



Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Algunas Contribuciones al Análisis Estadístico Bayesiano de Datos Soportados en Variedades Riemannianas

Andrés Marcelino Iturriaga Jofré

Director de la Tesis:
PROF. ALEJANDRO JARA

Tesis presentada en
cumplimiento de requisitos
para optar al grado
de Doctor en Estadística.

Miembros del Comité Examinador:
PROF. LUIS CASTRO - UNIVERSIDAD DE CONCEPCIÓN
PROF. LUIS GUTIÉRREZ - PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
PROF. FERNANDO QUINTANA - PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Diciembre, 2016

Derechos de autor ©2016 por Andrés Marcelino Iturriaga Jofré
Todos los derechos reservados. Ninguna parte de la tesis puede reproducirse o transmitirse bajo
ninguna forma o por ningún medio o procedimiento,
sin permiso por escrito del autor.

Agradecimientos

Me gustaría reconocer y agradecer el apoyo financiero de la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), a través del programa de “Becas para Estudios de Doctorado en Chile”. Esta investigación fue parcialmente apoyada por la infraestructura de supercómputo del NLHPC (ECM-02) y el Proyecto FONDECYT 1141193 otorgado a mi profesor guía, Alejandro Jara.

Andrés Iturriaga
Diciembre, 2016

Índice General

Agradecimientos	i
Resumen	vi
Lista de Figuras	ix
Lista de Tablas	x
Lista de Algoritmos	xi
1 Introducción	1
1.1 Antecedentes generales y revisión de la literatura	1
1.1.1 El contexto general	1
1.1.2 Variedades Riemannianas	3
1.1.3 Análisis estadístico de <i>formas</i>	5
1.1.4 Generalizaciones del concepto de media Euclidiana	9
1.1.5 Modelos Bayesianos noparamétricos	11
1.1.6 Modelos para medidas de probabilidad relacionadas	13
1.2 Motivación	15

1.2.1	Cálculo de la media en el espacio de las <i>formas</i> planares de Kendall desde una perspectiva Bayesiana paramétrica	15
1.2.2	Caracterización del soporte en modelos para medidas de probabilidad soportadas en espacios Polacos y relacionadas por predictores Euclidianos	17
1.3	Contribuciones de la tesis	18
2	Un Enfoque Bayesiano Paramétrico para la Estimación de la Media en el Espacio de las Formas Planares	21
2.1	Introducción	21
2.2	El modelo normal complejo	24
2.3	El modelo Bayesiano	31
2.4	Post-procesamiento y media de Fréchet a posteriori	38
2.5	Estudio de simulación	45
2.6	Conclusiones	51
3	Procesos de Dirichlet Dependientes en Espacios Polacos	53
3.1	Introducción	53
3.2	Definición del DDP	55
3.2.1	Definición general	56
3.2.2	Casos particulares	59
3.3	Propiedades del DDP	62
3.3.1	Continuidad	62
3.3.2	Estructura de asociación	64
3.3.3	Propiedades de soporte	66
3.4	Mezclas inducidas por un DDP	70
3.4.1	Definición	71
3.4.2	Propiedades de soporte	71
3.4.3	Propiedad de consistencia	79
3.5	Conclusiones	81

4	Conclusiones y Trabajo Futuro	83
4.1	Conclusiones	83
4.2	Trabajo futuro	84
A	Material Suplementario del Capítulo 2	86
A.1	Propiedades de la distribución normal compleja	86
A.2	Simulación de la distribución normal compleja	90
A.3	Deducción de distribuciones a posteriori	91
A.3.1	Distribución a posteriori para el parámetro de localización	91
A.3.2	Distribución a posteriori para el parámetro de escala	92
A.3.3	Distribución a posteriori para el parámetro de rotación	92
A.3.4	Distribución a posteriori para el representante de la forma media	93
A.3.5	Distribución a posteriori para el parámetro de variabilidad	94
A.3.6	Distribución a posteriori para hiperparámetros	94
A.4	Caracterización de la métrica de Procrustes parcial	95
A.5	Convergencia del promedio muestral	96
B	Material Suplementario del Capítulo 3	98
B.1	Teorema de consistencia de Kolmogorov generalizado	98
B.2	Demostración del Teorema 3.1	100
B.3	Demostración del Teorema 3.2	101
B.4	Demostración del Teorema 3.3	106
B.5	Demostración del Teorema 3.4	113
B.6	Demostración del Teorema 3.5	116
B.7	Demostración del Teorema 3.6	120
B.8	Demostración del Teorema 3.7	125
B.9	Demostración del Teorema 3.8	136
B.10	Demostración del Teorema 3.9	145
B.11	Demostración del Teorema 3.10	150
B.12	Demostración del Teorema 3.11	154

ÍNDICE GENERAL

B.13 Demostración del Teorema 3.12	157
B.14 Demostración del Teorema 3.13	161
B.15 Demostración del Teorema 3.14	166
B.16 Demostración del Teorema B.1	167
B.17 Demostración del Lema B.1	169
B.18 Demostración del Lema B.2	172
Bibliografía	174

Resumen

El análisis estadístico de *formas*, y de variedades diferenciables más generales, es un área de creciente interés en investigación, el que se explica por la necesidad de trabajar con datos que poseen una estructura geométrica compleja. Este tipo de datos surgen, por ejemplo, en morfometría, meteorología, arqueología y genética. Para analizar la información estadística contenida en ellos se requieren métodos estadísticos especiales, que tomen en cuenta la geometría de los espacios subyacentes. Pese a su gran relevancia, en la actualidad existe una escasez de metodologías Bayesianas para este tipo de datos debido, en parte, a la dificultad para construir medidas de probabilidad sobre una variedad general.

El objetivo principal de esta tesis es contribuir, desde una perspectiva Bayesiana, al análisis estadístico de datos soportados en una variedad Riemanniana. Los principales resultados se presentan en los Capítulos 2 y 3. En el Capítulo 2 se propone un procedimiento Bayesiano paramétrico para la estimación de la *forma* media, a partir de datos de *configuraciones* (un número finito de puntos en el plano cartesiano), conteniendo la información de *forma* de objetos de interés. En este contexto, se muestra el problema de falta de identificación de los parámetros de un modelo estadístico comúnmente utilizado, y se propone una estrategia de post-procesamiento que permite resolver dicho problema. En base a un criterio de optimali-

dad, se propone un estimador de la *forma* media que incorpora en su definición una estructura métrica de la variedad Riemanniana subyacente. Finalmente, se presentan los resultados de un estudio de simulación, donde se ilustra el comportamiento del estimador propuesto y se compara con otros estimadores existentes.

En el Capítulo 3 se propone y estudia las propiedades de un procedimiento Bayesiano noparamétrico para la estimación de medidas de probabilidad relacionadas por predictores definidas en un espacio Polaco general, el que incluye a las variedades Riemannianas completas. La propuesta se basa en la extensión de los procesos de Dirichlet dependientes y mezclas inducidos por estos, definidos originalmente en espacios Euclidianos, a espacios Polacos más generales. Para los procesos de Dirichlet dependientes se estudian las propiedades de continuidad, estructura de asociación y soporte. En el caso de mezclas inducidas por procesos de Dirichlet dependientes, se establecen condiciones suficientes para que el modelo tenga propiedades adecuadas de soporte y de consistencia débil de la distribución a posteriori.

El Capítulo 1 discute la literatura existente, donde se contextualiza este trabajo. En el Capítulo 4 se resumen los principales resultados de la tesis y se discuten algunas líneas de investigación futura.

Lista de Figuras

1.1	El panel (a) y (b) muestra las <i>configuraciones</i> asociadas a una misma mano, pero rotadas en un ángulo de π radianes en sentido antihorario. El panel (c) muestra el promedio Euclidiano de las <i>configuraciones</i> observadas en el panel (a) y (b).	8
1.2	A la izquierda, el espacio de <i>formas</i> planares y una <i>forma</i> $[\mu]$. A la derecha, el espacio de <i>configuraciones</i> y algunos representantes de la misma <i>forma</i>	9
2.1	En (a), (b) y (c) <i>configuraciones</i> medias reales para 15 hitos. En (d), (e) y (f) <i>configuraciones</i> medias reales para 30 hitos.	47
2.2	En (a), (b) y (c) <i>configuraciones</i> medias reales para 15 hitos. En (d), (e) y (f) <i>configuraciones</i> medias reales para 30 hitos.	47
2.3	Para un número de hitos $p = 15$ y un tamaño muestral $n = 30$: en (a) la configuración media real, en (b) la media a posteriori, en (c) el estimador de Procrustes generalizado y en (d) la media de Frechet a posteriori.	48

Lista de Tablas

2.1	Comparación de estimadores en términos de sesgo y desviación estándar (número entre paréntesis) de la distancia Riemanniana ρ a la <i>forma</i> media $[\mu]$, para distintos tamaños muestrales y número de hitos. La <i>forma</i> de referencia es la del número 3 y los estimadores analizados son: promedio muestral \bar{w} , media a posteriori $\hat{\mu}_p$, estimador de Procrustes generalizado $\hat{\mu}_{GPA}$, media de Fréchet a posteriori $\hat{\mu}$ y el estimador de Micheas et al. (2010) $\hat{\mu}_M$	49
2.2	Comparación de estimadores en términos de sesgo y desviación estándar (número entre paréntesis) de la distancia Riemanniana ρ a la <i>forma</i> media $[\mu]$, para distintos tamaños muestrales y número de hitos. La <i>forma</i> de referencia es la de una mano y los estimadores analizados son: promedio muestral \bar{w} , media a posteriori $\hat{\mu}_p$, estimador de Procrustes generalizado $\hat{\mu}_{GPA}$, media de Fréchet a posteriori $\hat{\mu}$ y el estimador de Micheas et al. (2010) $\hat{\mu}_M$	50

Lista de Algoritmos

2.1	Esquema de simulación del MCMC	33
2.2	Esquema de post-estimación	44

1.1 Antecedentes generales y revisión de la literatura

1.1.1 El contexto general

Un principio en el que se sustenta la teoría inferencial estadística es que los datos pueden ser modelados como realizaciones de elementos aleatorios. El problema inferencial surge porque hay desconocimiento respecto de dicha distribución, por lo que un experimento estadístico clásico se define como una familia de probabilidades de muestreo indexadas por parámetros (Fisher, 1922),

$$\mathcal{E}_c = \{(\mathcal{X}, \mathcal{X}), P_\theta : \theta \in \Theta\},$$

donde \mathcal{X} es el espacio muestral subyacente a los datos, \mathcal{X} es una σ -álgebra de subconjuntos de \mathcal{X} , θ es un índice llamado parámetro, Θ es un conjunto no vacío denominado espacio paramétrico (o espacio de parámetros) y

$$P_\bullet(\bullet) : \Theta \times \mathcal{X} \rightarrow [0, 1],$$

es una función, tal que $\forall \theta \in \Theta$, $P_\theta(\bullet)$ es una probabilidad definida sobre el espacio medible $(\mathcal{X}, \mathcal{X})$, denominada probabilidad de muestreo. El espacio paramétrico Θ puede ser de dimensión finita o infinita. En el primer caso se habla de un modelo estadístico paramétrico, mientras que en el segundo, de un modelo estadístico noparamétrico. Adicionalmente, un modelo estadístico clásico se dice globalmente identificado, si la función $\theta \rightarrow P_\theta(\bullet)$ es inyectiva (Koopmans & Reiersol, 1950).

En el contexto Bayesiano, la incertidumbre que existe respecto de θ se incorpora a través de la definición de una medida de probabilidad Π sobre Θ , denominada distribución a priori. En este caso es necesario dotar a Θ de una σ -álgebra de subconjuntos, que denotaremos por \mathcal{B} . Es importante notar que la σ -álgebra \mathcal{B} debe ser escogida de modo que la función $\theta \rightarrow P_\theta(A)$ sea \mathcal{B} -medible, $\forall A \in \mathcal{X}$. El modelo estadístico Bayesiano corresponde a la única medida de probabilidad

$$Q(B \times A) := \int_B P_\theta(A) \Pi(d\theta), \quad B \in \mathcal{B}, A \in \mathcal{X},$$

definida sobre el espacio producto $\Theta \times \mathcal{X}$, y que puede ser extendida de manera única a la σ -álgebra generada por $\mathcal{B} \times \mathcal{X}$, denotada $\mathcal{B} \otimes \mathcal{X}$ (ver, por ejemplo, Florens et al., 1990). Luego, el experimento estadístico Bayesiano queda definido a través del espacio de probabilidad dado por

$$\mathcal{E}_b = (\Theta \times \mathcal{X}, \mathcal{B} \otimes \mathcal{X}, Q).$$

Cuando el modelo Bayesiano es regular (ver, por ejemplo, Schervish, 1995; Rao, 2005), existe una desintegración dual para Q . Dicha desintegración es en términos de la denominada probabilidad predictiva P y la distribución a posteriori Π_x . Así, Π_x es una medida de probabilidad sobre el espacio medible (Θ, \mathcal{B}) , $\forall x \in \mathcal{X}$, y la función $x \rightarrow \Pi_x(B)$ es \mathcal{X} -medible, $\forall B \in \mathcal{B}$. Además,

$$Q(B \times A) = \int_B P_\theta(A) \Pi(d\theta) = \int_A \Pi_x(B) P(dx), \quad B \in \mathcal{B}, A \in \mathcal{X}.$$

Finalmente, cabe destacar que desde el punto de vista Bayesiano, todas las inferencias respecto

del parámetro θ quedan determinadas por la distribución a posteriori Π_x .

Los espacios muestrales que motivan esta tesis corresponden a variedades Riemannianas. En términos generales, una variedad corresponde a un espacio topológico que se parece localmente, pero no necesariamente globalmente, a un espacio vectorial topológico dado. Algunos ejemplos de espacios vectoriales topológicos usuales en el contexto de variedades son los espacios de Fréchet de dimensión infinita, los espacios de Banach de dimensión infinita, los espacios de Hilbert de dimensión infinita y los espacios Euclidianos (ver, por ejemplo, [Lang, 1995](#); [Klingenberg, 1995](#)). Dependiendo de la dimensión del espacio vectorial topológico subyacente, surgen variedades de dimensión finita, en el caso Euclidiano, e infinita, en los otros. Algunas disciplinas en las que surgen datos soportados sobre una variedad son astronomía, meteorología, geología, cartografía, biología y física (ver, por ejemplo, [Patrangenaru & Ellingson, 2015](#)).

1.1.2 Variedades Riemannianas

Una variedad diferenciable de dimensión finita, M , es un espacio topológico de Hausdorff, paracompacto y conexo, en el que pequeñas vecindades pueden ser mapeadas de manera homeomorfa a conjuntos abiertos de espacios Euclidianos, y tales mapeos pueden ser compuestos de manera diferenciable (para más detalles ver, por ejemplo, [do Carmo, 1992](#); [Klingenberg, 1995](#); [Lang, 1995](#); [Sakai, 1996](#); [Jost, 2008](#); [Petersen, 2016](#)). A través de curvas diferenciables,

$$\alpha : [0, 1] \rightarrow M,$$

es posible definir vectores tangentes en un punto $p \in M$, a partir de la derivada $\dot{\alpha}(t)$ evaluada en $t = 0$, con $\alpha(0) = p$. Denotaremos por $T_p(M)$ al conjunto de todos los vectores que son tangentes a M en un punto $p \in M$, el cual corresponde a un espacio vectorial de la misma dimensión que M , y sobre el cual se pueden realizar procedimientos estándar del análisis estadístico multivariado, tales como el análisis en componentes principales, el análisis discrimi-

nante y los modelos de regresión. Usando los espacios tangentes de cada punto de la variedad, es posible definir una estructura global sobre M como

$$TM := \bigcup_{p \in M} T_p(M),$$

denominada el “*tangent bundle*” de M .

Una métrica Riemanniana es un producto interno definido sobre los espacios tangentes de M , que cambia suavemente con $p \in M$. Así, una variedad Riemanniana corresponde a una variedad diferenciable premunida de una métrica Riemanniana. Dicha estructura provee de herramientas para el cálculo de derivadas, integrales y largo de curvas, entre otras cosas. Dados $v_1, v_2 \in T_p(M)$, denotamos por $\langle v_1, v_2 \rangle_p$ a su métrica Riemanniana. Sea $\alpha : [0, 1] \rightarrow M$ una curva de clase \mathcal{C}^1 sobre la variedad. Se define el largo de α como

$$L[\alpha] = \int_0^1 \langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle_{\alpha(t)}^{1/2} dt,$$

donde, por definición, $\dot{\alpha}(t) \in T_{\alpha(t)}(M)$. Luego, se define la distancia geodésica entre los puntos $p_1, p_2 \in M$ como

$$d(p_1, p_2) := \inf\{L[\alpha] : \alpha : [0, 1] \rightarrow M, \alpha(0) = p_1, \alpha(1) = p_2\},$$

la cual corresponde a una métrica intrínseca sobre M (ver, por ejemplo, el Capítulo 3 de [Dryden & Mardia, 2016](#)). Se puede demostrar que (M, d) es un espacio métrico (ver, por ejemplo, [Jost, 2008](#)), y si dicho espacio es completo, entonces la variedad Riemanniana se dice completa. De esta manera, una variedad Riemanniana completa es un caso particular de un espacio Polaco.

También es posible definir variedades infinito-dimensionales sobre espacios de Frechét, Banach y Hilbert (de dimensión infinita). Por ejemplo, dado un espacio de Hilbert separable H , una variedad de Hilbert es un espacio separable y metrizable, tal que cada punto de la variedad posee una vecindad que es homeomorfa a un subconjunto abierto de H (ver, por ejemplo, [Lang,](#)

1995). Sin embargo, el interés de este trabajo está en las variedades Riemannianas completas de dimensión finita, tales como el espacio de las *formas* planares de Kendall (Kendall, 1977).

1.1.3 Análisis estadístico de *formas*

Kendall (1977) definió la *forma* de un objeto, como la información geométrica de éste, una vez que han sido eliminados los efectos de rotación, traslación y escala. Los procedimientos estadísticos asociados al análisis estadístico de *formas* se pueden clasificar en dos grupos, de acuerdo con la manera en la que la respuesta se registra (Dryden & Mardia, 1998, 2016). Específicamente, existen procedimientos donde se asume que la información de los objetos se registra a través de curvas o superficies continuas, correspondientes al contorno del objeto. Por otra parte, existen procedimientos que asumen que la información corresponde a la localización de una cantidad finita de hitos o puntos claves de interés. El enfoque basado en hitos, a diferencia del basado en curvas o superficies, trabaja siempre con datos definidos en espacios de dimensión finita y ha sido el más estudiado.

Los hitos son puntos de correspondencia, entre objetos y dentro del mismo. Así, un hito puede estar relacionado a una función particular de una población en estudio (por ejemplo, el núcleo de una célula animal), y distintos hitos pueden estar relacionados de manera que el cambio de uno de ellos afecta al resto (por ejemplo, si se desplaza la punta del pulgar, el resto del dedo se mueve coherentemente también). Una *configuración* es un arreglo conteniendo la localización del conjunto de los hitos sobre un objeto en particular. En general, los hitos se pueden clasificar en: anatómicos, matemáticos y pseudo-hitos. Los hitos anatómicos son puntos que manifiestan alguna característica de tipo biológica o funcional. Algunos ejemplos son el encuentro de suturas en un cráneo y la esquina de un ojo. Los hitos matemáticos son puntos que poseen alguna propiedad de tipo matemática o geométrica. Algunos ejemplos son los puntos con alta curvatura y los puntos extremos. Los pseudo-hitos, finalmente, son puntos, típicamente equiespaciados, ubicados entre los hitos anatómicos y matemáticos. La clasificación anterior

no es única ni necesariamente excluyente, pero es la más utilizada (ver, por ejemplo, [Dryden & Mardia, 1998, 2016](#)).

Para lo que sigue, suponemos la existencia de $p \in \mathbb{N} \setminus \{0\}$ hitos en el plano cartesiano. De esta forma, un objeto se puede representar por un vector $\mathbf{w} \in \mathbb{C}^p$, donde \mathbb{C}^p es el espacio de *configuraciones*. La razón de trabajar con \mathbb{C}^p , en vez de $\mathcal{M}_{p \times 2}(\mathbb{R})$ (matrices a valores reales de p filas y 2 columnas), es que la transformación de rotación se puede representar de manera más sencilla usando elementos de variable compleja. En efecto, para rotar una figura en \mathbb{C}^p , basta con multiplicar por el número complejo $e^{i\theta}$, mientras que en $\mathcal{M}_{p \times 2}(\mathbb{R})$ se hace necesario multiplicar por una matriz de la forma

$$A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

En ambos casos θ es el ángulo de rotación del objeto de interés, medido típicamente en sentido antihorario.

Una manera de eliminar el efecto de traslación es multiplicando las *configuraciones*, por la izquierda, por la matriz

$$\mathbf{C} = \mathbf{I}_p - \frac{1}{p} \mathbf{1}_p \mathbf{1}_p^t \in \mathcal{M}_{p \times p}(\mathbb{R}),$$

donde $\mathbf{I}_p \in \mathcal{M}_{p \times p}(\mathbb{R})$ corresponde a la matriz diagonal de dimensión $p \times p$, y $\mathbf{1}_p \in \mathbb{R}^p$ es el vector de unos. Sea $\mathbf{w}_C := \mathbf{C}\mathbf{w} \in \mathbb{C}^p$. Entonces, la suma de las coordenadas de \mathbf{w}_C es nula y tiene centroide $\mathbf{0} \in \mathbb{C}^p$. Existen otras maneras de eliminar el efecto de traslación, multiplicando por otras matrices, tales como la sub-matriz de Helmert (para detalles ver [Dryden & Mardia, 1998, 2016](#)). Una manera de eliminar el efecto de escala es normalizando,

$$\mathbf{w}_S := \frac{\mathbf{w}_C}{\|\mathbf{w}_C\|} \in \mathbb{C}^p,$$

donde $\|\mathbf{w}\| := \mathbf{w}^* \mathbf{w}$, con \mathbf{w}^* siendo la traspuesta conjugada de \mathbf{w} . Al objeto resultante $\mathbf{w}_S \in$

$\mathcal{S}_2^p \subset \mathbb{C}^p$, se le denomina *preforma*, donde \mathcal{S}_2^p es el espacio de las *preformas*. Finalmente, el efecto de rotación se elimina fijando un ángulo de referencia arbitrario, y usando la representación polar de un número complejo. Luego, la *forma* asociada a una *preforma* $\mathbf{w}_S \in \mathcal{S}_2^p$, queda definida por

$$[\mathbf{w}_S] = \{e^{i\phi}\mathbf{w}_S : \phi \in [0, 2\pi)\} \in \Sigma_2^p,$$

donde Σ_2^p es el espacio de las *formas* planares. Luego, $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{C}^p$ tienen la misma *forma*, es decir, $[\mathbf{w}_1] = [\mathbf{w}_2]$, si existen $d \in \mathbb{C}$, $b \in \mathbb{R}_+$ y $\phi \in [0, 2\pi)$, tales que

$$\mathbf{w}_1 = d\mathbf{1}_p + b \exp\{i\phi\}\mathbf{w}_2.$$

Equivalentemente, la forma de una *configuración* $\boldsymbol{\mu} \in \mathbb{C}^p$ queda definida como

$$[\boldsymbol{\mu}] = \{d\mathbf{1}_p + b \exp\{i\phi\}\boldsymbol{\mu} : d \in \mathbb{C}, b \in \mathbb{R}_+, \phi \in [0, 2\pi)\}.$$

Una de las dificultades de trabajar con Σ_2^p , es que carece de una estructura de espacio vectorial (Kendall, 1984). Así, el promedio Euclidiano de dos *configuraciones* con *formas* parecidas, puede tener una *forma* completamente diferente, como se ilustra en la Figura 1.1. Una posible explicación de esto es que la transformación que a cada *configuración* le asocia su *forma* es no lineal. Luego, no es de esperar que una transformación lineal, como el promedio Euclidiano, sea capaz de capturar la información de *forma*, que es intrínsecamente no lineal. Así, dados $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{C}^p$, puede ocurrir que

$$\left[\frac{1}{2}\mathbf{w}_1 + \frac{1}{2}\mathbf{w}_2 \right] \neq \left[\frac{1}{2}[\mathbf{w}_1] + \frac{1}{2}[\mathbf{w}_2] \right],$$

donde el promedio de *formas* se calcula a través de representantes en el espacio de *configuraciones*.

En términos de la construcción de medidas de probabilidad, es más simple trabajar con \mathbb{C}^p que con Σ_2^p . Sin embargo, desde el punto de vista del análisis estadístico de *formas*, surgen al-

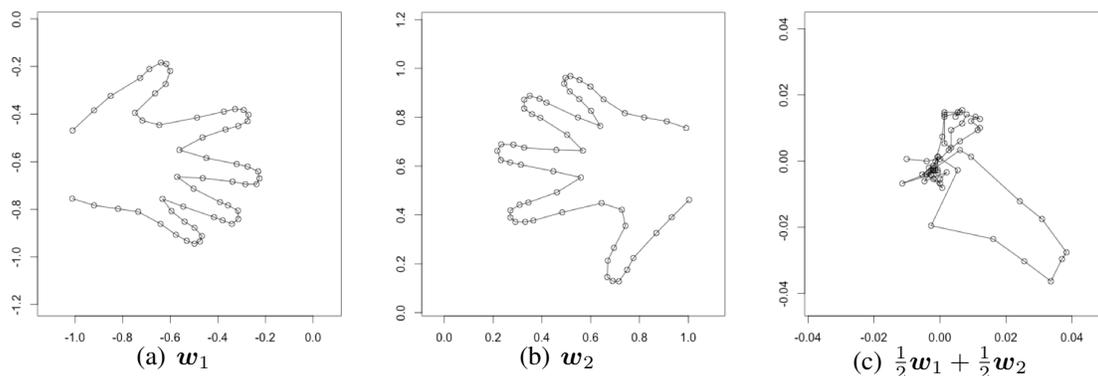


Figura 1.1: El panel (a) y (b) muestra las *configuraciones* asociadas a una misma mano, pero rotadas en un ángulo de π radianes en sentido antihorario. El panel (c) muestra el promedio Euclidiano de las *configuraciones* observadas en el panel (a) y (b).

gunas complicaciones al trabajar directamente con el espacio de *configuraciones*. La principal dificultad es que existen muchas *configuraciones* asociadas a la misma *forma*, como se ilustra en la Figura 1.2. Lo anterior frecuentemente trae consigo la aparición de problemas de falta de identificabilidad en los modelos estadísticos subyacentes.

Desde el punto de vista de la geometría diferencial, Σ_2^p tiene una estructura de variedad Riemanniana compacta de dimensión finita, con curvatura constante y positiva (ver, por ejemplo, [Small, 1996](#); [Kendall et al., 1999](#)). En particular, Σ_2^p es un espacio métrico compacto. Además, gracias a su curvatura positiva, sus vecindades geodésicas son pequeñas en comparación a las vecindades geodésicas de variedades con curvatura negativa. Lo anterior tiene repercusiones en lo que respecta a la unicidad global, de ciertas medidas de tendencia central que pueden ser definidas sobre Σ_2^p . Algunos ejemplos de medidas de tendencia central propuestos en la literatura son los baricentros Riemannianos y las medias de Fréchet (para detalles ver, por ejemplo, [Afsari, 2011](#)). En términos generales, dichos conceptos son extensiones del concepto de media Euclidiana.

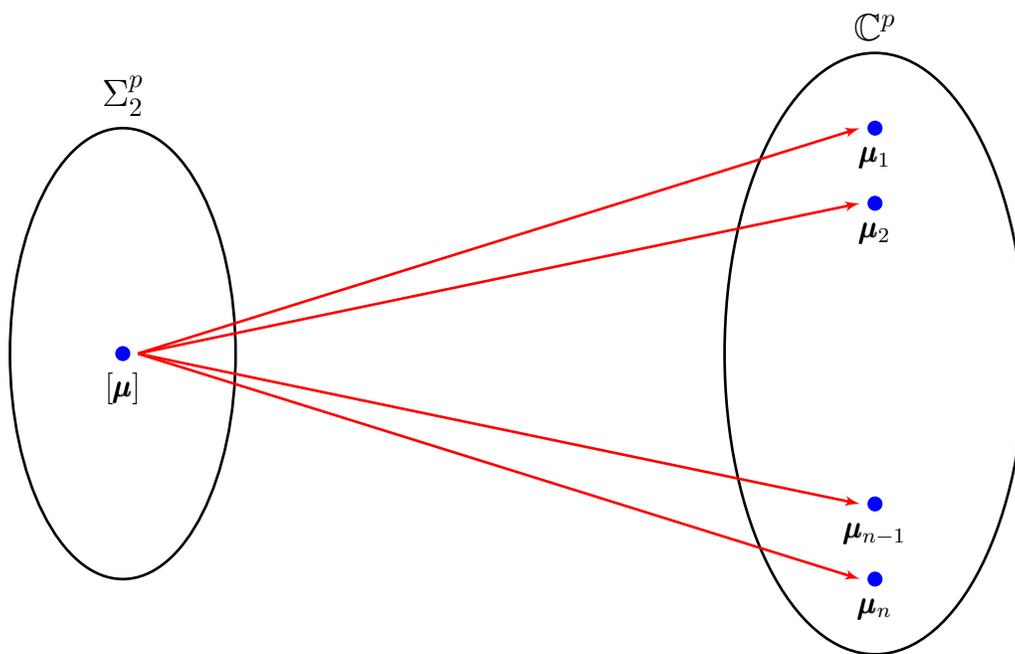


Figura 1.2: A la izquierda, el espacio de *formas* planares y una *forma* $[\mu]$. A la derecha, el espacio de *configuraciones* y algunos representantes de la misma *forma*.

1.1.4 Generalizaciones del concepto de media Euclidiana

Sea \mathcal{X} un espacio topológico, \mathcal{B} una σ -álgebra de Borel de subconjuntos de \mathcal{X} , y P una medida de probabilidad sobre el espacio medible $(\mathcal{X}, \mathcal{B})$. Cuando $\mathcal{X} = \mathbb{R}^n$, con $n \in \mathbb{N} \setminus \{0\}$, la media Euclidiana asociada a P está definida como

$$\mu := \int_{\mathbb{R}^n} x P(dx), \quad (1.1)$$

y $\mu \in \mathcal{X}$ (cuando existe). Cuando \mathcal{X} es un espacio de Banach separable de dimensión infinita, la expresión (1.1) no se puede aplicar sin supuestos adicionales. Lo anterior se debe a que la función $x \rightarrow \|x\|$ no es necesariamente P -integrable, donde $\|\cdot\|$ es una norma en \mathcal{X} . En este caso, la media μ de P queda definida por la expresión

$$\mu(l) := \int_{\mathcal{X}} l(x) P(dx), \quad (1.2)$$

donde $l \in \mathcal{X}^*$ es un elemento del dual topológico de \mathcal{X} (es decir, el espacio de todas las funciones lineales continuas de \mathcal{X} en \mathbb{R}). En este caso, sin supuestos adicionales, tales como que \mathcal{X} sea un espacio reflexivo, μ es un elemento del bidual topológico de \mathcal{X} (es decir, el dual topológico de \mathcal{X}^*). Para más detalles respecto del concepto de dualidad ver, por ejemplo, [Brezis \(2010\)](#). Cuando \mathcal{X} es un espacio de Hilbert separable y se cumple que

$$\int_{\mathcal{X}} \|x\| P(dx) < \infty,$$

entonces $\mu \in \mathcal{X}$. Además, μ es el único elemento satisfaciendo que

$$\langle \mu, y \rangle = \int_{\mathcal{X}} \langle x, y \rangle P(dx), \quad \forall y \in \mathcal{X}, \quad (1.3)$$

donde $\langle \cdot, \cdot \rangle$ es el producto punto de \mathcal{X} y $\|\cdot\|$ la norma asociada. Notar que la expresión (1.3) es equivalente a (1.2), debido a que $\mathcal{X} = \mathcal{X}^*$ por el teorema de representación de Riesz (ver, por ejemplo, [Aliprantis & Border, 2006](#)).

Las extensiones anteriores no funcionan cuando \mathcal{X} deja de ser un espacio vectorial topológico, que es lo que ocurre cuando \mathcal{X} es una variedad Riemanniana o un espacio métrico, donde la función de distancia no es inducida por una norma. En este contexto, una posible extensión del concepto de media Euclidiana a espacios métricos más generales, surge de la siguiente caracterización de la media Euclidiana μ ,

$$\int_{\mathbb{R}^n} \|x - \mu\|^2 P(dx) = \min_{a \in \mathbb{R}^n} \int_{\mathbb{R}^n} \|x - a\|^2 P(dx),$$

donde $\|\cdot\|$ es la norma Euclidiana usual. Así, si (\mathcal{X}, d) es un espacio métrico general, entonces se dice que $\mu \in \mathcal{X}$ es una media de Fréchet ([Fréchet, 1948](#)), si

$$\int_{\mathcal{X}} d^2(x, \mu) P(dx) = \min_{a \in \mathcal{X}} \int_{\mathcal{X}} d^2(x, a) P(dx).$$

Bajo supuestos adicionales, tales como la compacidad de \mathcal{X} , se puede demostrar la existen-

cia de medias de Fréchet, pero no necesariamente su unicidad. Por esta razón, se suele hablar del conjunto de las medias de Fréchet, en vez de la media de Fréchet. Además, existen otras generalizaciones del concepto de media Euclidiana, tales como las medias de Karcher y los baricentros Riemannianos. Más aún, se puede demostrar que la media de Fréchet es un caso particular del concepto de baricentro Riemanniano. En efecto, se dice que $\mu \in \mathcal{X}$ es un baricentro Riemanniano si

$$\frac{1}{p} \int_{\mathcal{X}} d^p(x, \mu) P(dx) = \min_{a \in \mathcal{X}} \frac{1}{p} \int_{\mathcal{X}} d^p(x, a) P(dx),$$

donde $1 \leq p < \infty$. Se prueba que mientras mayor sea p , menos resistente a la presencia de datos atípicos es la media anterior (ver, por ejemplo, [Afsari, 2011](#)). En esta tesis nos centraremos en las medias de Fréchet, como extensión del concepto de media Euclidiana.

1.1.5 Modelos Bayesianos noparamétricos

La especificación de un modelo Bayesiano noparamétrico (BNP) requiere de la especificación de una medida de probabilidad a priori sobre un espacio de parámetros de dimensión infinita. Algunos ejemplos incluyen el espacio de todas las medidas de probabilidad y el espacio de todas las funciones continuas, bajo las cuales un coeficiente de deriva garantiza la existencia de soluciones para una determinada ecuación diferencial estocástica.

Existe un interés creciente en el estudio de métodos BNP debido a su gran flexibilidad y a la existencia de mecanismos eficientes de computación de la distribución a posteriori ([Ghosh & Ramamoorthi, 2003](#); [Müller et al., 2004](#); [Hjort et al., 2010](#); [Müller et al., 2015](#); [Mitra & Müller, 2015](#)). Con respecto a esto último, algunos ejemplos incluyen algoritmos del tipo de cadenas de Markov Monte Carlo (MCMC) ([Metropolis et al., 1953](#); [Geman & Geman, 1984](#)), imputaciones secuenciales ([Kong et al., 1994](#); [Liu, 1996](#); [MacEachern et al., 1999](#)), recursiones predictivas ([Newton et al., 1998](#); [Newton & Zhang, 1999](#)) y métodos variacionales ([Opper & Saad, 2001](#); [Wainwright & Jordan, 2008](#); [Blei et al., 2006](#)).

La piedra angular de los modelos BNP es el proceso de Dirichlet (DP), definido por [Ferguson \(1973, 1974\)](#), y los modelos de mezclas inducidos por el DP (DPM) ([Ferguson, 1983](#); [Lo et al., 1984](#)). Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, $\mathcal{Y} \subseteq \mathbb{R}^p$, $\mathcal{B}(\mathcal{Y})$ la σ -álgebra de Borel de subconjuntos de \mathcal{Y} y $\mathcal{D}(\mathcal{Y})$ el espacio de todas las densidades con respecto a la medida de Lebesgue de \mathbb{R}^p . Entonces, un DPM es una función de densidad aleatoria definida $\forall \omega \in \Omega_0 \subseteq \Omega$, con $\mathbb{P}(\Omega_0) = 1$, como

$$f(\cdot | G(\omega)) := \int_{\Theta} \psi(\cdot, \theta) G(\omega)(d\theta) \in \mathcal{D}(\mathcal{Y}), \quad (1.4)$$

donde $\psi(\cdot, \theta) \in \mathcal{D}(\mathcal{Y})$, $\forall \theta \in \Theta \subseteq \mathbb{R}^q$, y G es un DP de parámetros (α, G^0) , con $\alpha \in \mathbb{R}_+$ y $G^0 \in \mathcal{P}(\Theta)$, donde $\mathcal{P}(\Theta)$ es el espacio de todas las medidas de probabilidad sobre el espacio medible $(\Theta, \mathcal{B}(\Theta))$ y $\mathcal{B}(\Theta)$ es la σ -álgebra de Borel de subconjuntos de Θ . Más aún, G es una medida de probabilidad aleatoria que puede ser representada mediante la siguiente representación tipo stick-breaking ([Sethuraman, 1994](#)),

$$G(\omega)(\cdot) = \sum_{i=1}^{\infty} \pi_i(\omega) \delta_{\theta_i(\omega)}(\cdot) \in \mathcal{P}(\Theta), \quad \forall \omega \in \Omega_0, \quad (1.5)$$

donde $\delta_{\theta}(\cdot)$ es la medida de Dirac en θ , $\pi_i = V_i \prod_{j < i} (1 - V_j)$, con $V_i | \alpha \stackrel{iid}{\sim} \text{Beta}(1, \alpha)$, y $\theta_i | G^0 \stackrel{iid}{\sim} G^0$. Propiedades y aplicaciones del DP pueden ser encontradas, por ejemplo, en [Ferguson \(1973, 1974\)](#), [Korwar & Hollander \(1973\)](#), [Antoniak \(1974\)](#), [Blackwell & MacQueen \(1973\)](#), [Cifarelli & Regazzini \(1990\)](#), [Hanson et al. \(2005\)](#), [Hjort & Ongaro \(2005\)](#) y [Hjort et al. \(2010\)](#).

Existen algunas alternativas al DP, tales como los modelos de muestreo de especies (SSM), introducidos por [Pitman \(1996\)](#). Los miembros de esta clase pueden ser representados como

$$G(B) = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}(B) + \left(1 - \sum_{i=1}^{\infty} w_i \delta_{\theta_i}(B) \right) G_0(B), \quad B \in \mathcal{B}(\Theta),$$

donde $\theta_i \stackrel{iid}{\sim} G_0$ son independientes de w_i , $i \geq 1$, y $\sum_{i=1}^{\infty} w_i \leq 1$ casi seguramente. Cuando

$\sum_{i=1}^{\infty} w_i = 1$ casi seguramente, el SSM se denomina propio y G es una medida de probabilidad discreta casi seguramente. Algunos ejemplos de SSM incluyen el DP, las medidas aleatorias normalizadas (Nieto-Barajas et al., 2004), el proceso Dirichlet-multinomial (Muliere & Secchi, 1995), ϵ -DP (Muliere & Tardella, 1998), el proceso Gaussiano inverso normalizado (Lijoi et al., 2005), el proceso beta de dos parámetros (Ishwaran & Zarepour, 2000), el proceso Poisson-Dirichlet de dos parámetros (Pitman & Yor, 1997) y el proceso stick-breaking (Ishwaran & James, 2001). El proceso stick-breaking es uno de los ejemplos más conocidos y satisface la siguiente expresión casi segura

$$G(B) = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}(B), \quad B \in \mathcal{B}(\Theta),$$

donde $w_i = V_i \prod_{j < i} (1 - V_j)$, $V_i \stackrel{ind.}{\sim} \text{Beta}(a_i, b_i)$, donde $\{a_i\}$ y $\{b_i\}$ son sucesiones de números positivos, y $\theta_i \stackrel{iid}{\sim} G_0$. El modelo resultante es propio si y sólo si $\sum_{i=1}^{\infty} E(\ln(1 - V_i)) = -\infty$.

En el contexto de espacios Euclidianos, se han caracterizado muchas de las propiedades de los DPM y sus extensiones, incluyendo soporte, consistencia a posteriori y tasas de concentración de la distribución a posteriori (ver, por ejemplo, Lo et al., 1984; Ghosal et al., 1999; Lijoi et al., 2005; Ghosal et al., 2007). Sin embargo, existe una escasez de resultados en el contexto de espacios muestrales más generales.

1.1.6 Modelos para medidas de probabilidad relacionadas

Motivado por el interés en la estimación de conjuntos de medidas de probabilidad relacionadas por predictores, se han propuesto varias extensiones a modelos del tipo (1.4) y (1.5), para incorporar la dependencia de predictores a los modelos, con la forma

$$f(y|x, G_x(\omega)) = \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta), \quad y \in \mathcal{Y}, \quad (1.6)$$

donde $f(y|x, G_x(\omega))$ es una densidad condicional indexada por los valores de un predictor $x \in \mathcal{X} \subseteq \mathbb{R}^s$. Así, la dependencia es introducida a través de la familia de medidas de probabilidad aleatorias G_x y el problema inferencial está relacionado con modelar $\{G_x : x \in \mathcal{X}\}$.

Algunos desarrollos tempranos en esta área aparecieron en [Cifarelli & Regazzini \(1978\)](#), quienes definieron la dependencia introduciendo un modelo de regresión sobre la medida de base de un DP. Un modelo más flexible fue propuesto por [MacEachern \(1999, 2000\)](#), denominado proceso de Dirichlet dependiente (DDP), el cual corresponde a un conjunto de procesos de Dirichlet marginales, con dependencia introducida a través de la siguiente modificación de la representación stick-breaking,

$$G_x(B) = \sum_{i=0}^{\infty} \pi_i(x) \delta_{\theta_i(x)}(B), \quad B \in \mathcal{B}(\Theta), \quad (1.7)$$

donde las masas puntuales $\theta_i(x)$, $i = 1, 2, \dots$ son procesos estocásticos independientes con conjunto de índices \mathcal{X} y distribuciones marginales G_x^0 , $x \in \mathcal{X}$, y los pesos toman la forma $\pi_i(x) = V_i(x) \prod_{j < i} (1 - V_j(x))$, donde $V_i(x)$, $i = 1, 2, \dots$ son procesos estocásticos independientes con conjunto de índices \mathcal{X} y distribuciones marginales $Beta(1, \alpha_x)$. Versiones del DDP con pesos independientes de predictores han sido aplicados de manera exitosa (ver, por ejemplo, [De Iorio et al., 2004](#); [Gelfand et al., 2005](#); [Jara et al., 2010](#)). Además, [Barrientos et al. \(2012\)](#) estudiaron propiedades de soporte de diversas versiones del DDP y procesos stick-breaking dependientes más generales.

Otras extensiones para tratar con medidas de probabilidad relacionadas incluyen mezclas DPM de normales ([Müller et al., 1996](#)), mezclas jerárquicas de DPM ([Müller et al., 2004](#)), DP jerárquico ([Teh et al., 2006](#)), DDP basado en orden ([Griffin & Steel, 2006](#)), DP anidado ([Rodríguez et al., 2008](#)), mezcla ponderada dependiente de predictores de DP ([Dunson et al., 2007](#)), kernel-stick breaking ([Dunson & Park, 2008](#)), matriz-stick breaking ([Dunson et al., 2008](#)), DP local ([Chung & Dunson, 2011](#)), logit-stick breaking ([Ren et al., 2011](#)), probit-stick breaking ([Chung & Dunson, 2009](#); [Rodríguez & Dunson, 2009](#)), modelo cluster-X ([Müller & Quintana,](#)

2010), modelo PPMx (Müller et al., 2011) y DP inclinado dependiente (Quintana, 2010). Procesos neutrales a la derecha dependientes y procesos de Poisson-Dirichlet de dos parámetros correlacionados han sido propuestos por Epifani & Lijoi (2010) y Leisen & Lijoi (2011), utilizando cópulas de Lévy. La clase general de medidas aleatorias completamente normalizadas dependientes ha sido estudiada por Nipoti (2011) y Lijoi et al. (2014). Basados sobre una formulación diferente del problema de estimación de densidades condicionales, Tokdar et al. (2010) y Jara & Hanson (2011) propusieron alternativas al enfoque de convoluciones de procesos stick-breaking dependientes.

