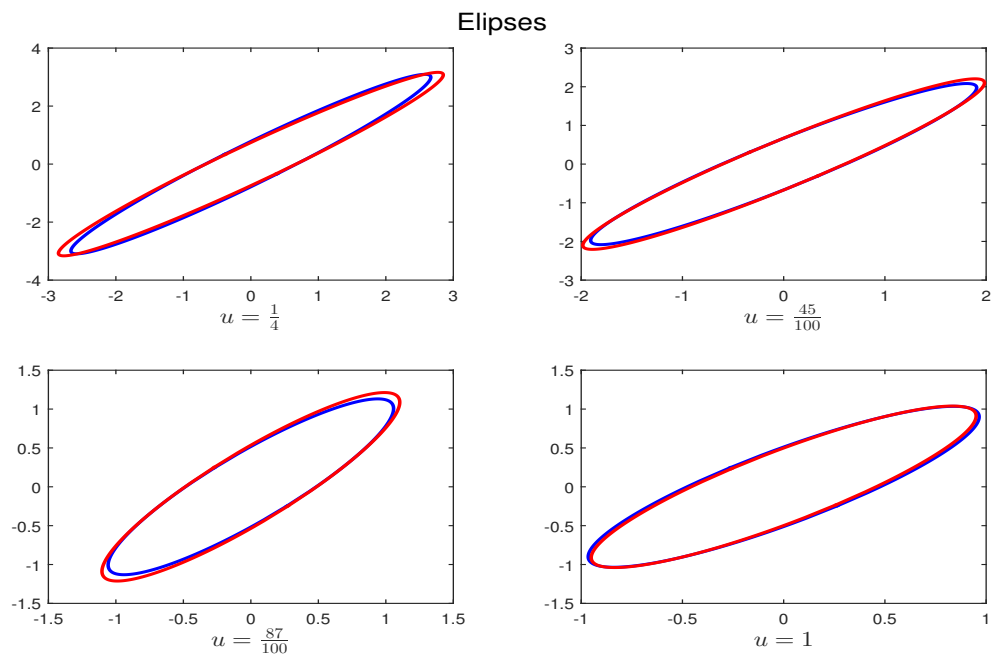


Figura 3.17

En la Figura 3.15 se observa una buena estimación del mayor valor propio, a pesar de que la dimensión es moderadamente alta.

La Figura 3.16 muestra el primer autovector de la matriz teórica $\Sigma(u)$ (en rojo) y de la matriz estimada $\hat{\Sigma}(u)$ (en azul) para diferentes valores de u . Observamos que la estimación presenta un buen comportamiento.

La Gráfica 3.17 muestra, para varios valores del ángulo $\theta = \theta(u) = 2\pi u$, las proyecciones en las dos primeras coordenadas de las elipses de radio 1 correspondientes a la matriz teórica $\Sigma(u)$ (en rojo) y a su estimación $\hat{\Sigma}(u)$ (en azul). Como se observa en este gráfico las elipses estimadas en general aproximan aceptablemente a las correspondientes poblacionales.

Aplicación a la caracterización de la maduración del EEG multi-canal

En este capítulo se aplica la metodología propuesta en el Capítulo 2 para caracterizar la maduración de la conectividad del electroencefalograma (EEG) multi-canal a diferentes frecuencias durante el primer año de vida.

4.1. Objetivo del estudio

El EEG es una técnica no invasiva de registro, a través del tiempo, del potencial eléctrico (voltaje) en un número p de electrodos colocados sobre la superficie de la cabeza. Matemáticamente, el registro de EEG en un sujeto puede representarse como una serie de tiempo p -variada estacionaria $X_t \in \mathbb{R}^p$, $t \in \mathbb{Z}$.

La matriz de densidad espectral $\Sigma(u; \omega)$ del registro de EEG X_t , a cada frecuencia ω , depende de la edad u del sujeto en el primer año de vida. Nuestro propósito es estimar la función de regresión $\Sigma(u; \omega)$ como función de la covariable edad u . Esto constituye un problema de regresión donde la función de regresión $\Sigma(u; \omega)$ toma valores en la variedad \mathcal{M}^C de las matrices d.n.n. complejas. Para esto utilizaremos una muestra de datos constituida por las matrices de densidad espectral muestrales $S(u_i; \omega)$ de n sujetos normales con edades (medida en años) $u_i \in [0, 1]$, ($i = 1, \dots, n$).

En la sección siguiente se describirá en detalle todos los aspectos involucrados en la obtención de los datos y en los procedimientos de estimación. La muestra de datos de $n = 123$ niños que se utilizó para el estudio que se presenta en este capítulo fue proporcionada por la Unidad de Investigación en Neurodesarrollo del Instituto de Neurobiología (IBN) de la Universidad Nacional Autónoma de México.

4.2. Datos de EEG

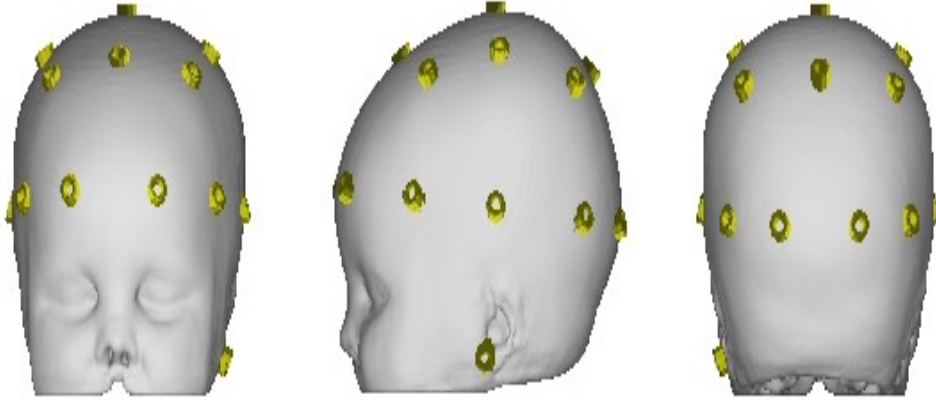


Figura 4.1: Vista 3D de la posiciones de los electrodos de registro del EEG.

Electrodos de registro

El registro de EEG de cada sujeto se realizó colocando $p = 19$ electrodos (canales) sobre el cuero cabelludo con acuerdo en el sistema internacional 10-20, como se muestra en las Figuras 4.1 y 4.2.

Montaje

En cada instante de tiempo t , cada electrodo de registro mide la diferencia de voltaje $V(t)$ entre dos electrodos, uno que se denomina "activo" y otro que se denomina "inactivo" ; a este último se le llama electrodo de referencia. La combinación de diferentes pares de electrodos activos e inactivos se denomina montaje. En este estudio se utilizó el montaje de referencia promedio ([Webster, 2010] y [Urrestarazu and Iriarte, 2005]).

Frecuencia de muestreo y segmentos de EEG

Se registran los datos con un período de muestreo de $\Delta t = 5$ milisegundos, lo que corresponde a una frecuencia de muestreo de $1/(\Delta t) = 200$ Hertz (i.e., en un segundo se tienen 200 instantes de registro). Para su análisis, el EEG se divide en un número de $N_s = 20$ segmentos, cada uno de tamaño $T = 2.56$ segundos. Por tanto, cada segmento contiene $N_T = 512$ instantes de registro. En el dominio de la frecuencia esto implica una frecuencia máxima de $\omega_{\text{máx}} = 100$ Hertz y una resolución en frecuencia de $\Delta\omega = 1/T = 0.39$ Hertz.

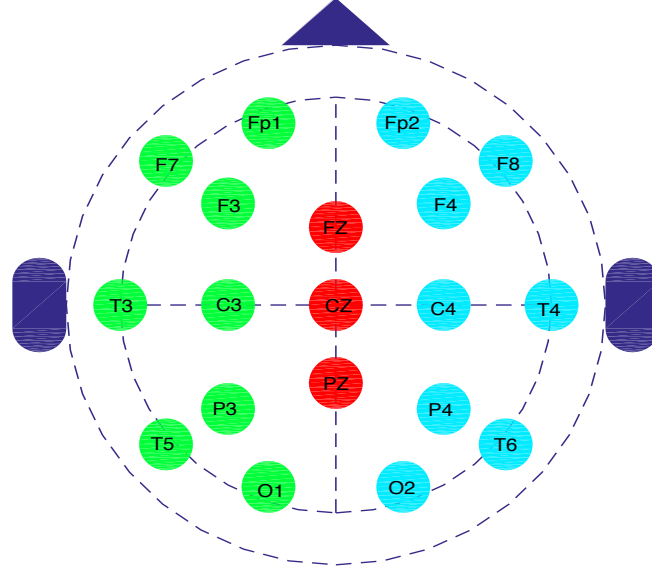


Figura 4.2: Sistema internacional 10-20 de colocación de electros de EEG. El color verde denota al hemisferio izquierdo, el color azul denota al hemisferio derecho y el rojo denota la línea media. La nomenclatura de letras significa: F: lóbulo frontal, T: lóbulo temporal, C: lóbulo central, P: lóbulo parietal, O: lóbulo occipital y Z: línea media.

Matrices de densidad espectral muestrales

Para cada sujeto con edad u_i , su correspondiente matriz de densidad espectral muestral $S(u_i; \omega_j)$ se calcula de la siguiente manera. Para cada segmento s ($s = 1, \dots, N_s$) de la serie de tiempo de EEG X_t^s ($t = 1 \dots, N_T$) se calcula su transformada de Fourier finita $X^s(\omega_j)$ en cada frecuencia discreta $\omega_j = j\Delta\omega$, $j = 1, \dots, N_\omega$. (Para detalles sobre conceptos de series de tiempo en el dominio de la frecuencia, ver Apéndice B.) Restringimos el análisis a un número de frecuencias discretas $N_\omega = 48$, lo que corresponde a trabajar con una frecuencia máxima de 50 Hertz. Finalmente, la matriz de densidad espectral muestral se calcula mediante

$$S(u_i; \omega_j) = \frac{1}{N_s} \sum_{s=1}^{N_s} \frac{X_t^s(\omega_j)(X_t^s(\omega_j))^*}{2\pi N_T}, \quad (4.1)$$

donde * indica transpuesta conjugada de matrices. Esta expresión (4.1) es el estimador de

Welch (ver B.18 en Apéndice B para más detalles).

En el cuadro 4.1 se resumen las características que acabamos de describir.