De la misma forma a lo observado en el contexto de medidas de probabilidad, existe una escasez de modelos para conjuntos de medidas de probabilidad dependientes en el contexto de espacios muestrales más generales.

1.2 Motivación

La principal motivación de esta tesis es contribuir al análisis estadístico de datos soportados sobre variedades Riemannianas. Las motivaciones particulares de cada capítulo se discuten a continuación.

1.2.1 Cálculo de la media en el espacio de las *formas* planares de Kendall desde una perspectiva Bayesiana paramétrica

Existen varios procedimientos estadísticos para la estimación de la media en el espacio de las formas planares de Kendall (ver, por ejemplo, Gower, 1975; Goodall, 1991; Kent & Mardia, 1997; Koschat & Swayne, 1991; Brignell et al., 2005). El más conocido es el método de Procrustes generalizado, propuesto originalmente por Gower (1975) y estudiado posteriormente en el contexto del análisis estadístico de formas por Goodall (1991). Este es un estimador noparamétrico que elimina los efectos de traslación, rotación y escala de los datos originales,

1.2. MOTIVACIÓN

para posteriormente promediar los datos transformados (también denominados coordenadas de Procrustes). [Kent & Mardia \(1997\)](#) muestran que si la matriz de covarianza de las *configuraciones* no es isotrópica, el estimador de Procrustes no es siempre consistente. En este contexto, se han propuesto extensiones del estimador, como los estimadores de Procrustes ponderados (ver, por ejemplo, [Koschat & Swayne, 1991](#); [Brignell et al., 2005](#)), que permiten trabajar con matrices de covarianza no isotrópicas.

Una característica no deseable del estimador de Procrustes y sus generalizaciones, es que es difícil obtener su distribución de muestreo, lo que limita los procedimientos inferenciales. En el contexto de test de hipótesis, se han propuesto test de permutaciones (ver, por ejemplo, [Salmaso & Brombin, 2013](#)), los cuales funcionan bajo el supuesto que los datos se encuentran concentrados en torno a la media. Esto permite trabajar con aproximaciones locales sobre los espacios tangentes y utilizar métodos inferenciales Euclidianos estándar. Sin embargo, dichos enfoques terminan dejando de lado las propiedades geométricas de los espacios muestrales subyacentes, lo que puede llevar a resultados inferenciales incorrectos. Por otra parte, [Bhattacharya & Patrangenaru \(2002, 2003, 2005\)](#) estudian el problema de estimación no paramétrica del parámetro de localización en una variedad Riemanniana completa, en particular sobre el espacio de las *formas* planares, a través del concepto de media de Fréchet ([Fréchet, 1948](#)), y analizan su distribución asintótica.

La literatura Bayesiana existente en este contexto, incluye los trabajos de [Theobald \(2009\)](#), [Theobald \(2012\)](#), [Micheas & Peng \(2010\)](#) y [Fox et al. \(2016\)](#). [Theobald \(2009\)](#) y [Theobald \(2012\)](#) asumen un modelo normal multivariado para las *configuraciones*, haciendo explícitos los parámetros de localización y rotación, y proponen estimadores máximos a posteriori y Bayesianos empíricos. Estos procedimientos, sin embargo, no consideran parámetros de escala, haciendo difícil su uso práctico para el problema de estimación de la *forma* media. Por otra parte, [Micheas & Peng \(2010\)](#) considera la versión compleja del modelo normal anterior, incluyendo parámetros de escala, y consideran una distribución a priori normal compleja para el parámetro de *forma* media del modelo. Además, proponen un estimador del parámetro de

forma media que intenta minimizar la suma de las distancias de Procrustes (completa) a cada *configuración* en la muestra. Este procedimiento es utilizado por [Fox et al. \(2016\)](#). Es importante destacar, que en ninguno de estos trabajos se estudia el problema de falta de identificación del modelo estadístico normal complejo.

1.2.2 Caracterización del soporte en modelos para medidas de probabilidad soportadas en espacios Polacos y relacionadas por predictores Euclidianos

El desarrollo de procedimientos estadísticos para datos no Euclidianos se ha centrado, además del problema de estimación de la media, en la estimación de densidades univariadas (ver, por ejemplo, [Pelletier, 2005](#); [Bhattacharya & Dunson, 2010, 2012b](#)) y en el problema de regresión para datos Euclidianos con predictores no Euclidianos (ver, por ejemplo, [Pelletier, 2006](#); [Bhattacharya & Dunson, 2012a](#)). [Pelletier \(2005\)](#) estudia el problema de estimación de densidades, adaptando técnicas de tipo *kernel*, sobre una variedad Riemanniana compacta. [Bhattacharya & Dunson \(2010, 2012b\)](#) estudian el problema de estimación de densidades soportadas sobre un espacio métrico compacto, a través de un modelo de mezclas y desde una perspectiva Bayesiana noparamétrica. Además, para el modelo propuesto, obtienen resultados de consistencia débil y fuerte. [Pelletier \(2006\)](#) estudia el problema de estimación noparamétrica de una función de regresión, con variable dependiente real y regresores soportados en una variedad Riemanniana cerrada. Finalmente, [Bhattacharya & Dunson \(2012a\)](#) estudian el problema de predicción de una variable categórica, con predictores soportados en una variedad general.

La revisión de la literatura muestra que existe una escasez de procedimientos para la estimación de medidas de probabilidad relacionadas por predictores, en el contexto de variables de respuesta soportadas en un espacio Polaco general, a diferencia de lo que sucede en el caso Euclideo (ver, por ejemplo, [Barrientos et al., 2012](#); [Pati et al., 2013](#)). [Barrientos et al. \(2012\)](#) definen formalmente el DDP (y algunas versiones más parsimoniosas) y obtienen propiedades

de soporte producto débil. Para el caso de mezclas inducidas por un DDP, obtienen propiedades de soporte producto Hellinger y soporte producto Kulback-Leibler (KL). Más aún, extienden dichos resultados para procesos stick-breaking dependientes más generales. Sin embargo, los resultados de [Barrientos et al. \(2012\)](#) no son directamente generalizables a un espacio Polaco no Euclidiano, debido a que su definición del DDP usa el concepto de función de distribución, el cual sólo es válido en espacios Euclidianos. [Pati et al. \(2013\)](#) caracterizan propiedades de soporte más fuertes que [Barrientos et al. \(2012\)](#), pero para modelos de mezclas de densidades Gaussianas muy específicos. Adicionalmente, obtienen condiciones para la consistencia fuerte de sus modelos. Sin embargo, su enfoque tampoco es generalizable al contexto de espacios Polacos generales, pues las medidas Gaussianas sólo son aplicables en el contexto de espacios vectoriales, y los espacios Polacos de interés para esta tesis no tienen necesariamente dicha estructura. Finalmente, el problema de caracterizar el soporte debe ser abordado, debido a que disponer de modelos Bayesianos con un soporte grande es una condición casi necesaria para que los modelos sean efectivamente noparamétricos y es una condición requerida para la consistencia de la distribución a posteriori.

1.3 Contribuciones de la tesis

Esta tesis consta de dos partes en las que se estudian métodos Bayesianos para el análisis estadístico de datos soportados en una variedad Riemanniana completa. Cada parte es presentada en capítulos independientes, autocontenidos y con notación independiente, que incluyen una introducción, desarrollo y conclusiones. Las contribuciones particulares de cada capítulo se describen a continuación.

En el Capítulo 2, se estudia el problema de estimación de la media de una muestra de objetos de interés, representados por un número finito de hitos en el plano cartesiano. Aprovechando la naturaleza de espacio cociente que tiene el espacio de las *formas* planares, se estudian propiedades de un modelo estadístico normal complejo definido directamente sobre el espa-

cio de *configuraciones*, que modela de manera explícita los parámetros de *forma* (traslación, rotación y escala) de cada objeto de interés. Así, mediante un análisis de identificabilidad se concluye la falta de identificabilidad global y local del modelo estadístico subyacente, y se analiza el efecto que esto tiene en la estimación del parámetro de *forma* media. A través de un post-procesamiento de las muestras a posteriori de los parámetros del modelo, se propone un estimador de la *forma* media que incorpora en su definición una estructura métrica del espacio de las *formas* planares. Luego, se muestra que el estimador propuesto es genuinamente Bayesiano, en el sentido que minimiza una determinada función de pérdida. Formalmente, el estimador Bayesiano propuesto corresponde a una media de Fréchet a posteriori, de una transformación del parámetro de *forma* media del modelo normal complejo. A través de un estudio de simulación de Monte Carlo, se muestra que el estimador propuesto es competitivo con el estimador de Procrustes y es mejor que las otras alternativas Bayesianas existentes.

En el Capítulo 3, se propone y estudia un procedimiento Bayesiano noparamétrico para la estimación de medidas de probabilidad soportadas sobre un espacio Polaco general, que incluye a las variedades Riemannianas completas, y relacionadas por predictores Euclidianos. Para esto, se generaliza al contexto de espacios Polacos la definición del DDP dada por [Barrientos et al. \(2012\)](#), la que no puede ser extendida directamente a espacios no Euclidianos, debido a que usa funciones de distribución acumulada y ese concepto sólo es válido en espacios Euclidianos. Además, la definición propuesta del DDP no limita al uso de cópulas la construcción de los procesos estocásticos asociados a los pesos y átomos. Después, debido a la utilidad de disponer de modelos más parsimoniosos, se hacen dos simplificaciones del DDP: en la primera sólo los átomos dependen de predictores, y en la segunda sólo los pesos dependen de predictores. Luego, para el DDP (y algunas de sus simplificaciones) se estudian propiedades de continuidad, estructura de asociación y soporte. En términos de soporte, se analizan distintas topologías, con el objeto de obtener propiedades de soporte más fuertes. Así, primero se considera la topología producto débil, después una topología tipo \mathbb{L}_1 débil y finalmente, una topología \mathbb{L}_∞ débil. Luego, para el problema de la estimación de densidades de probabilidad soportadas sobre un espacio métrico compacto y relacionadas por predictores Euclidianos, se

1.3. CONTRIBUCIONES DE LA TESIS

estudian modelos de mezclas inducidos por un DDP, en términos de sus propiedades de soporte. Similarmente, se consideran distintas topologías con el objeto de obtener propiedades de soporte más fuertes. Así, las topologías consideradas incluyen la topología producto Hellinger, la topología producto \mathbb{L}_∞ , la topología producto Kulback-Leibler (KL), la topología $\mathbb{L}_\infty \mathbb{L}_\infty$ (denominada topología \mathbb{L}_∞) y la topología \mathbb{L}_∞ KL. Finalmente, se caracteriza un resultado de consistencia débil de la distribución a posteriori bajo muestreo independiente.

Un Enfoque Bayesiano Paramétrico para la Estimación de la Media en el Espacio de las Formas Planares

2.1 Introducción

El análisis estadístico de la *forma* geométrica de objetos es esencial en una gran variedad de disciplinas, tales como biología, medicina, arqueología, geografía, geología, agricultura y genética (ver, por ejemplo, [Dryden & Mardia, 1998, 2016](#); [Micheas et al., 2007](#); [Shakeri et al., 2016](#)). [Kendall \(1977\)](#) define la *forma* de un objeto como toda la información geométrica que queda de él, una vez que han sido eliminados los efectos de traslación, rotación y escala. En muchas aplicaciones la información de *forma* de un objeto se registra a través de la localización espacial de

2.1. INTRODUCCIÓN

un conjunto de características importantes e identificables del objeto en estudio, denominadas *configuraciones*.

Un objetivo básico del análisis estadístico de *formas* es la estimación de la *forma* media a partir de la información de *configuraciones*. Este es un problema complejo debido a que el espacio de las *formas* carece de una estructura de espacio vectorial (Kendall, 1984), lo que genera la necesidad de extender el concepto de media Euclidiana (Fréchet, 1948). Se han propuesto diferentes metodologías para su estimación, dentro de los que se destaca el método de Procrustes generalizado, propuesto originalmente por Gower (1975) y estudiado posteriormente por Goodall (1991). Este es un método noparamétrico que se basa en la transformación de los datos originales, con el objeto de remover los efectos de traslación, rotación y escala. Se generan, así, las denominadas coordenadas de Procrustes, las que son promediadas para generar el estimador de la *forma* media.

Kent & Mardia (1997) demuestran que, bajo el supuesto que la matriz de covarianza de las *configuraciones* es isotrópica, el estimador de Procrustes generalizado es un estimador consistente de la *forma* media. No obstante, en casos no isotrópicos, dicho estimador no es necesariamente consistente. Debido a esto, se han propuesto varias extensiones del estimador de Procrustes, como el estimador de Procrustes ponderado, que permiten trabajar con modelos no isotrópicos (ver, por ejemplo, Lissitz et al., 1976; Koschat & Swayne, 1991; Dryden & Mardia, 1998; Brignell et al., 2005). En cualquier caso, el estimador de Procrustes generalizado y sus extensiones comparten los siguientes problemas: (i) es difícil obtener la distribución muestral del estimador de la *forma* media, (ii) no es posible hacer inferencias sobre los parámetros de localización, rotación y escala, y (iii) no es posible extender el enfoque a datos de regresión con predictores arbitrarios.

Una manera de resolver los problemas anteriores, es a través de la especificación de modelos de probabilidad de muestreo apropiados. En este contexto, y bajo una perspectiva Bayesiana, Theobald (2009) y Theobald (2012) asumen un modelo normal multivariado para las *configura-*

2.1. INTRODUCCIÓN

ciones, haciendo explícitos los parámetros de localización y rotación, y proponen estimadores máximos a posteriori y Bayesianos empíricos. Sin embargo, estos procedimientos no consideran parámetros de escala, haciendo difícil su uso práctico en el problema de estimación de la *forma* media. [Micheas & Peng \(2010\)](#), por otra parte, proponen un enfoque Bayesiano que si permite hacer análisis de *forma* al incorporar, en un modelo normal complejo parámetros de localización, traslación y escala. El estimador de la *forma* media propuesto, intenta minimizar la suma de las distancias de Procrustes (completa) a cada *configuración* en la muestra. Este procedimiento es utilizado por [Fox et al. \(2016\)](#). Sin embargo, el enfoque propuesto presenta problemas conceptuales, debido a que el esquema de simulación utilizado no genera una cadena de Markov, cuya distribución estacionaria es la distribución a posteriori del parámetro de interés. Además, el modelo considerado no se encuentra completamente especificado y es necesario fijar el valor de ciertos parámetros. De estos, el asociado a la matriz de covarianza isotrópica es el más delicado, debido a que tiene un efecto importante en la calidad de la estimación de la *forma* media. Finalmente, ninguno de los trabajos mencionados se hace cargo del problema de la falta de identificación del modelo estadístico subyacente, lo que puede afectar las estimaciones ([Gustafson, 2015](#)).

Los problemas de identificación de los modelos existentes pueden ser, en principio, solucionados a través del uso de restricciones de identificación sobre el espacio paramétrico ([Lindley, 1972](#)). Sin embargo, esta estrategia no es simple de implementar, debido a que tales restricciones del espacio paramétrico le confieren una geometría no Euclidiana, con la consecuente dificultad en la definición de modelos de probabilidad a priori con dicho soporte. Otra solución es utilizar la versión Bayesiana del método de expansión de parámetros de [Lawrence et al. \(2008\)](#), definido originalmente en su versión clásica por [Liu et al. \(1998\)](#). La idea de [Lawrence et al. \(2008\)](#), quienes estaban interesados en el problema de estimación de matrices de correlación, fue trabajar en el espacio de las matrices de covarianza, para después buscar una transformación que garantizara la pertenencia al espacio de las matrices de correlación.

En este trabajo, aprovechando la naturaleza de espacio cociente que tiene el espacio de las

formas planares, se estudian propiedades de un modelo estadístico normal complejo definido directamente sobre el espacio de *configuraciones*, que modela de manera explícita los parámetros de traslación, escala y rotación de cada objeto de interés. Así, mediante un análisis de identificabilidad se concluye la falta de identificabilidad global y local del modelo estadístico subyacente. Luego, se propone un método de expansión de parámetros que resuelve simultáneamente el problema de la falta de identificación y de estimación de la *forma* media. Para el estimador propuesto, se muestra que minimiza una función de pérdida, lo que permite interpretarlo como un estimador genuinamente Bayesiano. Finalmente, a través de un estudio de simulación de Monte Carlo, se muestra que el estimador propuesto es competitivo con el estimador de Procrustes y es mejor que las otras alternativas Bayesianas existentes.

La estructura del capítulo es la siguiente. En la Sección 2.2 se introduce el modelo estadístico y se analiza el problema de la falta de identificación. En la Sección 2.3 se presenta el modelo estadístico Bayesiano. En la Sección 2.4 se presenta la estrategia de post-procesamiento y el estimador de la *forma* media propuesto. En la Sección 2.5, a través de un estudio de simulación, se ilustra el comportamiento de la metodología propuesta. Finalmente, la Sección 2.6 contiene las conclusiones finales del trabajo.

2.2 El modelo normal complejo

Supongamos que para una muestra de n objetos observamos configuraciones para p hitos de interés en un plano Euclidiano, denotados por $\mathbf{w}_j \in \mathbb{C}^p$, $j = 1, \dots, n$. Asumiremos que, para $j = 1, \dots, n$,

$$\mathbf{w}_j \mid d_j, b_j, \phi_j, \boldsymbol{\mu}, \sigma^2 \stackrel{\text{ind.}}{\sim} \mathbb{CN}_p(d_j \mathbf{1}_p + b_j \exp\{i\phi_j\} \boldsymbol{\mu}, \sigma^2 \mathbf{I}_p), \quad (2.1)$$

donde $\mathbb{CN}_p(\boldsymbol{\lambda}, \boldsymbol{\Sigma})$ denota a la distribución p -variada normal compleja con localización $\boldsymbol{\lambda} \in \mathbb{C}^p$ y matriz Hermitiana de escala $\boldsymbol{\Sigma}$ (Wooding, 1956), $d_j \in \mathbb{C}$ es un parámetro de traslación,

2.2. EL MODELO NORMAL COMPLEJO

$\mathbf{1}_p$ es un vector p -dimensional de unos, $b_j \in \mathbb{R}_+$ es un parámetro de escala, $\phi_j \in [0, 2\pi)$ es un parámetro de rotación, $\boldsymbol{\mu} \in \mathbb{C}^p$ corresponde a un representante de la *forma* media de la población, $\sigma^2 > 0$ es un parámetro de escala de la distribución normal compleja y \mathbf{I}_p es la matriz diagonal p -dimensional. Notemos que si $\mathbf{w} \sim \mathcal{CN}_p(\boldsymbol{\lambda}, \boldsymbol{\Sigma})$, con $\boldsymbol{\lambda} \in \mathbb{C}^p$ y $\boldsymbol{\Sigma}$ una matriz Hermitiana, entonces

$$\begin{pmatrix} \text{Re}(\mathbf{w}) \\ \text{Im}(\mathbf{w}) \end{pmatrix} \sim \mathcal{N}_{2p} \left(\begin{pmatrix} \text{Re}(\boldsymbol{\lambda}) \\ \text{Im}(\boldsymbol{\lambda}) \end{pmatrix}, \frac{1}{2} \begin{pmatrix} \text{Re}(\boldsymbol{\Sigma}) & -\text{Im}(\boldsymbol{\Sigma}) \\ \text{Im}(\boldsymbol{\Sigma}) & \text{Re}(\boldsymbol{\Sigma}) \end{pmatrix} \right), \quad (2.2)$$

donde $\mathcal{N}_d(\boldsymbol{\gamma}, \boldsymbol{\Psi})$ denota a la distribución d -variada normal con media $\boldsymbol{\gamma}$ y matriz de (co)varianzas $\boldsymbol{\Psi}$. Referimos al lector al Apéndice A.1 y a Andersen (1995) para más propiedades sobre la distribución normal compleja.

Los supuestos del modelo implican que la distribución conjunta de los datos esta dada por

$$f(\mathbf{w}_1, \dots, \mathbf{w}_n) = \prod_{j=1}^n \pi^{-p} \sigma^{-1} \exp\{-\sigma^{-2}(\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu})^*(\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu})\}, \quad (2.3)$$

donde \mathbf{z}^* denota la transpuesta conjugada de \mathbf{z} . Una propiedad importante del modelo estadístico, que tiene profundas repercusiones al momento de hacer inferencias sobre $\boldsymbol{\mu}$, es que no es globalmente identificado desde un punto de vista frecuentista (Koopmans & Reiersol, 1950), como se prueba en la siguiente proposición.

Proposición 2.1. *Sea $\mathcal{M} = \{(\mathbb{C}^{np}, \mathcal{B}(\mathbb{C}^{np}), P_\theta) : \theta \in \Theta\}$ un modelo estadístico, donde $\mathcal{B}(\mathbb{C}^{np})$ es la σ -álgebra de los Borelianos en \mathbb{C}^{np} y P_θ es una familia de medidas de probabilidad sobre el espacio medible $(\mathbb{C}^{np}, \mathcal{B}(\mathbb{C}^{np}))$, indexadas por un parámetro θ y con densidades con respecto a la medida de Lebesgue, f , dada por la expresión (2.3), donde $\theta = (d_1, \dots, d_n, b_1, \dots, b_n, \phi_1, \dots, \phi_n, \boldsymbol{\mu}, \sigma) \in \Theta := \mathbb{C}^n \times \mathbb{R}_+^n \times [0, 2\pi)^n \times \Theta_0 \times \mathbb{R}_+$ con $\Theta_0 := \mathbb{C}^p \setminus \{\mathbf{u} = (u_1, \dots, u_p) \in \mathbb{C}^p : u_1 = u_2 = \dots = u_p\}$. En este contexto, la función $\theta \rightarrow P_\theta$ no es inyectiva.*

2.2. EL MODELO NORMAL COMPLEJO

DEMOSTRACIÓN: Es fácil verificar que para todo $\mu \in \Theta_0$,

$$[\mu] := \{\nu \in \Theta_0 : \nu = d\mathbf{1}_p + b \exp\{i\phi\}\mu, \text{ donde } d \in \mathbb{C}, b \in \mathbb{R}_+, \phi \in [0, 2\pi)\},$$

constituye una clase de equivalencia de objetos con la misma *forma* geométrica que μ . Ahora, tomemos $\theta_1 \in \Theta$ de la forma

$$\theta_1 = (d_1^{(1)}, \dots, d_n^{(1)}, b_1^{(1)}, \dots, b_n^{(1)}, \phi_1^{(1)}, \dots, \phi_n^{(1)}, \mu^{(1)}, \sigma),$$

y definamos

$$\lambda_j := d_j^{(1)} \mathbf{1}_p + b_j^{(1)} \exp\{i\phi_j^{(1)}\}\mu^{(1)} \in \mathbb{C}^p, \quad j = 1, \dots, n.$$

Luego, si tomamos $\mu^{(2)} \in [\lambda_1]$, con $\mu^{(2)} \neq \mu^{(1)}$, existen

$$(d_1^{(2)}, b_1^{(2)}, \phi_1^{(2)}) \in \mathbb{C} \times \mathbb{R}_+ \times [0, 2\pi),$$

tales que

$$d_1^{(1)} \mathbf{1}_p + b_1^{(1)} \exp\{i\phi_1^{(1)}\}\mu^{(1)} = d_1^{(2)} \mathbf{1}_p + b_1^{(2)} \exp\{i\phi_1^{(2)}\}\mu^{(2)}.$$

En forma similar, dado que $\lambda_2 \in [\lambda_1] = [\mu^{(2)}]$, existen

$$(d_2^{(2)}, b_2^{(2)}, \phi_2^{(2)}) \in \mathbb{C} \times \mathbb{R}_+ \times [0, 2\pi),$$

tales que

$$d_2^{(1)} \mathbf{1}_p + b_2^{(1)} \exp\{i\phi_2^{(1)}\}\mu^{(1)} = d_2^{(2)} \mathbf{1}_p + b_2^{(2)} \exp\{i\phi_2^{(2)}\}\mu^{(2)}.$$

Finalmente, dado que $\lambda_j \in [\lambda_1] = \dots = [\lambda_{j-1}] = [\mu^{(2)}]$, existen los parámetros

$$(d_j^{(2)}, b_j^{(2)}, \phi_j^{(2)}) \in \mathbb{C} \times \mathbb{R}_+ \times [0, 2\pi),$$

tales que

$$d_j^{(1)} \mathbf{1}_p + b_j^{(1)} \exp\{i\phi_j^{(1)}\}\mu^{(1)} = d_j^{(2)} \mathbf{1}_p + b_j^{(2)} \exp\{i\phi_j^{(2)}\}\mu^{(2)},$$

2.2. EL MODELO NORMAL COMPLEJO

$j = 3, \dots, n$. Ahora, definamos

$$\boldsymbol{\theta}_2 := (d_1^{(2)}, \dots, d_n^{(2)}, b_1^{(2)}, \dots, b_n^{(2)}, \phi_1^{(2)}, \dots, \phi_n^{(2)}, \boldsymbol{\mu}^{(2)}, \sigma).$$

Por construcción, $\boldsymbol{\theta}_2 \in \Theta$ y $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$. Sin embargo, $P_{\boldsymbol{\theta}_1} = P_{\boldsymbol{\theta}_2}$. Luego la función $\boldsymbol{\theta} \rightarrow P_{\boldsymbol{\theta}}$ no es inyectiva y se concluye que el modelo estadístico \mathcal{M} no está globalmente identificado. ■

Una consecuencia directa de la falta de identificación global del modelo estadístico es que cualquier estimador de $\boldsymbol{\mu}$ es inconsistente, como se demuestra en el siguiente corolario.

Corolario 2.1. *Bajo las condiciones de la proposición anterior queda garantizada la no existencia de estimadores asintóticamente insesgados y débilmente consistentes del parámetro $\boldsymbol{\mu}$.*

DEMOSTRACIÓN: Demostremos primero la no existencia de estimadores asintóticamente insesgados para $\boldsymbol{\mu}$. Razonando por contradicción, supongamos que $\{\mathbf{s}_n : n \in \mathbb{N}\}$ es un estimador asintóticamente insesgados de $\boldsymbol{\mu}$. Sean $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ tales que $P_{\boldsymbol{\theta}_1} = P_{\boldsymbol{\theta}_2}$ con

$$\boldsymbol{\theta}_i := (d_1^{(i)}, \dots, d_n^{(i)}, b_1^{(i)}, \dots, b_n^{(i)}, \phi_1^{(i)}, \dots, \phi_n^{(i)}, \boldsymbol{\mu}^{(i)}, \sigma), \quad i = 1, 2.$$

Replicando un argumento de [San Martín & Quintana \(2002\)](#) se tiene que

$$\boldsymbol{\mu}^{(1)} = \lim_{n \rightarrow \infty} E_{\boldsymbol{\theta}_1}(\mathbf{s}_n) = \lim_{n \rightarrow \infty} E_{\boldsymbol{\theta}_2}(\mathbf{s}_n) = \boldsymbol{\mu}^{(2)}.$$

Luego, $\forall j = 1, \dots, n$

$$\begin{aligned} d_j^{(1)} \mathbf{1}_p + b_j^{(1)} \exp\{i\phi_j^{(1)}\} \boldsymbol{\mu}^{(1)} &= d_j^{(2)} \mathbf{1}_p + b_j^{(2)} \exp\{i\phi_j^{(2)}\} \boldsymbol{\mu}^{(2)} \\ \Leftrightarrow \left(d_j^{(1)} - d_j^{(2)} \right) \mathbf{1}_p + \left(b_j^{(1)} \exp\{i\phi_j^{(1)}\} - b_j^{(2)} \exp\{i\phi_j^{(2)}\} \right) \boldsymbol{\mu}^{(1)} &= \mathbf{0} \\ \Rightarrow d_j^{(1)} = d_j^{(2)}, b_j^{(1)} = b_j^{(2)}, \phi_j^{(1)} = \phi_j^{(2)}, & \end{aligned}$$

pues $\boldsymbol{\mu}^{(1)}$ no puede tener todas sus coordenadas iguales y $b_j^{(1)} b_j^{(2)} \neq 0$. Así, $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ y el

2.2. EL MODELO NORMAL COMPLEJO

modelo estadístico está identificado, lo cual contradice la proposición anterior. Luego, no hay estimadores asintóticamente insesgados de μ .

Demostremos ahora la no existencia de estimadores débilmente consistentes para μ . Razonando de nuevo por contradicción, supongamos que $\{s_n : n \in \mathbb{N}\}$ es un estimador débilmente consistente de μ . Tomemos de nuevo $\theta_1, \theta_2 \in \Theta$ tales que $P_{\theta_1} = P_{\theta_2}$ con

$$\theta_i := (d_1^{(i)}, \dots, d_n^{(i)}, b_1^{(i)}, \dots, b_n^{(i)}, \phi_1^{(i)}, \dots, \phi_n^{(i)}, \mu^{(i)}, \sigma), \quad i = 1, 2.$$

Entonces $s_n \rightarrow \mu^{(i)}$ en probabilidad con respecto a P_{θ_i} para $i = 1, 2$. Luego, la parte real e imaginaria de cada coordenada del estimador s_n converge respectivamente a la parte real e imaginaria de cada coordenada del parámetro $\mu^{(i)}$, en probabilidad con respecto a P_{θ_i} , para $i = 1, 2$. Razonando como en la demostración de la Proposición 2 de [San Martín & Quintana \(2002\)](#) se concluye, por unicidad del límite en \mathbb{R} , que la parte real e imaginaria de cada coordenada de $\mu^{(1)}$ y $\mu^{(2)}$ son iguales. Entonces $\mu^{(1)} = \mu^{(2)}$ y razonando como en la primera parte de la demostración se concluye que $\theta_1 = \theta_2$, lo cual es una contradicción. Luego, no existen estimadores débilmente consistentes de μ . ■

A pesar de la falta de identificabilidad global, todavía es posible que el modelo estadístico sea localmente identificado ([Rothenberg, 1971](#)). En la siguiente proposición se demuestra, sin embargo, que esto no ocurre.

Proposición 2.2. *Sea $\mathcal{M} = \{(\mathbb{C}^{np}, \mathcal{B}(\mathbb{C}^{np}), P_\theta) : \theta \in \Theta\}$ un modelo estadístico, donde $\mathcal{B}(\mathbb{C}^{np})$ es la σ -álgebra de los Borelianos en \mathbb{C}^{np} y P_θ es una familia de medidas de probabilidad sobre el espacio medible $(\mathbb{C}^{np}, \mathcal{B}(\mathbb{C}^{np}))$, indexadas por un parámetro θ y con densidades con respecto a la medida de Lebesgue, f , dada por la expresión (2.3), donde $\theta = (d_1, \dots, d_n, b_1, \dots, b_n, \phi_1, \dots, \phi_n, \mu, \sigma) \in \Theta := \mathbb{C}^n \times \mathbb{R}_+^n \times [0, 2\pi)^n \times \Theta_0 \times \mathbb{R}_+$ con $\Theta_0 := \mathbb{C}^p \setminus \{\mathbf{u} = (u_1, \dots, u_p) \in \mathbb{C}^p : u_1 = u_2 = \dots = u_p\}$. Entonces para todo $\theta_0 \in \Theta$ la función $V_{\theta_0} \ni \theta \rightarrow P_\theta$ no es inyectiva, donde V_{θ_0} es una vecindad arbitraria de θ_0 .*

2.2. EL MODELO NORMAL COMPLEJO

DEMOSTRACIÓN: Notemos que para todo $\theta \in \Theta$,

$$[[\theta]] := \{\theta^* \in \Theta : P_{\theta^*} = P_{\theta}\},$$

forma una clase de equivalencia en Θ (Rothenberg, 1971; Hsiao, 1983). Luego, si tomamos

$$\theta = (d_1, \dots, d_n, b_1, \dots, b_n, \phi_1, \dots, \phi_n, \boldsymbol{\mu}, \sigma) \in \Theta,$$

entonces

$$\begin{aligned} [[\theta]] &= \{\theta^* = (d_1^*, \dots, d_n^*, b_1^*, \dots, b_n^*, \phi_1^*, \dots, \phi_n^*, \boldsymbol{\mu}_*, \sigma_*) \in \Theta : \sigma_* = \sigma, \\ &\quad d_j^* \mathbf{1}_p + b_j^* \exp\{i\phi_j^*\} \boldsymbol{\mu}_* = d_j \mathbf{1}_p + b_j \exp\{i\phi_j\} \boldsymbol{\mu}, j = 1, \dots, n\} \\ &= \mathbb{C}^n \times \mathbb{R}_+^n \times [0, 2\pi)^n \times [\boldsymbol{\mu}] \times \{\sigma\}, \end{aligned}$$

donde $[\boldsymbol{\mu}]$ es la clase de equivalencia de objetos con la misma *forma* geométrica que $\boldsymbol{\mu}$. De lo anterior se deduce que, para todo $\theta_0 \in \Theta$, no es posible determinar una vecindad V_{θ_0} de θ_0 , tales que la función $V_{\theta_0} \ni \theta \rightarrow P_{\theta}$ sea inyectiva. Así, concluimos que el modelo estadístico \mathcal{M} no está localmente identificado. ■

Observación 2.1. *De la expresión $[[\theta]] = \mathbb{C}^n \times \mathbb{R}_+^n \times [0, 2\pi)^n \times [\boldsymbol{\mu}] \times \{\sigma\}$ se puede deducir que el modelo estadístico no permite distinguir el parámetro $\boldsymbol{\mu}$, pero si su forma $[\boldsymbol{\mu}]$, lo que lo convierte en un modelo de interés para el análisis estadístico de forma. Más aún, esto explica por qué no debe interpretarse el parámetro $\boldsymbol{\mu}$ como la forma media poblacional, sino como un representante de ella.*

A pesar que el Corolario 2.1 afirma que cualquier estimador del parámetro $\boldsymbol{\mu}$ es inconsistente, nada impide que dicho estimador tenga la misma *forma* que $\boldsymbol{\mu}$, que es el parámetro de interés. En el siguiente ejemplo se muestra que, para datos provenientes del modelo (2.1), el promedio muestral puede ser un estimador insesgado de la *forma* media, bajo ciertos supuestos.

Ejemplo 2.1. Sea $\mathbf{w}_1, \dots, \mathbf{w}_n$ una muestra proveniente del modelo estadístico (2.1) y consideremos el promedio muestral:

$$\bar{\mathbf{w}} := \frac{1}{n} \sum_{j=1}^n \mathbf{w}_j.$$

Notemos que

$$\begin{aligned} E_{\boldsymbol{\theta}}(\bar{\mathbf{w}}) &= \frac{1}{n} \sum_{j=1}^n E_{\boldsymbol{\theta}}(\mathbf{w}_j) = \frac{1}{n} \sum_{j=1}^n (d_j \mathbf{1}_p + b_j \exp\{i\phi_j\} \boldsymbol{\mu}) \\ &= \left(\frac{1}{n} \sum_{j=1}^n d_j \right) \mathbf{1}_p + \left(\frac{1}{n} \sum_{j=1}^n b_j \exp\{i\phi_j\} \right) \boldsymbol{\mu}, \end{aligned}$$

donde $\frac{1}{n} \sum_{j=1}^n d_j, \frac{1}{n} \sum_{j=1}^n b_j \exp\{i\phi_j\} \in \mathbb{C}$. Luego, si suponemos que $\sum_{j=1}^n b_j \exp\{i\phi_j\} \neq 0 \in \mathbb{C}$, se tiene que $E_{\boldsymbol{\theta}}(\bar{\mathbf{w}}) \in [\boldsymbol{\mu}]$. Así, a pesar que $\bar{\mathbf{w}}$ no es un estimador insesgado de $\boldsymbol{\mu}$, si puede ser un estimador insesgado de su forma. Adicionalmente, se puede demostrar (ver el Apéndice A.5) que

$$\frac{1}{n} \sum_{i=1}^n (\mathbf{w}_j - E_{\boldsymbol{\theta}}(\mathbf{w}_j)) \xrightarrow{P_{\boldsymbol{\theta}}} \mathbf{0} \in \mathbb{C}^p, \quad \forall \boldsymbol{\theta} \in \Theta,$$

cuando $n \rightarrow \infty$. Sin embargo, no es evidente que el promedio muestral converja a un objeto que tenga la misma forma que $\boldsymbol{\mu}$. □

El ejemplo anterior nos dice que no es claro el comportamiento límite del promedio muestral. No obstante, en la Sección 2.5 se muestra a través de un estudio de simulación que, al menos en muestras pequeñas, $\bar{\mathbf{w}}$ puede ser un muy mal estimador de $[\boldsymbol{\mu}]$.

En relación al modelo estadístico normal complejo, cabe señalar que ha sido ocupado frecuentemente en el análisis estadístico de *forma*. Más aún, ha sido el punto de partida para modelos más complejos como la distribución de Bingham compleja (Kent, 1994). Esto se debe a que

$$f(\mathbf{w}) = f(e^{i\phi} \mathbf{w}), \quad \forall \mathbf{w} \in \mathcal{S}_2^p, \phi \in [0, 2\pi), \quad (2.4)$$

si $f(\cdot)$ es la función de densidad de probabilidad asociada al modelo $\mathbb{C}\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, donde \mathcal{S}_2^p

denota al espacio de las *preformas*. Así, cualquier modelo de probabilidad definido sobre \mathcal{S}_2^p debe satisfacer (2.4), si es que quiere definir un modelo de probabilidad válido sobre el espacio de las *formas* planares, denotado Σ_2^p .

Sin perjuicio de lo anterior, el enfoque utilizado en este capítulo es trabajar directamente con el modelo normal complejo sobre \mathbb{C}^p , más que con la distribución que dicho modelo induce sobre Σ_2^p . Esto no debiera ser un impedimento para estimar $[\boldsymbol{\mu}]$, pues gracias al Teorema 9.4 de Kendall et al. (1999) se cumple que cada w_j induce una distribución de probabilidad en Σ_2^p , cuya *forma* media coincide con $[\boldsymbol{\mu}]$.

2.3 El modelo Bayesiano

Para el modelo estadístico (2.1), Micheas & Peng (2010) proponen las siguientes distribuciones a priori:

$$\boldsymbol{\mu} \sim \mathbb{CN}_p(\boldsymbol{\mu}_0, k_0^2 \mathbf{I}_p), \quad (2.5)$$

$$d_1, \dots, d_n \stackrel{iid}{\sim} \mathbb{CN}(d_0, \tau^2), \quad (2.6)$$

$$b_1, \dots, b_n \stackrel{iid}{\sim} \mathcal{N}(0, 1) \text{ truncada desde la izquierda en cero}, \quad (2.7)$$

$$\phi_1, \dots, \phi_n \stackrel{iid}{\sim} U[0, 2\pi), \quad (2.8)$$

donde $\boldsymbol{\mu}_0 \in \mathbb{C}^k$, $k_0 \in \mathbb{R}_+$ y $d_0 \in \mathbb{C}$ son hiperparámetros por fijar. Micheas & Peng (2010) proponen calibrar el parámetro σ^2 y el hiperparámetro τ^2 , a través de un pequeño estudio de robustez. Para esto le asignan una grilla de valores tanto a σ^2 como a τ^2 , optando por aquellos valores que logran un mejor ajuste del modelo respecto a una configuración de referencia dada. Finalmente, la calidad del ajuste es evaluada a través de la distancia de Procrustes completa (ver, por ejemplo, Dryden & Mardia, 1998, 2016).

A modo de disponer de un modelo Bayesiano completamente especificado, completamos el

2.3. EL MODELO BAYESIANO

modelo de [Micheas & Peng \(2010\)](#) proponiendo las siguientes distribuciones a priori:

$$\sigma^2 \sim \text{Gamma} - \text{Inv}(\alpha_1, \beta_1), \quad (2.9)$$

$$\tau^2 \sim \text{Gamma} - \text{Inv}(\alpha_2, \beta_2), \quad (2.10)$$

donde $\alpha_1, \beta_1, \alpha_2, \beta_2 \in \mathbb{R}_+$ son hiperparámetros por determinar. En la especificación anterior hay dos propiedades imperantes al momento de elegir distribuciones a priori: soporte completo y conjugación. Además, se está haciendo un supuesto de independencia a priori entre los parámetros del modelo estadístico.

Para el modelo estadístico Bayesiano propuesto, se obtienen las siguientes distribuciones condicionales (para detalles, ver el Apéndice A.3):

$$\boldsymbol{\mu} | \cdot \sim \mathbb{C}\mathcal{N}_p \left(\frac{1}{\Delta} \boldsymbol{\eta}, \frac{1}{\Delta} \mathbf{I}_p \right), \quad (2.11)$$

$$\text{con } \Delta = \frac{1}{\sigma^2} \sum_{j=1}^n b_j^2 + \frac{1}{k_0^2} \text{ y } \boldsymbol{\eta} = \frac{1}{k_0^2} \boldsymbol{\mu}_0 + \frac{1}{\sigma^2} \sum_{j=1}^n b_j e^{-i\phi_j} (\mathbf{w}_j - d_j \mathbf{1}_p),$$

$$d_j | \cdot \sim \mathbb{C}\mathcal{N} \left(\frac{1}{V} \lambda_j, \frac{1}{V} \right), \quad j = 1, \dots, n, \quad (2.12)$$

$$\text{con } V = \frac{p}{\sigma^2} + \frac{1}{\tau^2} \text{ y } \lambda_j = \mathbf{1}_p^t (\mathbf{w}_j - b_j e^{i\phi_j} \boldsymbol{\mu}) \sigma^{-2} + d_0 \tau^{-2},$$

$$b_j | \cdot \sim \mathcal{N} \left(\frac{\gamma_j + \bar{\gamma}_j}{\xi}, \frac{\sigma^2}{\xi} \right) \text{ truncada desde la izquierda en cero, } \quad j = 1, \dots, n, \quad (2.13)$$

$$\text{con } \xi = \sigma^2 + 2\boldsymbol{\mu}^* \boldsymbol{\mu} \text{ y } \gamma_j = e^{-i\phi_j} \boldsymbol{\mu}^* (\mathbf{w}_j - d_j \mathbf{1}_p),$$

$$[\phi_j | \cdot] \propto \exp \left\{ -\sigma^{-2} (\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu})^* (\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu}) \right\} \mathbf{1}_{[0, 2\pi)}(\phi_j), \quad (2.14)$$

$$\sigma^2 | \cdot \sim \text{Gamma} - \text{Inv}(\alpha_1 + np, \beta_1 + q), \quad (2.15)$$

2.3. EL MODELO BAYESIANO

con $q = \sum_{j=1}^n (\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu})^* (\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu})$, y

$$\tau^2 | \cdot \sim \text{Gamma} - \text{Inv} \left(\alpha_2 + n, \beta_2 + \sum_{j=1}^n \overline{(d_j - d_0)(d_j - d_0)} \right). \quad (2.16)$$

La implementación computacional del modelo anterior se basa en el uso de métodos de cadenas de Markov Monte Carlo. Específicamente, se utiliza un esquema de Gibbs, el que se bosqueja en el Algoritmo 2.1. En este algoritmo, para actualizar el parámetro de rotación se utiliza un paso de Metrópolis, con una distribución de propuesta $U[0, 2\pi)$, al igual que en Micheas & Peng (2010). No obstante, en la implementación de ellos, al momento de actualizar el parámetro $\boldsymbol{\mu}$, se simulan varios candidatos y se quedan con el que minimiza la distancia de Procrustes completa a los datos. En nuestra opinión, esto termina invalidando su algoritmo de Gibbs propuesto, pues el mínimo de normales complejas no se distribuye como una normal compleja.

Algoritmo 2.1 Esquema de simulación del MCMC

- 1: Establecer valores iniciales para $k_0, d_0, \alpha_1, \beta_1, \alpha_2$ y β_2
 - 2: Simular valores iniciales para $\tau^{2(0)}, \sigma^{2(0)}, d_1^{(0)}, \dots, d_n^{(0)}, b_1^{(0)}, \dots, b_n^{(0)}, \phi_1^{(0)}, \dots, \phi_n^{(0)}$ y $\boldsymbol{\mu}^{(0)}$
 - 3: **for** $r = 1, \dots, K$ **do**
 - 4: **for** $j = 1, \dots, n$ **do**
 - 5: Simular $d_j^{(r+1)}$ desde $\pi(d_j^{(r)} | \mathbf{w}_1, \dots, \mathbf{w}_n, b_j^{(r)}, \boldsymbol{\mu}^{(r)}, \phi_j^{(r)}, \sigma^{2(r)}, \tau^{2(r)})$
 - 6: Simular $b_j^{(r+1)}$ desde $\pi(b_j^{(r)} | \mathbf{w}_1, \dots, \mathbf{w}_n, d_j^{(r+1)}, \boldsymbol{\mu}^{(r)}, \phi_j^{(r)}, \sigma^{2(r)}, \tau^{2(r)})$
 - 7: Simular $\phi_j^{(r+1)}$ desde $\pi(\phi_j^{(r)} | \mathbf{w}_1, \dots, \mathbf{w}_n, b_j^{(r+1)}, \boldsymbol{\mu}^{(r)}, d_j^{(r+1)}, \sigma^{2(r)}, \tau^{2(r)})$
 - 8: Simular $\boldsymbol{\mu}^{(r+1)}$ desde $\pi(\boldsymbol{\mu}^{(r)} | \mathbf{w}_1, \dots, \mathbf{w}_n, b_j^{(r+1)}, d_j^{(r+1)}, \phi_j^{(r+1)}, \sigma^{2(r)}, \tau^{2(r)})$
 - 9: Simular $\tau^{2(r+1)}$ desde $\pi(\tau^{2(r)} | \mathbf{w}_1, \dots, \mathbf{w}_n, b_j^{(r+1)}, \boldsymbol{\mu}^{(r+1)}, \phi_j^{(r+1)}, \sigma^{2(r)}, d_j^{(r+1)})$
 - 10: Simular $\sigma^{2(r+1)}$ desde $\pi(\sigma^{2(r)} | \mathbf{w}_1, \dots, \mathbf{w}_n, b_j^{(r+1)}, \boldsymbol{\mu}^{(r+1)}, \phi_j^{(r+1)}, \tau^{2(r+1)}, d_j^{(r+1)})$
 - 11: **end for**
 - 12: **end for**
 - 13: Eliminar las primeras B observaciones tras detectar convergencia
 - 14: **return** Muestra de tamaño $N = K - B$ para cada parámetro simulado: $\boldsymbol{\mu}^{(1)}, \dots, \boldsymbol{\mu}^{(N)}$;
 $j = 1, \dots, n$: $d_j^{(1)}, \dots, d_j^{(N)}, b_j^{(1)}, \dots, b_j^{(N)}, \phi_j^{(1)}, \dots, \phi_j^{(N)}$; $\sigma^{2(1)}, \dots, \sigma^{2(N)}$; $\tau^{2(1)}, \dots, \tau^{2(N)}$
-

2.3. EL MODELO BAYESIANO

Una vez obtenida la muestra a posteriori simulada $\boldsymbol{\mu}^{(1)}, \dots, \boldsymbol{\mu}^{(N)}$, hay que proponer una manera de estimar la *forma* de $\boldsymbol{\mu}$. En este contexto, la solución Bayesiana estándar sería considerar la media a posteriori de $\boldsymbol{\mu}$, esto es,

$$\hat{\boldsymbol{\mu}}_p := \frac{1}{N} \sum_{l=1}^N \boldsymbol{\mu}^{(l)}.$$

Es importante señalar, que a priori no es evidente que $\hat{\boldsymbol{\mu}}_p$ sea un buen o mal estimador de $[\boldsymbol{\mu}]$, pues no se dispone de una expresión analítica para $E(\boldsymbol{\mu} | \mathbf{w}_1, \dots, \mathbf{w}_n)$. Sin embargo, al menos desde un punto de vista de teoría de decisión, la función de pérdida cuadrática no pareciera ser apropiada en este contexto, debido a que la distancia Euclidiana ni siquiera es una métrica en Σ_2^p .

Para entender el comportamiento de $\hat{\boldsymbol{\mu}}_p$ como estimador de $[\boldsymbol{\mu}]$, en la Sección 2.5 se realiza un estudio de simulación en muestras pequeñas, donde se observa que $\hat{\boldsymbol{\mu}}_p$ puede tener un muy mal desempeño. En el siguiente ejemplo se muestra un caso donde $\hat{\boldsymbol{\mu}}_p$ es un mal estimador de $[\boldsymbol{\mu}]$.

Ejemplo 2.2. Sea $\boldsymbol{\mu} \in \mathcal{S}_2^p$ y supongamos que $\boldsymbol{\mu}^{(1)} = e^{i\phi} \boldsymbol{\mu}$ y $\boldsymbol{\mu}^{(2)} = e^{i(\phi+\pi)} \boldsymbol{\mu}$, con $\phi \in [0, \pi)$ conocido. Entonces

$$\hat{\boldsymbol{\mu}}_p = \frac{1}{2}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)}) = \mathbf{0} \in \mathbb{C}^p.$$

Así, aunque $[\boldsymbol{\mu}^{(1)}] = [\boldsymbol{\mu}^{(2)}] = [\boldsymbol{\mu}]$, se cumple que $\hat{\boldsymbol{\mu}}_p \notin [\boldsymbol{\mu}]$. □

Observación 2.2. Del ejemplo anterior se observa que aunque la muestra a posteriori de $\boldsymbol{\mu}$ contenga información de su forma, el promedio muestral podría no ser capaz de capturarla.

De la expresión 2.11 se observa que

$$E(\boldsymbol{\mu} | \mathbf{w}_1, \dots, \mathbf{w}_n, d_1, \dots, d_n, b_1, \dots, b_n, \phi_1, \dots, \phi_n, \sigma^2) = \frac{\frac{1}{k_0^2} \boldsymbol{\mu}_0 + \frac{1}{\sigma^2} \sum_{j=1}^n b_j e^{-i\phi_j} (\mathbf{w}_j - d_j \mathbf{1}_p)}{\frac{1}{\sigma^2} \sum_{j=1}^n b_j^2 + \frac{1}{k_0^2}},$$

y por propiedades de la esperanza condicional se tiene que

$$\begin{aligned} E(\boldsymbol{\mu}|\mathbf{w}_1, \dots, \mathbf{w}_n) &= E(E(\boldsymbol{\mu}|\mathbf{w}_1, \dots, \mathbf{w}_n, d_1, \dots, d_n, b_1, \dots, b_n, \phi_1, \dots, \phi_n, \sigma^2)|\mathbf{w}_1, \dots, \mathbf{w}_n) \\ &= E\left(\frac{\frac{1}{k_0^2}\boldsymbol{\mu}_0 + \frac{1}{\sigma^2} \sum_{j=1}^n b_j e^{-i\phi_j} (\mathbf{w}_j - d_j \mathbf{1}_p)}{\frac{1}{\sigma^2} \sum_{j=1}^n b_j^2 + \frac{1}{k_0^2}} \middle| \mathbf{w}_1, \dots, \mathbf{w}_n\right). \end{aligned} \quad (2.17)$$

La expresión anterior no se puede seguir desarrollando, pues no todos los parámetros tienen una forma explícita para su distribución a posteriori.

En el siguiente ejemplo se considera una simplificación de nuestro modelo, tras la cual es posible decir algo más respecto de la media a posteriori de $\boldsymbol{\mu}$.

Ejemplo 2.3. *Supongamos que los parámetros $d_1, \dots, d_n \in \mathbb{C}$, $b_1, \dots, b_n \in \mathbb{R}_+$, $\phi_1, \dots, \phi_n \in [0, 2\pi)$ y $\sigma \in \mathbb{R}_+$ son deterministas, y conocidos. Entonces, la media a posteriori de $\boldsymbol{\mu}$ estaría dada por*

$$\begin{aligned} E(\boldsymbol{\mu}|\mathbf{w}_1, \dots, \mathbf{w}_n) &= \frac{\frac{1}{k_0^2}\boldsymbol{\mu}_0 + \frac{1}{\sigma^2} \sum_{j=1}^n b_j e^{-i\phi_j} (\mathbf{w}_j - d_j \mathbf{1}_p)}{\frac{1}{\sigma^2} \sum_{j=1}^n b_j^2 + \frac{1}{k_0^2}} \\ &= \frac{\frac{1}{k_0^2}\boldsymbol{\mu}_0 + \frac{1}{\sigma^2} (\sum_{j=1}^n b_j^2) \boldsymbol{\mu} + \frac{1}{\sigma^2} \sum_{j=1}^n b_j e^{-i\phi_j} \boldsymbol{\epsilon}_j}{\frac{1}{\sigma^2} \sum_{j=1}^n b_j^2 + \frac{1}{k_0^2}}, \end{aligned}$$

y su valor esperado sería

$$\frac{\frac{1}{k_0^2}\boldsymbol{\mu}_0 + \frac{1}{\sigma^2} (\sum_{j=1}^n b_j^2) \boldsymbol{\mu}}{\frac{1}{\sigma^2} \sum_{j=1}^n b_j^2 + \frac{1}{k_0^2}},$$

expresión que no tiene necesariamente la misma forma que $\boldsymbol{\mu}$, salvo cuando $\boldsymbol{\mu}_0 \in [\boldsymbol{\mu}]$ o $\boldsymbol{\mu}_0 = \mathbf{0} \in \mathbb{C}^p$. Luego, aún en este caso idealizado, no necesariamente la media a posteriori de $\boldsymbol{\mu}$ es un estimador insesgado de su forma. Más aún, dado que en este caso

$$\boldsymbol{\mu}|\mathbf{w}_1, \dots, \mathbf{w}_n \sim \mathbb{CN}_p\left(\frac{\frac{1}{k_0^2}\boldsymbol{\mu}_0 + \frac{1}{\sigma^2} \sum_{j=1}^n b_j e^{-i\phi_j} (\mathbf{w}_j - d_j \mathbf{1}_p)}{\frac{1}{\sigma^2} \sum_{j=1}^n b_j^2 + \frac{1}{k_0^2}}, \frac{1}{\frac{1}{\sigma^2} \sum_{j=1}^n b_j^2 + \frac{1}{k_0^2}} \mathbf{I}_p\right),$$

se concluye, usando el Teorema 9.4 de [Kendall et al. \(1999\)](#), que la distribución a posteriori

2.3. EL MODELO BAYESIANO

de $\boldsymbol{\mu}$ induce una medida de probabilidad sobre Σ_2^p , que tiene por forma media, la forma de $E(\boldsymbol{\mu}|\mathbf{w}_1, \dots, \mathbf{w}_n)$. \square

En lo que sigue se analiza el problema de identificación desde el punto de vista Bayesiano, que denominaremos b-identificación. Así, en la siguiente proposición se demuestra que el modelo estadístico Bayesiano no es b-identificado.

Proposición 2.3. *El modelo estadístico Bayesiano no es b-identificado.*

DEMOSTRACIÓN: Razonando como en [San Martín & González \(2010\)](#), sea $\mathcal{S} = \sigma(\mathbf{w}_1, \dots, \mathbf{w}_n)$ la σ -álgebra asociada a los datos y

$$\mathcal{A} = \sigma(d_1) \vee \dots \vee \sigma(d_n) \vee \sigma(b_1) \vee \dots \vee \sigma(b_n) \vee \sigma(\phi_1) \vee \dots \vee \sigma(\phi_n) \vee \sigma(\boldsymbol{\mu}) \vee \sigma(\sigma^2)$$

la σ -álgebra asociada al espacio de parámetros. Definamos para todo $j = 1, \dots, n$

$$\boldsymbol{\lambda}_j := d_j \mathbf{1}_p + b_j \exp\{i\phi_j\} \boldsymbol{\mu}.$$

Notemos que para todo $j = 1, \dots, n$, $\boldsymbol{\lambda}_j = E(\mathbf{w}_j|\mathcal{A})$ y por lo tanto

$$\sigma(\boldsymbol{\lambda}_j) \subseteq \sigma\{E(f|\mathcal{A}) : f \in [\mathcal{S}]^+\} \subseteq \mathcal{A},$$

donde $[\mathcal{S}]^+$ denota el conjunto de las funciones no negativas y \mathcal{S} -medibles. Luego, dado que $\sigma(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n)$ es la σ -álgebra minimal que contiene a cada $\sigma(\boldsymbol{\lambda}_j)$ se tiene que

$$\sigma(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n) \subseteq \sigma\{E(f|\mathcal{A}) : f \in [\mathcal{S}]^+\}.$$

Similarmente, $\sigma^2 \mathbf{I}_p = E((\mathbf{w}_j - E(\mathbf{w}_j))(\mathbf{w}_j - E(\mathbf{w}_j))^*|\mathcal{A})$ y entonces

$$\sigma(\sigma^2) \subseteq \sigma\{E(f|\mathcal{A}) : f \in [\mathcal{S}]^+\}.$$

Luego, razonando como antes se cumple que

$$\sigma(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2) \subseteq \sigma\{E(f|\mathcal{A}) : f \in [\mathcal{S}]^+\}.$$

Es claro que para todo $j = 1, \dots, n$, $\sigma(d_j) \not\subseteq \sigma(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2)$, pues no existe una función medible tal que d_j sea función de $(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2)$. Por esa misma razón se verifica que $\sigma(b_j) \not\subseteq \sigma(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2)$, $\sigma(\phi_j) \not\subseteq \sigma(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2)$ y $\sigma(\boldsymbol{\mu}) \not\subseteq \sigma(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2)$. Obviamente se tiene que $\sigma(\sigma^2) \subseteq \sigma(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2)$. Así, \mathcal{A} no es b-identificado, ni d_j , ni b_j , ni ϕ_j , ni $\boldsymbol{\mu}$. ■

En la siguiente proposición se demuestra que la reparametrización $(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2)$ si es b-identificada.

Proposición 2.4. *El modelo estadístico Bayesiano con la reparametrización $(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2)$, donde $\boldsymbol{\lambda}_j := d_j \mathbf{1}_p + b_j \exp\{i\phi_j\} \boldsymbol{\mu}$, $j = 1, \dots, n$, es b-identificado.*

DEMOSTRACIÓN: Es directo que con la reparametrización $(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2)$ el modelo estadístico resultante es identificado. Siguiendo los comentarios de [San Martín & González \(2010\)](#), si \mathcal{A} es una σ -álgebra de Blackwell y \mathcal{S} es separable, entonces la identificación implica la b-identificación. Finalmente, se concluye el resultado porque tanto \mathcal{A} como \mathcal{S} son σ -álgebras de Borel asociadas a espacios polacos ([Florens et al., 1990](#)). ■

De acuerdo al Comentario 1 de [San Martín & González \(2010\)](#), al momento de actualizar un parámetro no identificado, sólo se actualizan los parámetros identificados (suficientes minimales), en nuestro caso, $(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n, \sigma^2)$. Luego, queda abierta la pregunta de si al momento de simular muestras a posteriori de $\boldsymbol{\mu}$, se está actualizando o no su *forma*. Esto no es evidente, pues $[\boldsymbol{\lambda}_1] = \dots = [\boldsymbol{\lambda}_n] = [\boldsymbol{\mu}]$, y queda como trabajo de investigación futuro.

Para establecer si es sólo la falta de identificabilidad del modelo estadístico subyacente lo que provoca un mal desempeño de $\hat{\boldsymbol{\mu}}_p$ como estimador de $[\boldsymbol{\mu}]$, como sugieren los resultados

que serán presentados en la Sección 2.5, surge la estrategia de intentar eliminar el problema de identificación, a través de un enfoque de expansión de parámetros. Para esto, siguiendo las ideas de Lawrence et al. (2008), se busca una transformación T sobre la muestra a posteriori $\boldsymbol{\mu}^{(1)}, \dots, \boldsymbol{\mu}^{(N)}$, que elimine el problema de falta de identificación. Además, se espera que T garantice que

$$\frac{1}{N} \sum_{l=1}^N T(\boldsymbol{\mu}^{(l)})$$

sea un mejor estimador de $[\boldsymbol{\mu}]$ que $\hat{\boldsymbol{\mu}}_p$.

2.4 Post-procesamiento y media de Fréchet a posteriori

En la siguiente proposición se demuestra que una manera de eliminar los problemas de identificación del modelo estadístico (2.1), es eligiendo siempre a un representante por cada clase de equivalencia $[\boldsymbol{\mu}]$.

Proposición 2.5. *Sea $\tilde{\mathcal{M}} = \{(\mathbb{C}^{np}, \mathcal{B}(\mathbb{C}^{np}), P_{\boldsymbol{\theta}}) : \boldsymbol{\theta} \in \tilde{\Theta}\}$ un modelo estadístico, donde $\mathcal{B}(\mathbb{C}^{np})$ es la σ -álgebra de los Borelianos en \mathbb{C}^{np} y $P_{\boldsymbol{\theta}}$ es una familia de medidas de probabilidad sobre el espacio medible $(\mathbb{C}^{np}, \mathcal{B}(\mathbb{C}^{np}))$, indexadas por un parámetro $\boldsymbol{\theta}$ y con densidades con respecto a la medida de Lebesgue, f , dada por la expresión (2.3), donde $\boldsymbol{\theta} = (d_1, \dots, d_n, b_1, \dots, b_n, \phi_1, \dots, \phi_n, \boldsymbol{\mu}, \sigma) \in \tilde{\Theta} := \mathbb{C}^n \times \mathbb{R}_+^n \times [0, 2\pi)^n \times \tilde{\Theta}_0 \times \mathbb{R}_+$ con $\tilde{\Theta}_0 \subseteq \Theta_0$ formado por un solo representante de cada clase de equivalencia $[\boldsymbol{\mu}] \subseteq \Theta_0$. En este contexto, la función $\boldsymbol{\theta} \rightarrow P_{\boldsymbol{\theta}}$ es inyectiva.*

DEMOSTRACIÓN: Sean $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \tilde{\Theta}$ tales que $P_{\boldsymbol{\theta}_1} = P_{\boldsymbol{\theta}_2}$ con

$$\boldsymbol{\theta}_i := (d_1^{(i)}, \dots, d_n^{(i)}, b_1^{(i)}, \dots, b_n^{(i)}, \phi_1^{(i)}, \dots, \phi_n^{(i)}, \boldsymbol{\mu}_i, \sigma_i), \quad i = 1, 2.$$

2.4. POST-PROCESAMIENTO Y MEDIA DE FRÉCHET A POSTERIORI

Entonces, $\sigma_1^2 = \sigma_2^2$ y para todo $j = 1, \dots, n$

$$d_j^{(1)} \mathbf{1}_p + b_j^{(1)} \exp\{i\phi_j^{(1)}\} \boldsymbol{\mu}_1 = d_j^{(2)} \mathbf{1}_p + b_j^{(2)} \exp\{i\phi_j^{(2)}\} \boldsymbol{\mu}_2.$$

Luego, $[\boldsymbol{\mu}_1] = [\boldsymbol{\mu}_2]$. Sin embargo, dado que sólo se tomó un representante por cada clase de equivalencia, entonces necesariamente $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. Finalmente, dado que $\boldsymbol{\mu}_1$ no puede tener todas sus coordenadas iguales y $b_j^{(1)} b_j^{(2)} \neq 0$, se concluye que para todo $j = 1, \dots, n$

$$d_j^{(1)} = d_j^{(2)}, \quad b_j^{(1)} = b_j^{(2)} \quad \text{y} \quad \phi_j^{(1)} = \phi_j^{(2)}.$$

Por lo tanto, $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ y el modelo estadístico $\widetilde{\mathcal{M}}$ está globalmente identificado. ■

En lo que sigue, siguiendo un esquema Bayesiano del método de expansión de parámetros, proponemos una transformación sobre la muestra a posteriori $\boldsymbol{\mu}^{(1)}, \dots, \boldsymbol{\mu}^{(N)}$ que elimina los problemas de falta de identificación, siguiendo las directrices de la Proposición 2.5.

Definamos las funciones

$$\begin{aligned} T_1 : \mathbb{C}^p &\rightarrow \mathbb{C}^p, \\ \boldsymbol{w} &\rightarrow T_1(\boldsymbol{w}) := \boldsymbol{C}\boldsymbol{w}, \end{aligned} \tag{2.18}$$

con $\boldsymbol{C} \in \mathcal{M}_{p \times p}(\mathbb{R})$ dado por

$$\boldsymbol{C} = \mathbf{I}_p - \frac{1}{p} \mathbf{1}_p \mathbf{1}_p^t,$$

y

$$\begin{aligned} T_2 : \mathbb{C}^p &\rightarrow \mathbb{C}^p, \\ \boldsymbol{w} &\rightarrow T_2(\boldsymbol{w}) := \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}, \end{aligned} \tag{2.19}$$

con $\|\boldsymbol{w}\| := \boldsymbol{w}^* \boldsymbol{w}$. Notemos que

$$\begin{aligned} T_2 \circ T_1 : \mathbb{C}^p &\rightarrow \mathcal{S}_2^p, \\ \boldsymbol{w} &\rightarrow T_2 \circ T_1(\boldsymbol{w}) = \frac{\boldsymbol{C}\boldsymbol{w}}{\|\boldsymbol{C}\boldsymbol{w}\|}, \end{aligned} \tag{2.20}$$

es una función que elimina los efectos de traslación y escala. Para eliminar el efecto de rotación, basta con escoger un ángulo de referencia con respecto al cual proyectar la muestra a posteriori transformada

$$T_2(T_1(\boldsymbol{\mu}^{(1)})), \dots, T_2(T_1(\boldsymbol{\mu}^{(N)})).$$

Definamos así, las funciones

$$\begin{aligned} T_3 : \mathcal{S}_2^p \times \mathcal{S}_2^p &\rightarrow \mathcal{S}_2^p, \\ (\boldsymbol{\nu}, \boldsymbol{\eta}) &\rightarrow T_3(\boldsymbol{\nu}, \boldsymbol{\eta}) := \exp\{i \arg(\boldsymbol{\nu}^* \boldsymbol{\eta})\} \boldsymbol{\nu}, \end{aligned} \quad (2.21)$$

donde $\arg(\boldsymbol{\nu}^* \boldsymbol{\eta})$ corresponde al argumento del número complejo $\boldsymbol{\nu}^* \boldsymbol{\eta}$, y

$$\begin{aligned} T : \mathbb{C}^p \times \mathcal{S}_2^p &\rightarrow \mathcal{S}_2^p, \\ (\boldsymbol{\mu}^{(l)}, \boldsymbol{\eta}) &\rightarrow T(\boldsymbol{\mu}^{(l)}, \boldsymbol{\eta}) := T_3(T_2 \circ T_1(\boldsymbol{\mu}^{(l)}), \boldsymbol{\eta}), \end{aligned} \quad (2.22)$$

donde $\boldsymbol{\eta} \in \mathcal{S}_2^p$ es un elemento fijo. Así, dado $\boldsymbol{\eta}$ arbitrario se tiene que T es una transformación que aplicada sobre la muestra a posteriori $\boldsymbol{\mu}^{(1)}, \dots, \boldsymbol{\mu}^{(N)}$, elimina el problema de falta de identificación. Para lo que sigue, definamos

$$d^2([\boldsymbol{\eta}], [\boldsymbol{\nu}]) := \min_{\alpha \in [0, 2\pi)} \|\boldsymbol{\eta} - e^{i\alpha} \boldsymbol{\nu}\|^2, \quad (2.23)$$

con $\boldsymbol{\eta}, \boldsymbol{\nu} \in \mathcal{S}_2^p$. Se puede demostrar que d es una métrica sobre Σ_p^2 (ver, por ejemplo, [Dryden & Mardia, 1998, 2016](#)). En la siguiente proposición se demuestra que existe una manera óptima de elegir el parámetro $\boldsymbol{\eta}$.

Proposición 2.6. *Una manera óptima de elegir el parámetro $\boldsymbol{\eta}$ es imponiendo que $[\boldsymbol{\eta}]$ sea una media de Fréchet empírica de $[T_2 \circ T_1(\boldsymbol{\mu}^{(1)})], \dots, [T_2 \circ T_1(\boldsymbol{\mu}^{(N)})]$ con respecto a la métrica $d(\cdot, \cdot)$, es decir, que*

$$[\boldsymbol{\eta}] = \arg \min_{[\boldsymbol{\nu}] \in \Sigma_p^2} \frac{1}{N} \sum_{l=1}^N d([\boldsymbol{\nu}], [T_2 \circ T_1(\boldsymbol{\mu}^{(l)})])^2.$$

2.4. POST-PROCESAMIENTO Y MEDIA DE FRÉCHET A POSTERIORI

DEMOSTRACIÓN: Dado $\boldsymbol{\eta} \in \mathcal{S}_2^p$, definamos $\boldsymbol{\nu}_l := T(\boldsymbol{\mu}^{(l)}, \boldsymbol{\eta})$. Se puede demostrar (ver el Apéndice A.4) que

$$\|\boldsymbol{\eta} - T_3(\boldsymbol{\nu}, \boldsymbol{\eta})\|^2 = d^2([\boldsymbol{\eta}], [\boldsymbol{\nu}]).$$

Luego,

$$\begin{aligned} \frac{1}{N} \sum_{l=1}^N \|\boldsymbol{\eta} - \boldsymbol{\nu}_l\|^2 &\geq \min_{\boldsymbol{\eta} \in \mathcal{S}_2^p} \frac{1}{N} \sum_{l=1}^N \|\boldsymbol{\eta} - T(\boldsymbol{\mu}^{(l)}, \boldsymbol{\eta})\|^2 \\ &= \min_{\boldsymbol{\eta} \in \mathcal{S}_2^p} \frac{1}{N} \sum_{l=1}^N d([\boldsymbol{\eta}], [T_2 \circ T_1(\boldsymbol{\mu}^{(l)})])^2 \\ &= \min_{[\boldsymbol{\eta}] \in \Sigma_2^p} \frac{1}{N} \sum_{l=1}^N d([\boldsymbol{\eta}], [T_2 \circ T_1(\boldsymbol{\mu}^{(l)})])^2. \end{aligned}$$

Así, una manera óptima de elegir $[\boldsymbol{\eta}]$ es escogiendo una media de Fréchet empírica (Fréchet, 1948) de $[T_2 \circ T_1(\boldsymbol{\mu}^{(1)})], \dots, [T_2 \circ T_1(\boldsymbol{\mu}^{(N)})]$, cuya existencia está garantizada debido a que Σ_2^p es un conjunto compacto (Kendall, 1984). ■

La proposición anterior no garantiza la unicidad de la media de Fréchet empírica. Al respecto sólo es posible garantizar unicidad en casos en los cuales las formas de $\boldsymbol{\mu}^{(1)}, \dots, \boldsymbol{\mu}^{(N)}$ están relativamente cerca. En efecto, gracias al Teorema 9.6 de Kendall et al. (1999) se tiene que una condición suficiente para la unicidad de la media de Fréchet empírica de $[T_2 \circ T_1(\boldsymbol{\mu}^{(1)})], \dots, [T_2 \circ T_1(\boldsymbol{\mu}^{(N)})]$, es que

$$d([T_2 \circ T_1(\boldsymbol{\mu}^{(l)})], [T_2 \circ T_1(\boldsymbol{\mu}^{(k)})]) < \frac{\pi}{4}, \quad (2.24)$$

para todo $l, k \in \{1, \dots, N\}$, con $l \neq k$, donde la métrica $d(\cdot, \cdot)$ está acotada por $\sqrt{2}$. En este caso, las formas de $T_2 \circ T_1(\boldsymbol{\mu}^{(1)}), \dots, T_2 \circ T_1(\boldsymbol{\mu}^{(N)})$ están en una bola geodésica de radio $\frac{\pi}{8}$, al igual que su única media de Fréchet empírica.

Queda pendiente la obtención de una fórmula de cómputo para la media de Fréchet empírica. En esta línea, los resultados de Kendall et al. (1999) nos dicen que dicha media, denotada

$\hat{\boldsymbol{\mu}}_{frechet}$, satisface la siguiente expresión:

$$\hat{\boldsymbol{\mu}}_{frechet} = T_2(\boldsymbol{\mu}_f), \quad (2.25)$$

donde

$$\boldsymbol{\mu}_f = \frac{1}{N} \sum_{l=1}^N T(\boldsymbol{\mu}^{(l)}, \boldsymbol{\mu}_f). \quad (2.26)$$

En lo que sigue no se hará distinción entre $\boldsymbol{\mu}_{frechet}$ y $\boldsymbol{\mu}_f$, puesto que ambos objetos tienen la misma forma, y los denominaremos media de Fréchet empírica a posteriori. Su fórmula de cálculo será la siguiente:

$$\hat{\boldsymbol{\mu}}_{frechet} = \frac{1}{N} \sum_{l=1}^N T(\boldsymbol{\mu}^{(l)}, \hat{\boldsymbol{\mu}}_{frechet}). \quad (2.27)$$

Un aspecto interesante del objeto anterior es que, a parte de resolver el problema de falta de identificación del modelo estadístico subyacente, entrega un estadístico que toma en cuenta la estructura métrica del espacio de las formas planares. Más aún, si consideramos la función de pérdida

$$L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = d^2([\boldsymbol{\mu}_1], [\boldsymbol{\mu}_2]), \quad \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{C}^p, \quad (2.28)$$

entonces podemos interpretar la media de Fréchet empírica a posteriori como un estimador genuinamente Bayesiano, en el sentido que minimiza una función de pérdida.

El siguiente ejemplo muestra que a diferencia de la media a posteriori, la media de Fréchet empírica a posteriori puede ser un estimador razonable de la *forma* de $\boldsymbol{\mu}$.

Ejemplo 2.4 (Continuación ejemplo 2.2). *La media de Fréchet empírica de $\boldsymbol{\mu}^{(1)}$ y $\boldsymbol{\mu}^{(2)}$ está dada por*

$$\hat{\boldsymbol{\mu}}_{frechet} = \frac{1}{2}(T_3(\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}) + T_3(\boldsymbol{\mu}^{(2)}, \boldsymbol{\mu})) = \frac{1}{2}(\boldsymbol{\mu} + \boldsymbol{\mu}) = \boldsymbol{\mu}.$$

Así, $[\hat{\boldsymbol{\mu}}_{frechet}] = [\boldsymbol{\mu}]$. □

También es posible definir las medias de Fréchet asociadas a la medida de probabilidad

2.4. POST-PROCESAMIENTO Y MEDIA DE FRÉCHET A POSTERIORI

inducida sobre Σ_2^p , por la distribución a posteriori $\boldsymbol{\mu}|\mathbf{w}_1, \dots, \mathbf{w}_n$. Denominaremos a dichas medias, las medias de Fréchet a posteriori. Gracias a la separabilidad del espacio métrico (Σ_2^p, d) (heredada de la propiedad de compacidad subyacente), se tiene que existe una relación entre las medias de Fréchet a posteriori y sus versiones empíricas, dada por la ley de los grandes números de [Ziezold \(1977\)](#). Así, si denotamos por M_μ el conjunto de las medias de Fréchet a posteriori y $M(\boldsymbol{\mu}^{(1)}, \dots, \boldsymbol{\mu}^{(N)})$ su versión empírica, entonces

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{N=n}^{\infty} M(\boldsymbol{\mu}^{(1)}, \dots, \boldsymbol{\mu}^{(N)})} \subseteq M_\mu \quad \text{c.s.}$$

La utilidad práctica del resultado anterior es cuando M_μ es un singleton, ya que en este caso la media de Fréchet empírica a posteriori converge casi seguramente, cuando $N \rightarrow \infty$, a la media de Fréchet a posteriori. En el caso general, sólo se puede argumentar que la media de Fréchet empírica a posteriori converge a alguna de las medias de Fréchet a posteriori. Cabe destacar, en base a lo anterior, que el estimador propuesto en (2.27) es la aproximación de Monte Carlo de la media de Fréchet asociada a la distribución a posteriori de $\boldsymbol{\mu}$, que denominaremos media de Fréchet a posteriori y denotaremos $\boldsymbol{\mu}_{\text{frechet}}$. Por construcción, este estimador es Bayesiano, pues minimiza la función de pérdida dada en (2.28). En efecto,

$$\int_{\mathbb{C}^p} d_s^2(\boldsymbol{\mu}, \boldsymbol{\mu}_{\text{frechet}}) \Pi(\boldsymbol{\mu}|\mathbf{w}_1, \dots, \mathbf{w}_n) = \inf_{\mathbf{a} \in \mathbb{C}^p} \int_{\mathbb{C}^p} d_s^2(\boldsymbol{\mu}, \mathbf{a}) \Pi(\boldsymbol{\mu}|\mathbf{w}_1, \dots, \mathbf{w}_n),$$

donde

$$d_s(\boldsymbol{\mu}, \mathbf{a}) = d(T_2 \circ T_1(\boldsymbol{\mu}), T_2 \circ T_1(\mathbf{a})),$$

es una pseudo-métrica y $\Pi(\boldsymbol{\mu}|\mathbf{w}_1, \dots, \mathbf{w}_n)$ es la distribución a posteriori de $\boldsymbol{\mu}$.

Para el cálculo de la media de Fréchet empírica a posteriori de $\boldsymbol{\mu}^{(1)}, \dots, \boldsymbol{\mu}^{(N)}$ hay que resolver el problema de punto fijo dado en (2.27). La propuesta de este trabajo para la estimación

de la forma de μ está dada por

$$\hat{\mu} = \frac{1}{N} \sum_{l=1}^N T(\mu^{(l)}, \hat{\mu}_{GPA}), \quad (2.29)$$

donde $\hat{\mu}_{GPA}$ el estimador de Procrustes generalizado de μ . Los resultados computacionales del estudio de simulación de la Sección 2.5 indican que en muestras finitas, $\hat{\mu}$ es un mejor estimador de la forma de μ , que la media a posteriori. Más aún, nuestra propuesta es un estimador competitivo cuando se compara con el estimador de Procrustes generalizado.

La implementación computacional del esquema de post-estimación se presenta en el Algoritmo 2.2.

Algoritmo 2.2 Esquema de post-estimación

- 1: Calcular el estimador de Procrustes generalizado: $\hat{\mu}_{GPA}$
 - 2: **for** $l = 1, \dots, N$ **do**
 - 3: Eliminar traslación y escala para $\mu^{(l)}$: $T_2 \circ T_1(\mu^{(l)})$
 - 4: Eliminar traslación y escala para $\hat{\mu}_{GPA}$: $T_2 \circ T_1(\hat{\mu}_{GPA})$
 - 5: Eliminar rotación para $\mu^{(l)}$, definiendo: $\nu_l := T_3(T_2 \circ T_1(\mu^{(l)}), T_2 \circ T_1(\hat{\mu}_{GPA}))$
 - 6: **end for**
 - 7: **return** $\hat{\mu} = \frac{1}{N} \sum_{l=1}^N \nu_l$
-

El esquema de post-estimación también tiene un efecto en el resto de los parámetros de forma, esto es:

$$d_j^{(1)}, \dots, d_j^{(N)}, \quad j = 1, \dots, n,$$

$$b_j^{(1)}, \dots, b_j^{(N)}, \quad j = 1, \dots, n,$$

$$\phi_j^{(1)}, \dots, \phi_j^{(N)}, \quad j = 1, \dots, n.$$

Para $l = 1, \dots, N$ y $j = 1, \dots, n$ se define

$$\mathbf{m}_j^{(l)} = d_j^{(l)} \mathbf{1}_p + b_j^{(l)} \exp\{i\phi_j^{(l)}\} \mu^{(l)}. \quad (2.30)$$

2.5. ESTUDIO DE SIMULACIÓN

Para obtener una transformación similar a T para los parámetros de forma anteriores, basta con proyectar $\mathbf{m}_j^{(l)}$ sobre $\boldsymbol{\nu}_l := T(\boldsymbol{\mu}^{(l)}, \widehat{\boldsymbol{\mu}}_{GPA})$. Una forma sencilla realizar lo anterior es proyectando $T_1(\mathbf{m}_j^{(l)})$ sobre $\boldsymbol{\nu}_l$. Así,

$$T_1(\mathbf{m}_j^{(l)}) = \left\| \boldsymbol{\nu}_l^* T_1(\mathbf{m}_j^{(l)}) \right\| \exp \left\{ i \arg \left(\boldsymbol{\nu}_l^* T_1(\mathbf{m}_j^{(l)}) \right) \right\} \boldsymbol{\nu}_l, \quad (2.31)$$

de donde se obtiene que

$$\mathbf{m}_j^{(l)} = d_j^{(l)} \mathbf{1}_p + b_j^{(l)} \exp \{ i \phi_j^{(l)} \} \boldsymbol{\nu}_l, \quad (2.32)$$

con

$$d_j^{(l)} = h_1(d_j^{(l)}, b_j^{(l)}, \phi_j^{(l)}, \boldsymbol{\mu}^{(l)}, \widehat{\boldsymbol{\mu}}_{GPA}) := \left(\frac{1}{p} \mathbf{1}_p \mathbf{1}_p^t \right) \mathbf{m}_j^{(l)}, \quad (2.33)$$

$$b_j^{(l)} = h_2(d_j^{(l)}, b_j^{(l)}, \phi_j^{(l)}, \boldsymbol{\mu}^{(l)}, \widehat{\boldsymbol{\mu}}_{GPA}) := \left\| \boldsymbol{\nu}_l^* T_1(\mathbf{m}_j^{(l)}) \right\|, \quad (2.34)$$

$$\phi_j^{(l)} = h_3(d_j^{(l)}, b_j^{(l)}, \phi_j^{(l)}, \boldsymbol{\mu}^{(l)}, \widehat{\boldsymbol{\mu}}_{GPA}) := \arg \left(\boldsymbol{\nu}_l^* T_1(\mathbf{m}_j^{(l)}) \right). \quad (2.35)$$

Notemos que $d_j^{(l)}$ no depende solamente de $d_j^{(l)}$. Una situación similar ocurre para $b_j^{(l)}$ y $\phi_j^{(l)}$.

2.5 Estudio de simulación

En esta sección, a través de un estudio de simulación en R ([R Core Team \(2016\)](#)), se analiza el desempeño de la metodología propuesta en las Secciones 2.3 y 2.4, para la estimación de la *forma* media de una muestra de *configuraciones*. En todos los escenarios considerados se conoce el valor real del parámetro $\boldsymbol{\mu}$, y se evalúa la diferencia entre esta cantidad y algunos estimadores de forma, utilizando la distancia Riemanniana ρ implementada en la librería `shapes` de [Dryden \(2016\)](#), del paquete estadístico R. La distancia anterior está dada por

$$\rho(\cdot, \cdot) = 2 \arcsin \left(\frac{d(\cdot, \cdot)}{2} \right),$$

2.5. ESTUDIO DE SIMULACIÓN

con d definida en (2.23), y satisface que $0 \leq \rho \leq \frac{\pi}{2}$ (ver, por ejemplo, Dryden & Mardia, 1998).

Las muestras simuladas se generan de acuerdo a un modelo normal complejo

$$\mathbf{w}_j \stackrel{ind}{\sim} \mathbb{CN}_p(d_j \mathbf{1}_p + b_j e^{i\phi_j} \boldsymbol{\mu}, \sigma^2 \mathbf{I}_p), \quad (\text{Modelo 1})$$

para $j = 1, \dots, n$, y una mezcla de normales complejas

$$\mathbf{w}_j \stackrel{ind}{\sim} \frac{1}{2} \mathbb{CN}_p(d_j \mathbf{1}_p + b_j e^{i\phi_j} \boldsymbol{\mu}_1, \sigma^2 \mathbf{I}_p) + \frac{1}{2} \mathbb{CN}_p(d_j \mathbf{1}_p + b_j e^{i\phi_j} \boldsymbol{\mu}_2, \sigma^2 \mathbf{I}_p), \quad (\text{Modelo 2})$$

para $j = 1, \dots, n$, donde

$$\boldsymbol{\mu} = \frac{1}{2} \boldsymbol{\mu}_1 + \frac{1}{2} \boldsymbol{\mu}_2, \quad (2.36)$$

es un parámetro conocido común, para los modelos 1 y 2. Los parámetros $d_1, \dots, d_n \in \mathbb{C}$, $b_1, \dots, b_n \in \mathbb{R}_+$, $\phi_1, \dots, \phi_n \in [0, 2\pi)$, $\sigma^2 \in \mathbb{R}_+$ y $p, n \in \mathbb{N}$ son conocidos. Los parámetros $\boldsymbol{\mu}_1$ y $\boldsymbol{\mu}_2$ utilizados se muestran en las Figuras 2.1 y 2.2, y fueron generados de manera que (a) y (b) fueran las restricciones de (d) y (e), respectivamente. Adicionalmente, $\boldsymbol{\mu}$ fue estandarizado en términos de su localización y escala, de manera que corresponde a una *preforma*.

En las Tablas 2.1 y 2.2 se reportan los resultados de un estudio de simulación sobre 200 muestras generadas bajo los modelos de probabilidad 1 y 2, respectivamente. En cada celda se muestra el promedio y la desviación estándar de la distancia Riemanniana ρ entre $\boldsymbol{\mu}$ y los siguientes estimadores: promedio muestral $\bar{\mathbf{w}}$, media a posteriori $\hat{\boldsymbol{\mu}}_p$, estimador de Procrustes generalizado $\hat{\boldsymbol{\mu}}_{GPA}$, la propuesta de este trabajo $\hat{\boldsymbol{\mu}}$ y el estimador de Micheas & Peng (2010) $\hat{\boldsymbol{\mu}}_M$. Se consideran distintos escenarios de simulación con tamaños muestrales $n \in \{30, 100\}$, número de hitos $p \in \{15, 30\}$, varianzas $\sigma^2 \in \{0,01, 0,000001\}$ y dos formas distintas: número 3 y manos. Los valores escogidos de σ^2 garantizan que las bases de datos simuladas se parecen “visualmente” a las configuraciones medias reales cuando $\sigma^2 = 0,000001$ y no cuando $\sigma^2 = 0,01$. Respecto del esquema de simulación del MCMC propuesto en este trabajo, se generaron cadenas de largo 50.000 y se eliminaron las primeras 10.000 observaciones, tras lo

2.5. ESTUDIO DE SIMULACIÓN

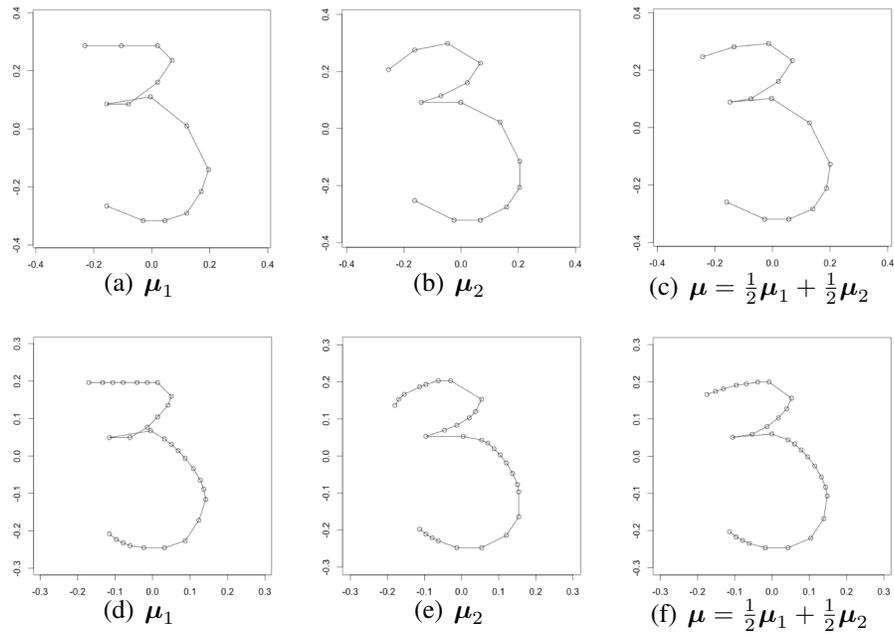


Figura 2.1: En (a), (b) y (c) *configuraciones* medias reales para 15 hits. En (d), (e) y (f) *configuraciones* medias reales para 30 hits.