$n = 123$	tamaño de la muestra
$p = 19$	dimensión de las matrices de densidad espectral
$N_s = 20$	número de segmentos
$T = 2.56$ segundos	tamaño del segmento
$\Delta t = 5$ milisegundos	período de muestreo
$T = 512$	instantes de registro
$1/(\Delta t) = 200$ Hertz	frecuencia de muestreo
$\omega_{\text{máx}} = 100$ Hertz	frecuencia máxima
$\Delta\omega = 0.39$ Hertz	resolución en frecuencia
$N_\omega = 48$	número de frecuencias consideradas en el estudio

Cuadro 4.1

4.3. Descriptores espectrales del EEG

En esta sección definiremos rasgos espectrales que se suelen utilizar en el análisis de EEG (para más detalles ver [Sameshima and (eds), 2014], [Sanei and Chambers, 2007], [Webster, 2010]).

Bandas de frecuencia

Para el análisis cuantitativo, el EEG se divide de manera estándar en cuatro bandas de frecuencias, llamadas delta (δ), theta (θ), alpha (α), beta (β) y gamma (γ).

- Banda delta [1.5, 3.5) Hertz.
- Banda theta [3.5, 7.5) Hertz.
- Banda alpha [7.5, 12.5) Hertz.
- Banda beta [12.5, 20) Hertz.
- Banda gamma [20, 50) Hertz.

Potencias relativas (RP)

Sea dada una matriz de densidad espectral del EEG $\Sigma(u; \omega_r) = (\sigma_{jk}(u; \omega_r))$ correspondiente a la frecuencia discreta ω_r y a un sujeto de edad u . Se definen los siguientes descriptores de potencia (energía) por bandas.

Para cada frecuencia discreta ω_r y para cada electrodo j ($j = 1, \dots, p$) se define la Potencia Relativa (RP) en el electrodo j y la frecuencia ω_r como

$$\sigma_{jj}^{\omega_r}(u) = \frac{\sigma_{jj}(u; \omega_r)}{\sum_{\omega_r=1}^{N_\omega} \sigma_{jj}(u; \omega_r)}.$$

Para cada banda b ($b = \delta, \theta, \alpha, \beta, \gamma$) y para electrodo j ($j = 1, \dots, p$) se define la Potencia Relativa (RP) en el electrodo j y la banda b como

$$\sigma_{jj}^b(u) = \frac{\sum_{\omega_r \in b} \sigma_{jj}(u; \omega_r)}{\sum_{\omega_r=1}^{N_\omega} \sigma_{jj}(u; \omega_r)}.$$

Coherencias

Para cada banda b ($b = \delta, \theta, \alpha, \beta, \gamma$) se define el módulo de la coherencia (o por simplicidad, coherencia) entre los electrodos j y k ($j, k = 1 \dots, p$) en la banda b como

$$\rho_{jk}^b(u) = \frac{\sum_{\omega_r \in b} \rho_{jk}(u; \omega_r)}{N_b},$$

donde N_b es el número de frecuencias discretas en la banda b , y $\rho_{jk}(u; \omega_r)$ es el módulo de la coherencia calculado a partir de la matriz espectral $\Sigma(u; \omega_r)$ (ver definición en Apéndice B).

Coherencias parciales

Para cada banda b ($b = \delta, \theta, \alpha, \beta, \gamma$) se define el módulo de la coherencia parcial (o por simplicidad, coherencia parcial) entre los electrodos j y k ($j, k = 1 \dots, p$) en la banda b como

$$\psi_{jk}^b(u) = \frac{\sum_{\omega_r \in b} \psi_{jk}(u; \omega_r)}{N_b},$$

donde $\psi_{jk}(u; \omega_r)$ es el módulo de la coherencia parcial calculado a partir de la matriz espectral $\Sigma(u; \omega_r)$ (ver definición en Apéndice B).

4.4. Procedimientos de estimación de curvas de regresión de matrices de conectividad del EEG

Aplicando el método de estimación elaborado en el Capítulo 2. Para cada una de las 48 frecuencias discretas ω_r ($r = 1, \dots, 48$) en las bandas δ, θ, α y β se calcularon las estimaciones de las siguientes funciones de regresión:

- 1) Regresión de la matriz de densidad espectral, $\Sigma(u; \omega_r)$, versus la covariable edad u :

$$\hat{\Sigma}(u; \omega_r) = \frac{\sum_{i=1}^n S(u_i; \omega_r) K\left(\frac{u_i - u}{h}\right)}{\sum_{i=1}^n K\left(\frac{u_i - u}{h}\right)}. \quad (4.2)$$

- 2) Regresión de la inversa de la matriz de densidad espectral, $B(u; \omega_r) = \Sigma^{-1}(u; \omega_r)$, versus la covariable edad u :

$$\hat{B}(u; \omega_r) = \frac{\sum_{i=1}^n \tilde{B}(u_i; \omega_r) K\left(\frac{u_i - u}{\hat{h}}\right)}{\sum_{i=1}^n K\left(\frac{u_i - u}{\hat{h}}\right)}, \quad (4.3)$$

donde

$$\tilde{B}(u_i; \omega_r) = S^{-1}(u_i; \omega_r). \quad (4.4)$$

- 3) Regresión de la matriz de coherencia $R(u; \omega_r) = D(u; \omega_r)^{-1/2} \Sigma(u; \omega_r) D(u; \omega_r)^{-1/2}$ versus la covariable edad u :

$$\hat{R}(u; \omega_r) = \frac{\sum_{i=1}^n \tilde{D}(u_i; \omega_r)^{-1/2} S(u_i; \omega_r) \tilde{D}(u_i; \omega_r)^{-1/2} K\left(\frac{u_i - u}{\hat{h}}\right)}{\sum_{i=1}^n K\left(\frac{u_i - u}{\hat{h}}\right)}, \quad (4.5)$$

donde $D(u; \omega_r) = \text{Diag}(\Sigma(u; \omega_r))$, y $\tilde{D}(u_i; \omega_r) = \text{Diag}(S(u_i; \omega_r))$.

- 4) Regresión de la matriz de coherencia parcial $\Psi(u; \omega_r) = D_B(u; \omega_r)^{-1/2} B(u; \omega_r) D_B(u; \omega_r)^{-1/2}$ versus la covariable edad u :

$$\hat{\Psi}(u; \omega_r) = \frac{\sum_{i=1}^n D_{\tilde{B}}(u_i; \omega_r)^{-1/2} \tilde{B}(u_i; \omega_r) D_{\tilde{B}}(u_i; \omega_r)^{-1/2} K\left(\frac{u_i - u}{\hat{h}}\right)}{\sum_{i=1}^n K\left(\frac{u_i - u}{\hat{h}}\right)}, \quad (4.6)$$

donde $D_B(u; \omega_r) = \text{Diag}(B(u; \omega_r))$, $D_{\tilde{B}}(u_i; \omega_r) = \text{Diag}(\tilde{B}(u_i; \omega_r))$, y $\tilde{B}(u_i; \omega_r)$ está definida por (4.4).

En todos los casos, la elección \hat{h} del ancho de banda h se efectúa por validación cruzada, análogamente a como se describió en la ecuación (3.4).

A continuación se describen los procedimientos de estimación de las curvas vs. edad de los diversos descriptores espectrales.

I) Estimación de las curvas de regresión de las potencias relativas (RP)

Se calcula $\hat{\Sigma}(u; \omega_r)$ como en (4.2). A partir de la estimación $\hat{\Sigma}(u; \omega)$ de $\Sigma(u; \omega)$ se calculan: estimaciones $\hat{\sigma}_{jj}^{\omega_r}(u)$ de las potencias relativas a alta resolución $\sigma_{jj}(u; \omega_r)$; y estimaciones $\hat{\sigma}_{jj}^b(u)$ de las potencias relativas por banda $\sigma_{jj}^b(u)$. Aquí, $j, k = 1, \dots, p$, denotan electrodos; $b = \delta, \theta, \alpha, \beta, \gamma$, denota bandas; y u denota edad.

II) Estimación de las curvas de regresión de las coherencias (C)

Comparamos dos procedimientos distintos para las estimaciones $\hat{\rho}_{jk}^b(u)$ de las coherencias por banda $\rho_{jk}^b(u)$, que son los siguientes:

Coh1) Se calcula la estimación de $\hat{\Sigma}(u; \omega_r)$ como en (4.2), y se toma como $\hat{\rho}_{jk}^b(u)$ las coherencias correspondientes a ella.

Coh2) Se calcula $\hat{R}(u; \omega_r)$ como en (4.5). A partir de esta matriz se calculan $\hat{\rho}_{jk}^b(u; \omega_r) = |\hat{R}_{jk}(u; \omega_r)|$. Y las estimaciones de las coherencias por banda $\hat{\rho}_{jk}^b(u)$ se toman como las correspondientes a las estimaciones de las coherencias de alta resolución $\hat{\rho}_{jk}^b(u; \omega_r)$.

III) Estimación de las curvas de regresión de las coherencias parciales (CP)

- CohP1) Se toman como estimaciones $\hat{\psi}_{jk}^b(u)$ de las coherencias parciales $\psi_{jk}^b(u)$ las coherencias parciales correspondientes a $\hat{\Sigma}(u; \omega_r)$, donde esta última se calcula como en (4.2).
- CohP2) Se calcula $\hat{B}(u; \omega_r)$ como en (4.3). Se toman como estimaciones $\hat{\psi}_{jk}^b(u)$ de las coherencias parciales $\psi_{jk}^b(u)$ las coherencias parciales correspondientes a $\hat{B}(u; \omega_r)$.
- CohP3) Se calcula $\hat{\Psi}(u; \omega_r)$ como en (4.6), y se toman $\hat{\psi}_{jk}(u; \omega_r) = |\hat{\Psi}_{jk}(u; \omega_r)|$ como estimaciones de las coherencias parciales a cada frecuencia ω_r . Las estimaciones de las coherencias parciales por banda $\hat{\psi}_{jk}^b(u)$ se toman como las correspondientes a las coherencias parciales de alta resolución $\hat{\psi}_{jk}(u; \omega_r)$.