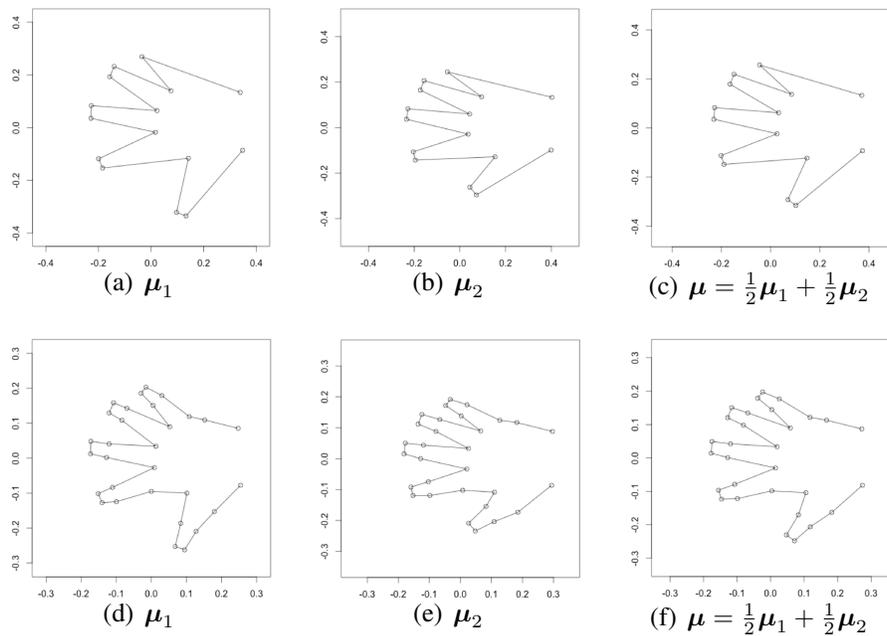


Figura 2.2: En (a), (b) y (c) *configuraciones* medias reales para 15 hits. En (d), (e) y (f) *configuraciones* medias reales para 30 hits.

2.5. ESTUDIO DE SIMULACIÓN

cual se verificó la estabilidad de las medias ergódicas. Los parámetros iniciales del MCMC fueron $\boldsymbol{\mu}_0 = \mathbf{0} \in \mathbb{C}^p$, $k_0^2 = 0,01$, $d_0 = 0 \in \mathbb{C}$, $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 100$. En relación al estimador propuesto en [Micheas & Peng \(2010\)](#), se generaron 1.000 cadenas de largo 1.000 y para cada una se eliminaron las primeras 200 observaciones. Los parámetros iniciales del esquema de simulación fueron $\boldsymbol{\mu}_0 = \mathbf{0} \in \mathbb{C}^p$, $k_0^2 = 0,01$, $d_0 = 0 \in \mathbb{C}$ y $\tau^2 = 0,01$. El parámetro σ^2 se calibró siguiendo las indicaciones de [Micheas & Peng \(2010\)](#) para cada escenario de simulación.

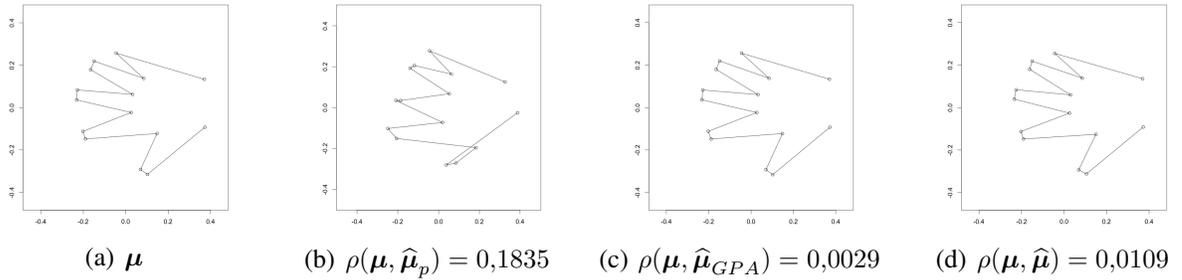


Figura 2.3: Para un número de hitos $p = 15$ y un tamaño muestral $n = 30$: en (a) la configuración media real, en (b) la media a posteriori, en (c) el estimador de Procrustes generalizado y en (d) la media de Frechet a posteriori.

De los resultados del estudio de simulación se desprenden varios hechos:

1. En general, $\hat{\boldsymbol{\mu}}_{GPA}$ posee el mejor rendimiento de los estimadores considerados. No obstante, aún en los escenarios en que $\hat{\boldsymbol{\mu}}_{GPA}$ es mejor que $\hat{\boldsymbol{\mu}}$ (por ejemplo, en el modelo 1 con $\sigma^2 = 0,000001$, $n = 30$ y $p = 15$ de la Tabla 2.2), la diferencia es apenas visualmente perceptible, como se aprecia en la Figura 2.3.
2. En todos los casos considerados $\hat{\boldsymbol{\mu}}$ tiene un mejor rendimiento que $\hat{\boldsymbol{\mu}}_p$. Más aún, $\hat{\boldsymbol{\mu}}_p$ es un muy mal estimador de $[\boldsymbol{\mu}]$ en algunos escenarios (por ejemplo, en los modelos 1 y 2 cuando $\sigma^2 = 0,01$). Así, la distribución a posteriori de $\boldsymbol{\mu}$ tiene más información de $[\boldsymbol{\mu}]$, que la incorporada por la media a posteriori.
3. La estimación de $[\boldsymbol{\mu}]$ es levemente mejor por parte de $\hat{\boldsymbol{\mu}}$ que de $\hat{\boldsymbol{\mu}}_{GPA}$, cuando $\sigma^2 = 0,01$, esto es, cuando la variabilidad es mayor, que es el caso de mayor interés, pues cuando

2.5. ESTUDIO DE SIMULACIÓN

Tabla 2.1: Comparación de estimadores en términos de sesgo y desviación estándar (número entre paréntesis) de la distancia Riemanniana ρ a la *forma* media $[\mu]$, para distintos tamaños muestrales y número de hitos. La *forma* de referencia es la del número 3 y los estimadores analizados son: promedio muestral \bar{w} , media a posteriori $\hat{\mu}_p$, estimador de Procrustes generalizado $\hat{\mu}_{GPA}$, media de Fréchet a posteriori $\hat{\mu}$ y el estimador de Micheas et al. (2010) $\hat{\mu}_M$.

n	30		100	
p	15	30	15	30
Modelo 1 con $\sigma^2 = 0,01$				
\bar{w}	0,2307(0,0344)	0,3540(0,0341)	0,5195(0,0769)	0,4152(0,0422)
$\hat{\mu}_p$	0,2505(0,0364)	0,4197(0,0380)	0,6223(0,0823)	0,4771(0,0453)
$\hat{\mu}_{GPA}$	0,0878(0,0126)	0,1279(0,0112)	0,0532(0,0075)	0,0773(0,0073)
$\hat{\mu}$	0,0713(0,0094)	0,1304(0,0126)	0,0495(0,0072)	0,0851(0,0090)
$\hat{\mu}_M$	0,0870(0,0119)	0,1318(0,0126)	0,0941(0,0128)	0,1045(0,0102)
Modelo 1 con $\sigma^2 = 0,000001$				
\bar{w}	0,0065(0,0009)	0,0123(0,0011)	0,0026(0,0004)	0,0141(0,0013)
$\hat{\mu}_p$	0,0686(0,0100)	0,0731(0,0070)	0,0186(0,0027)	0,0498(0,0050)
$\hat{\mu}_{GPA}$	0,0065(0,0009)	0,0090(0,0008)	0,0036(0,0005)	0,0053(0,0005)
$\hat{\mu}$	0,0120(0,0017)	0,0125(0,0013)	0,0063(0,0009)	0,0052(0,0005)
$\hat{\mu}_M$	0,0663(0,0100)	0,0807(0,0077)	0,0434(0,0059)	0,0797(0,0073)
Modelo 2 con $\sigma^2 = 0,01$				
\bar{w}	0,2658(0,0401)	0,2715(0,0268)	0,5079(0,0800)	0,4287(0,0395)
$\hat{\mu}_p$	0,3027(0,0440)	0,3122(0,0307)	0,6030(0,0849)	0,5705(0,0504)
$\hat{\mu}_{GPA}$	0,0953(0,0124)	0,1155(0,0110)	0,0532(0,0074)	0,0705(0,0066)
$\hat{\mu}$	0,0756(0,0116)	0,1167(0,0106)	0,0439(0,0064)	0,0633(0,0061)
$\hat{\mu}_M$	0,0947(0,0138)	0,1315(0,0120)	0,0871(0,0148)	0,1160(0,0109)
Modelo 2 con $\sigma^2 = 0,000001$				
\bar{w}	0,1100(0,0561)	0,1131(0,0613)	0,0762(0,0426)	0,0558(0,0289)
$\hat{\mu}_p$	0,2156(0,0766)	0,2219(0,0899)	0,1066(0,0502)	0,1097(0,0461)
$\hat{\mu}_{GPA}$	0,0167(0,0049)	0,0111(0,0067)	0,0077(0,0029)	0,0091(0,0029)
$\hat{\mu}$	0,0208(0,0076)	0,0303(0,0100)	0,0178(0,0069)	0,0209(0,0114)
$\hat{\mu}_M$	0,1278(0,0276)	0,1885(0,0419)	0,0562(0,0084)	0,0804(0,0077)

los datos están concentrados, sus *formas* se pueden aproximar por un espacio Euclidiano. Esto queda de manifiesto cuando $\sigma^2 = 0,000001$, pues $[\mu]$ puede ser estimada con poco error por \bar{w} .

2.5. ESTUDIO DE SIMULACIÓN

Tabla 2.2: Comparación de estimadores en términos de sesgo y desviación estándar (número entre paréntesis) de la distancia Riemanniana ρ a la *forma* media $[\mu]$, para distintos tamaños muestrales y número de hitos. La *forma* de referencia es la de una mano y los estimadores analizados son: promedio muestral \bar{w} , media a posteriori $\hat{\mu}_p$, estimador de Procrustes generalizado $\hat{\mu}_{GPA}$, media de Fréchet a posteriori $\hat{\mu}$ y el estimador de Micheas et al. (2010) $\hat{\mu}_M$.

n	30		100	
p	15	30	15	30
Modelo 1 con $\sigma^2 = 0,01$				
\bar{w}	0,3313(0,0480)	0,5516(0,0548)	0,2782(0,0408)	0,8102(0,0881)
$\hat{\mu}_p$	0,3482(0,0484)	0,5156(0,0482)	0,3252(0,0455)	1,1218(0,1067)
$\hat{\mu}_{GPA}$	0,0977(0,0136)	0,1297(0,0120)	0,0529(0,0070)	0,0708(0,0066)
$\hat{\mu}$	0,0819(0,0136)	0,1188(0,0118)	0,0502(0,0072)	0,0621(0,0060)
$\hat{\mu}_M$	0,1015(0,0150)	0,1590(0,0156)	0,0709(0,0098)	0,2012(0,0791)
Modelo 1 con $\sigma^2 = 0,000001$				
\bar{w}	0,0156(0,0022)	0,0033(0,0003)	0,0090(0,0013)	0,0039(0,0004)
$\hat{\mu}_p$	0,1568(0,0207)	0,0277(0,0027)	0,0895(0,0132)	0,0249(0,0025)
$\hat{\mu}_{GPA}$	0,0033(0,0005)	0,0056(0,0005)	0,0061(0,0009)	0,0034(0,0003)
$\hat{\mu}$	0,0122(0,0017)	0,0110(0,0010)	0,0069(0,0010)	0,0055(0,0006)
$\hat{\mu}_M$	0,0775(0,0110)	0,0742(0,0072)	0,4432(0,1148)	0,1002(0,0102)
Modelo 2 con $\sigma^2 = 0,01$				
\bar{w}	0,3001(0,0414)	0,4843(0,0461)	0,5482(0,0847)	0,4973(0,0517)
$\hat{\mu}_p$	0,3371(0,0449)	0,5006(0,0479)	0,5895(0,0838)	0,6133(0,0595)
$\hat{\mu}_{GPA}$	0,1212(0,0174)	0,1412(0,0139)	0,0590(0,0081)	0,0736(0,0077)
$\hat{\mu}$	0,1009(0,0155)	0,1481(0,0147)	0,0496(0,0083)	0,0689(0,0066)
$\hat{\mu}_M$	0,1123(0,0159)	0,1694(0,0165)	0,0941(0,0143)	0,1171(0,0117)
Modelo 2 con $\sigma^2 = 0,000001$				
\bar{w}	0,0818(0,0429)	0,0656(0,0347)	0,0732(0,0365)	0,1449(0,0783)
$\hat{\mu}_p$	0,1076(0,0452)	0,0797(0,0303)	0,1072(0,0494)	0,1686(0,0754)
$\hat{\mu}_{GPA}$	0,0124(0,0081)	0,0123(0,0062)	0,0075(0,0034)	0,0085(0,0037)
$\hat{\mu}$	0,0228(0,0088)	0,0272(0,0122)	0,0144(0,0071)	0,0159(0,0086)
$\hat{\mu}_M$	0,0623(0,0118)	0,0782(0,0102)	0,0531(0,0089)	0,0816(0,0086)

4. En todos los escenarios analizados, $\hat{\mu}$ tiene un mejor rendimiento que el estimador propuesto por [Micheas & Peng \(2010\)](#). Además, el procedimiento propuesto en este trabajo requiere simular sólo una cadena, por lo que es menos costoso computacionalmente.

5. El aumento en el número de hitos provoca, en general, un deterioro en todos los estimadores de *forma* comparados. Esto se explica por el aumento en la complejidad del modelo, expresado por el incremento en el número de parámetros.
6. La estimación de $[\mu]$ mejora con el aumento en el tamaño de muestra considerado, tanto para $\hat{\mu}$ como $\hat{\mu}_{GPA}$. Esto es de interés, pues se puede demostrar que el estimador de Procrustes generalizado es consistente para el modelo simulado, pero no se conoce el comportamiento límite de $\hat{\mu}$. Así, estos resultados sugieren un comportamiento cada vez mejor para $\hat{\mu}$, en la medida que se tomen más datos.

2.6 Conclusiones

En este capítulo se presentó un enfoque Bayesiano paramétrico para la estimación de la *forma* media, a partir de datos de *configuraciones*. Dicha metodología se obtuvo a través de un esquema de expansión de parámetros, imponiendo condiciones de identificabilidad sobre las muestras a posteriori del parámetro de interés. Además, el estimador propuesto considera en su construcción, la estructura métrica de la variedad Riemanniana subyacente, surgiendo de manera natural el concepto de la media de Fréchet empírica a posteriori. Más aún, el estimador propuesto se puede interpretar como un estimador genuinamente Bayesiano, debido a que minimiza una función de pérdida.

La metodología propuesta mejora el desempeño, en muestras finitas, de la media a posteriori para estimar la *forma* media, evidenciando así que el promedio no es capaz de recuperar eficientemente la información de *forma* contenida en la muestra a posteriori. Además, la media de Fréchet empírica a posteriori tiene un buen rendimiento cuando se compara incluso con el estimador de Procrustes generalizado, con la ventaja de permitir hacer inferencias sobre los parámetros de traslación, escala y rotación, y de la cuantificación de la incertidumbre en la estimación. Finalmente, la metodología propuesta puede ser extendida a otros modelos estadísticos

2.6. CONCLUSIONES

paramétricos o noparamétricos.

El trabajo de investigación futura, incluye (i) la extensión del procedimiento propuesto para datos de regresión, ya sea donde la respuesta es la *forma* o donde la información de *forma* es una covariable funcional del modelo, y (ii) la extensión del modelo paramétrico normal complejo, ya sea levantando el supuesto de varianza isotrópica, considerando un modelo de mezclas de normales complejas inducido por modelos Bayesianos noparamétricos o pasando del modelo planar al de hitos en tres dimensiones.

Procesos de Dirichlet Dependientes en Espacios Polacos

3.1 Introducción

En muchas situaciones prácticas, los datos están soportados en un espacio no Euclidiano. Algunos ejemplos surgen en morfometría ([Claude, 2008](#)), meteorología ([Mardia & Jupp, 2000](#)), arqueología ([Dryden & Mardia, 2016](#)) y genética ([Billera et al., 2001](#)). En estos contextos, considerar procedimientos estadísticos estándar, que no consideran las propiedades geométricas de los espacios subyacentes, puede llevar a inferencias estadísticas incorrectas. Debido a esto, ha existido un creciente interés por desarrollar procedimientos estadísticos en espacios más generales, tales como los espacios Polacos. Un espacio Polaco es un espacio topológico metrizable, completo y separable (ver, por ejemplo, [Bourbaki, 1998](#)). Algunos ejemplos de espacios

3.1. INTRODUCCIÓN

Polacos son los espacios de Hilbert separables, los espacios métricos compactos y las variedades Riemannianas completas.

Hasta la fecha, el desarrollo de procedimientos estadísticos para datos no Euclidianos se ha centrado en el problema de estimación de las medias (ver, por ejemplo, [Bhattacharya & Patrangenaru, 2002, 2003, 2005](#)), de estimación de densidades (ver, por ejemplo, [Pelletier, 2005](#); [Bhattacharya & Dunson, 2010, 2012b](#)) y en el problema de regresión para datos Euclidianos con predictores no Euclidianos (ver, por ejemplo, [Pelletier, 2006](#); [Bhattacharya & Dunson, 2012a](#)). [Bhattacharya & Patrangenaru \(2002, 2003, 2005\)](#) estudian el problema de estimación no paramétrica del parámetro de localización en una variedad Riemanniana, a través del concepto de media de Fréchet ([Fréchet, 1948](#)), y obtienen su distribución asintótica. [Pelletier \(2005\)](#) estudia el problema de estimación de densidades, adaptando técnicas de tipo *kernel*, sobre una variedad Riemanniana compacta. [Bhattacharya & Dunson \(2010, 2012b\)](#) estudian el problema de estimación de densidades soportadas sobre un espacio métrico compacto, a través de un modelo de mezclas y desde una perspectiva Bayesiana no paramétrica. Además, para el modelo propuesto, obtienen resultados de consistencia débil y fuerte. [Pelletier \(2006\)](#) estudia el problema de estimación no paramétrica de una función de regresión, con variable dependiente real y regresores soportados en una variedad Riemanniana cerrada. Finalmente, [Bhattacharya & Dunson \(2012a\)](#) estudian el problema de predicción de una variable categórica, con predictores soportados en una variedad general.

En este trabajo, el interés se centra en el problema de estimación de conjuntos de medidas de probabilidad con soporte en un espacio Polaco y relacionadas por predictores Euclidianos. Con este objeto, se generaliza el proceso de Dirichlet dependiente (DDP), y los modelos de mezclas inducidos por este, originalmente propuesto por [MacEachern \(1999, 2000\)](#), y estudiados posteriormente por [Barrientos et al. \(2012\)](#), en el contexto de espacios Euclidianos, para dar cuenta de respuestas soportadas en un espacio Polaco más general. Es importante destacar que la definición original del DDP dada por [MacEachern \(1999, 2000\)](#) y la definición alternativa dada por [Barrientos et al. \(2012\)](#) no se extienden de forma directa a espacios Polacos más

3.2. DEFINICIÓN DEL DDP

generales, debido a que estas construcciones hacen uso del concepto de función de distribución acumulada, sólo aplicable en espacios Euclidianos. En este trabajo, la existencia del DDP en espacios Polacos generales queda justificada por la extensión del teorema de existencia de Kolmogorov propuesta por [Neveu \(1965\)](#).

Adicionalmente, se establecen condiciones suficientes sobre la definición del DDP para que este tenga propiedades adecuadas de soporte, continuidad, estructura de asociación y consistencia de la distribución a posteriori bajo muestreo independiente. Además, se extienden los resultados de soporte reportados por [Barrientos et al. \(2012\)](#) y se caracteriza el soporte del proceso en topologías más fuertes, tomando algunas ideas de los resultados reportados por [Pati et al. \(2013\)](#), en el contexto de mezclas de densidades Gaussianas.

La estructura de este capítulo es la siguiente. En la Sección [3.2](#), se generaliza la definición del DDP al contexto de un espacio Polaco general y se definen algunas versiones más parsimoniosas. En la Sección [3.3](#) se estudian propiedades de continuidad, estructura de asociación y soporte del DDP, en el contexto de un espacio Polaco. En la Sección [3.4](#) se estudian propiedades soporte y consistencia débil para modelos de mezclas inducidos por un DDP, en el contexto de espacios métricos compactos. Finalmente, la Sección [3.5](#) contiene las conclusiones finales del trabajo.

3.2 Definición del DDP

Sea Θ un espacio Polaco y $\mathcal{P}(\Theta)$ el espacio de todas las medidas de probabilidad definidas sobre el espacio medible $(\Theta, \mathcal{B}(\Theta))$, donde $\mathcal{B}(\Theta)$ es la σ -álgebra de Borel de Θ . Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $\mathcal{X} \subseteq \mathbb{R}^q$. Estamos interesados en la definición de un DDP de la

3.2. DEFINICIÓN DEL DDP

forma

$$\begin{aligned} G_{\mathcal{X}} : \Omega &\rightarrow \mathcal{P}(\Theta)^{\mathcal{X}}, \\ \omega &\rightarrow G_{\mathcal{X}}(\omega) := \left\{ G_x(\omega)(\cdot) = \sum_{i=1}^{\infty} \pi_i(x, \omega) \delta_{\theta_i(x, \omega)}(\cdot) : x \in \mathcal{X} \right\}, \end{aligned}$$

donde $\mathcal{P}(\Theta)^{\mathcal{X}}$ es el espacio de todas las funciones de \mathcal{X} en $\mathcal{P}(\Theta)$,

$$\pi_i(x, \omega) = V_i(x, \omega) \prod_{j=1}^{i-1} [1 - V_j(x, \omega)],$$

con $V_i(x, \omega)$, $i \geq 1$, siendo procesos estocásticos con distribuciones marginales beta de parámetros $(1, \alpha_x)$, $\alpha_x \geq 0$, $\forall x \in \mathcal{X}$, y $\theta_i(x, \omega)$, $i \geq 1$, son procesos estocásticos definidos en Θ .

3.2.1 Definición general

La siguiente definición de un DDP generaliza de varias formas la construcción propuesta por [Barrientos et al. \(2012\)](#).

Definición 3.1. Sea $\alpha_{\mathcal{X}} := \{\alpha_x : x \in \mathcal{X}\} \subseteq \mathbb{R}_+$, $G_{\mathcal{X}}^0 := \{G_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^{\mathcal{X}}$ y $\{G_x : x \in \mathcal{X}\}$ un proceso estocástico definido sobre $(\Omega, \mathcal{F}, \mathbb{P})$, y a valores en $\mathcal{P}(\Theta)$, tal que:

- (a) $V_i : \mathcal{X} \times \Omega \rightarrow [0, 1]$, $i \geq 1$, son procesos estocásticos separables, independientes e idénticamente distribuidos, con ley caracterizada por un parámetro Ψ_V y con distribuciones marginales $Beta(1, \alpha_x)$, $x \in \mathcal{X}$.
- (b) $\theta_i : \mathcal{X} \times \Omega \rightarrow \Theta$, $i \geq 1$, son procesos estocásticos independientes e idénticamente distribuidos, con ley caracterizada por un parámetro Ψ_{θ} y con distribuciones marginales G_x^0 , $x \in \mathcal{X}$.

3.2. DEFINICIÓN DEL DDP

(c) Para todo $x \in \mathcal{X}$, $B \in \mathcal{B}(\Theta)$ y $\omega \in \Omega \setminus N$, con $\mathbb{P}(N) = 0$,

$$G_x(\omega)(B) := \sum_{i=1}^{\infty} \left\{ V_i(x, \omega) \prod_{j<i} [1 - V_j(x, \omega)] \right\} \delta_{\theta_i(x, \omega)}(B). \quad (3.1)$$

El proceso $G_{\mathcal{X}} := \{G_x : x \in \mathcal{X}\}$ será denominado proceso de Dirichlet dependiente de parámetros (Ψ_V, Ψ_{θ}) , y denotado $G_{\mathcal{X}} \sim \text{DDP}(\Psi_V, \Psi_{\theta})$.

En (a) de la Definición 3.1, la existencia de los procesos estocásticos V_1, V_2, \dots se puede garantizar por el teorema de consistencia de Kolmogorov y del uso de familias de cópulas (ver, por ejemplo, Barrientos et al., 2012). Específicamente, dada la colección de funciones de distribución acumulada de las distribuciones Beta(1, α_x), digamos $\{F_x : x \in \mathcal{X}\}$, podemos definir los procesos estocásticos $V_i : \mathcal{X} \times \Omega \rightarrow [0, 1]$, con distribuciones marginales deseadas, al especificar

$$\mathbb{P}\{\omega \in \Omega : V_i(x_1, \omega) \leq t_1, \dots, V_i(x_d, \omega) \leq t_d\} = C_{x_1, \dots, x_d}(F_{x_1}(t_1), \dots, F_{x_d}(t_d)),$$

para cualquier $t_1, \dots, t_d \in \mathbb{R}$, donde $\mathcal{C}_{\mathcal{X}}^V = \{C_{x_1, \dots, x_d} : x_1, \dots, x_d \in \mathcal{X}, d > 1\}$ es un conjunto de cópulas satisfaciendo las condiciones de consistencia de Kolmogorov. Específicamente, que para todo entero $d > 1$, $x_1, \dots, x_d \in \mathcal{X}$, $u_1, \dots, u_d \in [0, 1]$, $k \in \{1, \dots, d\}$ y una permutación cualquiera $\pi = (\pi_1, \dots, \pi_d)$ de $\{1, \dots, d\}$, se cumple que

$$C_{x_1, \dots, x_d}(u_1, \dots, u_d) = C_{x_{\pi_1}, \dots, x_{\pi_d}}(u_{\pi_1}, \dots, u_{\pi_d}),$$

$$C_{x_1, \dots, x_d}(u_1, \dots, u_{k-1}, 1, u_{k+1}, \dots, u_d) = C_{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_d}(u_1, \dots, u_{k-1}, u_{k+1}, \dots, u_d).$$

En (b) de la Definición 3.1, la existencia de los procesos estocásticos $\theta_1, \theta_2, \dots$ se puede garantizar por la generalización a espacios Polacos generales del teorema de consistencia de Kolmogorov, dada por Neveu (1965). La condición de separabilidad de los procesos, en este mismo punto, permite asegurar que el conjunto de los ω que hacen que el mapeo dado por la expresión (3.1) no sea una colección de medidas de probabilidad, sea un conjunto medible. En

3.2. DEFINICIÓN DEL DDP

efecto, en la representación casi segura dada en (3.1), el conjunto de medida nula depende de $x \in \mathcal{X}$ y está dado por

$$N_x = \left\{ \omega \in \Omega : \sum_{i=1}^{\infty} \left\{ V_i(x, \omega) \prod_{j<i} [1 - V_j(x, \omega)] \right\} < 1 \right\}.$$

Esto trae problemas en la representación (3.1) del DDP, puesto que se requiere que

$$\mathbb{P} \left(\bigcup_{x \in \mathcal{X}} N_x \right) = 0,$$

y el conjunto $\bigcup_{x \in \mathcal{X}} N_x$ no necesariamente pertenece a la σ -álgebra \mathcal{F} . Aquí, la condición de separabilidad de los procesos estocásticos $\{V_i(x, \omega) : x \in \mathcal{X}\}$, $i \geq 1$, garantiza que existen $x_l \in \mathcal{X}$, $l \geq 1$, tales que

$$\bigcup_{x \in \mathcal{X}} N_x = \bigcup_{l=1}^{\infty} N_{x_l},$$

y entonces $N := \bigcup_{x \in \mathcal{X}} N_x \in \mathcal{F}$. Luego,

$$\mathbb{P}(N) = \mathbb{P} \left(\bigcup_{x \in \mathcal{X}} N_x \right) = \mathbb{P} \left(\bigcup_{l=1}^{\infty} N_{x_l} \right) \leq \sum_{l=1}^{\infty} \mathbb{P}(N_{x_l}) = 0,$$

y para todo $\omega \in \Omega \setminus N$, $G_{\bullet}(\omega) \in \mathcal{P}(\Theta)^{\mathcal{X}}$. Finalmente, notemos también que, dado que $\mathcal{X} \subseteq \mathbb{R}^q$ es un conjunto separable, pues \mathbb{R}^q posee un subconjunto numerable denso, y $[0, 1] \subseteq \mathbb{R}$ es un conjunto compacto, entonces el Teorema 1, Capítulo 4, en [Gikhman & Skorokhod \(1969\)](#) asegura que siempre es posible encontrar versiones separables de los procesos estocásticos originales.

En relación a ejemplos de procesos estocásticos separables con marginales $\text{Beta}(1, \alpha_x)$, cualquier proceso estocástico con trayectorias continuas c.s. satisface la propiedad de separabilidad. Una manera de construir tales procesos, descrita por [MacEachern \(1999, 2000\)](#), se presenta en el siguiente ejemplo.

3.2. DEFINICIÓN DEL DDP

Ejemplo 3.1. Sea $\{Z_x : x \in \mathcal{X}\}$ un proceso estocástico a valores reales, con trayectorias continuas c.s. y función de distribución acumulada continua F_x , $x \in \mathcal{X}$. Sea B_x la función de distribución acumulada asociada a la Beta($1, \alpha_x$). Entonces

$$V_x(\omega) := B_x^{-1}(F_x(Z_x(\omega))), \quad x \in \mathcal{X},$$

es un proceso estocástico con marginales Beta($1, \alpha_x$) y trayectorias continuas c.s. En relación al proceso $\{Z_x : x \in \mathcal{X}\}$ se podría considerar, por ejemplo, un proceso de difusión o un proceso Gaussiano con trayectorias continuas.

3.2.2 Casos particulares

Es de interés buscar modelos más parsimoniosos que el propuesto en la Definición 3.1 y que posean propiedades adecuadas. En esta subsección se hacen dos simplificaciones respecto de la estructura de dependencia de los procesos estocásticos asociados a los pesos y átomos, respectivamente.

La siguiente versión del DDP simplifica la estructura de los procesos estocásticos separables, independientes e idénticamente distribuidos $\{V_i(x, \omega) : x \in \mathcal{X}\}$, $i \geq 1$, a variables aleatorias independientes e idénticamente distribuidas $V_1(\omega), V_2(\omega), \dots$, con ley Beta($1, \alpha$).

Definición 3.2. Sea $\alpha \in \mathbb{R}_+$, $G_{\mathcal{X}}^0 := \{G_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^{\mathcal{X}}$ y $\{G_x : x \in \mathcal{X}\}$ un proceso estocástico definido sobre $(\Omega, \mathcal{F}, \mathbb{P})$, y a valores en $\mathcal{P}(\Theta)$, tal que:

- (a) $V_i : \Omega \rightarrow [0, 1]$, $i \geq 1$, son variables aleatorias independientes e idénticamente distribuidas, con ley Beta($1, \alpha$).
- (b) $\theta_i : \mathcal{X} \times \Omega \rightarrow \Theta$, $i \geq 1$, son procesos estocásticos independientes e idénticamente distribuidos, con ley caracterizada por un parámetro Ψ_θ y con distribuciones marginales G_x^0 , $x \in \mathcal{X}$.

3.2. DEFINICIÓN DEL DDP

(c) Para todo $x \in \mathcal{X}$, $B \in \mathcal{B}(\Theta)$ y $\omega \in \Omega \setminus N$, con $\mathbb{P}(N) = 0$,

$$G_x(\omega)(B) := \sum_{i=1}^{\infty} \left\{ V_i(\omega) \prod_{j<i} [1 - V_j(\omega)] \right\} \delta_{\theta_i(x,\omega)}(B). \quad (3.2)$$

El proceso $G_{\mathcal{X}} := \{G_x : x \in \mathcal{X}\}$ será denominado proceso de Dirichlet dependiente con pesos comunes, de parámetros (α, Ψ_{θ}) , y denotado $G_{\mathcal{X}} \sim w\text{DDP}(\alpha, \Psi_{\theta})$.

En la definición anterior se verifica que para todo $\omega \in \Omega \setminus N$, con

$$N = \left\{ \omega \in \Omega : \sum_{i=1}^{\infty} \left\{ V_i(\omega) \prod_{j<i} [1 - V_j(\omega)] \right\} < 1 \right\} \in \mathcal{F},$$

y $\mathbb{P}(N) = 0$, $G_{\bullet}(\omega) \in \mathcal{P}(\Theta)^{\mathcal{X}}$.

La siguiente versión del DDP simplifica la estructura de los procesos estocásticos independientes $\theta_1(x, \omega), \theta_2(x, \omega), \dots$ a elementos aleatorios independientes $\theta_1(\omega), \theta_2(\omega), \dots$ con distribución común G^0 .

Definición 3.3. Sea $\alpha_{\mathcal{X}} := \{\alpha_x : x \in \mathcal{X}\} \subseteq \mathbb{R}_+$, $G^0 \in \mathcal{P}(\Theta)$ y $\{G_x : x \in \mathcal{X}\}$ un proceso estocástico definido sobre $(\Omega, \mathcal{F}, \mathbb{P})$, y a valores en $\mathcal{P}(\Theta)$, tal que:

(a) $V_i : \mathcal{X} \times \Omega \rightarrow [0, 1]$, $i \geq 1$, son procesos estocásticos separables, independientes e idénticamente distribuidos, con ley caracterizada por un parámetro Ψ_V y con distribuciones marginales $\text{Beta}(1, \alpha_x)$, $x \in \mathcal{X}$.

(b) $\theta_i : \Omega \rightarrow \Theta$, $i \geq 1$, son elementos aleatorios independientes e idénticamente distribuidos, con ley G^0 .

(c) Para todo $x \in \mathcal{X}$, $B \in \mathcal{B}(\Theta)$ y $\omega \in \Omega \setminus N$, con $\mathbb{P}(N) = 0$,

$$G_x(\omega)(B) := \sum_{i=1}^{\infty} \left\{ V_i(x, \omega) \prod_{j<i} [1 - V_j(x, \omega)] \right\} \delta_{\theta_i(\omega)}(B). \quad (3.3)$$

3.2. DEFINICIÓN DEL DDP

El proceso $G_{\mathcal{X}} := \{G_x : x \in \mathcal{X}\}$ será denominado proceso de Dirichlet dependiente con átomos comunes, de parámetros (Ψ_V, G^0) , y denotado $G_{\mathcal{X}} \sim \theta\text{DDP}(\Psi_V, G^0)$.

Respecto de la definición del DDP, y sus distintas versiones, la construcción de los procesos estocásticos $\{\theta_i(x, \omega) : x \in \mathcal{X}\}$, $i \geq 1$, puede llegar a ser una tarea complicada, dependiendo de la estructura que posea el espacio Θ . Lo anterior se puede complicar aún más, si es que se requieren de propiedades adicionales sobre el proceso, tales como la continuidad trayectorial. En este caso, se han propuesto algunas soluciones en el contexto de procesos de difusión, es decir, procesos Markovianos a trayectorias continuas.

Ejemplo 3.2. Sea Θ el espacio de las formas planares de [Kendall \(1977\)](#). Dicho conjunto puede ser dotado de una métrica, que le confiere el carácter de espacio métrico compacto (ver [Kendall, 1984](#)). Dada la naturaleza multivariada de Θ , no es claro cómo construir distribuciones en Θ^r , con $r \in \mathbb{N}$, a través de una familia Kolmogorov consistente. Al respecto, se han propuesto procesos de difusión desde dos enfoques diferentes: (a) directamente sobre cada hito, en el espacio de configuraciones, lo que se conoce como difusión Euclidiana de forma (ver, por ejemplo, [Kendall, 1977, 1988, 1990](#); [Le, 1991](#)), y (b) directamente sobre Θ , a través del análisis de generadores infinitesimales (ver, por ejemplo, [Le, 1994](#); [Kendall, 1998](#); [Ball et al., 2008](#); [Golalizadeh, 2010](#)). □

Una ventaja de trabajar con procesos de difusión, es que quedan completamente parametrizados por las funciones de deriva y el coeficiente de difusión, a través de ecuaciones diferenciales estocásticas (para una introducción de este tema ver, por ejemplo, [Karatzas & Shreve, 1991](#); [Oksendal, 2013](#)). En dicha construcción, la ecuación de Fokker-Plank relaciona la densidad de las distribuciones marginales del proceso con los coeficientes de deriva y difusión, lo que permite en principio construir difusiones con marginales dadas, vía la resolución de una ecuación en derivadas parciales (para detalles ver [Shiryayev, 1992](#)). Esto es de particular interés en el contexto de variedades Riemannianas, donde se puede definir una noción de derivada y, por

tanto, de ecuación diferencial (ver, por ejemplo, [Hsu, 2002](#)).

El punto (a) del Ejemplo 3.2 puede ser aplicado a una gran cantidad de variedades Riemannianas. Específicamente, algunas variedades pueden ser generadas vía el cuociente de un espacio de referencia localmente compacto \mathcal{Y} , por un subgrupo G de sus isometrías satisfaciendo que el conjunto $\{g \in G : gK \cap K \neq \emptyset\}$ sea finito, para todo compacto $K \subseteq \mathcal{Y}$. Esta condición técnica garantiza que el cuociente $\Theta = \mathcal{Y}/G$ sea una variedad. Sea $T : \mathcal{Y} \rightarrow \Theta$ la aplicación continua que a cada elemento del espacio de referencia \mathcal{Y} le asocia su clase de equivalencia. Suponiendo que somos capaces de construir un proceso estocástico $y(\cdot, \omega) \in \mathcal{Y}^{\mathcal{X}}$ con trayectorias continuas casi seguramente, esto es, salvo un conjunto $N \in \mathcal{F}$ con $\mathbb{P}(N) = 0$, entonces dado $\omega \in \Omega \setminus N$, se tiene que

$$\lim_{x \rightarrow x_0} T(y(x, \omega)) = T \left(\lim_{x \rightarrow x_0} y(x, \omega) \right) = T(y(x_0, \omega)),$$

y para el proceso estocástico $\theta(x, \omega) := T(y(x, \omega))$ queda garantizada su continuidad trayectorial casi seguramente. En el contexto del Ejemplo 3.2, T corresponde a la transformación que a cada configuración le asocia su forma (ver [Kendall, 1977](#)).

3.3 Propiedades del DDP

En esta sección se estudian propiedades de continuidad, estructura de asociación y soporte de un DDP.

3.3.1 Continuidad

Las propiedades de los procesos estocásticos que conforman el DDP, determinan importantes características de dicho proceso. En el siguiente resultado se muestra una propiedad de continuidad del DDP, bajo el supuesto de continuidad trayectorial de los procesos estocásticos $\{V_i(x, \omega) : x \in \mathcal{X}\}$ y $\{\theta_i(x, \omega) : x \in \mathcal{X}\}$, $i \geq 1$. Cabe destacar que propiedades de este

estilo son relevantes al momento de hacer inferencias, puesto que imponen una estructura de dependencia sobre las trayectorias del DDP.

Teorema 3.1. *Sea $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \text{DDP}(\Psi_V, \Psi_\theta)$, con (Ψ_V, Ψ_θ) tales que los procesos estocásticos $\{V_i(x, \omega) : x \in \mathcal{X}\}$ y $\{\theta_i(x, \omega) : x \in \mathcal{X}\}$, $i \geq 1$, tienen trayectorias continuas casi seguramente. Sean $x_0 \in \mathcal{X}$ y $\{x_n\}_{n \in \mathbb{N} \setminus \{0\}} \subseteq \mathcal{X}$. Entonces,*

$$\lim_{n \rightarrow \infty} x_n = x_0 \quad \Rightarrow \quad G_{x_n}(\omega) \xrightarrow{\mathcal{L}} G_{x_0}(\omega),$$

para todo $\omega \in \Omega \setminus N$, con $\mathbb{P}(N) = 0$.

La demostración del teorema anterior se encuentra en el Apéndice B.2 y es igual al caso Euclidiano, pues se basa en la caracterización de la convergencia débil de medidas de probabilidad, vistas como un subconjunto (convexo) del dual topológico del espacio de las funciones continuas y acotadas, la cual es válida en contexto de medidas de probabilidad soportadas en un espacio Polaco general.

En relación al Teorema 3.1, es de interés buscar condiciones que garanticen la existencia de procesos estocásticos $\{V_i(x, \omega) : x \in \mathcal{X}\}$ y $\{\theta_i(x, \omega) : x \in \mathcal{X}\}$, $i \geq 1$, con trayectorias continuas. Dado que $[0, 1]$ y Θ son espacios métricos completos, el Teorema 2.23 de Kallenberg (1997) entrega condiciones suficientes para la existencia de versiones continuas (casi seguramente) de los procesos estocásticos en cuestión. Estas condiciones son:

- Para $\{V(x) : x \in \mathcal{X}\}$: existen constantes positivas a_1, b_1, C_1 tales que

$$E(|V(x_1) - V(x_2)|^{a_1}) \leq C_1 \|x_1 - x_2\|^{b_1 + q}, \quad (3.4)$$

con $x_1, x_2 \in \mathcal{X} \subseteq \mathbb{R}^q$.

- Para $\{\theta(x) : x \in \mathcal{X}\}$: existen constantes positivas a_2, b_2, C_2 tales que

$$E([d_{\Theta}(\theta(x_1), \theta(x_2))]^{a_2}) \leq C_2 \|x_1 - x_2\|^{b_2 + a_2}, \quad (3.5)$$

con $x_1, x_2 \in \mathcal{X} \subseteq \mathbb{R}^q$, y donde d_{Θ} es una métrica en Θ .

Gracias al Teorema 2.3.1, Capítulo 5, en [Khoshnevisan \(2002\)](#) es posible generalizar la condición (3.4), bajo algunos supuestos adicionales, al caso en que \mathcal{X} es un espacio pseudométrico (o semimétrico) totalmente acotado. Sin embargo, este resultado no aplica para los procesos estocásticos $\{\theta_i(x, \omega) : x \in \mathcal{X}\}, i \geq 1$, cuando Θ es un espacio polaco general.

En el Ejemplo 3.1 se muestra una forma de construir procesos estocásticos con marginales Beta($1, \alpha_x$) y trayectorias continuas c.s., dado un proceso estocástico de referencia a valores reales y con trayectorias continuas c.s. Cuando dicho proceso de referencia es un proceso Gaussiano, las condiciones (3.4) y (3.5) caracterizan las propiedades que debe satisfacer la función de covarianza para que el proceso resultante posea trayectorias continuas c.s.

Una manera de construir procesos estocásticos con trayectorias continuas c.s. en el contexto que Θ es una variedad Riemanniana, es a través de la construcción de procesos de difusión. Para ejemplos ver [Hsu \(2002\)](#).

3.3.2 Estructura de asociación

En el siguiente teorema se obtiene una fórmula explícita para la función de covarianza del θ DDP. Además, se muestra que bajo ciertas condiciones sobre la estructura de dependencia de los procesos $\{V_i(x, \omega) : x \in \mathcal{X}\}, i \geq 1$, la correlación tiende a uno cuando los predictores se acercan.

Teorema 3.2. Sea $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \theta\text{DDP}(\Psi_V, G^0)$, con $\alpha_{\mathcal{X}} = \{\alpha_x : x \in \mathcal{X}\} \subseteq \mathbb{R}_+$ y $G^0 \in \mathcal{P}(\Theta)$. Entonces para todo $x, y \in \mathcal{X}$, $B \in \mathcal{B}(\Theta)$,

$$\text{Cov}(G_x(B), G_y(B)) = G^0(B) \left(\frac{r(x, y) + G^0(B) [m(x) - m(y) - r(x, y)]}{m(x) + m(y) - r(x, y)} \right),$$

donde

$$m(x) := E[V_1(x)] = \frac{1}{1 + \alpha_x}, \quad \forall x \in \mathcal{X},$$

$$r(x, y) := E[V_1(x)V_1(y)], \quad \forall x, y \in \mathcal{X}.$$

Más aún, si el parámetro Ψ_V es tal que las funciones $x \rightarrow \alpha_x$ y $(x, y) \rightarrow r(x, y)$ son continuas, entonces para todo $x_0 \in \mathcal{X}$, $B \in \mathcal{B}(\Theta)$,

$$\text{Corr}(G_x(B), G_{x_0}(B)) \rightarrow 1,$$

cuando $x \rightarrow x_0$.