4.5. Resultados del análisis

Para cada una de las 48 frecuencias discretas ω_r ($r = 1, \dots, 48$) se ajustó cada una de las regresiones (1)-(4) como se describió en la Sección 4.4. A partir de ellas se calcularon las estimaciones $\hat{\sigma}_{jj}^{\omega_r}(u)$, $\hat{\sigma}_{jj}^b(u)$, $\hat{\rho}_{jk}^b(u)$ y $\hat{\psi}_{jk}^b(u)$ de las potencias relativas a alta resolución, las potencias relativas por banda, las coherencias por banda y las coherencias parciales por banda, respectivamente.

En esta sección se muestran algunos aspectos de los resultados de este análisis estadístico.

En todos los gráficos donde aparecen puntos, estos corresponden a valores empíricos de un descriptor espectral de las matrices de densidad espectral muestral $S(u_i; \omega_r)$ — calculados como se describió en la Sección 4.3, pero sustituyendo en las fórmulas la matriz teórica $\Sigma(u; \omega_r)$ por una muestral $S(u_i; \omega_r)$.

4.5.1. Potencias relativas

Las potencias relativas mostraron el siguiente patrón al aumentar la edad: en la banda δ decrecen, y en las bandas θ , α y β aumentan. Este patrón es consistente con lo que ha sido reportado por diferentes autores (ver, e.g., [Otero et al., 2011]).

Esto se muestra en la Figura 4.3 para el electrodo T5. Allí se grafican las curvas de potencia relativa (RP) a cada frecuencia como función de la edad. El eje vertical representa RP (en escala logit); los colores azul, cyan, verde y magenta corresponden a frecuencias discretas de las bandas δ , θ , α y β , respectivamente. Es evidente que RP disminuye en la banda δ (curvas en azul), y aumenta en las restantes bandas θ , α y β .

T5

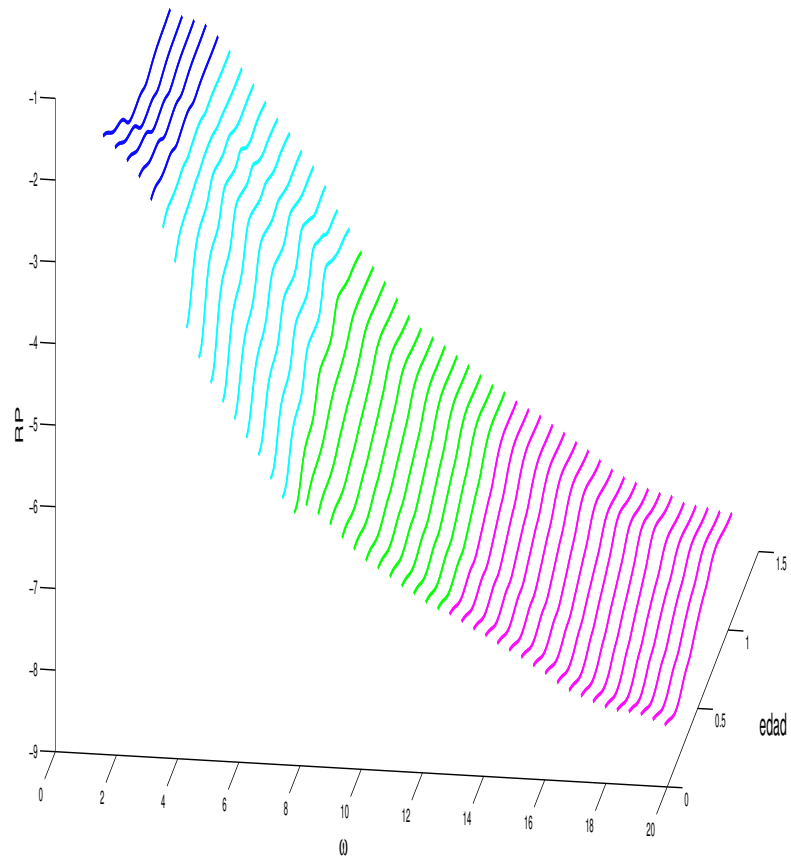
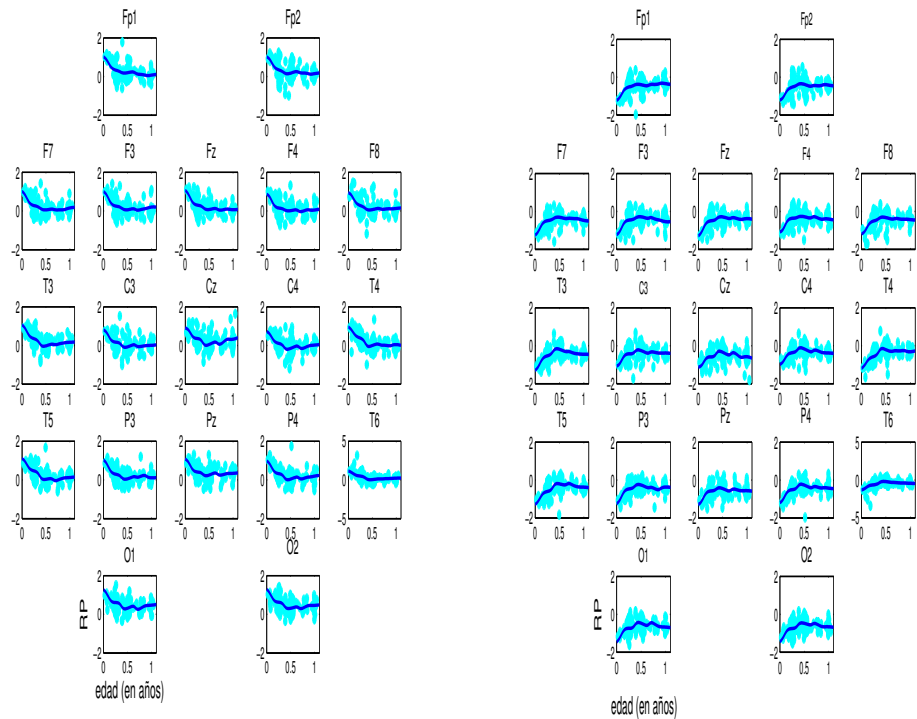


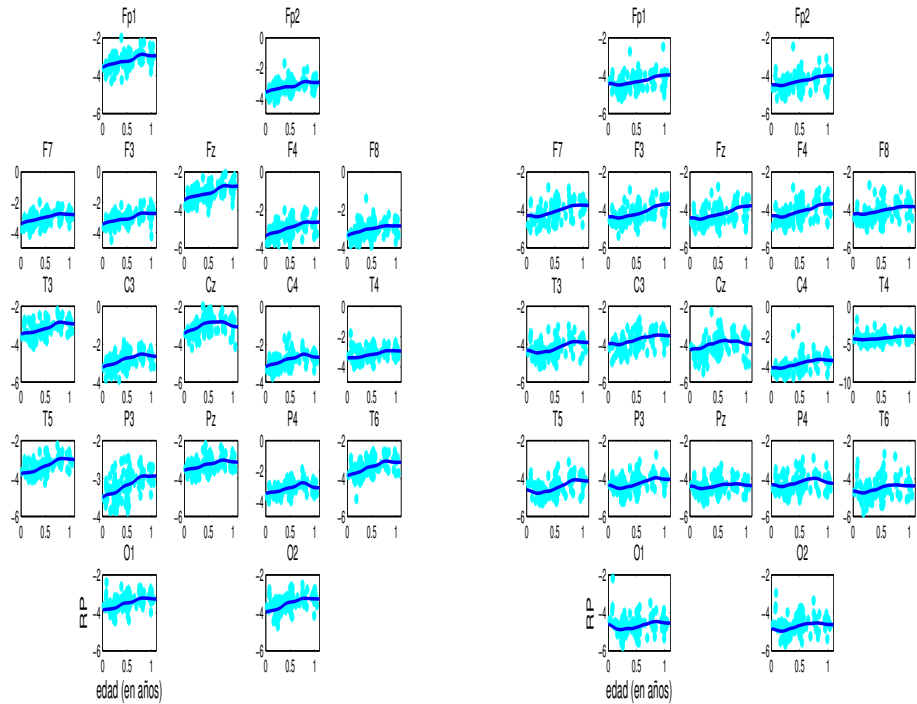
Figura 4.3: Potencia relativa (RP) en escala logit (eje vertical) en electrodo T5 al variar la edad para todas las frecuencias. Los colores azul, cian, verde y magenta corresponden a las frecuencias de las bandas δ , θ , α y β , respectivamente.

En particular, el mismo patrón de variación de las RP por bandas al aumentar la edad se presenta en todos los electrodos, como se puede observar en la Figura 4.4.



(a) Banda δ

(b) Banda θ

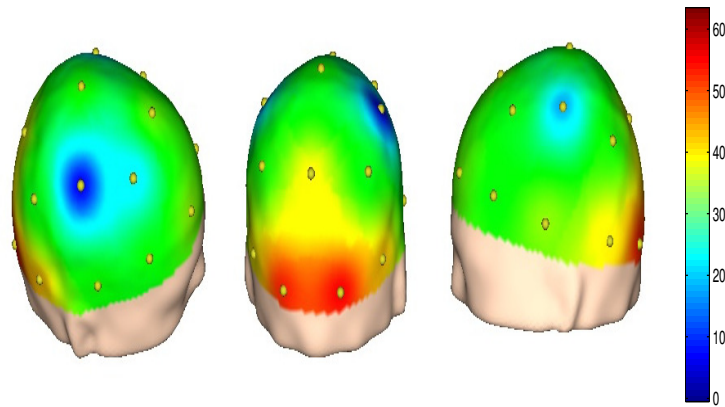


(c) Banda α

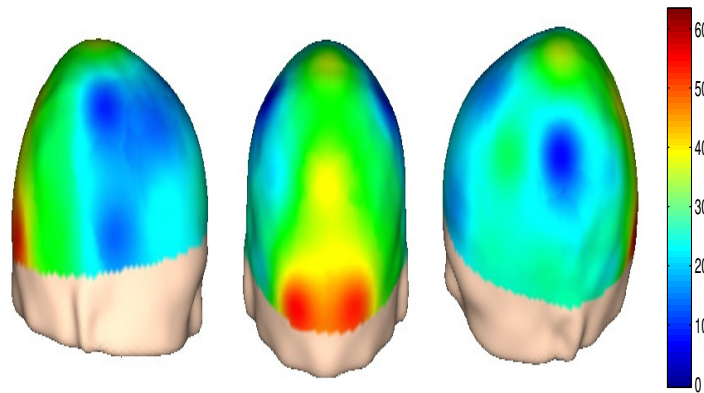
(d) Banda β

Figura 4.4: Curvas de RP por banda para cada electrodo. Cada regresión corresponde a un electrodo.