La demostración del teorema anterior se encuentra en el apéndice B.3 y toma algunas ideas de la demostración de la Proposición 2.1 de Gutiérrez et al. (2016). A pesar que G_x es una medida de probabilidad aleatoria soportada sobre un espacio Polaco Θ , para cualquier evento $B \in \mathcal{B}(\Theta)$ se cumple que $G_x(B)$ es una variable aleatoria a valores en el intervalo $[0, 1]$. De esta manera, el argumento de la demostración anterior es análogo al caso Euclidiano. La estrategia de demostración utilizada no puede ser aplicada al DDP sin supuestos adicionales, pues en este caso se necesita modelar la estructura de dependencia de las leyes de los procesos estocásticos $\{\theta_i(x) : x \in \mathcal{X}\}$, $i \geq 1$. En particular, hay que suponer que la función $x \rightarrow G_x^0(B)$ es continua, $\forall B \in \mathcal{B}(\Theta)$. Esto queda como trabajo de investigación futuro.

3.3.3 Propiedades de soporte

En esta sección se analizan propiedades de soporte del DDP. Dado un espacio topológico $(\mathcal{T}, \mathcal{B})$ y un elemento aleatorio $\mathbb{T} : \Omega \rightarrow \mathcal{T}$, el soporte $\text{supp}(\mathbb{T}) \subseteq \mathcal{T}$ corresponde al cerrado más pequeño (en términos de la inclusión), con respecto a la topología \mathcal{B} , tal que $\mathbb{P}\{\omega \in \Omega : \mathbb{T}(\omega) \in M\} = 1$.

Supongamos que el espacio Polaco de interés está premunido de una medida Boreliana localmente finita, estrictamente positiva y no atómica, λ_Θ . En esta sección se analizan propiedades de soporte topológico para el DDP y el θ DDP, bajo distintas topologías sobre $\mathcal{P}(\Theta)^\mathcal{X}$. Cabe destacar, que la separabilidad de Θ garantiza la existencia del soporte para medidas de probabilidad suportadas en Θ (ver, por ejemplo, [Ghosh & Ramamoorthi, 2003](#)).

En el siguiente resultado se dan condiciones suficientes para que $\mathcal{P}(\Theta_0)^\mathcal{X}$ sea el soporte del DDP, cuando dotamos a $\mathcal{P}(\Theta)^\mathcal{X}$ de la topología producto de la convergencia débil de medidas de probabilidad, y donde Θ_0 es el soporte común de $\{G_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^\mathcal{X}$.

Teorema 3.3. *Sea $G_\mathcal{X} = \{G_x : x \in \mathcal{X}\} \sim \text{DDP}(\Psi_V, \Psi_\theta)$, con $\alpha_\mathcal{X} = \{\alpha_x : x \in \mathcal{X}\} \subseteq \mathbb{R}_+$ y $G_\mathcal{X}^0 = \{G_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^\mathcal{X}$. Supongamos que los parámetros (Ψ_V, Ψ_θ) son tales que:*

- (i) $\forall x \in \mathcal{X}, \text{supp}(G_x^0) = \Theta_0$.
- (ii) $\forall (x_1, \dots, x_d) \in \mathcal{X}^d$, con $d \geq 1$, la distribución conjunta de $(V_i(x_1, \omega), \dots, V_i(x_d, \omega))$ tiene soporte completo en $[0, 1]^d$.
- (iii) $\forall (x_1, \dots, x_d) \in \mathcal{X}^d$, con $d \geq 1$, la distribución conjunta de $(\theta_i(x_1, \omega), \dots, \theta_i(x_d, \omega))$ tiene soporte completo en Θ_0^d .

Entonces $\mathcal{P}(\Theta_0)^\mathcal{X}$ es el soporte producto débil de $G_\mathcal{X}$.

La demostración del teorema anterior extiende las ideas de la demostración del Teorema 2.1 de [Barrientos et al. \(2012\)](#) al contexto de un espacio Polaco y se encuentra en el apéndice B.4.

3.3. PROPIEDADES DEL DDP

Desde un punto de vista más técnico, la clave de dicha extensión es la validez del teorema de Portmanteau en el contexto de un espacio Polaco y que dichos espacios son metrizablees.

En la búsqueda de versiones más parsimoniosas del DDP, que satisfagan adecuadas propiedades, es de interés analizar la propiedad de soporte producto débil para el θ DDP. Al respecto, en el siguiente teorema se dan condiciones suficientes para que el θ DDP tenga como soporte al conjunto $\mathcal{P}(\Theta_0)^{\mathcal{X}}$.

Teorema 3.4. *Sea $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \theta\text{DDP}(\Psi_V, G^0)$, con $\alpha_{\mathcal{X}} = \{\alpha_x : x \in \mathcal{X}\} \subseteq \mathbb{R}_+$ y $G^0 \in \mathcal{P}(\Theta)$. Supongamos que el parámetro Ψ_V es tal que:*

- (i) $\text{supp}(G^0) = \Theta_0$.
- (ii) $\forall (x_1, \dots, x_d) \in \mathcal{X}^d$, con $d \geq 1$, la distribución conjunta de $(V_i(x_1, \omega), \dots, V_i(x_d, \omega))$ tiene soporte completo en $[0, 1]^d$.

Entonces $\mathcal{P}(\Theta_0)^{\mathcal{X}}$ es el soporte producto débil de $G_{\mathcal{X}}$.

La demostración del teorema anterior extiende las ideas de la demostración del Teorema 2.3 de [Barrientos et al. \(2012\)](#) al contexto de un espacio Polaco y se encuentra en el apéndice B.5. Además, la demostración es lo suficientemente similar a la prueba del Teorema 3.3, como para que apliquen los mismos comentarios técnicos. Por otra parte, el Teorema 3.4 establece que, en términos de la propiedad de soporte producto débil, no se pierde nada en trabajar con un θ DDP en vez de un DDP. Esta situación ya había sido reportada, pero sólo en el contexto de espacios Euclidianos. Finalmente, es posible obtener resultados similares a los Teoremas 3.3 y 3.4, para el caso del w DDP (ver definición 3.2) y para procesos stick-breaking dependientes más generales. Esto queda como trabajo de investigación futuro.

En lo que sigue, estamos interesados en caracterizar nociones de soporte más fuertes, para lo cual hay que dotar al conjunto $\mathcal{P}(\Theta)^{\mathcal{X}}$ de topologías más fuertes. Así, en el siguiente teorema

3.3. PROPIEDADES DEL DDP

se caracteriza una propiedad de soporte de tipo \mathbb{L}_1 -débil para el θ DDP, bajo el supuesto que el espacio \mathcal{X} sea un conjunto compacto.

Teorema 3.5. *Sea $\mathcal{X} \subseteq \mathbb{R}^q$ un conjunto compacto, $q \in \mathcal{D}(\mathcal{X})$ y $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \theta$ DDP(Ψ_V, G^0), con $\alpha_{\mathcal{X}} = \{\alpha_x : x \in \mathcal{X}\} \subseteq \mathbb{R}_+$ y $G^0 \in \mathcal{P}(\Theta)$. Supongamos que el parámetro Ψ_V es, tal que:*

- (i) $\text{supp}(G^0) = \Theta$.
- (ii) $\{V_i(x, \omega) : x \in \mathcal{X}\}$, $i \geq 1$, son procesos estocásticos con trayectorias continuas casi seguramente.
- (iii) Para toda función continua $g : \mathcal{X} \rightarrow [0, 1]$,

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |V_i(x, \omega) - g(x)| < \epsilon \right\} > 0,$$

para $i \geq 1$ y $\forall \epsilon > 0$.

Sea $\mathcal{G}_{\mathcal{X}}^* \subseteq \mathcal{P}(\Theta)^{\mathcal{X}}$ una colección de medidas de probabilidad dependientes de predictores satisfaciendo:

- (iv) $x \rightarrow P_x(B)$ continua, $\forall B \in \mathcal{B}(\Theta)$.
- (v) Para toda sucesión $\{A_n\} \subseteq \mathcal{B}(\Theta)$, con $A_n \downarrow \emptyset$, se tiene que

$$\sup_{x \in \mathcal{X}} P_x(A_n) \downarrow 0.$$

Entonces para cualquier $\{P_x : x \in \mathcal{X}\} \in \mathcal{G}_{\mathcal{X}}^*$ y $\epsilon > 0$,

$$\mathbb{P} \left\{ \omega \in \Omega : \left| \int_{\mathcal{X}} \left(\int_{\Theta} g(\theta, x) G_x(\omega)(d\theta) - \int_{\Theta} g(\theta, x) P_x(d\theta) \right) q(x) dx \right| < \epsilon \right\} > 0,$$

donde $g \in \mathcal{C}_b(\Theta \times \mathcal{X})$ es una función continua y acotada, y $q(\cdot)$ es la función de densidad de probabilidad generadora de los predictores en \mathcal{X} .

3.3. PROPIEDADES DEL DDP

La demostración de este teorema se encuentra en el apéndice B.6 y adapta las ideas de la demostración del Teorema 5.6 de Pati et al. (2013), quienes trabajan con procesos probit stick-breaking dependientes en espacios Euclidianos, al contexto de un DDP en un espacio Polaco. Desde un punto de vista más técnico, los argumentos que permiten la extensión son el concepto de tensión de medidas de probabilidad, el teorema de Dini y el teorema de Egoroff, y su validez en el contexto de un espacio Polaco general.

De acuerdo con Pati et al. (2013), si se reemplaza el intervalo $[0, 1]$ por \mathbb{R} , entonces un proceso estocástico Gaussiano con función de media nula y kernel de covarianza $c_j(x, x') \propto \exp\{-A_j\|x - x'\|^2\}$, donde A_j tiene soporte completo sobre \mathbb{R}_+ , satisface la condición (iii) del Teorema 3.5. Usando dicho proceso y la construcción dada en el Ejemplo 3.1, es posible construir procesos estocásticos $\{V_i(x, \omega) : x \in \mathcal{X}\}$, $i \geq 1$, que cumplan las condiciones del Teorema 3.5.

En el siguiente teorema se caracteriza una propiedad de soporte de tipo \mathbb{L}_∞ -débil para el θ DDP, de nuevo, bajo el supuesto que el espacio \mathcal{X} sea un conjunto compacto.

Teorema 3.6. *Sea $\mathcal{X} \subseteq \mathbb{R}^q$ un conjunto compacto, $q \in \mathcal{D}(\mathcal{X})$ y $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \theta$ DDP(Ψ_V, G^0), con $\alpha_{\mathcal{X}} = \{\alpha_x : x \in \mathcal{X}\} \subseteq \mathbb{R}_+$ y $G^0 \in \mathcal{P}(\Theta)$. Supongamos que el parámetro Ψ_V es, tal que:*

- (i) $\text{supp}(G^0) = \Theta$.
- (ii) $\{V_i(x, \omega) : x \in \mathcal{X}\}$, $i \geq 1$, son procesos estocásticos con trayectorias continuas casi seguramente.
- (iii) Para toda función continua $g : \mathcal{X} \rightarrow [0, 1]$,

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |V_i(x, \omega) - g(x)| < \epsilon \right\} > 0,$$

para $i \geq 1$ y $\forall \epsilon > 0$.

Sea $\mathcal{G}_{\mathcal{X}}^* \subseteq \mathcal{P}(\Theta)^{\mathcal{X}}$ una colección de medidas de probabilidad dependientes de predictores satisfaciendo:

(iv) $x \rightarrow P_x(B)$ continua, $\forall B \in \mathcal{B}(\Theta)$.

(v) Para toda sucesión $\{A_n\} \subseteq \mathcal{B}(\Theta)$, con $A_n \downarrow \emptyset$, se tiene que

$$\sup_{x \in \mathcal{X}} P_x(A_n) \downarrow 0.$$

Entonces para cualquier $\{P_x : x \in \mathcal{X}\} \in \mathcal{G}_{\mathcal{X}}^*$ y $\epsilon > 0$,

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \left| \int_{\Theta} g(\theta) G_x(\omega)(d\theta) - \int_{\Theta} g(\theta) P_x(d\theta) \right| < \epsilon \right\} > 0,$$

donde $g \in \mathcal{C}_0(\Theta)$ es una función continua y acotada que se anula en infinito, y $q(\cdot)$ es la densidad de probabilidad generadora de los predictores en \mathcal{X} .

La demostración de este teorema se encuentra en el apéndice B.7 y adapta las ideas de la demostración del teorema 6.1 de [Pati et al. \(2013\)](#), quienes trabajan con procesos probit stick-breaking dependientes en espacios Euclidianos, al contexto de un DDP en un espacio Polaco general. Finalmente, la demostración del teorema anterior es lo suficientemente parecida a la demostración del Teorema 3.5, como para que apliquen los mismos comentarios técnicos.

3.4 Mezclas inducidas por un DDP

En esta sección estamos interesados en estudiar modelos para funciones de densidad de probabilidad dependientes de predictores. Debido a la naturaleza discreta del DDP, este no puede ser aplicado directamente a dicho problema.

3.4.1 Definición

Sea \mathcal{Y} un espacio métrico compacto, premunido de una medida estrictamente positiva, finita y no atómica, $\lambda_{\mathcal{Y}}$, y $\mathcal{D}(\mathcal{Y})$ el espacio de todas las funciones de densidad sobre \mathcal{Y} , con respecto a la medida $\lambda_{\mathcal{Y}}$. Queremos definir de una manera flexible, un modelo para densidades relacionadas por covariables. Sea $\psi(y, \mu, \sigma)$ una densidad de probabilidad sobre \mathcal{Y} , con localización $\mu \in \mathcal{Y}$ y escala $\sigma \in \mathbb{R}_+$, y $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\}$ un DDP (o algunas de sus variantes). Vamos a estudiar dos tipos de modelos de mezclas:

$$f(y|x, G_x(\omega)) = \int_{\mathcal{Y}} \psi(y, \mu, \sigma) G_x(\omega)(d\mu), \quad (3.6)$$

el cual corresponde a un modelo de mezclas de localizaciones, y

$$f(y|x, G_x(\omega)) = \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta), \quad (3.7)$$

con $\theta = (\mu, \sigma) \in \Theta := \mathcal{Y} \times \mathbb{R}_+$, el cual corresponde a un modelo de mezclas de localizaciones y escala. En el modelo (3.6), una distribución a priori sobre $\mathcal{D}(\mathcal{Y})^{\mathcal{X}}$ es inducida vía una distribución a priori $(\{G_x : x \in \mathcal{X}\}, \sigma) \sim \Pi_1$. En el modelo (3.7), una distribución a priori sobre $\mathcal{D}(\mathcal{Y})^{\mathcal{X}}$ es inducida vía una distribución a priori $\{G_x : x \in \mathcal{X}\} \sim \Pi_2$.

3.4.2 Propiedades de soporte

En esta sección estamos interesados en determinar propiedades de soporte para modelos de mezclas inducidos por un DDP. Al respecto, en el siguiente teorema se dan condiciones suficientes bajo las cuales un modelo de mezclas del tipo (3.6) simplificado, esto es, donde no hay parámetros de escala o el parámetro de escala es conocido, tiene un soporte producto Hellinger grande.

Teorema 3.7. *Supongamos que:*

3.4. MEZCLAS INDUCIDAS POR UN DDP

- (i) $\psi(\cdot, \mu) \in \mathcal{D}(\mathcal{Y}), \forall \mu \in \mathcal{Y}$.
- (ii) $(y, \mu) \rightarrow \psi(y, \mu)$ es una función continua sobre $\mathcal{Y} \times \mathcal{Y}$.
- (iii) $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \text{DDP}(\Psi_V, \Psi_\theta)$, con (Ψ_V, Ψ_θ) tales que $\mathcal{P}(\mathcal{Y})^{\mathcal{X}}$ sea su soporte producto débil.

Entonces el soporte producto Hellinger del proceso

$$\left\{ \int_{\mathcal{Y}} \psi(\cdot, \mu) G_x(\omega)(d\mu) : x \in \mathcal{X} \right\},$$

es el conjunto

$$\prod_{x \in \mathcal{X}} \left\{ \int_{\mathcal{Y}} \psi(\cdot, \mu) P_x(d\mu) : P_x \in \mathcal{P}(\mathcal{Y}) \right\}.$$

El teorema anterior sigue siendo válido si se cambia el DDP de la condición (iii), por un w DDP o un θ DDP que tenga como soporte producto débil al conjunto $\mathcal{P}(\mathcal{Y})^{\mathcal{X}}$. La demostración del teorema anterior se encuentra en el apéndice B.8 y sigue argumentos muy similares a la demostración del Teorema 2.4 de Barrientos et al. (2012). Sin embargo, en nuestro caso los espacios de interés no son necesariamente Euclidianos y los supuestos sobre el *kernel* ψ son diferentes. Específicamente, se reemplaza el supuesto de equicontinuidad de la familia de funciones $\{K_1 \ni \mu \rightarrow \psi(y, \mu) : y \in K_2\}$, con $K_1, K_2 \subseteq \mathcal{Y}$ compactos, por la continuidad conjunta de (ii). Desde un punto de vista técnico, la demostración del teorema anterior se basa en la extensión del teorema de Arzelá-Ascoli para funciones definidas sobre un espacio métrico compacto.

En el siguiente teorema se dan condiciones bajo las cuales un modelo de mezclas del tipo (3.7), tiene un soporte producto Hellinger grande.

Teorema 3.8. *Supongamos que:*

- (i) $\psi(\cdot, \mu, \sigma) \in \mathcal{D}(\mathcal{Y}), \forall \mu \in \mathcal{Y}, \sigma \in \mathbb{R}_+$.

3.4. MEZCLAS INDUCIDAS POR UN DDP

- (ii) $(y, \mu, \sigma) \rightarrow \psi(y, \mu, \sigma)$ es una función continua sobre $\mathcal{Y} \times \mathcal{Y} \times \mathbb{R}_+$.
- (iii) $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \text{DDP}(\Psi_V, \Psi_\theta)$, con (Ψ_V, Ψ_θ) tales que $\mathcal{P}(\mathcal{Y} \times \mathbb{R}_+)^{\mathcal{X}}$ sea su soporte producto débil.

Entonces el soporte producto Hellinger del proceso

$$\left\{ \int_{\mathcal{Y} \times \mathbb{R}_+} \psi(\cdot, \theta) G_x(\omega)(d\theta) : x \in \mathcal{X} \right\},$$

es el conjunto

$$\prod_{x \in \mathcal{X}} \left\{ \int_{\mathcal{Y} \times \mathbb{R}_+} \psi(\cdot, \theta) P_x(d\theta) : P_x \in \mathcal{P}(\mathcal{Y} \times \mathbb{R}_+) \right\},$$

donde $\theta = (\mu, \sigma) \in \mathcal{Y} \times \mathbb{R}_+$.

El teorema anterior sigue siendo válido si se cambia el DDP de la condición (iii), por un w DDP o un θ DDP que tenga como soporte producto débil al conjunto $\mathcal{P}(\mathcal{Y} \times \mathbb{R}_+)^{\mathcal{X}}$. La demostración del teorema anterior se encuentra en el apéndice B.9 y es muy similar a la demostración del Teorema 3.7, salvo porque ahora el conjunto Θ no es compacto, lo que lleva a ciertas dificultades técnicas para replicar la demostración de dicho teorema. Desde un punto de vista técnico, la clave de la demostración del teorema anterior es la posibilidad de utilizar argumentos de compacidad local para Θ .

En el siguiente teorema se dan condiciones bajo las cuales un modelo de mezclas del tipo (3.6), tiene un soporte producto Hellinger grande.

Teorema 3.9. *Supongamos que:*

- (i) $\psi(\cdot, \mu, \sigma) \in \mathcal{D}(\mathcal{Y})$, $\forall \mu \in \mathcal{Y}, \sigma \in \mathbb{R}_+$.
- (ii) $(y, \mu, \sigma) \rightarrow \psi(y, \mu, \sigma)$ es una función continua sobre $\mathcal{Y} \times \mathcal{Y} \times \mathbb{R}_+$.
- (iii) $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \text{DDP}(\Psi_V, \Psi_\theta)$, con (Ψ_V, Ψ_θ) tales que $\mathcal{P}(\mathcal{Y})^{\mathcal{X}}$ sea su soporte producto débil.

3.4. MEZCLAS INDUCIDAS POR UN DDP

(iv) $\sigma \sim \pi_\sigma$ con soporte completo sobre \mathbb{R}_+ .

Entonces el soporte producto Hellinger del proceso

$$\left\{ \int_{\mathcal{Y}} \psi(\cdot, \mu, \sigma(\omega)) G_x(\omega)(d\mu) : x \in \mathcal{X} \right\},$$

es el conjunto

$$\prod_{x \in \mathcal{X}} \left\{ \int_{\mathcal{Y}} \psi(\cdot, \mu, \sigma) P_x(d\mu) : P_x \in \mathcal{P}(\mathcal{Y}), \sigma \in \mathbb{R}_+ \right\}.$$

El teorema anterior sigue siendo válido si se cambia el DDP de la condición (iii), por un w DDP o un θ DDP que tenga como soporte producto débil al conjunto $\mathcal{P}(\mathcal{Y})^{\mathcal{X}}$. La demostración del teorema anterior se encuentra en el apéndice B.10 y sigue argumentos similares a las demostraciones de los Teoremas 3.7 y 3.8, salvo porque ahora hay que considerar una distribución de probabilidad para σ . Desde un punto de vista técnico, la demostración de este teorema se basa en una extensión del teorema de Arzelá-Ascoli a espacios métricos separables y la propiedad de compacidad local de \mathbb{R} .

En lo que sigue, estamos interesados en estudiar nociones de soporte más fuertes, para lo cual hay que dotar al conjunto $\mathcal{D}(\mathcal{Y})^{\mathcal{X}}$ de topologías más fuertes. Así, en el siguiente teorema se dan condiciones bajo las cuales un modelo de mezclas inducido por un DDP, tiene un soporte producto \mathbb{L}_∞ grande.

Teorema 3.10. *Supongamos que:*

- (i) $\psi(\cdot, \mu, \sigma) \in \mathcal{D}(\mathcal{Y}), \forall \mu \in \mathcal{Y}, \sigma \in \mathbb{R}_+$.
- (ii) $(y, \mu, \sigma) \rightarrow \psi(y, \mu, \sigma)$ es una función continua sobre $\mathcal{Y} \times \mathcal{Y} \times \mathbb{R}_+$.
- (iii) $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \text{DDP}(\Psi_V, \Psi_\theta)$, con (Ψ_V, Ψ_θ) tales que $\mathcal{P}(\mathcal{Y})^{\mathcal{X}}$ sea su soporte producto débil.

3.4. MEZCLAS INDUCIDAS POR UN DDP

(iv) $\sigma \sim \pi_\sigma$ con soporte completo sobre \mathbb{R}_+ .

Entonces el soporte producto \mathbb{L}_∞ del proceso

$$\left\{ \int_{\mathcal{Y}} \psi(\cdot, \mu, \sigma(\omega)) G_x(\omega)(d\mu) : x \in \mathcal{X} \right\},$$

es el conjunto

$$\prod_{x \in \mathcal{X}} \left\{ \int_{\mathcal{Y}} \psi(\cdot, \mu, \sigma) P_x(d\mu) : P_x \in \mathcal{P}(\mathcal{Y}), \sigma \in \mathbb{R}_+ \right\}.$$

El teorema anterior sigue siendo válido si se cambia el DDP de la condición (iii), por un w DDP o un θ DDP que tenga como soporte producto débil al conjunto $\mathcal{P}(\mathcal{Y})^{\mathcal{X}}$. La demostración del teorema anterior se encuentra en el apéndice B.11 y es muy similar a la demostración del Teorema 3.9. Desde un punto de vista técnico, la demostración de este teorema se basa en el hecho que las funciones continuas definidas sobre espacios métricos compactos alcanzan su máximo, más la propiedad de compacidad local de \mathbb{R} .

En el siguiente teorema se dan condiciones bajo las cuales un modelo de mezclas inducido por un DDP, tiene un soporte producto Kullback-Leibler (KL) grande.

Teorema 3.11. *Supongamos que:*

- (i) $\psi(\cdot, \mu, \sigma) \in \mathcal{D}(\mathcal{Y}), \forall \mu \in \mathcal{Y}, \sigma \in \mathbb{R}_+$.
- (ii) $(y, \mu, \sigma) \rightarrow \psi(y, \mu, \sigma)$ es una función continua sobre $\mathcal{Y} \times \mathcal{Y} \times \mathbb{R}_+$.
- (iii) $\psi(y, \mu, \sigma) > 0, \forall y \in \mathcal{Y}, \mu \in \mathcal{Y}, \sigma \in \mathbb{R}_+$.
- (iv) $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \text{DDP}(\Psi_V, \Psi_\theta)$, con (Ψ_V, Ψ_θ) tales que $\mathcal{P}(\mathcal{Y})^{\mathcal{X}}$ sea su soporte producto débil.
- (v) $\sigma \sim \pi_\sigma$ con soporte completo sobre \mathbb{R}_+ .

Entonces el soporte producto KL del proceso

$$\left\{ \int_{\mathcal{Y}} \psi(\cdot, \mu, \sigma(\omega)) G_x(\omega)(d\mu) : x \in \mathcal{X} \right\},$$

es el conjunto

$$\prod_{x \in \mathcal{X}} \left\{ \int_{\mathcal{Y}} \psi(\cdot, \mu, \sigma) P_x(d\mu) : P_x \in \mathcal{P}(\mathcal{Y}), \sigma \in \mathbb{R}_+ \right\}.$$

El teorema anterior sigue siendo válido si se cambia el DDP de la condición (iv), por un w DDP o un θ DDP que tenga como soporte producto débil al conjunto $\mathcal{P}(\mathcal{Y})^{\mathcal{X}}$. La demostración del teorema anterior se encuentra en el apéndice B.12 y se basa en el Teorema 3.10, la compacidad de \mathcal{Y} y el hecho que una función continua sobre un espacio métrico compacto alcanza su mínimo.

Siguiendo los argumentos de Barrientos et al. (2012) y los resultados obtenidos hasta el momento, es posible extender los resultados de soporte al contexto de modelos de mezclas inducidos por procesos stick-breaking dependientes. Esto queda como trabajo de investigación futuro.

En lo que sigue se busca caracterizar propiedades de soporte más fuertes. Una forma de hacer esto es cambiar la topología producto considerada hasta el momento, por topologías asociadas a la convergencia de funciones. Para esto, parece natural requerir que $\{G_x : x \in \mathcal{X}\}$, visto como una función $x \rightarrow G_x$, sea continua.

En el siguiente teorema se caracteriza el soporte $\mathbb{L}_\infty \mathbb{L}_\infty$, denominado simplemente soporte \mathbb{L}_∞ , para un modelo de mezclas inducido por un θ DDP, donde Θ es un espacio Polaco no necesariamente Euclidiano y \mathcal{X} es un conjunto compacto.

Teorema 3.12. *Supongamos que:*

3.4. MEZCLAS INDUCIDAS POR UN DDP

- (i) $\psi(\cdot, \theta) \in \mathcal{D}(\mathcal{Y}), \forall \theta \in \Theta$.
- (ii) $(y, \theta) \rightarrow \psi(y, \theta)$ es una función continua sobre $\mathcal{Y} \times \Theta$.
- (iii) \mathcal{Y}, Θ y \mathcal{X} son espacios métricos compactos, con $\mathcal{X} \subseteq \mathbb{R}^q$.
- (iv) Sea $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \theta\text{DDP}(\Psi_V, G^0)$, con $\alpha_{\mathcal{X}} = \{\alpha_x : x \in \mathcal{X}\} \subseteq \mathbb{R}_+$ y $G^0 \in \mathcal{P}(\Theta)$. Supongamos que el parámetro Ψ_V es tal que:
 - (iv.1) $\text{supp}(G^0) = \Theta$.
 - (iv.2) $\{V_i(x, \omega) : x \in \mathcal{X}\}, i \geq 1$, son procesos estocásticos con trayectorias continuas casi seguramente.
 - (iv.3) Para toda función continua $g : \mathcal{X} \rightarrow [0, 1]$,

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |V_i(x, \omega) - g(x)| < \epsilon \right\} > 0,$$

para $i \geq 1$ y $\forall \epsilon > 0$.

Sea $\mathcal{G}_{\mathcal{X}}^* \subseteq \mathcal{P}(\Theta)^{\mathcal{X}}$ una colección de medidas de probabilidad dependientes de predictores satisfaciendo:

- (v) $x \rightarrow P_x(B)$ continua, $\forall B \in \mathcal{B}(\Theta)$.
- (vi) Para toda sucesión $\{A_n\} \subseteq \mathcal{B}(\Theta)$, con $A_n \downarrow \emptyset$, se tiene que

$$\sup_{x \in \mathcal{X}} P_x(A_n) \downarrow 0.$$

Entonces para cualquier $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{G}_{\mathcal{X}}^*$ y $\epsilon > 0$,

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right| < \epsilon \right\} > 0.$$

La demostración del teorema anterior se encuentra en el apéndice B.13 y usa el Teorema 3.6, más una versión del teorema de Arzelá-Ascoli, para funciones definidas sobre un espacio

métrico compacto. El teorema anterior nos permite caracterizar un importante resultado de soporte \mathbb{L}_∞ KL, que enunciamos a continuación.

Teorema 3.13. *Supongamos que:*

- (i) $\psi(\cdot, \theta) \in \mathcal{D}(\mathcal{Y}), \forall \theta \in \Theta$.
- (ii) $(y, \theta) \rightarrow \psi(y, \theta)$ es una función continua sobre $\mathcal{Y} \times \Theta$.
- (iii) \mathcal{Y}, Θ y \mathcal{X} son espacios métricos compactos, con $\mathcal{X} \subseteq \mathbb{R}^q$.
- (iv) Sea $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \theta\text{DDP}(\Psi_V, G^0)$, con $\alpha_{\mathcal{X}} = \{\alpha_x : x \in \mathcal{X}\} \subseteq \mathbb{R}_+$ y $G^0 \in \mathcal{P}(\Theta)$. *Supongamos que el parámetro Ψ_V es tal que:*
 - (iv.1) $\text{supp}(G^0) = \Theta$.
 - (iv.2) $\{V_i(x, \omega) : x \in \mathcal{X}\}, i \geq 1$, son procesos estocásticos con trayectorias continuas casi seguramente.
 - (iv.3) Para toda función continua $g : \mathcal{X} \rightarrow [0, 1]$,

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |V_i(x, \omega) - g(x)| < \epsilon \right\} > 0,$$

para $i \geq 1$ y $\forall \epsilon > 0$.

Sea $\mathcal{G}_{\mathcal{X}}^* \subseteq \mathcal{P}(\Theta)^{\mathcal{X}}$ una colección de medidas de probabilidad dependientes de predictores satisfaciendo:

- (v) $x \rightarrow P_x(B)$ continua, $\forall B \in \mathcal{B}(\Theta)$.
- (vi) Para toda sucesión $\{A_n\} \subseteq \mathcal{B}(\Theta)$, con $A_n \downarrow \emptyset$, se tiene que

$$\sup_{x \in \mathcal{X}} P_x(A_n) \downarrow 0.$$

Entonces para cualquier $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{G}_{\mathcal{X}}^*$ y $\epsilon > 0$,

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \int_{\mathcal{Y}} f^0(y|x) \ln \left(\frac{f^0(y|x)}{f(y|x, G_x(\omega))} \right) \lambda_{\mathcal{Y}}(dy) < \epsilon \right\} > 0,$$

donde

$$f^0(y|x) := \int_{\Theta} \psi(y, \theta) P_x^0(d\theta),$$

y

$$f(y|x, G_x(\omega)) := \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta).$$

La demostración del teorema anterior se encuentra en el apéndice [B.14](#). Dicha demostración usa los Teoremas [3.6](#) y [3.13](#), la compacidad de los espacios \mathcal{X} e \mathcal{Y} , y una versión del teorema de Arzelá-Ascoli, para funciones definidas sobre un espacio métrico compacto. El teorema anterior es fundamental para obtener un resultado de consistencia débil, que se presenta en la subsección siguiente.

En esta subsección hemos estudiado propiedades de soporte para distintos modelos de mezclas inducidos por un DDP. En esta línea, los teoremas [3.12](#) y [3.13](#) pueden ser obtenidos para otras versiones del modelo de mezclas. Esto queda como trabajo de investigación futuro.

3.4.3 Propiedad de consistencia

En el siguiente teorema se obtiene una propiedad de soporte débil para un modelo de mezclas inducido por un θ DDP.

Teorema 3.14. *Supongamos que:*

- (i) $\psi(\cdot, \theta) \in \mathcal{D}(\mathcal{Y})$, $\forall \theta \in \Theta$.
- (ii) $(y, \theta) \rightarrow \psi(y, \theta)$ es una función continua sobre $\mathcal{Y} \times \Theta$.

3.4. MEZCLAS INDUCIDAS POR UN DDP

(iii) \mathcal{Y} , Θ y \mathcal{X} son espacios métricos compactos, con $\mathcal{X} \subseteq \mathbb{R}^q$.

(iv) Sea $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\} \sim \theta\text{DDP}(\Psi_V, G^0)$, con $\alpha_{\mathcal{X}} = \{\alpha_x : x \in \mathcal{X}\} \subseteq \mathbb{R}_+$ y $G^0 \in \mathcal{P}(\Theta)$. Supongamos que el parámetro Ψ_V es tal que:

(iv.1) $\text{supp}(G^0) = \Theta$.

(iv.2) $\{V_i(x, \omega) : x \in \mathcal{X}\}$, $i \geq 1$, son procesos estocásticos con trayectorias continuas casi seguramente.

(iv.3) Para toda función continua $g : \mathcal{X} \rightarrow [0, 1]$,

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |V_i(x, \omega) - g(x)| < \epsilon \right\} > 0,$$

para $i \geq 1$ y $\forall \epsilon > 0$.

Sea $\mathcal{G}_{\mathcal{X}}^* \subseteq \mathcal{P}(\Theta)^{\mathcal{X}}$ una colección de medidas de probabilidad dependientes de predictores satisfaciendo:

(v) $x \rightarrow P_x(B)$ continua, $\forall B \in \mathcal{B}(\Theta)$.

(vi) Para toda sucesión $\{A_n\} \subseteq \mathcal{B}(\Theta)$, con $A_n \downarrow \emptyset$, se tiene que

$$\sup_{x \in \mathcal{X}} P_x(A_n) \downarrow 0.$$

Entonces la distribución a posteriori asociada con la distribución conjunta aleatoria inducida por el θDDP , $m^{(\cdot)}(y, x) = q(x)f(y|x, G_x(\cdot))$, donde q es la densidad generadora de los predictores y

$$f(y|x, G_x(\omega)) = \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta),$$

es débilmente consistente, bajo muestreo independiente, en cualquier densidad conjunta de la forma $m_0(y, x) = q(x)f^0(y|x)$ con

$$f^0(y|x) = \int_{\Theta} \psi(y, \theta) P_x^0(d\theta),$$

y $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{G}_{\mathcal{X}}^*$.

La demostración del teorema anterior se encuentra en el apéndice [B.15](#). Dicha demostración se basa en la caracterización de soporte \mathbb{L}_{∞} KL del Teorema [3.13](#) y en un resultado de [Schwartz \(1965\)](#).

Es importante señalar que la propiedad de soporte anterior, a pesar de ser interesante, no es del todo satisfactoria, pues el soporte débil no nos dice mucho cuando trabajamos con densidades. Por esta razón es recomendable buscar una propiedad de consistencia fuerte. Esto queda como trabajo de investigación futuro.

3.5 Conclusiones

En este capítulo generalizamos la definición formal del DDP dada por [Barrientos et al. \(2012\)](#), al contexto de espacios Polacos y predictores soportados en un subconjunto del espacio Euclidiano. Adicionalmente, se consideraron dos simplicaciones de dicho proceso: w DDP, en la cual los pesos no dependen de predictores; y el θ DDP, en el cual los átomos no dependen de covariables. En completa generalidad, esto es, para un espacio Polaco cualquiera, probamos que el DDP posee adecuadas propiedades de continuidad, estructura de asociación y soporte. Adicionalmente, bajo el supuesto de compacidad del espacio que soporta a las covariables y usando el θ DDP, se caracterizaron dos nociones de soporte fuerte: \mathbb{L}_1 débil y \mathbb{L}_{∞} débil.

Para modelos de mezclas inducidos por el DDP o el θ DDP, se caracterizan el soporte producto Hellinger, el soporte producto \mathbb{L}_{∞} y el soporte producto Kullback-Leibler. Bajo supuestos de compacidad, se caracterizan además el soporte $\mathbb{L}_{\infty} \mathbb{L}_{\infty}$ y el soporte \mathbb{L}_{∞} Kullback-Leibler, para modelos de mezclas inducidos por un θ DDP con soporte completo. En este último caso se demuestra, además, una propiedad de consistencia débil para el modelo de mezclas subyacente.

3.5. CONCLUSIONES

El trabajo futuro incluye el estudio de propiedades de soporte fuerte y tasas de convergencia para modelos de mezclas inducidos por el θ DDP, bajo el supuesto de compacidad de los espacios asociados. Para el caso de la consistencia fuerte se podría utilizar el enfoque de [Pati et al. \(2013\)](#). Sin embargo, el problema de tasas de convergencia es mucho más complicado. Gracias a que una variedad Riemanniana completa es un caso particular de un espacio Polaco, la investigación futura comsidera la implementación y aplicación del enfoque Bayesiano noparamétrico propuesto a variedades particulares como el espacio de las formas planares de Kendall, una hiperesfera, espacios Euclidianos compactos, las variedades de Hadamard, las variedades de Stiefel y las variedades de Hilbert, entre otros ejemplos. Para los tres primeros casos, se podrían utilizar los kernel propuestos por [Bhattacharya & Dunson \(2012b\)](#).

Conclusiones y Trabajo Futuro

En esta tesis, hemos estudiado dos tópicos diferentes en el contexto de modelos Bayesianos sobre variedades Riemannianas. Este capítulo resume los principales resultados de la tesis y discute potenciales líneas de investigación futura.

4.1 Conclusiones

En el Capítulo 2, propusimos un estimador de la media basado en un modelo Bayesiano paramétrico definido en el espacio de las *formas* planares de [Kendall \(1977\)](#), que corresponde a una variedad Riemanniana de dimensión finita. El estimador propuesto incorpora en su definición la estructura métrica de la variedad subyacente y tiene un mejor rendimiento en muestras finitas que otros estimadores Bayesianos, en especial, que el estimador propuesto por [Micheas & Peng \(2010\)](#). La idea clave detrás de la construcción del estimador, fue resolver el problema de identificación que surge al intentar hacer inferencia sobre el espacio de las *formas* planares, a través del espacio de *configuraciones* correspondiente.

En el Capítulo 3, generalizamos los procesos de Dirichlet dependientes, originalmente definidos en espacios Euclidianos, a espacios métricos Polacos, premunidos de una medida de referencia estrictamente positiva y localmente finita. En particular, dicha extensión nos permite definir un DDP en el contexto de una variedad Riemanniana completa dotada de la medida de volumen Riemanniana, asociada a una métrica Riemanniana. Se establecieron condiciones necesarias para que los procesos tengan propiedades adecuadas de soporte, continuidad y estructura de asociación. El énfasis se centró en la versión del modelo donde los puntos de soporte son independientes de los predictores, θ DDP. Utilizando dicho proceso y modelos de mezclas inducidos por él, se caracterizaron propiedades de soportes más fuertes, bajo el supuesto de compacidad de los espacios subyacentes. Así, por ejemplo, para el θ DDP se obtienen condiciones suficientes para que este tenga soporte completo \mathbb{L}_1 -débil y \mathbb{L}_∞ -débil. Para modelos de mezclas inducido por el θ DDP se obtienen condiciones suficientes para que éste tenga soporte completo \mathbb{L}_∞ y \mathbb{L}_∞ Kulback-Leibler. Finalmente, se caracteriza una propiedad de consistencia débil para un modelo de mezclas inducido por el θ DDP.

4.2 Trabajo futuro

Los resultados presentados en esta tesis pueden ser aplicados y extendidos en diferentes direcciones, algunas de las cuales son descritas a continuación. En el Capítulo 2, planeamos extender el modelo paramétrico de diversas maneras. Una primera opción es levantar el supuesto de varianza isotrópica, el cual puede ser demasiado reductivo en algunas ocasiones. En este caso se podría utilizar una distribución Wishart compleja inversa de [Andersen \(1995\)](#) para la matriz de varianza-covarianza, manteniendo así la propiedad de conjugación del modelo Bayesiano subyacente. Una segunda opción es considerar un modelo de mezclas de normales complejas inducido por modelos Bayesianos noparamétricos discretos, estudiar propiedades de identificabilidad, soporte, consistencia y tasas de convergencia del modelo subyacente, y explorar esquemas eficientes de simulación de cadenas de Markov Monte Carlo. Una tercera opción es extender el modelo planar al de hitos en tres dimensiones. Para esto se podría explorar cambiar

4.2. TRABAJO FUTURO

la normal compleja a una normal multivariada apropiada, teniendo en cuenta que en este caso el espacio de las *formas* de Kendall posee singularidades, esto es, el espacio pierde su carácter homogéneo. En todos los casos, la idea es utilizar el esquema de expansión de parámetros propuesto en este capítulo, para la estimación de la media en el espacio de las *formas* de Kendall.

En el Capítulo 3, planeamos aplicar la metodología desarrollada para algunas variedades Riemannianas particulares. Una primera opción es trabajar sobre el espacio de las *formas* planares de Kendall (1977). En este caso se puede utilizar como kernel una distribución Watson compleja (Mardia & Dryden, 1999). También se pueden explorar distribuciones elípticas complejas (Micheas et al., 2006). Para el modelo resultante hay que examinar esquemas eficientes de simulación de cadenas de Markov Monte Carlo. Una segunda opción es trabajar sobre una hiperesfera y un espacio Euclidiano compacto utilizando como kernel una distribución von Mises-Fisher y una distribución Gaussiana multivariada, respectivamente, siguiendo las ideas de Bhattacharya & Dunson (2012a). Una tercera línea de investigación tiene que ver con la extensión de los resultados de soporte débil del modelo de mezclas inducido por un DDP a un resultado de soporte fuerte, siguiendo las ideas de Pati et al. (2013) para mezclas Gaussianas.

Material Suplementario del Capítulo 2

A.1 Propiedades de la distribución normal compleja

Las siguientes definiciones fueron tomadas de [Andersen \(1995\)](#).

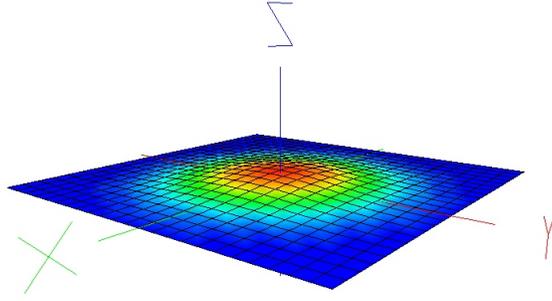
Definición A.1 (Distribución normal compleja estándar). $\mathbf{X} \sim \mathbb{CN}(0, 1)$ si $[\mathbf{X}] \sim \mathcal{N}_2(\mathbf{0}, \frac{1}{2}\mathbf{I}_2)$.

Propiedad A.1 (Invarianza rotacional). Si $\mathbf{X} \sim \mathbb{CN}(0, 1)$ y $c \in \mathbb{C}$ con $|c| = 1$, entonces $c\mathbf{X} \sim \mathbb{CN}(0, 1)$.

Propiedad A.2 (Función de densidad de $\mathbb{CN}(0, 1)$). Sea $\mathbf{X} \sim \mathbb{CN}(0, 1)$. Entonces la función de densidad de \mathbf{X} c/r a la medida de Lebesgue en \mathbb{C} es:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\pi} \exp\{-\bar{\mathbf{x}}\mathbf{x}\}, \quad \mathbf{x} \in \mathbb{C}. \quad (\text{A.1})$$

Gráficamente:



Definición A.2 (Distribución normal compleja). Sean $\mathbf{Z} \sim \mathbb{CN}(0, 1)$, $\boldsymbol{\theta} \in \mathbb{C}$ y $\sigma \in \bar{\mathbb{R}}_+$. Entonces $\mathbf{X} = \boldsymbol{\theta} + \sigma \mathbf{Z} \sim \mathbb{CN}(\boldsymbol{\theta}, \sigma^2)$.