En la Figura 4.5 se muestra una perspectiva espacialmente 3D de tal patrón de decrecimiento de la RP en la banda δ , comparando dos edades.



(a) RP en banda delta. Edad 1.08 meses.



(b) RP en banda delta. Edad 12 meses.

Figura 4.5

Es de destacar que algunas curvas de regresión presentan comportamiento altamente no lineal. Esto se ilustra en el electrodo Pz para el RP en una frecuencia de la banda θ

en la Figura 4.6. Nuestro método de estimación no paramétrico permite captar estos comportamientos, mientras que ajustes tradicionales lineales y cuadráticos no serían capaces de captarlo adecuadamente.

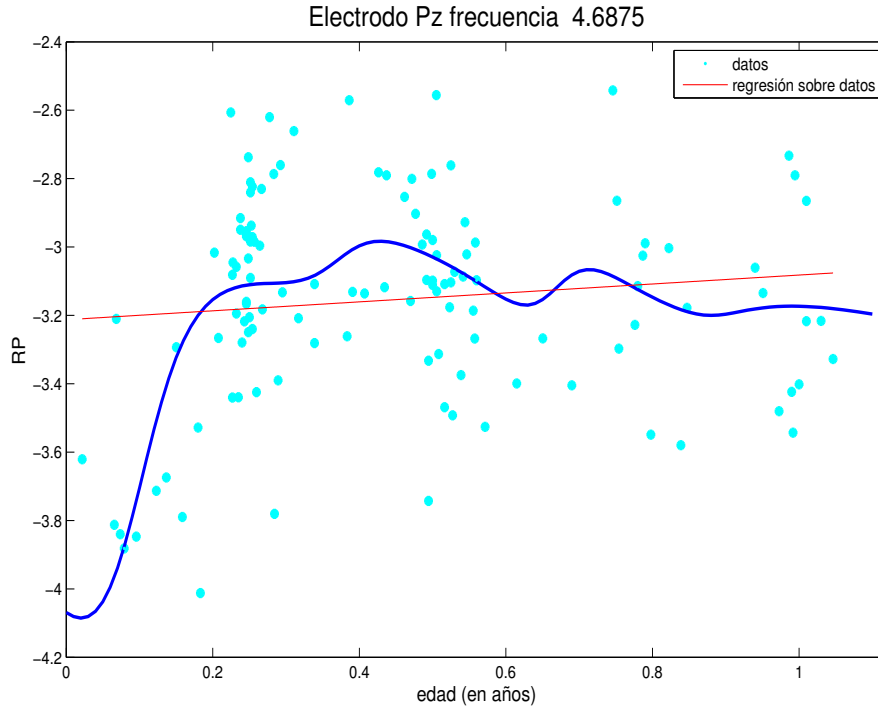


Figura 4.6

4.5.2. Coherencias

La Figura 4.7 muestra, en la banda δ y tomando como referencia el electrodo $Fp2$, los resultados de estimaciones de las coherencias (curvas) y los correspondientes valores empíricos (puntos). La curva en color magenta corresponde al procedimiento Coh1 y la curva azul al procedimiento Coh2, ambos descritos en la Sección 4.4. Se puede observar que ambos procedimientos muestran ajustes similares, siendo Coh2 ligeramente mejor. El mismo comportamiento se observó en general en todos los ajustes de coherencias.

Se puede observar en la Figura 4.7 que hay a lo sumo ligeras variaciones del patrón de coherencia en la banda δ al aumentar la edad. Lo mismo ocurre en otras bandas. En la Figura 4.8 esto se ilustra para la banda α mostrando la imagen de los pares de coherencias entre todos los electrodos en dos diferentes edades.

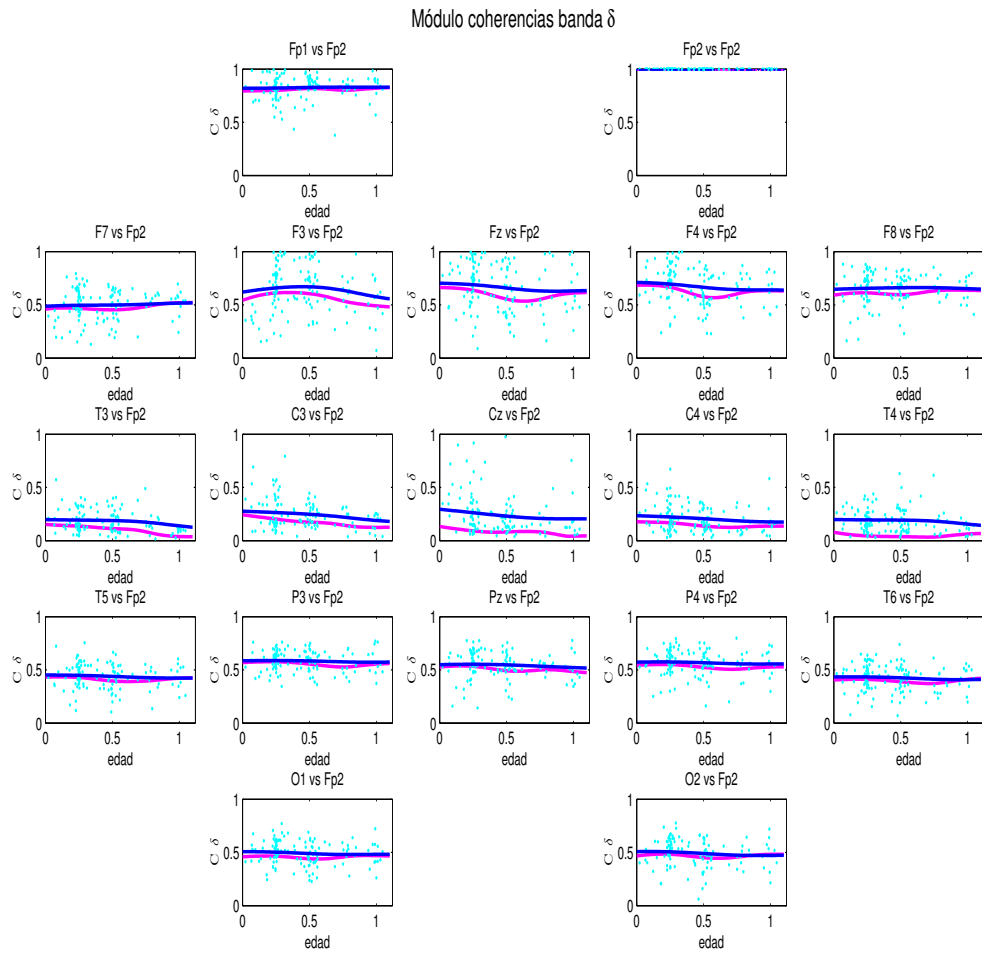


Figura 4.7: Para la banda δ , curvas de coherencia entre el electrodo $Fp2$ y los restantes. Azul corresponde al método de ajuste Coh2 y magenta al método de ajuste Coh1.

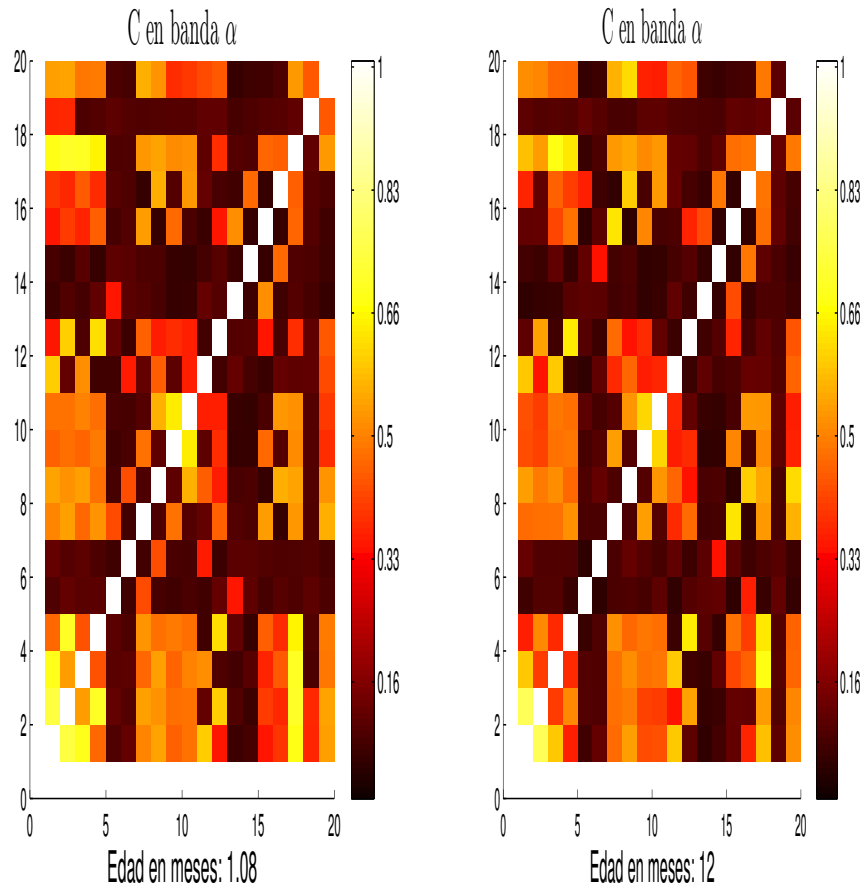


Figura 4.8: Ejes horizontal y vertical corresponden a los 19 electrodos. El color en cada punto de la imagen codifica la magnitud de la coherencia entre los electrodos correspondientes en la banda α .

4.5.3. Coherencias parciales

La Figura 4.9 muestra, en la banda δ y tomando como referencia el electrodo $F8$, los resultados de estimaciones de las coherencias parciales (curvas) y los correspondientes valores empíricos (puntos). La curva en color magenta corresponde al procedimiento CohP1, la curva en verde corresponde al procedimiento CohP2 y la curva en azul al CohP3, todos descritos en la Sección 4.4. Se puede observar que el procedimiento CohP3 claramente muestra mejor ajuste a los puntos empíricos. Este mismo comportamiento se observó en general en todos los ajustes de coherencias parciales. Sobre la base de estos resultados se justifica recomendar que el procedimiento CohP3 sea el más adecuado para la estimación de coherencias.

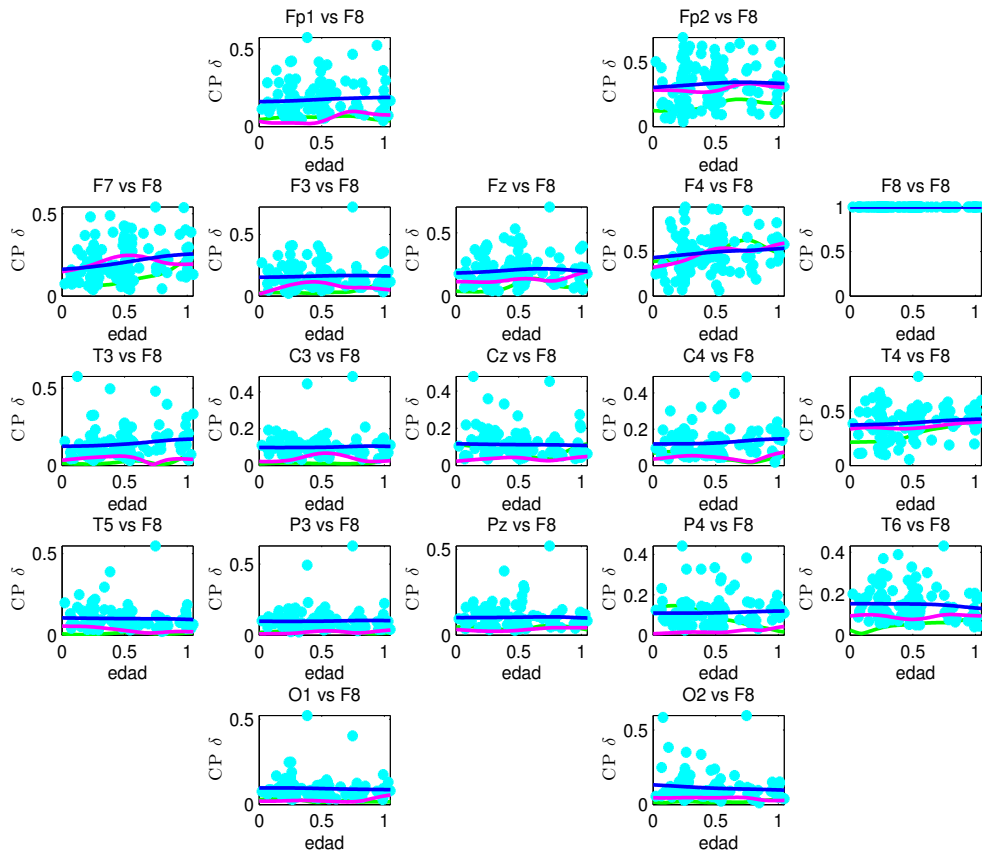
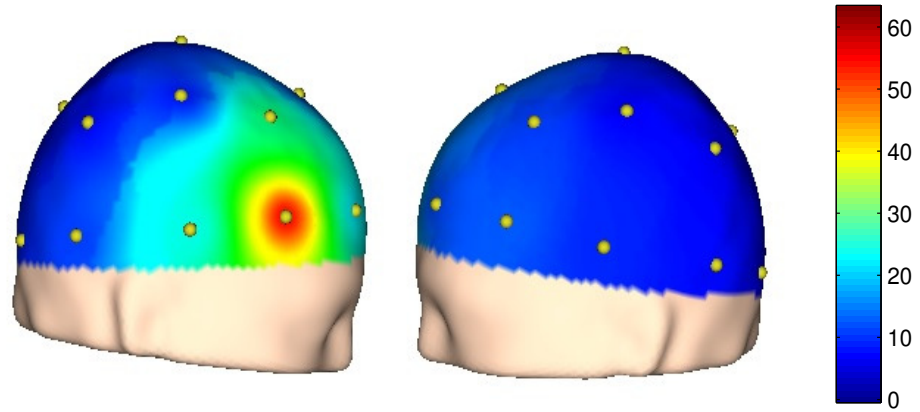


Figura 4.9: En la banda δ y tomando como referencia el electrodo $F8$, se muestran los resultados de estimaciones de las coherencias parciales (curvas) y los correspondientes valores empíricos (puntos). La curva en color magenta corresponde al procedimiento CohP1, la curva en verde corresponde al procedimiento CohP2 y la curva en azul al CohP3.

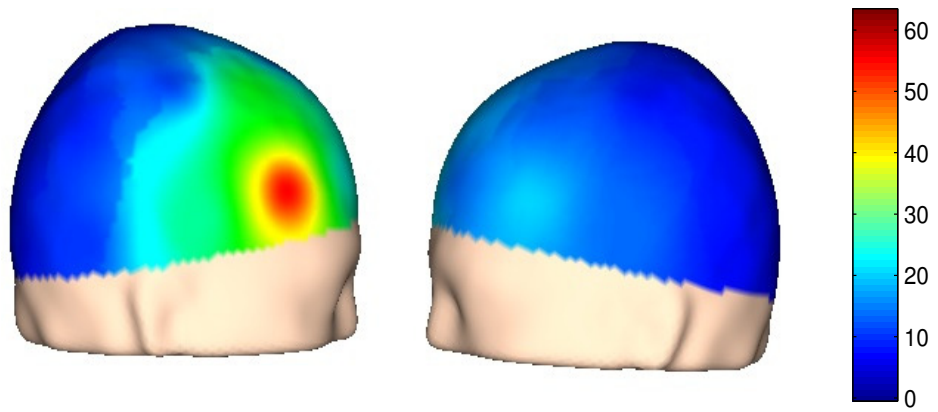
A diferencia de las curvas de regresión de las coherencias no parciales, las coherencias parciales sí muestran apreciables cambios con la edad del sujeto durante el primer año de vida, como se puede observar en la Figura 4.9. En general, el patrón que parece emerger del análisis exploratorio de las curvas ajustadas es el siguiente: muchas coherencias parciales de rango espacial corto tienden a disminuir con la edad, mientras que algunas de rango mayor tienden a aumentar. Esto es coherente con hallazgos e hipótesis sobre la maduración de la conectividad que han publicado diversos autores.

Otra ilustración de este hecho se muestra en la Figura 4.10. Allí se muestran las coherencias parciales entre el electrodo $F8$ (frontal derecho) contra todos los restantes elec-

trodos, en la banda δ para dos edades distintas. Nótese que hay un cambio con la edad de la coloración (de azul a blanco) en la zona frontal del hemisferio izquierdo. Esto representa un incremento en la coherencia parcial entre zonas frontales de largo rango, específicamente entre las zonas frontales de ambos hemisferios.



(a) Edad 1.08 meses



(b) Edad 12 meses

Figura 4.10: Se muestran las coherencias parciales entre el electrodo *F8* (frontal derecho) contra todos los restantes electrodos, en la banda δ para dos edades distintas.

Durante el trabajo de investigación se abordó y se dio solución a la problemática de la estimación no paramétrica de funciones de regresión con valores en una variedad convexa, haciendo énfasis en el caso de la variedad de matrices definidas no negativas (d.n.n.), debido a que la motivación práctica principal de este trabajo era describir la maduración de la conectividad del electroencefalograma (EEG) a través del análisis de regresión de sus matrices de densidad espectral como funciones de la edad del sujeto. En el primer año de vida este tema cobra relevancia debido a que la detección de anomalías de conectividad a temprana edad puede tener un impacto decisivo en la rehabilitación de niños con patologías asociadas a riesgos perinatales. ([Bosch-Bayard et al., 2012], [Otero et al., 2011]). De manera tal que para cada sujeto, i con edad u_i , el tipo de dato relevante para describir la maduración de la conectividad en EEG, es la matriz de covarianza $S(u_i) = (s_{jk}(u_i))$. Por lo cual el interés principal es estudiar la correspondiente matriz de covarianza poblacional $\Sigma(u) = (\sigma_{jk}(u))$ como función de regresión con respecto a la edad u del sujeto como covariable. En tal situación, el parámetro que se deseaba estimar era la correspondiente matriz de densidad espectral $\Sigma(u_i)$. Es conocido que una matriz de densidad espectral muestral tiene distribución asintótica Wishart compleja $W_p^C(l, \Sigma(u_i))$ (ver, e.g., [Brillinger, 2001]) y que por lo tanto es una matriz de covarianza. Por tanto, se cumple (asintóticamente) que $\Sigma(u_i) = E(S(u_i))$, donde el valor esperado representa la media extrínseca asociada a la métrica euclídeana del espacio ambiente, \mathbb{R}^{p^2} , de la variedad \mathcal{M} . De modo que, para nuestro propósito práctico, se impuso considerar la función de regresión $\Sigma(u)$ en el sentido de una media extrínseca según el encajamiento de inclusión o identidad dado por $id : A \in \mathcal{M} \subset \mathbb{R}^{p^2} \rightarrow A \in \mathbb{R}^{p^2}$.

Habiendo planteado el problema, se estudiaron las actuales metodologías utilizadas para describir la maduración de la conectividad, algunos ejemplos de ellas son las aplicadas en [Bosch-Bayard et al., 2012], [Otero et al., 2011], [Bosch-Bayard et al., 2001], y [Szava et al., 1994] las cuales no resuelven el problema de fondo, debido a que las estimaciones así obtenidas no son necesariamente matrices de covarianza (por lo cual no son

d.n.n.). Es así como nos vimos en la necesidad de considerar la estimación de la matriz de regresión $\Sigma(u)$ como un problema de estimación en la variedad de las matrices d.n.n. \mathcal{M} . Existe una creciente literatura estadística que aborda la definición e inferencia de medias y funciones de regresión con valores en esta variedad ([Bhattacharya and Bhattacharya, 2015], [Turaga and Srivastava, 2016], [Chikuse, 2003], [Bhattacharya and Patrangenaru, 2003], [Bhattacharya and Patrangenaru, 2005], [Abhishek Bhattacharya, 2008],[Yuan et al., 2012], [Zhu et al., 2009], [Pérez and González-Farias, 2013], [Shi et al., 2009]). Es importante aclarar que la pertinencia de usar una u otra media intrínseca, o una media extrínseca, depende del modelo generador de los datos. El dominio de aplicación motivó a considerar la función de regresión en el sentido de media extrínseca en una variedad según el encajamiento de inclusión en un espacio euclideo.