Propiedad A.3. Sean $\boldsymbol{\theta} \in \mathbb{C}$ y $\sigma \in \bar{\mathbb{R}}_+$. Entonces $\mathbf{X} \sim \mathbb{CN}(\boldsymbol{\theta}, \sigma^2)$ ssi $[\mathbf{X}] \sim \mathcal{N}_2\left([\boldsymbol{\theta}], \frac{\sigma^2}{2} \mathbf{I}_2\right)$.

Propiedad A.4 (Función de densidad de $\mathbb{CN}(\boldsymbol{\theta}, \sigma^2)$). Sean $\boldsymbol{\theta} \in \mathbb{C}$, $\sigma \in \bar{\mathbb{R}}_+$ y $\mathbf{X} \sim \mathbb{CN}(\boldsymbol{\theta}, \sigma^2)$. Entonces la función de densidad de \mathbf{X} c/r a la medida de Lebesgue en \mathbb{C} es:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\pi\sigma^2} \exp\left\{-\frac{1}{\sigma^2} \overline{(\mathbf{x} - \boldsymbol{\theta})}(\mathbf{x} - \boldsymbol{\theta})\right\}, \quad \mathbf{x} \in \mathbb{C}. \quad (\text{A.2})$$

Propiedad A.5 (Función característica de $\mathbb{CN}(\boldsymbol{\theta}, \sigma^2)$). Sean $\boldsymbol{\theta} \in \mathbb{C}$, $\sigma \in \bar{\mathbb{R}}_+$ y $\mathbf{X} \sim \mathbb{CN}(\boldsymbol{\theta}, \sigma^2)$. Entonces la función característica de \mathbf{X} es:

$$\varphi_{\mathbf{X}}(\boldsymbol{\xi}) = \exp\left\{i\text{Re}(\bar{\boldsymbol{\xi}}\boldsymbol{\theta}) - \frac{\sigma^2}{4}\bar{\boldsymbol{\xi}}\boldsymbol{\xi}\right\}, \quad \boldsymbol{\xi} \in \mathbb{C}. \quad (\text{A.3})$$

Propiedad A.6. Sean $\mathbf{X}_1, \dots, \mathbf{X}_p$ v.a's complejas independientes tales que $\mathbf{X}_i \sim \mathbb{CN}(\boldsymbol{\theta}_i, \sigma_i^2)$ con $\boldsymbol{\theta}_i \in \mathbb{C}$ y $\sigma_i \in \bar{\mathbb{R}}_+$, para $i = 1, \dots, p$. Sean $\mathbf{c}_i, \mathbf{d}_i \in \mathbb{C}$, $i = 1, \dots, p$. Entonces

$$\sum_{i=1}^p (\mathbf{c}_i \mathbf{X}_i + \mathbf{d}_i) \sim \mathbb{CN}\left(\sum_{i=1}^p (\mathbf{c}_i \boldsymbol{\theta}_i + \mathbf{d}_i), \sum_{i=1}^p \mathbf{c}_i \bar{\mathbf{c}}_i \sigma_i^2\right). \quad (\text{A.4})$$

Definición A.3 (Distribución normal compleja multivariada). $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_p)$ se distribuye como una normal compleja p -variada si $\forall \mathbf{c} = (\mathbf{c}_1, \dots, \mathbf{c}_p) \in \mathbb{C}^p$, $\langle \mathbf{X}, \mathbf{c} \rangle = \sum_{i=1}^p \bar{\mathbf{c}}_i \mathbf{X}_i = \mathbf{c}^* \mathbf{X}$ se distribuye como una normal compleja (univariada).

Propiedad A.7 (Función característica de $\mathbb{CN}(\boldsymbol{\theta}, \mathbf{H})$). Sea \mathbf{X} un vector aleatorio complejo p -dimensional con $\mathcal{L}(\mathbf{X}) = \mathbb{CN}_p(\boldsymbol{\theta}, \mathbf{H})$, donde $\boldsymbol{\theta} \in \mathbb{C}^p$ y $\mathbf{H} \in \mathbb{C}_S^{p \times p}$. La función característica

A.1. PROPIEDADES DE LA DISTRIBUCIÓN NORMAL COMPLEJA

de \mathbf{X} está dada por

$$\varphi_{\mathbf{X}}(\boldsymbol{\xi}) = \exp\left(i\operatorname{Re}(\boldsymbol{\xi}^* \boldsymbol{\theta}) - \frac{\boldsymbol{\xi}^* \mathbf{H} \boldsymbol{\xi}}{4}\right), \quad \boldsymbol{\xi} \in \mathbb{C}^p.$$

Propiedad A.8 (Propiedades de normal compleja multivariada). *Sea \mathbf{X} un vector aleatorio complejo p -dimensional.*

1. Si existe $\boldsymbol{\theta} \in \mathbb{C}^p$ y $\mathbf{H} \in \mathbb{C}_S^{p \times p}$ tal que para todo $\mathbf{c} \in \mathbb{C}^p$ se cumple que

$$\mathcal{L}(\mathbf{c}^* \mathbf{X}) = \mathbb{CN}(\mathbf{c}^* \boldsymbol{\theta}, \mathbf{c}^* \mathbf{H} \mathbf{c}),$$

entonces

$$\mathcal{L}(\mathbf{X}) = \mathbb{CN}_p(\boldsymbol{\theta}, \mathbf{H}).$$

2. Sea $\mathbf{D} \in \mathbb{C}^{q \times p}$ y $\mathbf{d} \in \mathbb{C}^q$. Si $\mathcal{L}(\mathbf{X}) = \mathbb{CN}_p(\boldsymbol{\theta}, \mathbf{H})$, donde $\boldsymbol{\theta} \in \mathbb{C}^p$ y $\mathbf{H} \in \mathbb{C}_S^{p \times p}$, entonces

$$\mathcal{L}(\mathbf{D}\mathbf{X} + \mathbf{d}) = \mathbb{CN}_p(\mathbf{D}\boldsymbol{\theta} + \mathbf{d}, \mathbf{D}\mathbf{H}\mathbf{D}^*).$$

Propiedad A.9 (Relación con la distribución normal real multivariada). *Sea \mathbf{X} un vector aleatorio complejo p -dimensional. Entonces*

$$\mathcal{L}(\mathbf{X}) = \mathbb{CN}_p(\boldsymbol{\theta}, \mathbf{H})$$

si y solo si

$$\mathcal{L}([\mathbf{X}]) = \mathcal{N}_{2p}\left([\boldsymbol{\theta}], \frac{1}{2}\{\mathbf{H}\}\right),$$

donde $\boldsymbol{\theta} \in \mathbb{C}^p$ y $\mathbf{H} \in \mathbb{C}_S^{p \times p}$.

Propiedad A.10 (Función de densidad de $\mathbb{CN}_p(\boldsymbol{\theta}, \mathbf{H})$). *Sea \mathbf{X} un vector aleatorio complejo p -dimensional con $\mathcal{L}(\mathbf{X}) = \mathbb{CN}_p(\boldsymbol{\theta}, \mathbf{H})$, donde $\boldsymbol{\theta} \in \mathbb{C}^p$ y $\mathbf{H} \in \mathbb{C}_S^{p \times p}$. La función de densidad de probabilidad de \mathbf{X} con respecto a la medida de Lebesgue sobre \mathbb{C}^p está dada por*

$$f_{\mathbf{X}}(\mathbf{x}) = \pi^{-p} \det(\mathbf{H})^{-1} \exp\left\{-\mathbf{x}^* \mathbf{H}^{-1} (\mathbf{x} - \boldsymbol{\theta})\right\}, \quad \mathbf{x} \in \mathbb{C}^p.$$

A.1. PROPIEDADES DE LA DISTRIBUCIÓN NORMAL COMPLEJA

Propiedad A.11. Sean $\mathbf{X}_1, \dots, \mathbf{X}_n$ vectores aleatorios complejos p -dimensionales independientes con $\mathcal{L}(\mathbf{X}_j) = \mathbb{C}\mathcal{N}_p(\boldsymbol{\theta}_j, \mathbf{H}_j)$, donde $\boldsymbol{\theta}_j \in \mathbb{C}^p$ y $\mathbf{H}_j \in \mathbb{C}_S^{p \times p}$, y sean $c_j \in \mathbb{C}$ y $\mathbf{d}_j \in \mathbb{C}^p$, para $j = 1, 2, \dots, n$. Entonces

$$\mathcal{L}\left(\sum_{j=1}^n (c_j \mathbf{X}_j + \mathbf{d}_j)\right) = \mathbb{C}\mathcal{N}_p\left(\sum_{j=1}^n (c_j \boldsymbol{\theta}_j + \mathbf{d}_j), \sum_{j=1}^n c_j \bar{c}_j \mathbf{H}_j\right).$$

Propiedad A.12. Sea \mathbf{X} un vector aleatorio complejo p -dimensional con $\mathcal{L}(\mathbf{X}) = \mathbb{C}\mathcal{N}_p(\boldsymbol{\theta}, \mathbf{H})$, donde $\boldsymbol{\theta} \in \mathbb{C}^p$ y $\mathbf{H} \in \mathbb{C}_S^{p \times p}$. Sean \mathbf{X} , $\boldsymbol{\theta}$ y \mathbf{H} particionados como

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix} \quad \text{y} \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix},$$

donde \mathbf{X}_j y $\boldsymbol{\theta}_j$ son $p_j \times 1$ y \mathbf{H}_{jk} es $p_j \times p_k$, para $j, k = 1, 2$, y $p = p_1 + p_2$. Entonces \mathbf{X}_1 y \mathbf{X}_2 son independientes ssi $\mathbf{H}_{12} = \mathbf{O}$.

Propiedad A.13. Sea \mathbf{X} un vector aleatorio complejo p -dimensional con $\mathcal{L}(\mathbf{X}) = \mathbb{C}\mathcal{N}_p(\boldsymbol{\theta}, \mathbf{H})$, donde $\boldsymbol{\theta} \in \mathbb{C}^p$ y $\mathbf{H} \in \mathbb{C}_S^{p \times p}$, y sean $\mathbf{C} \in \mathbb{C}^{m \times p}$ y $\mathbf{D} \in \mathbb{C}^{q \times p}$. Entonces $\mathbf{C}\mathbf{X}$ y $\mathbf{D}\mathbf{X}$ son independientes ssi $\mathbf{C}\mathbf{H}\mathbf{D}^* = \mathbf{O}$.

Propiedad A.14 (Distribución marginal). Sea \mathbf{X} un vector aleatorio complejo p -dimensional con $\mathcal{L}(\mathbf{X}) = \mathbb{C}\mathcal{N}_p(\boldsymbol{\theta}, \mathbf{H})$, donde $\boldsymbol{\theta} \in \mathbb{C}^p$ y $\mathbf{H} \in \mathbb{C}_S^{p \times p}$. Sean \mathbf{X} , $\boldsymbol{\theta}$ y \mathbf{H} particionados como

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix} \quad \text{y} \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix},$$

donde \mathbf{X}_j y $\boldsymbol{\theta}_j$ son $p_j \times 1$ y \mathbf{H}_{jk} es $p_j \times p_k$, para $j, k = 1, 2$, y $p = p_1 + p_2$. Entonces para $j = 1, 2$ se cumple que

$$\mathcal{L}(\mathbf{X}_j) = \mathbb{C}\mathcal{N}_{p_j}(\boldsymbol{\theta}_j, \mathbf{H}_{jj}).$$

Propiedad A.15. Sea \mathbf{X} un vector aleatorio complejo p -dimensional con $\mathcal{L}(\mathbf{X}) = \mathbb{C}\mathcal{N}_p(\boldsymbol{\theta}, \mathbf{H})$,

A.2. SIMULACIÓN DE LA DISTRIBUCIÓN NORMAL COMPLEJA

donde $\theta \in \mathbb{C}^p$ y $\mathbf{H} \in \mathbb{C}_S^{p \times p}$. Sean \mathbf{X} , $\boldsymbol{\theta}$ y \mathbf{H} particionados como

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix} \quad \text{y} \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix},$$

donde \mathbf{X}_j y $\boldsymbol{\theta}_j$ son $p_j \times 1$ y \mathbf{H}_{jk} es $p_j \times p_k$, para $j, k = 1, 2$, y $p = p_1 + p_2$. Si $\mathbf{H}_{22} > \mathbf{O}$, entonces se cumple que

$$\mathcal{L}(\mathbf{X}_1 - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{X}_2) = \mathbb{CN}_{p_1}(\boldsymbol{\theta}_1 - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\boldsymbol{\theta}_2, \mathbf{H}_{11} - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{21})$$

y

$$\mathbf{X}_1 - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{X}_2 \perp \mathbf{X}_2.$$

Propiedad A.16. Sea \mathbf{X} un vector aleatorio complejo p -dimensional con $\mathcal{L}(\mathbf{X}) = \mathbb{CN}_p(\boldsymbol{\theta}, \mathbf{H})$, donde $\theta \in \mathbb{C}^p$ y $\mathbf{H} \in \mathbb{C}_S^{p \times p}$. Sean \mathbf{X} , $\boldsymbol{\theta}$ y \mathbf{H} particionados como

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix} \quad \text{y} \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix},$$

donde \mathbf{X}_j y $\boldsymbol{\theta}_j$ son $p_j \times 1$ y \mathbf{H}_{jk} es $p_j \times p_k$, para $j, k = 1, 2$, y $p = p_1 + p_2$. Si $\mathbf{H}_{22} > \mathbf{O}$, entonces la distribución condicional de \mathbf{X}_1 dado \mathbf{X}_2 es

$$\mathcal{L}(\mathbf{X}_1|\mathbf{X}_2) = \mathbb{CN}_{p_1}(\boldsymbol{\theta}_1 + \mathbf{H}_{12}\mathbf{H}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\theta}_2), \mathbf{H}_{11} - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{21}).$$

A.2 Simulación de la distribución normal compleja

Si $\mathbf{w} \sim \mathbb{CN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces su función de densidad de probabilidad está dada por:

$$f(\mathbf{w}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \pi^{-p} \det(\boldsymbol{\Sigma})^{-1} \exp \left\{ -(\mathbf{w} - \boldsymbol{\mu})^* \boldsymbol{\Sigma}^{-1} (\mathbf{w} - \boldsymbol{\mu}) \right\}, \quad \mathbf{w} \in \mathbb{C}^p,$$

A.3. DEDUCCIÓN DE DISTRIBUCIONES A POSTERIORI

donde $\boldsymbol{\mu} \in \mathbb{C}^p$ y $\boldsymbol{\Sigma} \in \mathcal{M}_{p \times p}(\mathbb{C})$ (matrices de $p \times p$ a coeficientes complejos).

¿Cómo simular $\mathbb{CN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$?

Sea $\boldsymbol{w} \sim \mathbb{CN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{w} = \text{Re}(\boldsymbol{w}) + i\text{Im}(\boldsymbol{w})$, $\boldsymbol{\mu} = \text{Re}(\boldsymbol{\mu}) + i\text{Im}(\boldsymbol{\mu})$ y $\boldsymbol{\Sigma} = \text{Re}(\boldsymbol{\Sigma}) + i\text{Im}(\boldsymbol{\Sigma})$. Entonces

$$\begin{pmatrix} \text{Re}(\boldsymbol{w}) \\ \text{Im}(\boldsymbol{w}) \end{pmatrix} \sim \mathcal{N}_{2p} \left(\begin{pmatrix} \text{Re}(\boldsymbol{\mu}) \\ \text{Im}(\boldsymbol{\mu}) \end{pmatrix}, \frac{1}{2} \begin{pmatrix} \text{Re}(\boldsymbol{\Sigma}) & -\text{Im}(\boldsymbol{\Sigma}) \\ \text{Im}(\boldsymbol{\Sigma}) & \text{Re}(\boldsymbol{\Sigma}) \end{pmatrix} \right).$$

Entonces, si $\tilde{\boldsymbol{w}} = \begin{pmatrix} \text{Re}(\boldsymbol{w}) \\ \text{Im}(\boldsymbol{w}) \end{pmatrix}$, $\tilde{\boldsymbol{\mu}} = \begin{pmatrix} \text{Re}(\boldsymbol{\mu}) \\ \text{Im}(\boldsymbol{\mu}) \end{pmatrix}$ and $\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \text{Re}(\boldsymbol{\Sigma}) & -\text{Im}(\boldsymbol{\Sigma}) \\ \text{Im}(\boldsymbol{\Sigma}) & \text{Re}(\boldsymbol{\Sigma}) \end{pmatrix}$ sigue que la función de densidad de probabilidad de $\tilde{\boldsymbol{w}}$ está dada por:

$$f(\tilde{\boldsymbol{w}}|\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}) = \pi^{-p} |\tilde{\boldsymbol{\Sigma}}|^{-1/2} \exp\{-(\tilde{\boldsymbol{w}} - \tilde{\boldsymbol{\mu}})^t \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{w}} - \tilde{\boldsymbol{\mu}})\}.$$

A.3 Deducción de distribuciones a posteriori

A.3.1 Distribución a posteriori para el parámetro de localización

Para $j = 1, \dots, n$, notar que

$$\begin{aligned} [d_j|\cdot] &\propto [\boldsymbol{w}_j|\boldsymbol{\mu}, d_j, b_j, \phi_j, \sigma^2][d_j] \\ &\propto \exp\{-\sigma^{-2}(\boldsymbol{w}_j - d_j\mathbf{1}_p - b_j e^{i\phi_j}\boldsymbol{\mu})^*(\boldsymbol{w}_j - d_j\mathbf{1}_p - b_j e^{i\phi_j}\boldsymbol{\mu})\} \\ &\quad \cdot \exp\{-\tau^{-2}(\overline{d_j - d_0})(d_j - d_0)\} \\ &\propto \exp\{-\sigma^{-2}[\overline{d_j}\mathbf{1}_p^t \mathbf{1}_p d_j - 2\text{Re}(\overline{d_j}\mathbf{1}_p^t(\boldsymbol{w}_j - b_j e^{i\phi_j}\boldsymbol{\mu}))]\} \exp\{-\tau^2[\overline{d_j}d_j - \text{Re}(\overline{d_j}d_0)]\} \\ &= \exp\left\{-\left(\frac{p}{\sigma^2} + \frac{1}{\tau^2}\right)\overline{d_j}d_j + 2\text{Re}(\overline{d_j}[\mathbf{1}_p^t(\boldsymbol{w}_j - b_j e^{i\phi_j}\boldsymbol{\mu})\sigma^{-2} + d_0\tau^{-2}])\right\}. \end{aligned}$$

Se concluye que

$$d_j | \cdot \sim \mathbb{CN} \left(\frac{1}{V} \lambda_j, \frac{1}{V} \right), \quad j = 1, \dots, n,$$

con $V = \frac{p}{\sigma^2} + \frac{1}{\tau^2}$ y $\lambda_j = \mathbf{1}_p^t (\mathbf{w}_j - b_j e^{i\phi_j} \boldsymbol{\mu}) \sigma^{-2} + d_0 \tau^{-2}$.

A.3.2 Distribución a posteriori para el parámetro de escala

Para $j = 1, \dots, n$, notar que

$$\begin{aligned} [b_j | \cdot] &\propto [\mathbf{w}_j | \boldsymbol{\mu}, d_j, b_j, \phi_j, \sigma^2][b_j] \\ &\propto \exp \left\{ -\frac{1}{\sigma^2} (\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu})^* (\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu}) \right\} \exp \left\{ -\frac{b_j^2}{2} \right\} \mathbb{1}_{[0, \infty)}(b_j) \\ &\propto \exp \left\{ -\frac{1}{\sigma^2} (b_j^2 \boldsymbol{\mu}^* \boldsymbol{\mu} - b_j e^{-i\phi_j} \boldsymbol{\mu}^* (\mathbf{w}_j - d_j \mathbf{1}_p) - b_j (\mathbf{w}_j - d_j \mathbf{1}_p)^* e^{i\phi_j} \boldsymbol{\mu}) \right\} \\ &\cdot \exp \left\{ -\frac{b_j^2}{2} \right\} \mathbb{1}_{[0, \infty)}(b_j) \\ &= \exp \left\{ -\frac{1}{2\sigma^2} (b_j^2 [2\boldsymbol{\mu}^* \boldsymbol{\mu} + \sigma^2] - 2b_j e^{-i\phi_j} \boldsymbol{\mu}^* (\mathbf{w}_j - d_j \mathbf{1}_p) - 2b_j (\mathbf{w}_j - d_j \mathbf{1}_p)^* e^{i\phi_j} \boldsymbol{\mu}) \right\} \\ &\cdot \mathbb{1}_{[0, \infty)}(b_j). \end{aligned}$$

Se concluye que

$$b_j | \cdot \sim \mathcal{N} \left(\frac{\gamma_j + \bar{\gamma}_j}{\xi}, \frac{\sigma^2}{\xi} \right) \quad \text{truncada desde la izquierda en cero,} \quad j = 1, \dots, n,$$

con $\xi = \sigma^2 + 2\boldsymbol{\mu}^* \boldsymbol{\mu}$ y $\gamma_j = e^{-i\phi_j} \boldsymbol{\mu}^* (\mathbf{w}_j - d_j \mathbf{1}_p)$.

A.3.3 Distribución a posteriori para el parámetro de rotación

Para $j = 1, \dots, n$, notar que

$$\begin{aligned} [\phi_j | \cdot] &\propto [\mathbf{w}_j | \boldsymbol{\mu}, d_j, b_j, \phi_j, \sigma^2][\phi_j] \\ &\propto \exp \left\{ -\frac{1}{\sigma^2} (\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu})^* (\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu}) \right\} \mathbb{1}_{[0, 2\pi]}(\phi_j). \end{aligned}$$

A.3.4 Distribución a posteriori para el representante de la forma media

Notar que

$$\begin{aligned}
 [\boldsymbol{\mu}|\cdot] &\propto \prod_{j=1}^n [\mathbf{w}_j | \boldsymbol{\mu}, d_j, b_j, \phi_j, \sigma^2] [\boldsymbol{\mu}] \\
 &\propto \prod_{j=1}^n \exp \left\{ -\frac{1}{\sigma^2} (\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu})^* (\mathbf{w}_j - d_j \mathbf{1}_p - b_j e^{i\phi_j} \boldsymbol{\mu}) \right\} \\
 &\quad \cdot \exp \left\{ -\frac{1}{k_0^2} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^* (\boldsymbol{\mu} - \boldsymbol{\mu}_0) \right\} \\
 &\propto \exp \left\{ -\left(\frac{1}{\sigma^2} \sum_{j=1}^n b_j^2 + \frac{1}{k_0^2} \right) \boldsymbol{\mu}^* \boldsymbol{\mu} \right. \\
 &\quad \left. + 2Re \left(\boldsymbol{\mu}^* \left[\frac{1}{k_0^2} \boldsymbol{\mu}_0 + \frac{1}{\sigma^2} \sum_{j=1}^n b_j e^{-i\phi_j} (\mathbf{w}_j - d_j \mathbf{1}_p) \right] \right) \right\},
 \end{aligned}$$

y se concluye que

$$\boldsymbol{\mu}|\cdot \sim \mathbb{CN}_p \left(\frac{1}{\Delta} \boldsymbol{\eta}, \frac{1}{\Delta} \mathbf{I}_p \right).$$

$$\text{con } \Delta = \frac{1}{\sigma^2} \sum_{j=1}^n b_j^2 + \frac{1}{k_0^2} \text{ y } \boldsymbol{\eta} = \frac{1}{k_0^2} \boldsymbol{\mu}_0 + \frac{1}{\sigma^2} \sum_{j=1}^n b_j e^{-i\phi_j} (\mathbf{w}_j - d_j \mathbf{1}_p).$$

A.3.5 Distribución a posteriori para el parámetro de variabilidad

Notar que

$$\begin{aligned}
 [\sigma^2|\cdot] &\propto \prod_{j=1}^n [\mathbf{w}_j|\boldsymbol{\mu}, d_j, b_j, \phi_j, \sigma^2][\sigma^2] \\
 &\propto \prod_{j=1}^n \pi^{-p} \left(\frac{1}{\sigma^2}\right)^p \exp\left\{-\frac{1}{\sigma^2}(\mathbf{w}_j - d_j\mathbf{1}_p - b_j e^{i\phi_j}\boldsymbol{\mu})^*(\mathbf{w}_j - d_j\mathbf{1}_p - b_j e^{i\phi_j}\boldsymbol{\mu})\right\} \\
 &\quad \cdot (\sigma^2)^{-(\alpha_1+1)} \exp\left\{-\frac{\beta_1}{\sigma^2}\right\} \\
 &\propto (\sigma^2)^{-np} \exp\left\{-\frac{1}{\sigma^2} \sum_{j=1}^n (\mathbf{w}_j - d_j\mathbf{1}_p - b_j e^{i\phi_j}\boldsymbol{\mu})^*(\mathbf{w}_j - d_j\mathbf{1}_p - b_j e^{i\phi_j}\boldsymbol{\mu})\right\} \\
 &\quad \cdot (\sigma^2)^{-(\alpha_1+1)} \exp\left\{-\frac{\beta_1}{\sigma^2}\right\} \\
 &= (\sigma^2)^{-(\alpha_1+np+1)} \exp\left\{-\frac{1}{\sigma^2} \left[\beta_1 + \sum_{j=1}^n (\mathbf{w}_j - d_j\mathbf{1}_p - b_j e^{i\phi_j}\boldsymbol{\mu})^*(\mathbf{w}_j - d_j\mathbf{1}_p - b_j e^{i\phi_j}\boldsymbol{\mu}) \right]\right\},
 \end{aligned}$$

y se concluye que

$$\sigma^2|\cdot \sim \text{Gamma} - \text{Inv} \left(\alpha_1 + np, \beta_1 + \sum_{j=1}^n (\mathbf{w}_j - d_j\mathbf{1}_p - b_j e^{i\phi_j}\boldsymbol{\mu})^*(\mathbf{w}_j - d_j\mathbf{1}_p - b_j e^{i\phi_j}\boldsymbol{\mu}) \right).$$

A.3.6 Distribución a posteriori para hiperparámetros

Notar que

$$\begin{aligned}
 [\tau^2|\cdot] &\propto \prod_{j=1}^n [d_j|\tau^2][\tau^2] \\
 &\propto \prod_{j=1}^n \tau^{-2} \exp\{-\tau^{-2} \overline{(d_j - d_0)}(d_j - d_0)\} (\tau^2)^{-(\alpha_2+1)} \exp\left\{-\frac{\beta_2}{\tau^2}\right\} \\
 &= (\tau^2)^{-(\alpha_2+n+1)} \exp\left\{-\frac{1}{\tau^2} \left[\beta_2 + \sum_{j=1}^n \overline{(d_j - d_0)}(d_j - d_0) \right]\right\}.
 \end{aligned}$$

Se concluye que

$$\tau^2 | \cdot \sim \text{Gamma} - \text{Inv} \left(\alpha_2 + n, \beta_2 + \sum_{j=1}^n \overline{(d_j - d_0)}(d_j - d_0) \right).$$

A.4 Caracterización de la métrica de Procrustes parcial

Notemos que

$$d^2([\boldsymbol{\eta}], [\boldsymbol{\nu}]) = \min_{\alpha \in [0, 2\pi)} \|\boldsymbol{\eta} - e^{i\alpha} \boldsymbol{\nu}\|^2.$$

Definamos $h(\alpha) = (\boldsymbol{\eta} - e^{i\alpha} \boldsymbol{\nu})^* (\boldsymbol{\eta} - e^{i\alpha} \boldsymbol{\nu})$. Entonces

$$h(\alpha) = \boldsymbol{\eta}^* \boldsymbol{\eta} - e^{i\alpha} \boldsymbol{\eta}^* \boldsymbol{\nu} - e^{-i\alpha} \boldsymbol{\nu}^* \boldsymbol{\eta} + \boldsymbol{\nu}^* \boldsymbol{\nu}.$$

Sea $\boldsymbol{\eta}^* \boldsymbol{\nu} = \gamma e^{i\phi}$. Luego

$$\begin{aligned} e^{i\alpha} \boldsymbol{\eta}^* \boldsymbol{\nu} + e^{-i\alpha} \boldsymbol{\nu}^* \boldsymbol{\eta} &= \gamma \left(e^{i(\alpha+\phi)} + e^{-i(\alpha+\phi)} \right) \\ &= 2\gamma \cos(\alpha + \phi). \end{aligned}$$

Por lo tanto, basta maximizar $2\gamma \cos(\alpha + \phi)$ para concluir que el α óptimo está dado por

$$\alpha_{opt} = -\phi = -\arg(\boldsymbol{\eta}^* \boldsymbol{\nu}) = \arg(\boldsymbol{\nu}^* \boldsymbol{\eta}),$$

de donde se concluye lo pedido. ■

A.5 Convergencia del promedio muestral

Sea w_1, \dots, w_n una muestra proveniente del modelo estadístico (2.1). Se quiere demostrar que

$$\frac{1}{n} \sum_{j=1}^n (w_j - E_{\theta}(w_j)) \xrightarrow{P_{\theta}} \mathbf{0} \in \mathbb{C}^p, \quad \forall \theta \in \Theta, \quad (\text{A.5})$$

cuando $n \rightarrow \infty$. Introduzcamos la siguiente notación: $w_j = (w_{1j}, \dots, w_{pj})$, $\forall j = 1, \dots, n$, donde $w_{lj} \in \mathbb{C}$, $\forall l = 1, \dots, p$. Adicionalmente, sea $w_{lj}^0 = \text{Re}(w_{lj})$ la parte real de w_{lj} y $w_{lj}^1 = \text{Im}(w_{lj})$ la parte imaginaria de w_{lj} . Notemos que $w_{l1}^k, \dots, w_{ln}^k$ son variables aleatorias independientes $\forall k = 0, 1$. Más aún, $w_{lj}^k \sim \mathcal{N}\left(E(w_{lj}^k), \frac{\sigma^2}{2}\right)$. Así, para demostrar (A.5) basta con demostrar que para cualquier $l = 1, \dots, p$, $k = 0, 1$ se tiene que

$$\frac{1}{n} \sum_{i=1}^n (w_{lj}^k - E_{\theta}(w_{lj}^k)) \xrightarrow{P_{\theta}} 0 \in \mathbb{R}, \quad \forall \theta \in \Theta,$$

cuando $n \rightarrow \infty$. En efecto, para $l = 1, \dots, p$ y $k = 0, 1$ arbitrarios, definamos

$$X_n := \frac{1}{n} \sum_{i=1}^n (w_{lj}^k - E_{\theta}(w_{lj}^k)).$$

Para concluir, basta probar que X_n converge a cero en probabilidad. En efecto, aplicando la desigualdad de Chebyshev se tiene que $\forall \epsilon > 0$

$$P_{\theta}(|X_n| > \epsilon) \leq \frac{E_{\theta}(X_n^2)}{\epsilon^2}.$$

A.5. CONVERGENCIA DEL PROMEDIO MUESTRAL

Luego, basta demostrar que X_n converge en \mathbb{L}_2 . En efecto,

$$\begin{aligned} E_{\theta}(X_n^2) &= \frac{1}{n^2} E_{\theta} \left(\left(\sum_{i=1}^n w_{ij}^k - E_{\theta} \left(\sum_{i=1}^n w_{ij}^k \right) \right)^2 \right) \\ &= \frac{1}{n^2} \text{Var}_{\theta} \left(\sum_{i=1}^n w_{ij}^k \right) \\ &= \frac{1}{n^2} \sum_{j=1}^n \text{Var}_{\theta}(w_{ij}^k) \\ &= \frac{1}{n^2} \sum_{j=1}^n \frac{\sigma^2}{2} \\ &= \frac{\sigma^2}{2n} \rightarrow 0, \end{aligned}$$

cuando $n \rightarrow \infty$ y se concluye (A.5). ■

Material Suplementario del Capítulo 3

B.1 Teorema de consistencia de Kolmogorov generalizado

Las definiciones de esta sección fueron sacadas de la sección 15 de [Aliprantis & Border \(2006\)](#) y las damos para que el documento esté autocontenido.

Sean $\{(X_t, \Sigma_t) : t \in T\}$ una familia de espacios medibles. Para todo $H \subseteq T$, con $H \neq \emptyset$, se definen

$$X_H = \prod_{t \in H} X_t \quad \text{y} \quad X_{-H} = \prod_{t \notin H} X_t.$$

Para $H \subseteq G \subseteq T$, sea P_{GH} la proyección natural de X_G sobre X_H .

Definición B.1 (Familia Kolmogorov consistente). *La familia de medidas de probabilidad $\{\mu_F\}$ se denomina **Kolmogorov consistente** si para todo subconjuntos finitos F y G , con $F \subseteq G \subseteq T$, la proyección $P_{GF} : X_G \rightarrow X_F$ satisface que $\mu_G \circ P_{GF}^{-1} = \mu_F$.*

B.1. TEOREMA DE CONSISTENCIA DE KOLMOGOROV GENERALIZADO

Para cada subconjunto finito $F \subseteq T$, $\Sigma_F = \bigotimes_{t \in F} \Sigma_t$ denota la σ -álgebra producto. En este caso, dado $A \in \Sigma_F$, un conjunto de la forma $A \times X_{-F}$ es un F -cilindro. La colección de todos los F -cilindros genera una álgebra de subconjuntos de X_T . La σ -álgebra generada por dicha álgebra se denomina σ -álgebra producto infinita y se denota por $\Sigma_T = \bigotimes_{t \in T} \Sigma_t$.

Una familia Kolmogorov consistente $\{\mu_F\}$ posee una **extensión de Kolmogorov**, si existe una medida de probabilidad μ sobre (X_T, Σ_T) tal que para todo subconjunto finito $F \subseteq T$, $\mu \circ P_{TF}^{-1} = \mu_F$.

Una familia \mathcal{C} de subconjuntos de X posee la propiedad de la **intersección finita** si para cada sub-familia finita y no vacía de \mathcal{C} , la intersección de sus elementos es no vacía. Una familia \mathcal{C} de subconjuntos de X es una **clase compacta** si cada secuencia $\{C_n\}$ en \mathcal{C} con la propiedad de la intersección finita tiene una intersección no vacía.

Propiedad B.1 (Teorema de consistencia de Kolmogorov generalizado). *Sea $\{(X_t, \Sigma_t) : t \in T\}$ una familia de espacios medibles y $\{\mu_F\}$ una familia de medidas de probabilidad Kolmogorov consistente. Supongamos que para cada $t \in T$ hay una clase compacta $\mathcal{C}_t \subseteq \Sigma_t$ tales que*

$$\mu_t(A) = \sup\{\mu_t(C) : C \subseteq A \text{ y } C \in \mathcal{C}_t\},$$

para cada $A \in \Sigma_T$. Entonces hay una única medida de probabilidad μ sobre (X_T, Σ_T) que extiende a cada μ_F , con $F \subseteq T$ finito.

Propiedad B.2 (Corolario del resultado anterior). *Sea $\{X_t : t \in T\}$ una familia de espacios polacos equipados con sus σ -álgebras de Borel y $\{\mu_F\}$ una familia de distribuciones Kolmogorov consistentes. Entonces hay una única medida de probabilidad μ sobre (X_T, Σ_T) que extiende a cada μ_F , con $F \subseteq T$ finito.*

B.2 Demostración del Teorema 3.1

Sean $\Omega_i^\theta, \Omega_i^V \in \mathcal{F}$, con $i \geq 1$, los conjuntos de eventos en los cuales los procesos estocásticos $\{\theta_i(x) : x \in \mathcal{X}\}$ y $\{V_i(x) : x \in \mathcal{X}\}$ tienen trayectorias continuas, respectivamente. Por hipótesis, $\forall i \geq 1, \mathbb{P}(\Omega_i^\theta) = \mathbb{P}(\Omega_i^V) = 1$. Definamos $W := \bigcap_{i=1}^{\infty} (\Omega_i^V \cap \Omega_i^\theta)$. Notemos que $W \in \mathcal{F}$ y $\mathbb{P}(W) = 1$. En efecto,

$$\begin{aligned} \mathbb{P}(W) &= 1 - \mathbb{P}(W^c) \\ &= 1 - \mathbb{P}\left(\bigcup_{i=1}^{\infty} (\Omega_i^V \cap \Omega_i^\theta)^c\right) \\ &\geq 1 - \sum_{i=1}^{\infty} \mathbb{P}\left((\Omega_i^V \cap \Omega_i^\theta)^c\right) \\ &\geq 1 - \sum_{i=1}^{\infty} (\mathbb{P}\left((\Omega_i^V)^c\right) + \mathbb{P}\left((\Omega_i^\theta)^c\right)) \\ &\geq 1. \end{aligned}$$

Luego, $\forall \omega \in W$ las funciones $V_i(\cdot, \omega) \in \mathcal{C}(\mathcal{X}; [0, 1]) := \{V : \mathcal{X} \rightarrow [0, 1] : \text{fn. continua}\}$ y $\theta_i(\cdot, \omega) \in \mathcal{C}(\mathcal{X}; \Theta) := \{\theta : \mathcal{X} \rightarrow \Theta : \text{fn. continua}\}$, con $i \geq 1$. Definamos $N := W^c$. Entonces $\mathbb{P}(N) = 0$. Sean $f_1, \dots, f_k \in C_b(\Theta) := \{f : \Theta \rightarrow \mathbb{R} : \text{fn. continua y acotada}\}$. Entonces $\forall j = 1, \dots, k$,

$$\int_{\Theta} f_j(\theta) G_{x_n}(\omega)(d\theta) = \sum_{i=1}^{\infty} \left\{ V_i(x_n, \omega) \prod_{l < i} [1 - V_l(x_n, \omega)] \right\} f_j(\theta_i(x_n, \omega)),$$

y

$$\int_{\Theta} f_j(\theta) G_{x_0}(\omega)(d\theta) = \sum_{i=1}^{\infty} \left\{ V_i(x_0, \omega) \prod_{l < i} [1 - V_l(x_0, \omega)] \right\} f_j(\theta_i(x_0, \omega)).$$

Notemos que las series anteriores están bien definidas, pues $f_j \circ \theta_i(\cdot)$ es una función acotada, $\forall j = 1, \dots, k$, y los números entre llaves suman 1. En el lema B.2 del apéndice B.18 se demuestra, usando el teorema de la convergencia dominada (ver, por ejemplo, el teorema 16.4

B.3. DEMOSTRACIÓN DEL TEOREMA 3.2

de Billingsley (1995) o el teorema 1.4.49 de Tao (2011)), que

$$\lim_{n \rightarrow \infty} \int_{\Theta} f_j(\theta) G_{x_n}(\omega)(d\theta) = \int_{\Theta} f_j(\theta) G_{x_0}(\omega)(d\theta), \quad \forall j = 1, \dots, k.$$

Así, dado $\epsilon > 0$, $\exists N_j \in \mathbb{N}$ tal que $\forall n \geq N_j$,

$$\left| \int_{\Theta} f_j(\theta) G_{x_n}(\omega)(d\theta) - \int_{\Theta} f_j(\theta) G_{x_0}(\omega)(d\theta) \right| < \epsilon, \quad \forall j = 1, \dots, k.$$

Entonces, tomando $N_0 := \max\{N_1, \dots, N_k\}$ se concluye que G_{x_n} está en una vecindad débil de G_{x_0} , para todo $n \geq N_0$, c.s. y se concluye el teorema.

En la demostración anterior faltó considerar el conjunto $\Omega^* \in \mathcal{F}$, con $P(\Omega^*) = 1$, en el cual es válida la representación stick-breaking

$$G_x(\omega)(B) = \sum_{i=1}^{\infty} \left\{ V_i(x, \omega) \prod_{j < i} [1 - V_j(x, \omega)] \right\} \delta_{\theta_i(x, \omega)}(B),$$

y trabajar con $W \cap \Omega^*$. Sin embargo, se verifica que $\mathbb{P}(W \cap \Omega^*) = 1$, de manera que la omisión de Ω^* no genera ningún problema en la validez del teorema 3.1. ■

B.3 Demostración del Teorema 3.2

Sea

$$G_x(B) = \sum_{i=1}^{\infty} \left\{ V_i(x) \prod_{j < i} [1 - V_j(x)] \right\} \delta_{\theta_i}(B),$$

con $x \in \mathcal{X}$ y $B \in \mathcal{B}(\Theta)$. Entonces para $x, y \in \mathcal{X}$ se tiene que

$$\begin{aligned} G_x(B)G_y(B) &= \sum_{i=1}^{\infty} W_i(x)W_i(y)\delta_{\theta_i}(B) \\ &\quad + \sum_{i \neq j} W_i(x)W_j(y)\delta_{\theta_i}(B)\delta_{\theta_j}(B), \end{aligned}$$

B.3. DEMOSTRACIÓN DEL TEOREMA 3.2

con $W_i(x) = V_i(x) \prod_{j<i} [1 - V_j(x)]$ para $i \geq 2$ y $W_1(x) = V_1(x)$. Luego, tomando esperanza condicional resulta que

$$\begin{aligned} & E(G_x(B)G_y(B)|V_i(x), i = 1, \dots, V_j(y), j = 1, \dots) \\ &= \sum_{i=1}^{\infty} W_i(x)W_i(y)G^0(B) + \sum_{i \neq j} W_i(x)W_j(y)[G^0(B)]^2. \end{aligned}$$

Entonces, por propiedad de las esperanzas iteradas e independencia se tiene que

$$E(G_x(B)G_y(B)) = \sum_{i=1}^{\infty} E[W_i(x)W_i(y)]G_0(B) + \sum_{i \neq j} E[W_i(x)W_j(y)][G_0(B)]^2.$$

Calculemos

$$\begin{aligned} \sum_{i=1}^{\infty} E[W_i(x)W_i(y)] &= \sum_{i=1}^{\infty} E \left[V_i(x) \prod_{j<i} [1 - V_j(x)] V_i(y) \prod_{j<i} [1 - V_j(y)] \right] \\ &= \sum_{i=1}^{\infty} E[V_i(x)V_i(y)] \prod_{j<i} E[(1 - V_j(x))(1 - V_j(y))] \\ &= \sum_{i=1}^{\infty} r(x, y) [1 - m(x) - m(y) + r(x, y)]^{i-1}. \end{aligned}$$

Nos gustaría afirmar que

$$1 - m(x) - m(y) + r(x, y) < 1 \quad \Leftrightarrow \quad r(x, y) < m(x) + m(y).$$

En efecto, notemos que $\forall \omega \in \Omega \setminus N$, con $\mathbb{P}(N) = 0$, $V_1(x, \omega) \leq 1$. Por lo tanto, $V_1(x)V_1(y) \leq V_1(y)$ c.s. y $V_1(x)V_1(y) \leq V_1(x)$ c.s. Así, $E[V_1(x)V_1(y)] \leq E[V_1(x)]$ y $E[V_1(x)V_1(y)] \leq E[V_1(y)]$, y entonces

$$r(x, y) \leq \frac{m(x) + m(y)}{2} < m(x) + m(y).$$

B.3. DEMOSTRACIÓN DEL TEOREMA 3.2

Por tanto,

$$\begin{aligned}
 \sum_{i=1}^{\infty} E[W_i(x)W_i(y)] &= \sum_{i=1}^{\infty} r(x, y)[1 - m(x) - m(y) + r(x, y)]^{i-1} \\
 &= r(x, y) \left(\frac{1}{1 - [1 - m(x) - m(y) + r(x, y)]} \right) \\
 &= \frac{r(x, y)}{m(x) + m(y) - r(x, y)}.
 \end{aligned}$$

Calculemos ahora

$$\begin{aligned}
 \sum_{i \neq j} E[W_i(x)W_j(y)] &= 2 \sum_{i < j} E[W_i(x)W_j(y)] \\
 &= 2 \left\{ \sum_{j=2}^{\infty} E[W_1(x)W_j(y)] + \sum_{j=3}^{\infty} E[W_2(x)W_j(y)] + \dots \right\}.
 \end{aligned}$$

Notemos que

$$\begin{aligned}
 \sum_{j=2}^{\infty} E[W_1(x)W_j(y)] &= \sum_{j=2}^{\infty} E \left[V_1(x)V_j(y) \prod_{l < j} [1 - V_l(y)] \right] \\
 &= \sum_{j=2}^{\infty} E[V_1(x)(1 - V_1(y))] \prod_{l=2}^{j-1} E[1 - V_l(y)] E[V_j(y)] \\
 &= \sum_{j=2}^{\infty} m(y)(m(x) - r(x, y))(1 - m(y))^{j-2}.
 \end{aligned}$$

Dado que

$$1 - m(y) = \frac{\alpha_y}{1 + \alpha_y} < 1,$$

se cumple que

$$\sum_{j=2}^{\infty} E[W_1(x)W_j(y)] = m(x) - r(x, y).$$

Similarmente,

$$\begin{aligned}
 \sum_{j=3}^{\infty} E[W_2(x)W_j(y)] &= \sum_{j=3}^{\infty} E \left[V_2(x)(1 - V_1(x))V_j(y) \prod_{l<j} [1 - V_l(y)] \right] \\
 &= \sum_{j=3}^{\infty} (1 - m(x) - m(y) + r(x, y))(m(x) - r(x, y))(1 - m(y))^{j-3}m(y) \\
 &= (m(x) - r(x, y))(1 - m(x) - m(y) + r(x, y)),
 \end{aligned}$$

y

$$\sum_{j=4}^{\infty} E[W_3(x)W_j(y)] = (m(x) - r(x, y))(1 - m(x) - m(y) + r(x, y))^2.$$

Luego,

$$\begin{aligned}
 \sum_{i \neq j} E[W_i(x)W_j(y)] &= 2 \sum_{i=1}^{\infty} (m(x) - r(x, y))(1 - m(x) - m(y) + r(x, y))^{i-1} \\
 &= 2(m(x) - r(x, y)) \left(\frac{1}{1 - (1 - m(x) - m(y) + r(x, y))} \right) \\
 &= \frac{2(m(x) - r(x, y))}{m(x) + m(y) - r(x, y)}.
 \end{aligned}$$

Así,

$$\begin{aligned}
 E(G_x(B)G_y(B)) &= \left(\frac{r(x, y)}{m(x) + m(y) - r(x, y)} \right) G^0(B) \\
 &\quad + \left(\frac{2[m(x) - r(x, y)]}{m(x) + m(y) - r(x, y)} \right) [G^0(B)]^2.
 \end{aligned}$$

B.3. DEMOSTRACIÓN DEL TEOREMA 3.2

Dado que $E(G_x(B)) = G^0(B)$ se tiene que

$$\begin{aligned} Cov(G_x(B), G_y(B)) &= \left(\frac{r(x, y)}{m(x) + m(y) - r(x, y)} \right) G^0(B) \\ &\quad + \left(\frac{m(x) - m(y) - r(x, y)}{m(x) + m(y) - r(x, y)} \right) [G^0(B)]^2 \\ &= G^0(B) \left(\frac{r(x, y) + G^0(B) [m(x) - m(y) - r(x, y)]}{m(x) + m(y) - r(x, y)} \right). \end{aligned}$$

Notemos que

$$\begin{aligned} r(x, x) &= E[V_1^2(x)] = Var(V_1(x)) + E^2[V_1(x)] \\ &= \frac{\alpha_x}{(1 + \alpha_x)^2(2 + \alpha_x)} + m^2(x) = m(x) \left(\frac{2}{2 + \alpha_x} \right). \end{aligned}$$

Por lo tanto,

$$Var(G_x(B)) = \left(\frac{1}{1 + \alpha_x} \right) G^0(B)(1 - G^0(B)),$$

que es una expresión conocida, pues el DDP es marginalmente (es decir, para todo $x \in \mathcal{X}$) un DP. Finalmente, por hipótesis las funciones $x \rightarrow m(x)$ y $(x, y) \rightarrow r(x, y)$ son continuas. Luego, por álgebra de funciones continuas se concluye que la función $(x, y) \rightarrow Cov(G_x(B), G_y(B))$ es continua para todo $B \in \mathcal{B}(\Theta)$. Así, tomando $y = x_0$, se tiene que $Corr(G_x(B), G_{x_0}(B))$ es una función continua en x y además,

$$\lim_{x \rightarrow x_0} Corr(G_x(B), G_{x_0}(B)) = Corr(G_{x_0}(B), G_{x_0}(B)) = 1,$$

de donde se concluye el teorema. ■

B.4 Demostración del Teorema 3.3

Notemos que $\mathcal{P}(\Theta_0)^{\mathcal{X}}$ es un conjunto cerrado. En efecto, sea $\{\mathcal{P}_n\}_{n \geq 1} \subseteq \mathcal{P}(\Theta_0)^{\mathcal{X}}$ una sucesión, con $\mathcal{P}_n = \{P_x^n : x \in \mathcal{X}\}$, tales que $P_x^n \xrightarrow{d} P_x^0$ cuando $n \rightarrow \infty$, $\forall x \in \mathcal{X}$, con $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^{\mathcal{X}}$. Dado que Θ_0 es un conjunto cerrado, entonces por el teorema de Portmanteau (ver, por ejemplo, teorema 2.1 de [Billingsley, 1999](#)) se tiene que

$$P_x^0(\Theta_0) \geq \limsup_n P_x^n(\Theta_0) = 1, \quad \forall x \in \mathcal{X}.$$

Luego, $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta_0)^{\mathcal{X}}$ y se concluye que $\mathcal{P}(\Theta_0)^{\mathcal{X}}$ es un subconjunto cerrado de $\mathcal{P}(\Theta)^{\mathcal{X}}$, con la topología producto de la convergencia débil de medidas de probabilidad.