Fue así como para la inferencia de funciones de regresión propusimos un método no paramétrico para funciones de regresión con valores en una variedad convexa basado en núcleos positivos, pues la positividad del núcleo garantiza que nuestra propuesta tome valores solamente en la variedad de las matrices d.n.n. y el ser paramétrico le da características de flexibilidad lo cual nos permiten modelar patrones complejos y es computacionalmente eficiente. Se demostró que es consistente y asintóticamente normal, las cuales son características deseables para un estimador, además se obtuvo su orden de convergencia en media cuadrática,

Se exploró computacionalmente el comportamiento del estimador en diversas condiciones a través de datos simulados, del estimador, lo cual nos llevo a concluir de manera empírica que el estimador tenía un buen desempeño.

Finalmente, aplicamos la metodología propuesta para obtener una caracterización de la maduración de la conectividad del EEG multi-canal a diferentes frecuencias durante el primer año de vida en datos de voltaje. Los resultados obtenidos bajo estas estimaciones concuerdan con resultados anteriores y además parecen soportar algunas de las hipótesis empíricas que se tienen en el área de maduración de conectividad.

Algunas líneas futuras de investigación son: expandir el enfoque planteado al caso de regresión polinomial local, para así tener un estimador más general; aplicar la metodología propuesta en datos de individuos neonatales con riesgos perinatales para poder comparar sus normas con aquellas que obtuvimos para sujetos neonatales saludables; aplicar la metodología propuesta a datos de fuente lo que nos permitiría tener un entendimiento complementario de los resultados; y finalmente, desarrollar técnicas que nos permitan elegir de manera óptima y con bases teóricas bien fundamentadas los anchos de banda.

APÉNDICE A

Variedades diferenciales

En este apéndice se compendian conceptos básicos de geometría diferencial. Para más detalles, ver e.g. [Carmo, 1992].

Definición A.1. Sea $(\mathbf{M}, \mathcal{T})$ un espacio topológico de Hausdorff con base numerable. Decimos que \mathbf{M} es una **variedad topológica** si existe un número natural d tal que para cada punto $p \in \mathbf{M}$ existe una vecindad abierta U con centro en p , y una función continua $y : U \rightarrow \mathbb{R}^d$ que es un homeomorfismo en $y(U)$, donde U es un subconjunto abierto de \mathbf{R}^d .

La pareja (U, y) es llamada la **carta**, o **sistema coordinado**, en \mathbf{M} y el número natural d es llamado la **dimensión** de \mathbf{M} .

Se usara la notación \mathbf{M}^d para hacer evidente la dimensión de \mathbf{M} .

Definición A.2. Sea \mathbf{M} una variedad topológica. Entonces un C^r -**atlas** para \mathbf{M} es una colección de cartas

$$\mathbf{A} = \{(U_\alpha, y_\alpha) \mid \alpha \in I\}$$

tal que \mathbf{A} cubre por completo a \mathbf{M} , es decir

$$\mathbf{M} = \cup_\alpha U_\alpha.$$

Para toda $\alpha, \beta \in I$, también es posible definir la correspondiente **carta de transición**,

$$y_\beta \circ y_\alpha^{-1} \big|_{y_\alpha(U_\alpha \cap U_\beta)} : y_\alpha(U_\alpha \cap U_\beta) \rightarrow \mathbb{R}^d,$$

la cual es d -veces continuamente diferenciable.

Una carta (U, y) en \mathbf{M} se dice que es **compatible** con un C^r -atlas \mathbf{A} en \mathbf{M} , si $\mathbf{A} \cup \{(U, y)\}$ es un C^r -atlas. Un C^r -atlas $\tilde{\mathbf{A}}$ se dice **maximal** si contiene todas las cartas que

son compatibles con él. Un atlas maximal $\tilde{\mathbf{A}}$ en \mathbf{M} también es llamado una C^r -**estructura** en \mathbf{M} . La pareja $(\mathbf{M}, \tilde{\mathbf{A}})$ es nombrada una C^r -variedad o una **variedad diferenciable** de clase C^r , si \mathbf{M} es una variedad topológica y \mathbf{A} es una C^r -estructura en \mathbf{M} . Decimos que una variedad diferenciable es **suave** si sus cartas de transición son C^∞ .

Definición A.3. Sean d, n números naturales tales que $n \geq m$, $n \geq 1$ y $(N^n, \tilde{\mathbf{B}})$ una C^r -variedad. Un subconjunto M de N se dice que es una **subvariedad** de N si para cada punto $p \in M$ existe un carta $(U_p, y_p) \in \tilde{\mathbf{B}}$ tal que $p \in U_p$ en $y_p : U_p \rightarrow \mathbb{R}^d \times \mathbb{R}^{n-d}$ satisface

$$y_p(U_p \cap M) = y_p(U_p) \cap (\mathbb{R}^d \times \{0\}). \quad (\text{A.1})$$

El número natural $(n - d)$ es llamado la **codimensión** de M en N .

Sean $(\mathbf{M}^d, \tilde{\mathbf{A}})$ y $(N^n, \tilde{\mathbf{B}})$ dos C^r -variedades. Un mapeo $\phi : \mathbf{M} \rightarrow N$ se dice **diferenciable** de clase C^r si para todos los mapeos $(U, y) \in \tilde{\mathbf{A}}$ y $(V, z) \in \tilde{\mathbf{B}}$, la carta

$$z \circ \phi \circ x^{-1}|_{x(U \cap \phi^{-1}(V))} : x(U \cap \phi^{-1}(V)) \subset \mathbb{R}^d \rightarrow \mathbb{R}^n \quad (\text{A.2})$$

es de clase C^r .

Un mapeo diferenciable $\gamma : I \rightarrow \mathbf{M}$ definido en un intervalo abierto de \mathbb{R} es llamado una **curva** diferenciable en \mathbf{M} . Un mapeo diferenciable $f : \mathbf{M} \rightarrow \mathbb{R}$ es llamado una **función** en \mathbf{M} .

El conjunto de funciones suaves definidas en \mathbf{M} es denotado por $C^\infty(M)$.

Definición A.4. Sean dos variedades $(\mathbf{M}_1, \tilde{\mathbf{A}}_1)$ y $(\mathbf{M}_2, \tilde{\mathbf{A}}_2)$ de clase C^r . Se dice que son **difeomorfas** si existe un C^r -mapeo $\phi : \mathbf{M}_1 \rightarrow \mathbf{M}_2$ tal que la inversa $\phi^{-1} : \mathbf{M}_2 \rightarrow \mathbf{M}_1$ es de clase C^r . El mapeo ϕ se dice entonces que es un **difeomorfismo** entre $(\mathbf{M}_1, \tilde{\mathbf{A}}_1)$ y $(\mathbf{M}_2, \tilde{\mathbf{A}}_2)$.

Si \mathbf{M} es una variedad diferenciable y p es un punto en la variedad \mathbf{M} , un **vector tangente** X_p en p es un mapeo $X_p : \mathcal{E}(p) \rightarrow \mathbb{R}$ tal que

- I. $X_p(\lambda \cdot f + \mu \cdot g) = \lambda \cdot X_p(f) + \mu \cdot X_p(g)$,
- II. $X_p(f \cdot g) = X_p(f) \cdot g(p) + f(p) \cdot X_p(g)$

para toda $\lambda, \mu \in \mathbb{R}$ y $f, g \in \mathcal{E}(p)$. Entonces el conjunto de vectores tangentes en p se denomina el **espacio tangente** en p y es denotado por $T_p\mathbf{M}$

Sea $\phi : \mathbf{M} \rightarrow N$ un mapeo diferenciable entre variedades. El **diferencial** $d\phi_p$ de ϕ en el punto $p \in \mathbf{M}$ es el mapeo $d\phi_p : T_p\mathbf{M} \rightarrow T_{\phi(p)}N$ tal que para toda $X_p \in T_p\mathbf{M}$ y $f \in \mathcal{E}(\phi(p))$ se cumple que

$$(d\phi_p(X_p))(f) = X_p(f \circ \phi).$$

Un mapeo diferenciable $\phi : \mathbf{M} \rightarrow N$ entre variedades se dice que es una **inmersión** si para cada $p \in \mathbf{M}$ el diferencial $d\phi_p : T_p\mathbf{M} \rightarrow T_{\phi(p)}N$ es inyectivo. Por otra parte, un **encajamiento** es una inmersión $\phi : \mathbf{M} \rightarrow N$ que es un homeomorfismo en su imagen $\phi(\mathbf{M})$.

Dada una variedad suave M , definimos $C^\infty(\mathbf{M})$ como el anillo conmutativo de funciones suaves en \mathbf{M} y $C^\infty(TM)$ como el conjunto de campos vectoriales suaves en \mathbf{M} (que forma un módulo sobre $C^\infty(M)$). Definimos $C_0^\infty(TM) = C^\infty(\mathbf{M})$. Para cada $k \in \mathbb{Z}^+$, consideremos el k -fold producto tensor de $C^\infty(TM)$.

$$C_k^\infty(TM) = C^\infty(TM) \otimes \dots \otimes C^\infty(TM) \quad (\text{A.3})$$

Un **campo vectorial** B en \mathbf{M} del tipo (r, s) es una función $B : C_r^\infty(TM) \rightarrow C_s^\infty(TM)$ que satisface

$$\begin{aligned} & B(X_1 \otimes \dots \otimes X_{l-1} \otimes (f \cdot Y + g \cdot Z) \otimes X_{l+1} \otimes \dots \otimes X_r) \\ &= f \cdot B(X_1 \otimes \dots \otimes X_{l-1} \otimes Y \otimes X_{l+1} \otimes \dots \otimes X_r) + g \cdot B(X_1 \otimes \dots \otimes X_{l-1} \otimes Z \otimes X_{l+1} \otimes \dots \otimes X_r) \end{aligned}$$

para toda $X_1, \dots, X_r, Y, Z \in C^\infty(TM)$, $f, g \in C^\infty(\mathbf{M})$ y $l = 1, \dots, r$. Para acortar la notación denotaremos a $B(X_1 \otimes \dots \otimes X_r)$ como $B(X_1, \dots, X_r)$.

Se denotará por B_p a la restricción multi-lineal del campo tensorial B al producto tensor r -fold

$$\bigotimes_{l=1}^r T_p \mathbf{M} \quad (\text{A.4})$$

del espacio vectorial $T_p \mathbf{M}$, dado por

$$B_p : ((X_1)_p, \dots, (X_r)_p) \mapsto B(X_1, \dots, X_r)(p). \quad (\text{A.5})$$

Definición A.5. Sea \mathbf{M} una variedad suave. Una **métrica riemanniana** g en \mathbf{M} es un campo tensor

$$g : C_2^\infty(TM) \rightarrow C_0^\infty(TM)$$

tal que, para cada $p \in \mathbf{M}$, la restricción

$$g_p = g|_{T_p \mathbf{M} \otimes T_p \mathbf{M}} : T_p \mathbf{M} \otimes T_p \mathbf{M} \rightarrow \mathbb{R}, \quad (\text{A.6})$$

con

$$g_p : (X_p, Y_p) \mapsto g(X \cdot Y)(p), \quad (\text{A.7})$$

es un producto interior en el espacio tangente $T_p \mathbf{M}$. La pareja (\mathbf{M}, g) es llamada una **variedad Riemanniana**.

Decimos que \mathbf{M} es una **variedad diferenciable riemanniana** si es una variedad Riemanniana y diferenciable.

El producto interior en el espacio vectorial \mathbb{R}^d , dado por

$$\langle u, v \rangle_{\mathbb{R}^d} = \sum_{k=1}^d (u_k v_k)$$

define una **métrica Riemanniana**.

Una métrica riemanniana sobre una variedad \mathcal{M} define una longitud para cada curva en \mathcal{M} .

Se define la **distancia geodésica** (o distancia intrínseca) entre los puntos p y q en \mathcal{M} como la longitud mínima de las curvas suaves de $p = \gamma(a)$ a $q = \gamma(b)$.

[Apéndice A]

Teoría espectral de series de tiempo multivariadas

Este apéndice comprende de conceptos y resultados básicos de análisis espectral de series de tiempo multivariadas. Para más detalles, ver e.g. los libros ([Brillinger, 2001]) y ([Brockwell and Davis, 2006]).

Sea $\mathbf{X}_t = (X_{t1}, \dots, X_{tp})^T$ una serie p -valuada, con componentes X_{ti} ($t = 0, \pm 1, \dots$ y $i = 1, \dots, p$), y segundo momento finito, es decir $EX_{ti}^2 < \infty$.

Se definen los **vectores de medias**, como

$$\boldsymbol{\mu}_t := E\mathbf{X}_t = (\mu_{t1}, \dots, \mu_{tp})^T, \quad (\text{B.1})$$

y las **matrices de covarianza** como

$$\Gamma(t+h, t) := E[(\mathbf{X}_{t+h} - \boldsymbol{\mu}_t)(\mathbf{X}_t - \boldsymbol{\mu}_t)^T] = [\gamma_{ij}(t+h, t)]_{i,j=1}^p. \quad (\text{B.2})$$

Definición B.1. Decimos que la serie de tiempo \mathbf{X}_t , con vector de medias (B.1) y matrices de covarianza (B.2), es **estacionaria de segundo orden** si $\boldsymbol{\mu}_t$ y $\Gamma(t+h, t)$, para $h = 0, \pm 1, \dots$, son independientes de t .

En el caso de series estacionarias se suele usar la siguiente notación:

$$\boldsymbol{\mu} := E\mathbf{X}_t = (\mu_1, \dots, \mu_p)^T, \quad \text{y} \quad (\text{B.3})$$

$$\Gamma(h) = E[(\mathbf{X}_{t+h} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})^T] = [\gamma_{ij}(h)]_{i,j=1}^p. \quad (\text{B.4})$$

A (B.4) se le suele denominar **función de matriz de covarianza**.

Definición B.2. Se define la **función de matriz de correlación**, $R(\cdot)$, mediante

$$R(h) := \left[\frac{\gamma_{ij}(h)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}} \right]_{i,j=1}^p = [\rho(h)]_{i,j=1}^p. \quad (\text{B.5})$$

Teorema B.3. $\Gamma(\cdot)$ es la función de matriz de covarianza de un procesos p -valuado estacionario $\{X_t, t = 0 \pm 1, \dots\}$ si y solo si

$$\Gamma(h) = \int_{[-\pi, \pi]} \exp\{ihv\} dF(v), \quad (\text{B.6})$$

donde $h = 0, \pm 1, \dots$, y $F(\cdot)$ es una función de $[-\pi, \pi]$ en $\mathbb{R}^{p \times p}$ que cumple: $F(-\pi) = 0$, $F(\cdot)$ es continua por la derecha y $F(\omega) - F(\lambda)$ es matriz d.n.n. para todo $\lambda \leq \omega$. F es llamada la matriz de distribución espectral de $\{X_t\}$ o de $\Gamma(\cdot)$.

Teorema B.4. (Teorema de representación espectral). Si $\{X_t\}$ es una secuencia estacionaria con media cero y matriz de distribución espectral $F(\cdot)$, entonces existe un proceso continuo por la derecha y con incrementos ortogonales $\{Z(\lambda), -\pi \leq \lambda \leq \pi\}$ tal que:

I. para $-\pi \leq \lambda \leq \pi$

$$E[(Z(\lambda) - Z(-\pi))(Z(\lambda) - Z(-\pi))^*] = F(\lambda) \quad (\text{B.7})$$

II. con probabilidad 1

$$X_t = \int_0^{2\pi} \exp\{-i\omega t\} d\mathbf{Z}_X(\omega) \quad (\text{B.8})$$

para todo $t = 0, \pm 1, \dots$, * denota transpuesto conjugada.

Supongamos que, para todo $i, j = 1, \dots, p$

$$\sum_{h=-\infty}^{\infty} |\gamma_{ij}(h)| < \infty. \quad (\text{B.9})$$

Entonces existe la derivada $f(\omega)$ de la distribución espectral $F(\omega)$ y se cumple la relación

$$f(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \exp\{-i\omega h\} \Gamma(h) \quad (\text{B.10})$$

Denotemos $f(\omega) = (f_{jk}(\omega))$. El elemento diagonal $f_{jj}(\omega)$ de la matriz de densidad espectral $f(\omega)$ es conocido como el **espectro de energía** de la serie X_{tj} , con $j = 1, \dots, p$.

Dado que la función de covarianza cruzada de dos series de tiempo $\{X_{ti}\}$ y $\{X_{tj}\}$ en general no es simétrica respecto a cero, la matriz de espectros cruzados usualmente es una matriz compleja.

Definición B.5. Se define la matriz de **coherencia** (coherency) $R(\omega) = (R_{kj}(\omega))$ mediante

$$R_{kj}(\omega) = \frac{f_{kj}(\omega)}{[f_{kk}(\omega)f_{jj}(\omega)]^{1/2}} \quad (\text{B.11})$$

para $-\infty < \omega < \infty$.

A la magnitud $\rho_{kj}(\omega) = |R_{kj}(\omega)|$ le llamaremos el módulo de la coherencia, o simplemente coherencia, entre las series X_{tk} y X_{tj} en la frecuencia ω .

Definición B.6. Sea $B(\omega) = f^{-1}(\omega)$. Se define la matriz de **coherencias parciales** $\Psi(\omega) = (\Psi_{kj}(\omega))$ mediante

$$\Psi_{kj}(\omega) = \frac{B_{kj}(\omega)}{[B_{kk}(\omega)B_{jj}(\omega)]^{1/2}}$$

para $-\infty < \omega < \infty$.

A la magnitud $\psi_{kj}(\omega) = |\Psi_{kj}(\omega)|$ le llamaremos el módulo de la coherencia parcial, o simplemente coherencia parcial, entre las series X_{tk} y X_{tj} en la frecuencia ω .

Estimaciones

Supongamos que contamos con un segmento X_t ($t = 0, \dots, T-1$) de T valores consecutivos de una serie de tiempo p -valuada con media 0 y matriz de densidad espectral $f(\omega)$. La **transformada de Fourier finita** se define por

$$d^{(T)}(\omega) = \left[\sum_t h\left(\frac{1}{T}\right) \right] X_t \exp\{-i\omega t\} \quad (\text{B.12})$$

para $-\infty < \omega < \infty$, donde $h(t)$ es una función de *estrechamiento* que converge a cero para $|t| \rightarrow \infty$.

Un estimador de $f(\omega)$ es el periodograma definido por

$$\begin{aligned} \mathbf{I}^{(T)}(\omega) &= [\mathbf{I}_{ij}^{(T)}(\omega)] \\ &= \{[2\pi\mathbf{H}_{ij}^{(T)}(0)]^{-1} d_i^{(T)}(\omega) d_j^{(T)\bar{}}(\omega)\} \end{aligned} \quad (\text{B.13})$$

para $\omega \neq 0, \pm 2\pi, \dots$

El estimador anterior tiene las mismas propiedades de simetría y periodicidad que tiene $f(\omega)$.

La distribución normal multivariada con media μ_X y matriz de covarianza Σ la denotaremos como $N_p(\mu, \Sigma)$.

Se dice que un vector aleatorio X con componentes complejas, $X \in \mathbb{C}^p$, tiene distribución **normal multivariada compleja** con media μ y matriz de covarianza Σ , lo cual denotaremos por $N_p^C(\mu, \Sigma)$, si

$$\begin{bmatrix} \text{Re}(X) \\ \text{Im}(X) \end{bmatrix}$$

se distribuye como

$$N_{2p} \left(\begin{bmatrix} \text{Re}(X) \\ \text{Im}(X) \end{bmatrix}, \begin{bmatrix} \text{Re}(\Sigma) & -\text{Im}(\Sigma) \\ \text{Im}(\Sigma) & \text{Re}(\Sigma) \end{bmatrix} \right). \quad (\text{B.14})$$

En tal caso,

$$\begin{aligned} E(X) &= \mu \\ E[(X - \mu)(X - \mu)^*] &= \Sigma. \end{aligned}$$

Si X_1, \dots, X_n son independientes $\mathcal{N}_p^C(0, \Sigma)$ entonces la matriz aleatoria $p \times p$

$$W = \sum_{j=1}^n X_j X_j^* \quad (\text{B.15})$$

se dice que tiene **distribución Wishart compleja** de *dimensión p* y *grados de libertad n*.