Por otra parte, $\mathbb{P}(\omega \in \Omega : \theta_i(x, \omega) \in \Theta_0) = 1, \forall x \in \mathcal{X}, i = 1, 2, \dots$ y entonces

$$\mathbb{P}(\omega \in \Omega : \theta_i(x, \omega) \in \Theta_0, i = 1, 2, \dots) = 1, \quad \forall x \in \mathcal{X}.$$

Así,

$$\mathbb{P} \left(\omega \in \Omega : \sum_{i=1}^{\infty} \left\{ V_i(x, \omega) \prod_{j < i} [1 - V_j(x, \omega)] \right\} \delta_{\theta_i(x, \omega)}(\cdot) \in \mathcal{P}(\Theta_0) \right) = 1, \quad \forall x \in \mathcal{X}.$$

Definamos

$$\Omega_x = \left\{ \omega \in \Omega : \sum_{i=1}^{\infty} \left\{ V_i(x, \omega) \prod_{j < i} [1 - V_j(x, \omega)] \right\} \delta_{\theta_i(x, \omega)}(\cdot) \in \mathcal{P}(\Theta_0) \right\}.$$

La existencia de versiones separables de $\{V_i(x) : x \in \mathcal{X}\}, i \geq 1$, ya fue analizada, y puede ser verificada incluso cuando \mathcal{X} es un espacio métrico separable. La existencia de versiones separables de $\{\theta_i(x) : x \in \mathcal{X}\}, i \geq 1$, puede ser verificada bajo la condición (3.5) o aplicando el Teorema 2, Capítulo 4, de [Gikhman & Skorokhod \(1969\)](#), cuando Θ es un espacio separable localmente compacto¹. En este caso es posible generalizar \mathcal{X} a un espacio métrico separable.

¹Este caso puede ser un poco restrictivo, debido a que la compacidad local en el contexto de espacio vectoriales

B.4. DEMOSTRACIÓN DEL TEOREMA 3.3

Así, la separabilidad de los procesos estocásticos subyacentes garantiza que

$$\bigcap_{x \in \mathcal{X}} \Omega_x = \bigcap_{l=1}^{\infty} \Omega_{x_l},$$

con

$$\bigcap_{x \in \mathcal{X}} \Omega_x = \left\{ \omega \in \Omega : \sum_{i=1}^{\infty} \left\{ V_i(\cdot, \omega) \prod_{j < i} [1 - V_j(\cdot, \omega)] \right\} \delta_{\theta_i(\cdot, \omega)}(\cdot) \in \mathcal{P}(\Theta_0)^{\mathcal{X}} \right\}.$$

Por lo tanto, $\mathbb{P}(\omega \in \Omega : G(\cdot, \omega) \in \mathcal{P}(\Theta_0)^{\mathcal{X}}) = 1$.

Para concluir el teorema hay que probar que $\mathcal{P}(\Theta_0)^{\mathcal{X}}$ es el cerrado más pequeño de $\mathbb{P} \circ G_{\mathcal{X}}^{-1}$ -medida uno. En efecto, sea $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta_0)^{\mathcal{X}}$. Una base de $\{P_x^0 : x \in \mathcal{X}\}$ para la topología producto de la convergencia de débil de medidas de probabilidad está dada por

$$\prod_{i=1}^T \left\{ P_{x_i} \in \mathcal{P}(\Theta_0) : \left| \int_{\Theta_0} f_{ij} dP_{x_i} - \int_{\Theta_0} f_{ij} dP_{x_i}^0 \right| < \epsilon_i, j = 1, \dots, K_i \right\} \times \mathcal{P}(\Theta_0)^{\mathcal{X} \setminus \{x_1, \dots, x_T\}},$$

donde $x_1, \dots, x_T \in \mathcal{X}$; $T, K_1, \dots, K_T \in \mathbb{N} \setminus \{0\}$; $f_{ij} \in \mathcal{C}_b(\Theta_0) := \{f : \Theta_0 \rightarrow \mathbb{R} : \text{fn. continua acotada}\}$ para $i = 1, \dots, T$, $j = 1, \dots, K_i$ y $\epsilon_1, \dots, \epsilon_T \in (0, \infty)$. Así, basta demostrar que la pre-imagen de

$$\prod_{i=1}^T \left\{ P_{x_i} \in \mathcal{P}(\Theta_0) : \left| \int_{\Theta_0} f_{ij}(\theta) dP_{x_i}(\theta) - \int_{\Theta_0} f_{ij}(\theta) dP_{x_i}^0(\theta) \right| < \epsilon_i, j = 1, \dots, K_i \right\}, \quad (\text{B.1})$$

tiene una \mathbb{P} medida positiva para concluir que $\mathcal{P}(\Theta_0)^{\mathcal{X}}$ es el soporte producto débil del DDP. Para esto, en lo que sigue se demostrará que (B.1) contiene un conjunto cuya pre-imagen tiene \mathbb{P} medida positiva. La construcción de a continuación sigue los mismos pasos de la demostración de [Barrientos et al. \(2012\)](#), salvo porque en vez de trabajar con la medida de Lebesgue en \mathbb{R}^q , se considera una medida Boreliana σ -finita con soporte Θ arbitraria $\lambda(\cdot)$. Notemos que λ es localmente finita, por ser σ -finita, y estrictamente positiva, por tener a Θ como soporte. Luego, gracias al Teorema B.1 (demostrado en el apéndice B.16) y al teorema de Portmanteau se tiene

topológicos Hausdorff implica dimensión finita (ver corolario 2, teorema 3, capítulo I, de [Weil, 1967](#))

que el conjunto definido como

$$U(Q_{x_1}, \dots, Q_{x_T}, \{A_{ij}\}, \epsilon_0) \\ = \prod_{i=1}^T \left\{ P_{x_i} \in \mathcal{P}(\Theta_0) : \left| P_{x_i}(A_{ij}) - Q_{x_i}(A_{ij}) \right| < \epsilon_0, j = 1, \dots, m_i \right\}.$$

está contenido en (B.1) donde $Q_{x_i} \in \mathcal{P}(\Theta_0)$ tal que $Q_{x_i} \ll G_{x_i}^0$ y $G_{x_i}^0 \ll \lambda$, $i = 1, \dots, T$; $m_1, \dots, m_T \in \mathbb{N} \setminus \{0\}$; $A_{ij} \subseteq \Theta_0$, $j = 1, \dots, m_i$, $i = 1, \dots, T$, son conjuntos medibles tales que $Q_{x_i}(\partial A_{ij}) = 0$ y $\epsilon_0 > 0$. Sean $\nu_{i,j} \in \{0, 1\}$, $j = 1, \dots, m_i$, $i = 1, \dots, T$, y

$$B_{\nu_{1,1}, \dots, \nu_{T, m_T}} := \bigcap_{i=1}^T \bigcap_{j=1}^{m_i} A_{ij}^{\nu_{i,j}},$$

donde $A_{ij}^1 = A_{ij}$ y $A_{ij}^0 = A_{ij}^c$, con el complemento tomado con respecto a Θ_0 . Entonces, $\{B_{\nu_{1,1}, \dots, \nu_{T, m_T}}\}_{\nu_{i,j} \in \{0,1\}}$ es una partición medible de Θ_0 y sigue que

$$\prod_{i=1}^T \left\{ P_{x_i} \in \mathcal{P}(\Theta_0) : \left| P_{x_i}(B_{\nu_{1,1}, \dots, \nu_{T, m_T}}) - Q_{x_i}(B_{\nu_{1,1}, \dots, \nu_{T, m_T}}) \right| < \epsilon_0 2^{-\sum_{i=1}^T m_i}, \nu_{i,j} \in \{0, 1\} \right\},$$

está contenido en $U(Q_{x_1}, \dots, Q_{x_T}, \{A_{ij}\}, \epsilon_0)$. Simplificando la notación, sea

$$J := \left\{ \boldsymbol{\nu} = (\nu_{1,1}, \dots, \nu_{T, m_T}) : G_{x_i}^0(B_{\nu_{1,1}, \dots, \nu_{T, m_T}}) > 0 \right\}.$$

Notar que el conjunto J no depende de i , debido al soporte común de los $G_{x_i}^0$. Por otra parte, sea $M : J \rightarrow \{0, \dots, k\}$ una biyección con $k = |J| - 1$ y definamos $E_{M(\boldsymbol{\nu})} := B_{\boldsymbol{\nu}}$, $\forall \boldsymbol{\nu} \in J$. Para $i = 1, \dots, T$ se define

$$\mathbf{s}_{x_i} = (w_{(x_i,0)}, \dots, w_{(x_i,k)}) := (Q_{x_i}(E_0), \dots, Q_{x_i}(E_k)) \in \Delta_k,$$

con

$$\Delta_k = \left\{ (w_0, \dots, w_k) : w_i \geq 0, i = 0, \dots, k, \sum_{i=0}^k w_i = 1 \right\},$$

B.4. DEMOSTRACIÓN DEL TEOREMA 3.3

y

$$B(\mathbf{s}_{x_i}, \epsilon) := \{(w_0, \dots, w_k) \in \Delta_k : w_{(x_i, j)} - \epsilon < w_j < w_{(x_i, j)} + \epsilon, j = 0, \dots, k\},$$

con $\epsilon = \epsilon_0 2^{-\sum_{l=1}^T m_l}$. Notemos que

$$\begin{aligned} & \{\omega \in \Omega : (G_{x_i}(\omega)(E_0), \dots, G_{x_i}(\omega)(E_k)) \in B(\mathbf{s}_{x_i}, \epsilon), i = 1, \dots, T\} \\ & \subseteq \{\omega \in \Omega : (G_{x_1}(\omega), \dots, G_{x_T}(\omega)) \in U(Q_{x_1}, \dots, Q_{x_T}, \{A_{ij}\}, \epsilon_0)\}. \end{aligned}$$

Luego, para concluir el teorema basta demostrar que

$$\mathbb{P}(\omega \in \Omega : (G_{x_i}(\omega)(E_0), \dots, G_{x_i}(\omega)(E_k)) \in B(\mathbf{s}_{x_i}, \epsilon), i = 1, \dots, T) > 0. \quad (\text{B.2})$$

Definamos por $\Omega_0 \subseteq \Omega$ al conjunto medible para el cual se verifican las siguientes condiciones:

(a.1) Para $j = 0, \dots, k$,

$$\theta_{j+1}(x_1, \omega), \dots, \theta_{j+1}(x_T, \omega) \in E_j.$$

(a.2) Para $i = 1, \dots, T$,

$$w_{(x_i, 0)} - \frac{\epsilon}{2} < V_1(x_i, \omega) < w_{(x_i, 0)} + \frac{\epsilon}{2}.$$

(a.3) Para $i = 1, \dots, T$ y $j = 1, \dots, k$,

$$\frac{w_{(x_i, j)} - \frac{\epsilon}{2}}{\prod_{l < j+1} [1 - V_l(x_i, \omega)]} < V_{j+1}(x_i, \omega) < \frac{w_{(x_i, j)} + \frac{\epsilon}{2}}{\prod_{l < j+1} [1 - V_l(x_i, \omega)]}.$$

(a.4) Para $i = 1, \dots, T$,

$$\begin{aligned} & \frac{1 - \sum_{j=1}^k \left\{ V_j(x_i, \omega) \prod_{l < j} [1 - V_l(x_i, \omega)] \right\} - \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]} < V_{k+1}(x_i, \omega) \\ & < \frac{1 - \sum_{j=1}^k \left\{ V_j(x_i, \omega) \prod_{l < j} [1 - V_l(x_i, \omega)] \right\} + \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]}. \end{aligned}$$

B.4. DEMOSTRACIÓN DEL TEOREMA 3.3

Notemos que las condiciones anteriores no son únicas y surgen de imponer que

$$w_{(x_i, s)} - \epsilon < \sum_{j=1}^{\infty} \left\{ V_j(x_i, \omega) \prod_{l < j} [1 - V_l(x_i, \omega)] \right\} \delta_{\theta_j(x_i, \omega)}(E_s) < w_{(x_i, s)} + \epsilon,$$

para $i = 1, \dots, T$ y $s = 0, \dots, k$. Así, por ejemplo, la condición (a.4) se obtiene de obligar que

$$-\frac{\epsilon}{2} < \sum_{j=k+2}^{\infty} \left\{ V_j(x_i, \omega) \prod_{l < j} [1 - V_l(x_i, \omega)] \right\} < \frac{\epsilon}{2},$$

para $i = 1, \dots, T$. Se observa que si $\omega \in \Omega_0$, entonces

$$(G_{x_i}(\omega)(E_0), \dots, G_{x_i}(\omega)(E_k)) \in B(\mathbf{s}_{x_i}, \epsilon), \quad i = 1, \dots, T.$$

Sean

$$Q_1 := \prod_{i=1}^T \left[w_{(x_i, 0)} - \frac{\epsilon}{2}, w_{(x_i, 0)} + \frac{\epsilon}{2} \right],$$

para $j = 1, \dots, k-1$,

$$\begin{aligned} Q_{j+1}(\omega) &= Q_{j+1}(V_1(x_1, \omega), \dots, V_j(x_T, \omega)) \\ &:= \prod_{i=1}^T \left[\frac{w_{(x_i, j)} - \frac{\epsilon}{2}}{\prod_{l < j+1} [1 - V_l(x_i, \omega)]}, \frac{w_{(x_i, j)} + \frac{\epsilon}{2}}{\prod_{l < j+1} [1 - V_l(x_i, \omega)]} \right], \end{aligned}$$

y

$$\begin{aligned}
 Q_{k+1}(\omega) &= Q_{k+1}(V_1(x_1, \omega), \dots, V_k(x_T, \omega)) \\
 &:= \prod_{i=1}^T \left[\max \left\{ \frac{1 - \sum_{j=1}^k \left\{ V_j(x_i, \omega) \prod_{l < j} [1 - V_l(x_i, \omega)] \right\} - \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]}, \frac{w(x_i, k) - \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]} \right\}, \right. \\
 &\quad \left. \min \left\{ \frac{1 - \sum_{j=1}^k \left\{ V_j(x_i, \omega) \prod_{l < j} [1 - V_l(x_i, \omega)] \right\} + \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]}, \frac{w(x_i, k) + \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]} \right\} \right].
 \end{aligned}$$

Entonces, por independencia de los procesos estocásticos subyacentes al DDP se tiene que

$$\begin{aligned}
 &\mathbb{P}(\omega \in \Omega : (G_{x_i}(\omega)(E_0), \dots, G_{x_i}(\omega)(E_k)) \in B(\mathbf{s}_{x_i}, \epsilon), i = 1, \dots, T) \\
 &\geq \mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in Q_j(\omega), j = 1, \dots, k+1) \times \\
 &\quad \prod_{j=1}^{k+1} \mathbb{P}(\omega \in \Omega : (\theta_j(x_1, \omega), \dots, \theta_j(x_T, \omega)) \in E_{j-1}^T) \times \\
 &\quad \prod_{j=k+2}^{\infty} \mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in [0, 1]^T) \times \\
 &\quad \prod_{j=k+2}^{\infty} \mathbb{P}(\omega \in \Omega : (\theta_j(x_1, \omega), \dots, \theta_j(x_T, \omega)) \in \Theta_0^T) \\
 &= \mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in Q_j(\omega), j = 1, \dots, k+1) \times \\
 &\quad \prod_{j=1}^{k+1} \mathbb{P}(\omega \in \Omega : (\theta_j(x_1, \omega), \dots, \theta_j(x_T, \omega)) \in E_{j-1}^T),
 \end{aligned}$$

pues por construcción se verifican las igualdades

$$\mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in [0, 1]^T) = 1,$$

y

$$\mathbb{P}(\omega \in \Omega : (\theta_j(x_1, \omega), \dots, \theta_j(x_T, \omega)) \in \Theta_0^T) = 1,$$

B.4. DEMOSTRACIÓN DEL TEOREMA 3.3

para $j = k + 2, \dots, \infty$. Luego, para demostrar (B.2) y concluir el teorema basta probar que

$$\mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in Q_j(\omega), j = 1, \dots, k + 1) > 0,$$

y que para $j = 1, \dots, k + 1$,

$$\mathbb{P}(\omega \in \Omega : (\theta_j(x_1, \omega), \dots, \theta_j(x_T, \omega)) \in E_{j-1}^T) > 0.$$

En efecto, por la condición (ii) del teorema se tiene que $(V_j(x_1, \omega), \dots, V_j(x_T, \omega))$ tiene soporte completo en $[0, 1]^T$. Definamos los eventos

$$\Omega_j^* := \{\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in Q_j(\omega)\}.$$

Por la condición de soporte se tiene que $\mathbb{P}(\Omega_1^*) > 0$. Similarmente,

$$\mathbb{P}\{\omega \in \Omega_1^* : (V_2(x_1, \omega), \dots, V_2(x_T, \omega)) \in Q_2(\omega)\} > 0,$$

y así sucesivamente. Luego, se cumple que

$$\mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in Q_j(\omega), j = 1, \dots, k + 1) > 0.$$

Finalmente, para todo $j = 1, \dots, k + 1$,

$$\mathbb{P}(\omega \in \Omega : (\theta_j(x_1, \omega), \dots, \theta_j(x_T, \omega)) \in E_{j-1}^T) > 0,$$

por las condiciones (i) y (iii) del teorema, y debido a que $\emptyset \neq E_{j-1} \subseteq \Theta_0$, para todo $j = 1, \dots, k + 1$. Esto completa la demostración del teorema. ■

B.5 Demostración del Teorema 3.4

De manera análoga a la demostración del Teorema 3.3, basta probar (B.2), esto es,

$$\mathbb{P}(\omega \in \Omega : (G_{x_i}(\omega)(E_0), \dots, G_{x_i}(\omega)(E_k)) \in B(s_{x_i}, \epsilon), i = 1, \dots, T) > 0. \quad (\text{B.3})$$

En efecto, sea $\Omega_0 \subseteq \Omega$ tales que se verifican las siguientes condiciones:

(b.1) Para $j = 0, \dots, k$,

$$\theta_{j+1}(\omega) \in E_j.$$

(b.2) Para $i = 1, \dots, T$,

$$w_{(x_i,0)} - \frac{\epsilon}{2} < V_1(x_i, \omega) < w_{(x_i,0)} + \frac{\epsilon}{2}.$$

(b.3) Para $i = 1, \dots, T$ y $j = 1, \dots, k$,

$$\frac{w_{(x_i,j)} - \frac{\epsilon}{2}}{\prod_{l < j+1} [1 - V_l(x_i, \omega)]} < V_{j+1}(x_i, \omega) < \frac{w_{(x_i,j)} + \frac{\epsilon}{2}}{\prod_{l < j+1} [1 - V_l(x_i, \omega)]}.$$

(b.4) Para $i = 1, \dots, T$,

$$\begin{aligned} \frac{1 - \sum_{j=1}^k \left\{ V_j(x_i, \omega) \prod_{l < j} [1 - V_l(x_i, \omega)] \right\} - \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]} &< V_{k+1}(x_i, \omega) \\ &< \frac{1 - \sum_{j=1}^k \left\{ V_j(x_i, \omega) \prod_{l < j} [1 - V_l(x_i, \omega)] \right\} + \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]}. \end{aligned}$$

Es claro que si $\omega \in \Omega_0$, entonces $(G_{x_i}(\omega)(E_0), \dots, G_{x_i}(\omega)(E_k)) \in B(s_{x_i}, \epsilon)$, para $i = 1, \dots, T$.

Así, basta demostrar que $\mathbb{P}(\Omega_0) > 0$ para concluir el teorema. Definiendo

$$Q_1 := \prod_{i=1}^T \left[w_{(x_i,0)} - \frac{\epsilon}{2}, w_{(x_i,0)} + \frac{\epsilon}{2} \right],$$

B.5. DEMOSTRACIÓN DEL TEOREMA 3.4

para $j = 1, \dots, k - 1$,

$$\begin{aligned} Q_{j+1}(\omega) &= Q_{j+1}(V_1(x_1, \omega), \dots, V_j(x_T, \omega)) \\ &:= \prod_{i=1}^T \left[\frac{w_{(x_i, j)} - \frac{\epsilon}{2}}{\prod_{l < j+1} [1 - V_l(x_i, \omega)]}, \frac{w_{(x_i, j)} + \frac{\epsilon}{2}}{\prod_{l < j+1} [1 - V_l(x_i, \omega)]} \right], \end{aligned}$$

y

$$\begin{aligned} Q_{k+1}(\omega) &= Q_{k+1}(V_1(x_1, \omega), \dots, V_k(x_T, \omega)) \\ &:= \prod_{i=1}^T \left[\max \left\{ \frac{1 - \sum_{j=1}^k \left\{ V_j(x_i, \omega) \prod_{l < j} [1 - V_l(x_i, \omega)] \right\} - \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]}, \frac{w_{(x_i, k)} - \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]} \right\}, \right. \\ &\quad \left. \min \left\{ \frac{1 - \sum_{j=1}^k \left\{ V_j(x_i, \omega) \prod_{l < j} [1 - V_l(x_i, \omega)] \right\} + \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]}, \frac{w_{(x_i, k)} + \frac{\epsilon}{2}}{\prod_{l < k+1} [1 - V_l(x_i, \omega)]} \right\} \right], \end{aligned}$$

se tiene que

$$\begin{aligned} &\mathbb{P}(\omega \in \Omega : (G_{x_i}(\omega)(E_0), \dots, G_{x_i}(\omega)(E_k)) \in B(s_{x_i}, \epsilon), i = 1, \dots, T) \\ &\geq \mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in Q_j(\omega), j = 1, \dots, k + 1) \times \\ &\quad \prod_{j=1}^{k+1} \mathbb{P}(\omega \in \Omega : \theta_j(\omega) \in E_{j-1}) \times \\ &\quad \prod_{j=k+2}^{\infty} \mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in [0, 1]^T) \times \\ &\quad \prod_{j=k+2}^{\infty} \mathbb{P}(\omega \in \Omega : \theta_j(\omega) \in \Theta_0) \\ &= \mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in Q_j(\omega), j = 1, \dots, k + 1) \times \\ &\quad \prod_{j=1}^{k+1} \mathbb{P}(\omega \in \Omega : \theta_j(\omega) \in E_{j-1}), \end{aligned}$$

B.5. DEMOSTRACIÓN DEL TEOREMA 3.4

pues por construcción se verifican las igualdades

$$\mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in [0, 1]^T) = 1,$$

y $\mathbb{P}(\omega \in \Omega : \theta_j(\omega) \in \Theta_0) = 1$, para $j = k + 2, \dots, \infty$. Luego, basta demostrar que

$$\mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in Q_j(\omega), j = 1, \dots, k + 1) > 0,$$

y que $\mathbb{P}(\omega \in \Omega : \theta_j(\omega) \in E_{j-1}) > 0$ para $j = 1, \dots, k + 1$, para concluir el teorema.

En efecto, razonando como en la demostración del teorema anterior, por la condición (ii) del teorema se tiene que $(V_j(x_1, \omega), \dots, V_j(x_T, \omega))$ tiene soporte completo en $[0, 1]^T$. Definamos los eventos

$$\Omega_j^* := \{\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in Q_j(\omega)\}.$$

Por la condición de soporte se tiene que $\mathbb{P}(\Omega_1^*) > 0$. Similarmente,

$$\mathbb{P}\{\omega \in \Omega_1^* : (V_2(x_1, \omega), \dots, V_2(x_T, \omega)) \in Q_2(\omega)\} > 0,$$

y así sucesivamente. Luego, se cumple que

$$\mathbb{P}(\omega \in \Omega : (V_j(x_1, \omega), \dots, V_j(x_T, \omega)) \in Q_j(\omega), j = 1, \dots, k + 1) > 0.$$

Finalmente, dado que Θ_0 es el soporte topológico de G^0 , por la condición (i) del teorema, se tiene que

$$\mathbb{P}(\omega \in \Omega : \theta_j(\omega) \in E_{j-1}) > 0,$$

para $j = 1, \dots, k + 1$, lo cual completa la demostración del teorema. ■

B.6 Demostración del Teorema 3.5

Siguiendo las ideas de [Pati et al. \(2013\)](#), sea $\{P_x : x \in \mathcal{X}\} \in \mathcal{G}_{\mathcal{X}}^*$ y $\epsilon > 0$. Por demostrar que

$$\mathbb{P} \left\{ \omega \in \Omega : \left| \int_{\mathcal{X}} \int_{\Theta} g(\theta, x) G_x(\omega)(d\theta) q(x) dx - \int_{\mathcal{X}} \int_{\Theta} g(\theta, x) P_x(d\theta) q(x) dx \right| < \epsilon \right\} > 0,$$

con $g : \Theta \times \mathcal{X} \rightarrow [0, 1]$ una función uniformemente continua. Sin pérdida de generalidad se puede suponer que $\exists \Theta_1 \subseteq \Theta$ (compacto) tal que $\sup_{x \in \mathcal{X}} |g(\theta, x)| < \epsilon, \forall \theta \in \Theta_1^c$. Por otro lado, se puede demostrar gracias a la propiedad de tensión de medidas de probabilidad sobre un espacio Polaco, la continuidad de $x \rightarrow P_x(\Theta_1)$ y la compacidad de \mathcal{X} , que $\exists \Theta_2 \subseteq \Theta$ (compacto) tal que $P_x(\Theta_2) > 1 - \epsilon, \forall x \in \mathcal{X}$. Definamos $\Theta_0 = \Theta_1 \cup \Theta_2$. Claramente dicho conjunto es compacto y satisface las condiciones anteriores. Entonces

$$\begin{aligned} & \int_{\mathcal{X}} \left(\int_{\Theta} g(\theta, x) G_x(\omega)(d\theta) - \int_{\Theta} g(\theta, x) P_x(d\theta) \right) q(x) dx \\ & \leq \int_{\mathcal{X}} \left(\sum_{i=1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega), x) - \int_{\Theta_0} g(\theta, x) P_x(d\theta) \right) q(x) dx \\ & + \int_{\mathcal{X}} \left(\int_{\Theta_0^c} g(\theta, x) P_x(d\theta) \right) q(x) dx \\ & \leq \int_{\mathcal{X}} \left(\sum_{i=1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega), x) - \int_{\Theta_0} g(\theta, x) P_x(d\theta) \right) q(x) dx + \epsilon, \end{aligned}$$

pues

$$\begin{aligned} \int_{\mathcal{X}} \left(\int_{\Theta_0^c} g(\theta, x) P_x(d\theta) \right) q(x) dx & \leq \int_{\mathcal{X}} P_x(\Theta_0^c) q(x) dx \\ & < \epsilon \int_{\mathcal{X}} q(x) dx = \epsilon, \end{aligned}$$

donde $\pi_i(x, \omega) = V_i(x, \omega) \prod_{l < i} [1 - V_l(x, \omega)]$.

Sean $\{A_{k,n} : k = 1, \dots, n\}$ una sucesión de particiones medibles de Θ_0 con refinamiento creciente cuando $n \rightarrow \infty$ y tales que $\max_{1 \leq k \leq n} \text{diam}(A_{k,n}) \rightarrow 0$ cuando $n \rightarrow \infty$. Fijando $\tilde{\theta}_{k,n} \in$

B.6. DEMOSTRACIÓN DEL TEOREMA 3.5

$A_{k,n}$, $k = 1, \dots, n$, entonces por el teorema de convergencia dominada

$$\int_{\mathcal{X}} \left(\sum_{k=1}^n g(\tilde{\theta}_{k,n}, x) P_x(A_{k,n}) \right) q(x) dx \rightarrow \int_{\mathcal{X}} \int_{\Theta_0} g(\theta, x) P_x(d\theta) q(x) dx,$$

cuando $n \rightarrow \infty$, pues por (iv) $\sup_{x \in \mathcal{X}} P_x(A_{k,n}) \rightarrow 0$ cuando $n \rightarrow \infty$ y entonces $h_n(x) = \sum_{k=1}^n g(\tilde{\theta}_{k,n}, x) P_x(A_{k,n}) \rightarrow h(x) = \int_{\Theta_0} g(\theta, x) P_x(d\theta)$, $\forall x \in \mathcal{X}$, por ser una suma de Riemann con g Lebesgue integrable, y finalmente $h_n(x)$ es una función uniformemente acotada.

Así, $\exists n_1 \in \mathbb{N}$ tal que $\forall n \geq n_1$,

$$\begin{aligned} & \left| \int_{\mathcal{X}} \left(\int_{\Theta} g(\theta, x) G_x(\omega)(d\theta) - \int_{\Theta} g(\theta, x) P_x(d\theta) \right) q(x) dx \right| \\ & \leq \left| \int_{\mathcal{X}} \left(\sum_{i=1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega), x) - \sum_{k=1}^n g(\tilde{\theta}_{k,n}, x) P_x(A_{k,n}) \right) q(x) dx \right| \\ & + \left| \int_{\mathcal{X}} \left(\sum_{k=1}^n g(\tilde{\theta}_{k,n}, x) P_x(A_{k,n}) - \int_{\Theta_0} g(\theta, x) P_x(d\theta) \right) q(x) dx \right| + \epsilon \\ & \leq \left| \int_{\mathcal{X}} \left(\sum_{i=1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega), x) - \sum_{k=1}^n g(\tilde{\theta}_{k,n}, x) P_x(A_{k,n}) \right) q(x) dx \right| + 2\epsilon. \end{aligned}$$

Considere el conjunto

$$\Omega_1 = \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |\pi_1(x, \omega) - P_x(A_{1,n_1})| < \frac{\epsilon}{n_1}, \dots, \sup_{x \in \mathcal{X}} |\pi_{n_1}(x, \omega) - P_x(A_{n_1,n_1})| < \frac{\epsilon}{n_1} \right\}.$$

Por el lema B.1, que es demostrado en el apéndice B.17, se tiene que $\mathbb{P}(\Omega_1) > 0$. Dado que $\sum_{i=1}^{\infty} \pi_i(x, \omega) = 1$ c.s., existe $\Omega_2 \subseteq \Omega$ con $\mathbb{P}(\Omega_2) = 1$ tal que $\forall \omega \in \Omega_1 \cap \Omega_2$, $g_n(x, \omega) := \sum_{i=1}^n \pi_i(x, \omega) \rightarrow 1$ cuando $n \rightarrow \infty$, $\forall x \in \mathcal{X}$ (claramente $\mathbb{P}(\Omega_1 \cap \Omega_2) > 0$). La convergencia anterior es uniforme gracias a que se verifican las hipótesis del teorema de Dini, esto es, $\{g_n\}_{n \geq 1}$ son funciones continuas definidas en un compacto y monótonas crecientes, que convergen a la función continua idénticamente igual a 1. Gracias a el teorema de Egoroff existe $\Omega_3 \subseteq \Omega_2 \cap \Omega_1$, con $\mathbb{P}(\Omega_3) > 0$, tal que $g_n(x, \omega) \rightarrow 1$ uniformemente en $x \in \mathcal{X}$ y en $\omega \in \Omega_3$. Así, $\exists n_\epsilon \geq n_1$

B.6. DEMOSTRACIÓN DEL TEOREMA 3.5

(independiente de x y ω) tal que $\sum_{i=n_\epsilon+1}^{\infty} \pi_i(x, \omega) < \epsilon, \forall \omega \in \Omega_3, x \in \mathcal{X}$.

Se define el conjunto

$$\Omega_4 = \Omega_3 \cap \{\omega \in \Omega : \theta_{n_1+1}(\omega) \in \Theta_0, \dots, \theta_{n_\epsilon-1}(\omega) \in \Theta_0\}.$$

Para $\omega \in \Omega_4$ se tiene que

$$\begin{aligned} & \left| \int_{\mathcal{X}} \left(\int_{\Theta} g(\theta, x) G_x(\omega)(d\theta) - \int_{\Theta} g(\theta, x) P_x(d\theta) \right) q(x) dx \right| \\ & \leq \int_{\mathcal{X}} \left(\sum_{k=1}^{n_1} \left| \pi_k(x, \omega) g(\theta_k(\omega), x) - g(\tilde{\theta}_{k, n_1}, x) P_x(A_{k, n_1}) \right| \right) q(x) dx + 4\epsilon, \end{aligned}$$

pues

$$\begin{aligned} & \left| \int_{\mathcal{X}} \left(\sum_{i=1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega), x) - \sum_{k=1}^{n_1} g(\tilde{\theta}_{k, n_1}, x) P_x(A_{k, n_1}) \right) q(x) dx \right| \\ & = \left| \int_{\mathcal{X}} \left(\sum_{i=1}^{n_1} \pi_i(x, \omega) g(\theta_i(\omega), x) - \sum_{k=1}^{n_1} g(\tilde{\theta}_{k, n_1}, x) P_x(A_{k, n_1}) \right) q(x) dx \right. \\ & \quad \left. + \int_{\mathcal{X}} \sum_{i=n_1+1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega), x) q(x) dx \right| \\ & \leq \int_{\mathcal{X}} \left(\sum_{k=1}^{n_1} \left| \pi_k(x, \omega) g(\theta_k(\omega), x) - g(\tilde{\theta}_{k, n_1}, x) P_x(A_{k, n_1}) \right| \right) q(x) dx \\ & \quad + \int_{\mathcal{X}} \sum_{i=n_1+1}^{n_\epsilon} \pi_i(x, \omega) g(\theta_i(\omega), x) q(x) dx \\ & \quad + \int_{\mathcal{X}} \sum_{i=n_\epsilon+1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega), x) q(x) dx, \end{aligned}$$

$$\begin{aligned} \int_{\mathcal{X}} \sum_{i=n_1+1}^{n_\epsilon} \pi_i(x, \omega) g(\theta_i(\omega), x) q(x) dx & < \epsilon \int_{\mathcal{X}} \sum_{i=n_1+1}^{n_\epsilon} \pi_i(x, \omega) q(x) dx \\ & \leq \epsilon \int_{\mathcal{X}} q(x) dx = \epsilon, \end{aligned}$$

y

$$\begin{aligned} \int_{\mathcal{X}} \sum_{i=n_\epsilon+1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega), x) q(x) dx &\leq \int_{\mathcal{X}} \sum_{i=n_\epsilon+1}^{\infty} \pi_i(x, \omega) q(x) dx \\ &< \epsilon \int_{\mathcal{X}} q(x) dx = \epsilon. \end{aligned}$$

Por otra parte,

$$\begin{aligned} &\int_{\mathcal{X}} \left(\sum_{k=1}^{n_1} \left| \pi_k(x, \omega) g(\theta_k(\omega), x) - g(\tilde{\theta}_{k, n_1}, x) P_x(A_{k, n_1}) \right| \right) q(x) dx \\ &\leq \sum_{k=1}^{n_1} \int_{\mathcal{X}} \pi_k(x, \omega) \left| g(\theta_k(\omega), x) - g(\tilde{\theta}_{k, n_1}, x) \right| q(x) dx + \epsilon. \end{aligned}$$

En efecto,

$$\begin{aligned} &\sum_{k=1}^{n_1} \left| \pi_k(x, \omega) g(\theta_k(\omega), x) - g(\tilde{\theta}_{k, n_1}, x) P_x(A_{k, n_1}) \right| \\ &= \sum_{k=1}^{n_1} \left| \pi_k(x, \omega) g(\theta_k(\omega), x) - \pi_k(x, \omega) g(\tilde{\theta}_{k, n_1}, x) \right. \\ &\quad \left. + \pi_k(x, \omega) g(\tilde{\theta}_{k, n_1}, x) - g(\tilde{\theta}_{k, n_1}, x) P_x(A_{k, n_1}) \right| \\ &\leq \sum_{k=1}^{n_1} \pi_k(x, \omega) |g(\theta_k(\omega), x) - g(\tilde{\theta}_{k, n_1}, x)| \\ &\quad + \sum_{k=1}^{n_1} g(\tilde{\theta}_{k, n_1}, x) |\pi_k(x, \omega) - P_x(A_{k, n_1})|, \end{aligned}$$

y

$$\begin{aligned} \sum_{k=1}^{n_1} g(\tilde{\theta}_{k, n_1}, x) |\pi_k(x, \omega) - P_x(A_{k, n_1})| &\leq \sum_{k=1}^{n_1} |\pi_k(x, \omega) - P_x(A_{k, n_1})| \\ &< \sum_{k=1}^{n_1} \frac{\epsilon}{n_1} = \epsilon. \end{aligned}$$

B.7. DEMOSTRACIÓN DEL TEOREMA 3.6

Por otro lado, dada la continuidad uniforme de $g(\cdot, \cdot)$ existen conjuntos B_k , $k = 1, \dots, n_1$, independientes de x tales que si

$$(\theta_1(\omega), \dots, \theta_{n_1}(\omega)) \in B_1 \times \dots \times B_{n_1},$$

entonces $|g(\theta_k(\omega), x) - g(\tilde{\theta}_{k,n_1}, x)| < \epsilon$, $k = 1, \dots, n_1$. Más aún,

$$\begin{aligned} \int_{\mathcal{X}} \sum_{k=1}^{n_1} \pi_k(x, \omega) |g(\theta_k(\omega), x) - g(\tilde{\theta}_{k,n_1}, x)| q(x) dx &< \epsilon \int_{\mathcal{X}} \sum_{k=1}^{n_1} \pi_k(x, \omega) q(x) dx \\ &\leq \epsilon \int_{\mathcal{X}} q(x) dx = \epsilon. \end{aligned}$$

Luego, para $\omega \in \Omega_5 = \Omega_4 \cap \{\omega \in \Omega : \theta_1(\omega) \in B_1, \dots, \theta_{n_1}(\omega) \in B_{n_1}\}$ se tiene que

$$\left| \int_{\mathcal{X}} \int_{\Theta} g(\theta, x) G_x(\omega)(d\theta) q(x) dx - \int_{\mathcal{X}} \int_{\Theta} g(\theta, x) P_x(d\theta) q(x) dx \right| < 5\epsilon.$$

Finalmente, dado que $\mathbb{P}(\Omega_3) > 0$ y $\{\omega \in \Omega : \theta_{n_1+1}(\omega) \in \Theta_0, \dots, \theta_{n_e-1}(\omega) \in \Theta_0\}$, $\{\omega \in \Omega : \theta_1(\omega) \in B_1, \dots, \theta_{n_1}(\omega) \in B_{n_1}\}$ son eventos independientes de probabilidad positiva sigue que $\mathbb{P}(\Omega_5) > 0$ y se concluye el teorema. ■

B.7 Demostración del Teorema 3.6

Siguiendo las ideas de [Pati et al. \(2013\)](#), sea $\{P_x : x \in \mathcal{X}\} \in \mathcal{G}_{\mathcal{X}}^*$ con soporte compacto común y $\epsilon > 0$. Por demostrar que

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \left| \int_{\Theta} g(\theta) G_x(\omega)(d\theta) - \int_{\Theta} g(\theta) P_x(d\theta) \right| < \epsilon \right\} > 0,$$

B.7. DEMOSTRACIÓN DEL TEOREMA 3.6

con $g : \Theta \rightarrow [0, 1]$ una función uniformemente continua que se anula en infinito. Así, $\exists \Theta_1 \subseteq \Theta$ (compacto) tal que $|g(\theta)| < \epsilon, \forall \theta \in \Theta_1^c$. Por otro lado, se puede demostrar gracias a la propiedad de tensión de medidas de probabilidad sobre un espacio polaco, la continuidad de $x \rightarrow P_x(\Theta_1)$ y la compacidad de \mathcal{X} , que $\exists \Theta_2 \subseteq \Theta$ (compacto) tal que $P_x(\Theta_2) > 1 - \epsilon, \forall x \in \mathcal{X}$. Definamos $\Theta_0 = \Theta_1 \cup \Theta_2$. Claramente dicho conjunto es compacto y satisface las condiciones anteriores. Entonces

$$\begin{aligned} & \sup_{x \in \mathcal{X}} \left| \int_{\Theta} g(\theta) G_x(\omega)(d\theta) - \int_{\Theta} g(\theta) P_x(d\theta) \right| \\ & \leq \sup_{x \in \mathcal{X}} \left| \sum_{i=1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega)) - \int_{\Theta_0} g(\theta) P_x(d\theta) \right| + \sup_{x \in \mathcal{X}} \int_{\Theta_0^c} g(\theta) P_x(d\theta) \\ & \leq \sup_{x \in \mathcal{X}} \left| \sum_{i=1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega)) - \int_{\Theta_0} g(\theta) P_x(d\theta) \right| + \epsilon, \end{aligned}$$

pues

$$\sup_{x \in \mathcal{X}} \int_{\Theta_0^c} g(\theta) P_x(d\theta) \leq \sup_{x \in \mathcal{X}} P_x(\Theta_0^c) < \epsilon,$$

donde $\pi_i(x, \omega) = V_i(x, \omega) \prod_{l < i} [1 - V_l(x, \omega)]$.