Para los siguientes teoremas supondremos que la serie $X(t)$, con $t = 0, \pm 1, \dots$ es estacionaria y que la dependencia a lapsos es suficientemente pequeña como

$$\sum_{u_1, \dots, u_k = -\infty}^{\infty} |c_{a_1, \dots, a_k}(u_1, \dots, u_k)| < \infty, \quad (\text{B.16})$$

donde c_{a_1, \dots, a_k} son los cumulantes de la serie de tiempo.

Teorema B.7. *Sea $X(t)$, $t = 0, \pm 1, \dots$ una serie real p multivariada. Supongamos que $h_i(t)$, $-\infty < t < \infty$ es acotada, de variación acotada y se anula para $|t| > 1$. Sea $I^{(T)}(\omega)$ como arriba. Supongamos que $2\omega_j \neq 0 \pmod{2\pi}$ y $\omega_j \pm \omega_k \neq 0 \pmod{2\pi}$ para $1 \leq j < k \leq J$. Entonces $I^{(T)}(\omega_j)$, para $j = 1, \dots, J$ son asintóticamente independientes $W_p^C(1, f(\lambda_j))$.*

En el caso de función de estrechamiento h constante se cumple el siguiente teorema

Teorema B.8. *Sea $X(t)$, $t = 0, \pm 1, \dots$ una serie real p -variada. Definamos*

$$I^{(T)}(\omega) = (2\pi T)^{-1} \left(\sum_{t=0}^{T-1} X(t) \exp\{-i\omega t\} \right) \left(\sum_{t=0}^{T-1} X(t) \exp\{-i\omega t\} \right)^* \quad (\text{B.17})$$

para $-\infty < \omega < \infty$. Sea $s_j(T)$ un entero tal que $\omega_j(T) = 2\pi s_j(T)/T$ tienda a ω_j cuando $T \rightarrow \infty$ para $j = 1, \dots, J$. Suponga $2\omega_j(T)$, con $2\omega_j \neq 0 \pmod{2\pi}$ y $\omega_j \pm \omega_k \neq 0 \pmod{2\pi}$ para $1 \leq j < k \leq J$ y T suficientemente grande. Entonces $I^{(T)}(\omega_j(T))$, $j = 1, \dots, J$ son asintóticamente independientes $W_p^C(1, f(\omega_j))$, para $j = 1, \dots, J$.

Estos teoremas son importantes pues nos hablan de la convergencia asintótica que esperamos tener de los periodogramas.

Definimos el estimador de **Welch** mediante

$$\hat{f}(\omega) = \frac{1}{N} \sum_{s=1}^N \frac{X_t^s(\omega_j)(X_t^s(\omega_j))^*}{2\pi T}. \quad (\text{B.18})$$

Probabilidad en variedades diferenciales

Sea \mathbf{M} una variedad *diferenciable de dimensión d* . Sea dado un *encajamiento* inyectivo diferenciable de \mathbf{M} en algún espacio euclideo E^D , $J : \mathbf{M} \rightarrow E^D$, cuyo diferencial es también inyectivo.

El encajamiento J induce la distancia

$$\rho(x, y) = \|J(x) - J(y)\|$$

en \mathbf{M} , donde $\|\cdot\|$ denota la norma en el espacio E^D . La distancia ρ es conocida como **distancia extrínseca** en \mathbf{M} . Dada una distribución de probabilidad Q en \mathbf{M} , la función de Fréchet $F : \mathcal{M} \rightarrow \mathbb{R}_+$ se define por

$$F(x) = \int_{\mathbf{M}} \rho^2(x, y) Q(dy).$$

Un punto x de la variedad se dice que es la *media de Fréchet*, si en él alcanza su valor mínimo la función de Fréchet F .

Definición C.1. Sean (\mathbf{M}, ρ) y J como arriba. Sea Q en \mathcal{M} una distribución de probabilidad con función de Fréchet finita. El conjunto de medias de Fréchet de Q (con respecto a la métrica ρ) es conocido como **conjunto de medias extrínsecas** de Q (asociado al encajamiento J).

Diremos que Q tiene **media extrínseca** μ_E si el conjunto de medias extrínsecas de Q es un conjunto unitario.

Sea (\mathbf{M}, g) una variedad Riemanniana completamente conectada de dimensión d con tensor métrico g . Sea d_g la distancia geodésica en \mathbf{M} . Dada una distribución de probabilidad Q en \mathbf{M} , consideremos la función de Fréchet F_g definida por

$$F_g(p) = \int_{\mathbf{M}} d_g^2(p, m) Q(dm).$$

Un punto x de la variedad \mathcal{M} se dice que es la *media de Fréchet* con respecto a F_g si en él alcanza su valor mínimo la función F_g .

Definición C.2. *El conjunto de medias de Fréchet de Q es llamado **conjunto de medias intrínsecas** de Q .*

Bibliografía

- [Abhishek Bhattacharya, 2008] Abhishek Bhattacharya, R. B. (2008). Statistics on riemannian manifolds: Asymptotic distribution and curvature. *Proceedings of the American Mathematical Society*, 136(8):2959–2967.
- [Bhattacharya and Bhattacharya, 2015] Bhattacharya, A. and Bhattacharya, K. (2015). *Nonparametric Inference on Manifolds*. Cambridge University Press.
- [Bhattacharya and Patrangenaru, 2003] Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds -1. *The Annals of Statistics*, 31(1):1–29.
- [Bhattacharya and Patrangenaru, 2005] Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds -2. *The Annals of Statistics*, 33(3):1225–1259.
- [Bosch-Bayard et al., 2001] Bosch-Bayard, J., Valdés-Sosa, P., Virues-Alba, T., Aubert-Vázquez, E., John, E. R., Harmony, T., Riera-Díaz, J., and Trujillo-Barreto, N. (2001). 3d statistical parametric mapping of eeg source spectra by means of variable resolution electromagnetic tomography (vareta). *Clinical Electroencephalography*, 32(2):47–61.
- [Bosch-Bayard et al., 2012] Bosch-Bayard, J., Valdés-Sosa, P. A., Fernandez, T., Otero, G., Rivero, B. P., Ricardo-Garcell, J., González-Frankenberger, B., Galán-García, L., Fernandez-Bouzas, A., Aubert-Vazquez, E., Lage-Castellanos, A., Rodríguez-Valdés, R., and Harmony, T. (2012). 3d statistical parametric mapping of quiet sleep eeg in the first year of life. *NeuroImage*, 59(4):3297 – 3308.
- [Brillinger, 2001] Brillinger, D. R. (2001). *Time Series. Data Analysis and Theory*. SIAM: Society for Industrial and Applied Mathematics, San Francisco, second edition.
- [Brockwell and Davis, 2006] Brockwell, P. J. and Davis, R. A. (2006). *Time Series: Theory and Methods*. Springer, Nueva York, second edition.
- [Carmo, 1992] Carmo, M. D. (1992). *Riemannian Geometry*. Birkhauser Boston.

- [Chikuse, 2003] Chikuse, Y. (2003). *Statistics on Special Manifolds*. Springer.
- [Lin et al., 2015] Lin, L., Thomas, B. S., Zhu, H., and Dunson, D. B. (2015 (preprint)). Extrinsic local regression on manifold-valued data.
- [Lütkepohl, 1996] Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons.
- [Nadaraya, 1989] Nadaraya, E. A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves*. Kluwer Academic Publishers.
- [Otero et al., 2011] Otero, G., Harmony, T., Pliego-Rivero, F., Ricardo-Garcell, J., Bosch-Bayard, J., Porcayo-Mercado, R., Fernández-Bouzas, A., Díaz-Comas, L., Galán, L., Vieyra-Reyes, P., and Fernández, T. (2011). Qeeg norms for the first year of life. *Early Human Development*, 87:691–703.
- [Pascual-Marqui et al., 2014] Pascual-Marqui, R. D., Biscay, R. J., Bosch-Bayard, J., Lehmann, D., Kochi, K., Kinoshita, T., Yamada, N., and Sadato, N. (2014). Assessing direct paths of intracortical causal information flow of oscillatory activity with the isolated effective coherence (icoh). *Frontiers in Human Neuroscience*, 8(448).
- [Pérez and González-Farías, 2013] Pérez, R. A. and González-Farías, G. (2013). Partial Least Squares Regression on Symmetric Positive-Definite Matrices. *Revista Colombiana de Estadística*, 36:177 – 192.
- [Sameshima and (eds), 2014] Sameshima, K. and (eds), L. A. B. (2014). *Methods in Brain Connectivity Inference through Multivariate Time Series Analysis*. CRC Press Taylor & Francis Group.
- [Sanei and Chambers, 2007] Sanei, S. and Chambers, J. (2007). *EEG signal processing*. John Wiley & Sons, Ltd.
- [Shi et al., 2009] Shi, X., Styner, M., Lieberman, J., Ibrahim, J., Lin, W., and Zhu, H. (2009). Intrinsic regression models for manifold-valued data. In Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., and Taylor, C., editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2009*, volume 5762 of *Lecture Notes in Computer Science*, pages 192–199. Springer Berlin Heidelberg.
- [Szava et al., 1994] Szava, S., Valdes, P., Biscay, R., Galan, L., Bosch, J., Clark, I., and Jimenez, J. (1994). High resolution quantitative eeg analysis. *Brain Topography*, 6(3):211–219.
- [Tsybakov, 2009] Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer.
- [Turaga and Srivastava, 2016] Turaga, P. K. and Srivastava, A. (2016). *Riemannian Computing in Computer Vision*. Springer International Publishing.
- [Urrestarazu and Iriarte, 2005] Urrestarazu, E. and Iriarte, J. (2005). Análisis matemáticos en el estudio de señales electroencefalográficas. In *Revista Neurología*, volume 5762, pages 423–434. Viguera.
- [Webster, 2010] Webster, J. G. (2010). *Medical Instrumentation Application and Design*. John Wiley & Sons, Inc, San Francisco, fourth edition.

- [Worsley et al., 2005] Worsley, K. J., Chen, J.-I., Lerch, J., and Evans, A. C. (2005). Comparing functional connectivity via thresholding correlations and singular value decomposition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1457):913–920.
- [Yuan et al., 2012] Yuan, Y., Zhu, H., Lin, W., and Marron, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(4):697–719.
- [Zhu et al., 2009] Zhu, H., Chen, Y., Ibrahim, J. G., Li, Y., Hall, C., and Lin, W. (2009). Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *Journal of the American Statistical Association*, 104(487):pp. 1203–1212.