Sean $\{A_{k,n} : k = 1, \dots, n\}$ una sucesión de particiones medibles de Θ_0 con refinamiento creciente cuando $n \rightarrow \infty$ y tales que $\max_{1 \leq k \leq n} \text{diam}(A_{k,n}) \rightarrow 0$ cuando $n \rightarrow \infty$. Tomando $\tilde{\theta}_{k,n} = \arg \max_{\theta \in A_{k,n}} g(\theta) \in A_{k,n}, k = 1, \dots, n$, entonces por el teorema de Dini se tiene que

$$\sup_{x \in \mathcal{X}} \left| \sum_{k=1}^n g(\tilde{\theta}_{k,n}) P_x(A_{k,n}) - \int_{\Theta_0} g(\theta) P_x(d\theta) \right| \rightarrow 0,$$

cuando $n \rightarrow \infty$, pues $h_n(x) = \sum_{k=1}^n g(\tilde{\theta}_{k,n}) P_x(A_{k,n}) \rightarrow h(x) = \int_{\Theta_0} g(\theta) P_x(d\theta), \forall x \in \mathcal{X}$, $\{h_n\}$ es monótona decreciente, \mathcal{X} es compacto y $h(\cdot)$ es una función continua gracias a (iii).

B.7. DEMOSTRACIÓN DEL TEOREMA 3.6

Así, $\exists n_1 \in \mathbb{N}$, tal que $\forall n \geq n_1$,

$$\begin{aligned}
& \sup_{x \in \mathcal{X}} \left| \int_{\Theta} g(\theta) G_x(\omega)(d\theta) - \int_{\Theta} g(\theta) P_x(d\theta) \right| \\
& \leq \sup_{x \in \mathcal{X}} \left| \sum_{i=1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega)) - \sum_{k=1}^n g(\tilde{\theta}_{k,n}) P_x(A_{k,n}) \right| \\
& + \sup_{x \in \mathcal{X}} \left| \sum_{k=1}^n g(\tilde{\theta}_{k,n}) P_x(A_{k,n}) - \int_{\Theta_0} g(\theta) P_x(d\theta) \right| + \epsilon \\
& \leq \sup_{x \in \mathcal{X}} \left| \sum_{i=1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega)) - \sum_{k=1}^n g(\tilde{\theta}_{k,n}) P_x(A_{k,n}) \right| + 2\epsilon.
\end{aligned}$$

Considere el conjunto

$$\Omega_1 = \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |\pi_1(x, \omega) - P_x(A_{1,n_1})| < \frac{\epsilon}{n_1}, \dots, \sup_{x \in \mathcal{X}} |\pi_{n_1}(x, \omega) - P_x(A_{n_1,n_1})| < \frac{\epsilon}{n_1} \right\}.$$

Por el lema B.1, que es demostrado en el apéndice B.17, se tiene que $\mathbb{P}(\Omega_1) > 0$. Dado que $\sum_{i=1}^{\infty} \pi_i(x, \omega) = 1$ c.s., existe $\Omega_2 \subseteq \Omega$ con $\mathbb{P}(\Omega_2) = 1$ tal que $\forall \omega \in \Omega_1 \cap \Omega_2$, $g_n(x, \omega) := \sum_{i=1}^n \pi_i(x, \omega) \rightarrow 1$ cuando $n \rightarrow \infty$, $\forall x \in \mathcal{X}$ (claramente $\mathbb{P}(\Omega_1 \cap \Omega_2) > 0$). La convergencia anterior es uniforme gracias a que se verifican las hipótesis del teorema de Dini, esto es, $\{g_n\}_{n \geq 1}$ son funciones continuas definidas en un compacto y monótonas crecientes, que convergen a la función continua idénticamente igual a 1. Gracias a el teorema de Egoroff existe $\Omega_3 \subseteq \Omega_2 \cap \Omega_1$, con $\mathbb{P}(\Omega_3) > 0$, tal que $g_n(x, \omega) \rightarrow 1$ uniformemente en $x \in \mathcal{X}$ y en $\omega \in \Omega_3$. Así, $\exists n_\epsilon \geq n_1$ (independiente de x y ω) tal que $\sum_{i=n_\epsilon+1}^{\infty} \pi_i(x, \omega) < \epsilon$, $\forall \omega \in \Omega_3, x \in \mathcal{X}$.

Se define el conjunto

$$\Omega_4 = \Omega_3 \cap \{ \omega \in \Omega : \theta_{n_1+1}(\omega) \in \Theta_0, \dots, \theta_{n_\epsilon-1}(\omega) \in \Theta_0 \}.$$

B.7. DEMOSTRACIÓN DEL TEOREMA 3.6

Para $\omega \in \Omega_4$ se tiene que

$$\begin{aligned} & \sup_{x \in \mathcal{X}} \left| \int_{\Theta} g(\theta) G_x(\omega)(d\theta) - \int_{\Theta} g(\theta) P_x(d\theta) \right| \\ & \leq \sup_{x \in \mathcal{X}} \left(\sum_{k=1}^{n_1} \left| \pi_k(x, \omega) g(\theta_k(\omega), x) - g(\tilde{\theta}_{k, n_1}, x) P_x(A_{k, n_1}) \right| \right) + 4\epsilon, \end{aligned}$$

pues

$$\begin{aligned} & \sup_{x \in \mathcal{X}} \left| \sum_{i=1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega)) - \sum_{k=1}^{n_1} g(\tilde{\theta}_{k, n_1}) P_x(A_{k, n_1}) \right| \\ & \leq \sup_{x \in \mathcal{X}} \left| \sum_{k=1}^{n_1} \pi_k(x, \omega) g(\theta_k(\omega)) - \sum_{k=1}^{n_1} g(\tilde{\theta}_{k, n_1}) P_x(A_{k, n_1}) \right| \\ & + \sup_{x \in \mathcal{X}} \sum_{i=n_1+1}^{n_\epsilon} \pi_i(x, \omega) g(\theta_i(\omega)) + \sup_{x \in \mathcal{X}} \sum_{i=n_\epsilon+1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega)), \end{aligned}$$

$$\sup_{x \in \mathcal{X}} \sum_{i=n_1+1}^{n_\epsilon} \pi_i(x, \omega) g(\theta_i(\omega)) < \epsilon \sup_{x \in \mathcal{X}} \sum_{i=n_1+1}^{n_\epsilon} \pi_i(x, \omega) \leq \epsilon,$$

y

$$\sup_{x \in \mathcal{X}} \sum_{i=n_\epsilon+1}^{\infty} \pi_i(x, \omega) g(\theta_i(\omega)) \leq \sup_{x \in \mathcal{X}} \sum_{i=n_\epsilon+1}^{\infty} \pi_i(x, \omega) < \epsilon.$$

Por otra parte,

$$\begin{aligned} & \sup_{x \in \mathcal{X}} \left(\sum_{k=1}^{n_1} \left| \pi_k(x, \omega) g(\theta_k(\omega)) - g(\tilde{\theta}_{k, n_1}) P_x(A_{k, n_1}) \right| \right) \\ & \leq \sup_{x \in \mathcal{X}} \sum_{k=1}^{n_1} \pi_k(x, \omega) \left| g(\theta_k(\omega)) - g(\tilde{\theta}_{k, n_1}) \right| + \epsilon. \end{aligned}$$

En efecto,

$$\begin{aligned}
 & \sum_{k=1}^{n_1} \left| \pi_k(x, \omega) g(\theta_k(\omega)) - g(\tilde{\theta}_{k,n_1}) P_x(A_{k,n_1}) \right| \\
 &= \sum_{k=1}^{n_1} \left| \pi_k(x, \omega) g(\theta_k(\omega)) - \pi_k(x, \omega) g(\tilde{\theta}_{k,n_1}) \right. \\
 & \quad \left. + \pi_k(x, \omega) g(\tilde{\theta}_{k,n_1}, x) - g(\tilde{\theta}_{k,n_1}, x) P_x(A_{k,n_1}) \right| \\
 &\leq \sum_{k=1}^{n_1} \pi_k(x, \omega) |g(\theta_k(\omega)) - g(\tilde{\theta}_{k,n_1})| \\
 & \quad + \sum_{k=1}^{n_1} g(\tilde{\theta}_{k,n_1}) |\pi_k(x, \omega) - P_x(A_{k,n_1})|,
 \end{aligned}$$

y

$$\begin{aligned}
 \sum_{k=1}^{n_1} g(\tilde{\theta}_{k,n_1}, x) |\pi_k(x, \omega) - P_x(A_{k,n_1})| &\leq \sum_{k=1}^{n_1} |\pi_k(x, \omega) - P_x(A_{k,n_1})| \\
 &< \sum_{k=1}^{n_1} \frac{\epsilon}{n_1} = \epsilon.
 \end{aligned}$$

Por otro lado, dada la continuidad uniforme de $g(\cdot)$ existen conjuntos B_k , $k = 1, \dots, n_1$, tales que si

$$(\theta_1(\omega), \dots, \theta_{n_1}(\omega)) \in B_1 \times \dots \times B_{n_1},$$

entonces $|g(\theta_k(\omega)) - g(\tilde{\theta}_{k,n_1})| < \epsilon$, $k = 1, \dots, n_1$. Más aún,

$$\sup_{x \in \mathcal{X}} \sum_{k=1}^{n_1} \pi_k(x, \omega) |g(\theta_k(\omega)) - g(\tilde{\theta}_{k,n_1})| < \epsilon \sup_{x \in \mathcal{X}} \sum_{k=1}^{n_1} \pi_k(x, \omega) \leq \epsilon.$$

Luego, para $\omega \in \Omega_5 = \Omega_4 \cap \{\omega \in \Omega : \theta_1(\omega) \in B_1, \dots, \theta_{n_1}(\omega) \in B_{n_1}\}$ se tiene que

$$\sup_{x \in \mathcal{X}} \left| \int_{\Theta} g(\theta) G_x(\omega)(d\theta) - \int_{\Theta} g(\theta) P_x(d\theta) \right| < 5\epsilon.$$

Finalmente, dado que $\mathbb{P}(\Omega_3) > 0$ y $\{\omega \in \Omega : \theta_{n_1+1}(\omega) \in \Theta_0, \dots, \theta_{n_\epsilon-1}(\omega) \in \Theta_0\}$, $\{\omega \in \Omega : \theta_1(\omega) \in B_1, \dots, \theta_{n_1}(\omega) \in B_{n_1}\}$ son eventos independientes de probabilidad positiva, sigue que $\mathbb{P}(\Omega_5) > 0$ y se concluye el teorema. ■

B.8 Demostración del Teorema 3.7

Siguiendo los argumentos de [Barrientos et al. \(2012\)](#), sea $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^\mathcal{X}$ y $\{f_x^0 : x \in \mathcal{X}\}$ con

$$f_x^0(\cdot) = \int_{\Theta} \psi(\cdot, \theta) P_x^0(d\theta), \quad x \in \mathcal{X}.$$

Notar que $\forall x \in \mathcal{X}$ se tiene que $f_x^0(\cdot)$ es una función de densidad de probabilidad en \mathcal{Y} . En efecto, dado $x \in \mathcal{X}$ se tiene que

$$\begin{aligned} \int_{\mathcal{Y}} f_x^0(y) \lambda_{\mathcal{Y}}(dy) &= \int_{\mathcal{Y}} \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\ &= \int_{\Theta} \int_{\mathcal{Y}} \psi(y, \theta) \lambda_{\mathcal{Y}}(dy) P_x^0(d\theta) \\ &= \int_{\Theta} P_x^0(d\theta) = P_x^0(\Theta) = 1, \end{aligned}$$

donde la segunda igualdad se satisface gracias a que la función ψ es medible y positiva, de manera que se tienen las hipótesis del teorema de Tonelli (ver página 147 de [Cohn, 2013](#)). La tercera igualdad se tiene por (i). En todo lo que sigue, se considera la topología producto inducida por la métrica de Hellinger.

Sean $\epsilon > 0$ y $x_1, \dots, x_T \in \mathcal{X}$ puntos arbitrarios. Para concluir el teorema basta con demostrar que el modelo de mezclas inducido por el DDP asigna probabilidad positiva a un con-

junto de la forma

$$\prod_{i=1}^T \left\{ \int_{\Theta} \psi(\cdot, \theta) P_{x_i}(d\theta) : \int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) < \epsilon, P_{x_i} \in \mathcal{P}(\Theta) \right\}. \quad (\text{B.4})$$

Dado que el DDP tiene soporte completo, basta construir una vecindad débil de $\{P_x^0 : x \in \mathcal{X}\}$, tal que sus elementos satisfagan (B.4).

El carácter de espacio Polaco de \mathcal{Y} garantiza que la medida de probabilidad definida por f_x^0 es tensa, para todo $x \in \mathcal{X}$. Así, existe un compacto $K_{x_i} \subseteq \mathcal{Y}$, $i = 1, \dots, T$, tal que

$$\int_{K_{x_i}^c} f_{x_i}^0(y) \lambda_{\mathcal{Y}}(dy) < \frac{\epsilon}{8},$$

y $\lambda_{\mathcal{Y}}(K_{x_i}) < \infty$.

Para $i = 1, \dots, T$ se define

$$h_{i,1}^0(\theta) = \int_{K_{x_i}^c} \psi(y, \theta) \lambda_{\mathcal{Y}}(dy).$$

Notar que $h_{i,1}^0$ es una función acotada y continua. En efecto, $\forall \theta \in \Theta$,

$$|h_{i,1}^0(\theta)| \leq \int_{\mathcal{Y}} \psi(y, \theta) \lambda_{\mathcal{Y}}(dy) = 1,$$

y para $\theta_1, \theta_2 \in \Theta$,

$$\begin{aligned}
 |h_{i,1}^0(\theta_1) - h_{i,1}^0(\theta_2)| &= \left| \int_{K_{x_i}^c} \psi(y, \theta_1) \lambda_{\mathcal{Y}}(dy) - \int_{K_{x_i}^c} \psi(y, \theta_2) \lambda_{\mathcal{Y}}(dy) \right| \\
 &= \left| \int_{K_{x_i}} \psi(y, \theta_1) \lambda_{\mathcal{Y}}(dy) - \int_{K_{x_i}} \psi(y, \theta_2) \lambda_{\mathcal{Y}}(dy) \right| \\
 &\leq \int_{K_{x_i}} |\psi(y, \theta_1) - \psi(y, \theta_2)| \lambda_{\mathcal{Y}}(dy) \\
 &\leq \left(\sup_{y \in K_{x_i}} |\psi(y, \theta_1) - \psi(y, \theta_2)| \right) \lambda_{\mathcal{Y}}(K_{x_i}).
 \end{aligned}$$

La equicontinuidad de $\psi(y, \cdot)$, $\forall y \in K_{x_i}$, y la finitud de $\lambda_{\mathcal{Y}}(K_{x_i})$, garantizan la continuidad de $h_{i,1}^0$.

Dado $\rho > 0$, similarmente, por el carácter de espacio polaco de Θ , existe un compacto $R_{x_i} \subseteq \Theta$, $i = 1, \dots, T$, tal que $P_{x_i}^0(R_{x_i}^c) < \rho$. Se define $R := \bigcup_{i=1}^T R_{x_i}$. Notar que dicho conjunto es cerrado y que $\forall i = 1, \dots, T$, se tiene que

$$P_{x_i}^0(R^c) = P_{x_i}^0\left(\bigcap_{i=1}^T R_{x_i}^c\right) \leq P_{x_i}^0(R_{x_i}^c) < \rho.$$

Se define

$$R_{x_i,r} := \{\theta \in \Theta : d(\theta, R_{x_i}) < r\},$$

con $r > 0$. Notar que este conjunto es abierto y por tanto, $R_{x_i,r}^c$ es un conjunto cerrado. Similarmente se define $R_r := \bigcup_{i=1}^T R_{x_i,r}$. Este conjunto es abierto y por lo tanto R_r^c es un conjunto cerrado. Luego, las funciones $d(\cdot, R)$ y $d(\cdot, R_r^c)$ son continuas y, por álgebra de funciones continuas, las funciones

$$h_{i,2}^0(\theta) = \frac{d(\theta, R)}{d(\theta, R) + d(\theta, R_r^c)},$$

para $i = 1, \dots, T$, son continuas y acotadas por 1.

Sea \bar{R}_r la adherencia de R_r (con respecto a Θ). Claramente \bar{R}_r es cerrado y compacto.

B.8. DEMOSTRACIÓN DEL TEOREMA 3.7

Notar que para $i = 1, \dots, T$, la familia de funciones $\{\bar{R}_r \ni \theta \rightarrow \psi(y, \theta) : y \in K_{x_i}\}$ es (uniformemente) equicontinua (por (ii)) y $\sup_{y \in K_{x_i}} \psi(y, \theta) < \infty, \forall \theta \in \bar{R}_r$, dado que $\psi(\cdot, \theta)$ es continua y K_{x_i} es compacto. Luego por el teorema de Arzela-Ascoli (ver página 167 de [Royden et al., 1988](#)) se concluye que dado $\eta_i > 0$, existe una partición $A_{i,1}, \dots, A_{i,n_i}$ de K_{x_i} y puntos $z_{i,1} \in A_{i,1}, \dots, z_{i,n_i} \in A_{i,n_i}$, tales que

$$\sup_{y \in A_{i,j}} \sup_{\theta \in \bar{R}_r} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| < \eta_i,$$

para $i = 1, \dots, T, j = 1, \dots, n_i$. Luego, para $i = 1, \dots, T$ y $j = 1, \dots, n_i$ se define

$$h_{i,j}^1(\theta) = \psi(z_{i,j}, \theta).$$

Claramente $h_{i,j}^1$ es una función continua y acotada.

Se considera el conjunto

$$\prod_{i=1}^T \left\{ P_{x_i} \in \mathcal{P}(\Theta) : \left| \int_{\Theta} h_{i,j_i}^l(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_{i,j_i}^l(\theta) P_{x_i}^0(d\theta) \right| < \nu, \right. \\ \left. l = 0, 1, j_0 = 1, 2, 1 \leq j_1 \leq n_i \right\}, \quad (\text{B.5})$$

para $\nu > 0$. En lo que sigue se muestra que para adecuadas elecciones de ν, η, r y ρ , cada elemento $(P_{x_1}, \dots, P_{x_T})$ que satisfaga (B.5), verifica que

$$\int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) < \epsilon,$$

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

Notar que

$$\begin{aligned} \int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) &= \int_{K_{x_i}^c} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) \\ &+ \int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy), \end{aligned}$$

para $i = 1, \dots, T$. Dado que

$$\left| \int_{\Theta} h_{i,1}^0(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_{i,1}^0(\theta) P_{x_i}^0(d\theta) \right| < \nu,$$

entonces

$$\int_{\Theta} h_{i,1}^0(\theta) P_{x_i}(d\theta) < \nu + \int_{\Theta} h_{i,1}^0(\theta) P_{x_i}^0(d\theta) < \nu + \frac{\epsilon}{8},$$

pues

$$\begin{aligned} \int_{\Theta} h_{i,1}^0(\theta) P_{x_i}^0(d\theta) &= \int_{\Theta} \int_{K_{x_i}^c} \psi(y, \theta) \lambda_{\mathcal{Y}}(dy) P_{x_i}^0(d\theta) \\ &= \int_{K_{x_i}^c} \int_{\Theta} \psi(y, \theta) P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\ &= \int_{K_{x_i}^c} f_{x_i}^0(y) \lambda_{\mathcal{Y}}(dy), \end{aligned}$$

gracias al teorema de Tonelli. Luego,

$$\begin{aligned}
 \int_{K_{x_i}^c} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) &\leq \int_{K_{x_i}^c} \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &+ \int_{K_{x_i}^c} f_{x_i}^0(y) \lambda_{\mathcal{Y}}(dy) \\
 &= \int_{\Theta} \int_{K_{x_i}^c} \psi(y, \theta) \lambda_{\mathcal{Y}}(dy) P_{x_i}(d\theta) \\
 &+ \int_{K_{x_i}^c} f_{x_i}^0(y) \lambda_{\mathcal{Y}}(dy) \\
 &= \int_{\Theta} h_{i,1}^0(\theta) P_{x_i}(d\theta) + \int_{K_{x_i}^c} f_{x_i}^0(y) \lambda_{\mathcal{Y}}(dy) \\
 &< \nu + \frac{\epsilon}{8} + \frac{\epsilon}{8} = \nu + \frac{\epsilon}{4}.
 \end{aligned}$$

En lo que sigue se busca acotar

$$\begin{aligned}
 &\int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) \\
 &= \int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy).
 \end{aligned}$$

En efecto, notar que gracias a la desigualdad triangular

$$\int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) \leq I_{i,1} + I_{i,2} + I_{i,3},$$

donde

$$I_{i,1} = \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy),$$

$$I_{i,2} = \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy),$$

e

$$I_{i,3} = \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}^0(d\theta) - \int_{\Theta} \psi(y, \theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy).$$

Lo siguiente es acotar $I_{i,1}$, $I_{i,2}$ e $I_{i,3}$ respectivamente. En efecto,

$$\begin{aligned}
I_{i,1} &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
&= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} [\psi(y, \theta) - \psi(z_{i,j}, \theta)] P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
&\leq \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\Theta} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
&= \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{R}_r} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
&\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{R}_r^c} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
&\leq \sum_{j=1}^{n_i} \int_{A_{i,j}} \sup_{\theta \in \bar{R}_r} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}(\bar{R}_r) \lambda_{\mathcal{Y}}(dy) \\
&\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{R}_r^c} [\psi(y, \theta) + \psi(z_{i,j}, \theta)] P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
&\leq \sum_{j=1}^{n_i} \sup_{y \in A_{i,j}} \sup_{\theta \in \bar{R}_r} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}(\bar{R}_r) \lambda_{\mathcal{Y}}(A_{i,j}) \\
&\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{R}_r^c} \psi(y, \theta) P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{R}_r^c} \psi(z_{i,j}, \theta) P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
&< \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + \int_{\bar{R}_r^c} \int_{K_{x_i}} \psi(y, \theta) \lambda_{\mathcal{Y}}(dy) P_{x_i}(d\theta) \\
&\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \sup_{\theta \in \bar{R}_r^c} \psi(z_{i,j}, \theta) P_{x_i}(\bar{R}_r^c) \lambda_{\mathcal{Y}}(dy) \\
&\leq \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + P_{x_i}(\bar{R}_r^c) + \left(\max_{j \in \{1, \dots, n_i\}} \sup_{\theta \in \bar{R}_r^c} \psi(z_{i,j}, \theta) \right) P_{x_i}(\bar{R}_r^c) \lambda_{\mathcal{Y}}(K_{x_i}) \\
&\leq \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + P_{x_i}(\bar{R}_r^c) + M_i P_{x_i}(\bar{R}_r^c) \lambda_{\mathcal{Y}}(K_{x_i}),
\end{aligned}$$

donde

$$M_i = \max_{j \in \{1, \dots, n_i\}} \max_{\theta \in \Theta} \psi(z_{i,j}, \theta), \quad i = 1, \dots, T.$$

Notar que la compacidad de Θ y continuidad de $\psi(y, \cdot)$, garantizan que el máximo anterior se

alcanza $\forall y \in \mathcal{Y}$. Se sabe que

$$\left| \int_{\Theta} h_{i,2}^0(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_{i,2}^0(\theta) P_{x_i}^0(d\theta) \right| < \nu,$$

y entonces

$$\int_{\Theta} h_{i,2}^0(\theta) P_{x_i}(d\theta) < \nu + \int_{\Theta} h_{i,2}^0(\theta) P_{x_i}^0(d\theta),$$

que es equivalente a

$$\int_{R^c} h_{i,2}^0(\theta) P_{x_i}(d\theta) < \nu + \int_{R^c} h_{i,2}^0(\theta) P_{x_i}^0(d\theta),$$

pues $h_{i,2}^0(\theta) = 0$, para todo $\theta \in R$. Además,

$$P_{x_i}(\overline{R}_r^c) = \int_{\overline{R}_r^c} h_{i,2}^0(\theta) P_{x_i}(d\theta) \leq \int_{R^c} h_{i,2}^0(\theta) P_{x_i}(d\theta),$$

para $r > 0$, lo suficientemente grande, pues $\overline{R}_r^c \subseteq R_r^c$ y $h_{i,2}^0(\theta) = 1$, para todo $\theta \in R_r^c$. Por otro lado,

$$\int_{R^c} h_{i,2}^0(\theta) P_{x_i}^0(d\theta) \leq P_{x_i}^0(R^c) < \rho.$$

Luego,

$$P_{x_i}(\overline{R}_r^c) < \nu + \rho,$$

y por lo tanto,

$$I_{i,1} < \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + (\nu + \rho) + M_i(\nu + \rho) \lambda_{\mathcal{Y}}(K_{x_i}).$$

De manera similar se tiene que

$$\begin{aligned}
 I_{i,3} &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}^0(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} [\psi(y, \theta) - \psi(z_{i,j}, \theta)] P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\Theta} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{R}_r} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{R}_r^c} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^{n_i} \int_{A_{i,j}} \sup_{\theta \in \bar{R}_r} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}^0(\bar{R}_r) \lambda_{\mathcal{Y}}(dy) \\
 &\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{R}_r^c} [\psi(y, \theta) + \psi(z_{i,j}, \theta)] P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^{n_i} \sup_{y \in A_{i,j}} \sup_{\theta \in \bar{R}_r} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}^0(\bar{R}_r) \lambda_{\mathcal{Y}}(A_{i,j}) \\
 &\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{R}_r^c} \psi(y, \theta) P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{R}_r^c} \psi(z_{i,j}, \theta) P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &< \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + \int_{\bar{R}_r^c} \int_{K_{x_i}} \psi(y, \theta) \lambda_{\mathcal{Y}}(dy) P_{x_i}^0(d\theta) \\
 &\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \sup_{\theta \in \bar{R}_r^c} \psi(z_{i,j}, \theta) P_{x_i}^0(\bar{R}_r^c) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + P_{x_i}^0(\bar{R}_r^c) + \left(\max_{j \in \{1, \dots, n_i\}} \sup_{\theta \in \bar{R}_r^c} \psi(z_{i,j}, \theta) \right) P_{x_i}^0(\bar{R}_r^c) \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &\leq \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + P_{x_i}^0(\bar{R}_r^c) + M_i P_{x_i}^0(\bar{R}_r^c) \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &< \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + \rho + M_i \rho \lambda_{\mathcal{Y}}(K_{x_i}),
 \end{aligned}$$

con

$$M_i = \max_{j \in \{1, \dots, n_i\}} \max_{\theta \in \Theta} \psi(z_{i,j}, \theta), \quad i = 1, \dots, T,$$

al igual que antes.

Por otra parte,

$$\begin{aligned}
 I_{i,2} &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} h_{i,j}^1(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_{i,j}^1(\theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &< \sum_{j=1}^{n_i} \int_{A_{i,j}} \nu \lambda_{\mathcal{Y}}(dy) \\
 &= \nu \lambda_{\mathcal{Y}}(K_{x_i}).
 \end{aligned}$$

Por lo tanto,

$$\begin{aligned}
 \int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) &\leq I_{i,1} + I_{i,2} + I_{i,3} \\
 &< \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + (\nu + \rho) + M_i(\nu + \rho) \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &\quad + \nu \lambda_{\mathcal{Y}}(K_{x_i}) + \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + \rho + M_i \rho \lambda_{\mathcal{Y}}(K_{x_i}).
 \end{aligned}$$

Así,

$$\begin{aligned}
 \int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) &= \int_{K_{x_i}^c} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) \\
 &+ \int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) \\
 &< \nu + \frac{\epsilon}{4} + \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + (\nu + \rho) \\
 &+ M_i(\nu + \rho) \lambda_{\mathcal{Y}}(K_{x_i}) + \nu \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &+ \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + \rho + M_i \rho \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &= \frac{\epsilon}{4} + \nu[2 + \lambda_{\mathcal{Y}}(K_{x_i})(1 + M_i)] \\
 &+ 2\eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + 2\rho[1 + M_i \lambda_{\mathcal{Y}}(K_{x_i})].
 \end{aligned}$$

Finalmente, tomando

$$\nu = \frac{\epsilon}{4 \max\{2 + \lambda_{\mathcal{Y}}(K_{x_i})(1 + M_i) : i = 1, \dots, T\}},$$

$$\eta_i = \eta = \frac{\epsilon}{8 \max\{\lambda_{\mathcal{Y}}(K_{x_i}) : i = 1, \dots, T\}},$$

y

$$\rho = \frac{\epsilon}{8 \max\{1 + M_i \lambda_{\mathcal{Y}}(K_{x_i}) : i = 1, \dots, T\}},$$

se concluye que

$$\int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) < \epsilon,$$

y como $\epsilon > 0$ es arbitrario, se concluye la demostración del teorema. ■

B.9 Demostración del Teorema 3.8

La demostración es muy similar a la del teorema 3.7, de manera que sólo se discutirá los aspectos diferentes.

Sean $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^{\mathcal{X}}$ y $\{f_x^0 : x \in \mathcal{X}\}$ con

$$f_x^0(\cdot) = \int_{\Theta} \psi(\cdot, \theta) P_x^0(d\theta), \quad x \in \mathcal{X},$$

$\epsilon > 0$ y $x_1, \dots, x_T \in \mathcal{X}$ puntos arbitrarios. Existen compactos $K_{x_i} \subseteq \mathcal{Y}$, $i = 1, \dots, T$, tal que

$$\int_{K_{x_i}^c} f_{x_i}^0(y) \lambda_{\mathcal{Y}}(dy) < \frac{\epsilon}{8}.$$

Para $i = 1, \dots, T$ se define

$$h_{i,1}^0(\theta) = \int_{K_{x_i}^c} \psi(y, \theta) \lambda_{\mathcal{Y}}(dy),$$

una función acotada y continua.

Sea $\rho > 0$. Dado que Θ no es un espacio compacto, no es clara la existencia de compactos $R_{x_i} \subseteq \Theta$, $i = 1, \dots, T$, tal que $P_{x_i}^0(R_{x_i}^c) < \rho$.

Afirmación. Existen compactos $R_{x_i} \subseteq \Theta$, $i = 1, \dots, T$, tal que $P_{x_i}^0(R_{x_i}^c) < \rho$.

Dem. de la afirmación: dado que Θ es un espacio métrico separable, entonces para cada $k \in \mathbb{N}$ hay una sucesión A_{k1}, A_{k2}, \dots de bolas de radio $\frac{1}{k}$ que cubren Θ . Sea $i \in \{1, \dots, T\}$. Se elige $n_k \in \mathbb{N}$ lo suficientemente grande tales que $P_{x_i}^0(\bigcup_{i \leq n_k} A_{ki}) > 1 - \frac{\rho}{2^k}$. Se define

$$R_{x_i} := \overline{\bigcap_{k \in \mathbb{N}} \bigcup_{i \leq n_k} A_{ki}} \subseteq \bigcap_{k \in \mathbb{N}} \bigcup_{i \leq n_k} \overline{A_{ki}}.$$

Notar que A_{ki} tiene la forma

$$A_{ki} = B_{\mathcal{Y}}\left(\mu_i, \frac{1}{k}\right) \times B_{\mathbb{R}_+}\left(\sigma_i, \frac{1}{k}\right) \Rightarrow \overline{A_{ki}} = \overline{B_{\mathcal{Y}}\left(\mu_i, \frac{1}{k}\right)} \times \overline{B_{\mathbb{R}_+}\left(\sigma_i, \frac{1}{k}\right)},$$

con $\mu_i \in \mathcal{Y}$, $\sigma_i \in \mathbb{R}_+$, y $B_{\mathcal{Y}}(\cdot, \cdot)$ y $B_{\mathbb{R}_+}(\cdot, \cdot)$ la bolas abiertas en \mathcal{Y} y \mathbb{R}_+ respectivamente. De inmediato $\overline{B_{\mathcal{Y}}(\mu_i, \frac{1}{k})}$ y $\overline{B_{\mathbb{R}_+}(\sigma_i, \frac{1}{k})}$ son compactos, y por lo tanto, $\overline{A_{ki}}$ es un conjunto compacto. Dado que la unión finita de compactos es compacta y que la unión numerable de cerrados es cerrada, se concluye que

$$\bigcap_{k \in \mathbb{N}} \bigcup_{i \leq n_k} \overline{A_{ki}},$$

es un cerrado que está contenido en un compacto, y por tanto, es un conjunto compacto. Gracias a esto, R_{x_i} es un cerrado contenido en un conjunto compacto. Se concluye así que R_{x_i} es un conjunto compacto. Por otro lado,

$$P_{x_i}^0\left(\bigcup_{k \in \mathbb{N}} \bigcap_{i \leq n_k} A_{ki}^c\right) \leq \sum_{k \in \mathbb{N}} P_{x_i}^0\left(\bigcap_{i \leq n_k} A_{ki}^c\right) < \rho \sum_{k=1}^{\infty} \frac{1}{2^k} = \rho.$$

Luego,

$$P_{x_i}^0(R_{x_i}) \geq P_{x_i}^0\left(\bigcap_{k \in \mathbb{N}} \bigcup_{i \leq n_k} A_{ki}\right) > 1 - \rho \Rightarrow P_{x_i}^0(R_{x_i}^c) < \rho,$$

y se concluye la afirmación.

Se define $R := \bigcup_{i=1}^T R_{x_i}$. Dicho conjunto es cerrado y compacto $\forall i = 1, \dots, T$,

$$P_{x_i}^0(R^c) = P_{x_i}^0\left(\bigcap_{i=1}^T R_{x_i}^c\right) \leq P_{x_i}^0(R_{x_i}^c) < \rho.$$

Se define

$$R_{x_i, r} := \{\theta \in \Theta : d(\theta, R_{x_i}) < r\},$$

con $r > 0$. Este conjunto es abierto. Similarmente se define $R_r := \bigcup_{i=1}^T R_{x_i, r}$. Este conjunto es abierto. Claramente $R \subseteq R_r$.

Afirmación. Dado que Θ es un espacio localmente compacto, existe un abierto V con adherencia \bar{V} compacta, tales que $R \subseteq V \subseteq \bar{V} \subseteq R_r$.

Dem. de la afirmación: sea $x \in R$. Dado que R_r es una vecindad de x , entonces existe otra vecindad V_x de x tal que \bar{V}_x es compacta y $\bar{V}_x \subseteq R_r$ (gracias al teorema 29.2, página 185, de [Munkres \(2000\)](#)). Repitiendo este procedimiento para cada $x \in R$ se obtiene un recubrimiento abierto de R dado por $\bigcup_{x \in R} V_x$. Como R es compacto, entonces existe un sub-recubrimiento abierto finito

$$V = \bigcup_{i=1}^n V_{x_i},$$

de R , con $n \in \mathbb{N}$ y $x_i \in R, i = 1, \dots, n$. Claramente,

$$\bar{V} = \overline{\bigcup_{i=1}^n V_{x_i}} = \bigcup_{i=1}^n \bar{V}_{x_i},$$

es un conjunto compacto, que por construcción está contenido en R_r . Así, $R \subseteq V \subseteq \bar{V} \subseteq R_r$.

Luego, las funciones $d(\cdot, R)$ y $d(\cdot, V^c)$ son continuas y, por álgebra de funciones continuas, las funciones

$$h_{i,2}^0(\theta) = \frac{d(\theta, R)}{d(\theta, R) + d(\theta, V^c)},$$

para $i = 1, \dots, T$, son continuas y acotadas por 1.

Para $i = 1, \dots, T$, la familia de funciones $\{\bar{V} \ni \theta \rightarrow \psi(y, \theta) : y \in K_{x_i}\}$ es (uniformemente) equicontinua (por (ii)) y $\sup_{y \in K_{x_i}} \psi(y, \theta) < \infty, \forall \theta \in \bar{V}$. Por el teorema de Arzela-Ascoli, dado $\eta_i > 0$, existe una partición $A_{i,1}, \dots, A_{i,n_i}$ de K_{x_i} y puntos $z_{i,1} \in A_{i,1}, \dots, z_{i,n_i} \in A_{i,n_i}$ tales que

$$\sup_{y \in A_{i,j}} \sup_{\theta \in \bar{V}} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| < \eta_i,$$

B.9. DEMOSTRACIÓN DEL TEOREMA 3.8

para $i = 1, \dots, T$, $j = 1, \dots, n_i$. Para $i = 1, \dots, T$ y $j = 1, \dots, n_i$ se define

$$h_{i,j}^1(\theta) = \psi(z_{i,j}, \theta),$$

una función continua y acotada.

Sea $\nu > 0$ y el conjunto

$$\prod_{i=1}^T \left\{ P_{x_i} \in \mathcal{P}(\Theta) : \left| \int_{\Theta} h_{i,j_l}^l(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_{i,j_l}^l(\theta) P_{x_i}^0(d\theta) \right| < \nu, \right. \\ \left. l = 0, 1, j_0 = 1, 2, 1 \leq j_1 \leq n_i \right\}.$$

Notar que

$$\int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) = \int_{K_{x_i}^c} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) \\ + \int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy),$$

para $i = 1, \dots, T$. Dado que

$$\left| \int_{\Theta} h_{i,1}^0(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_{i,1}^0(\theta) P_{x_i}^0(d\theta) \right| < \nu,$$

entonces

$$\int_{\Theta} h_{i,1}^0(\theta) P_{x_i}(d\theta) < \nu + \int_{\Theta} h_{i,1}^0(\theta) P_{x_i}^0(d\theta) < \nu + \frac{\epsilon}{8}.$$

Luego,

$$\int_{K_{x_i}^c} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) \leq \int_{K_{x_i}^c} \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\ + \int_{K_{x_i}^c} f_{x_i}^0(y) \lambda_{\mathcal{Y}}(dy) \\ < \nu + \frac{\epsilon}{4}.$$

Gracias a la desigualdad triangular

$$\int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) \leq I_{i,1} + I_{i,2} + I_{i,3},$$

donde

$$I_{i,1} = \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy),$$

$$I_{i,2} = \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy),$$

e

$$I_{i,3} = \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}^0(d\theta) - \int_{\Theta} \psi(y, \theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy).$$

Notar que

$$\begin{aligned}
 I_{i,1} &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} [\psi(y, \theta) - \psi(z_{i,j}, \theta)] P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\Theta} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{V}} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{V}^c} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^{n_i} \int_{A_{i,j}} \sup_{\theta \in \bar{V}} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}(\bar{V}) \lambda_{\mathcal{Y}}(dy) \\
 &\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{V}^c} [\psi(y, \theta) + \psi(z_{i,j}, \theta)] P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &< \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + \int_{\bar{V}^c} \int_{K_{x_i}} \psi(y, \theta) \lambda_{\mathcal{Y}}(dy) P_{x_i}(d\theta) \\
 &\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \sup_{\theta \in \bar{V}^c} \psi(z_{i,j}, \theta) P_{x_i}(\bar{V}^c) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + P_{x_i}(\bar{V}^c) + \left(\max_{j \in \{1, \dots, n_i\}} \sup_{\theta \in \bar{V}^c} \psi(z_{i,j}, \theta) \right) P_{x_i}(\bar{V}^c) \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &\leq \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + P_{x_i}(\bar{V}^c) + M_i P_{x_i}(\bar{V}^c) \lambda_{\mathcal{Y}}(K_{x_i}),
 \end{aligned}$$

donde

$$M_i = \max_{j \in \{1, \dots, n_i\}} \max_{\theta \in \Theta} \psi(z_{i,j}, \theta), \quad i = 1, \dots, T.$$

Notar que la compacidad de Θ y continuidad de $\psi(y, \cdot)$, garantizan que el máximo anterior se alcanza $\forall y \in \mathcal{Y}$. Se sabe que

$$\left| \int_{\Theta} h_{i,2}^0(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_{i,2}^0(\theta) P_{x_i}^0(d\theta) \right| < \nu,$$

y entonces

$$\int_{\Theta} h_{i,2}^0(\theta) P_{x_i}(d\theta) < \nu + \int_{\Theta} h_{i,2}^0(\theta) P_{x_i}^0(d\theta),$$

que es equivalente a

$$\int_{R^c} h_{i,2}^0(\theta) P_{x_i}(d\theta) < \nu + \int_{R^c} h_{i,2}^0(\theta) P_{x_i}^0(d\theta),$$

pues $h_{i,2}^0(\theta) = 0$ para todo $\theta \in R$. Además,

$$P_{x_i}(\bar{V}^c) = \int_{\bar{V}^c} h_{i,2}^0(\theta) P_{x_i}(d\theta) \leq \int_{R^c} h_{i,2}^0(\theta) P_{x_i}(d\theta),$$

pues $\bar{V}^c \subseteq V^c$ y $h_{i,2}^0(\theta) = 1$ para todo $\theta \in V^c$. Por otro lado,

$$\int_{R^c} h_{i,2}^0(\theta) P_{x_i}^0(d\theta) \leq P_{x_i}^0(R^c) < \rho.$$

Luego,

$$P_{x_i}(\bar{V}^c) < \nu + \rho,$$

y por lo tanto,

$$I_{i,1} < \eta_i \lambda(K_{x_i}) + (\nu + \rho) + M_i(\nu + \rho) \lambda(K_{x_i}).$$

De manera similar se tiene que

$$\begin{aligned}
 I_{i,3} &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}^0(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} [\psi(y, \theta) - \psi(z_{i,j}, \theta)] P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\Theta} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{V}} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{V}^c} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^{n_i} \int_{A_{i,j}} \sup_{\theta \in \bar{V}} |\psi(y, \theta) - \psi(z_{i,j}, \theta)| P_{x_i}^0(\bar{V}) \lambda_{\mathcal{Y}}(dy) \\
 &\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\bar{V}^c} [\psi(y, \theta) + \psi(z_{i,j}, \theta)] P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &< \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + \int_{\bar{V}^c} \int_{K_{x_i}} \psi(y, \theta) \lambda_{\mathcal{Y}}(dy) P_{x_i}^0(d\theta) \\
 &\quad + \sum_{j=1}^{n_i} \int_{A_{i,j}} \sup_{\theta \in \bar{V}^c} \psi(z_{i,j}, \theta) P_{x_i}^0(\bar{V}^c) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + P_{x_i}^0(\bar{V}^c) + \left(\max_{j \in \{1, \dots, n_i\}} \sup_{\theta \in \bar{V}^c} \psi(z_{i,j}, \theta) \right) P_{x_i}^0(\bar{V}^c) \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &\leq \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + P_{x_i}^0(\bar{V}^c) + M_i P_{x_i}^0(\bar{V}^c) \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &< \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + \rho + M_i \rho \lambda_{\mathcal{Y}}(K_{x_i}),
 \end{aligned}$$

con

$$M_i = \max_{j \in \{1, \dots, n_i\}} \max_{\theta \in \Theta} \psi(z_{i,j}, \theta), \quad i = 1, \dots, T,$$

al igual que antes.

Por otra parte,

$$\begin{aligned}
 I_{i,2} &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} h_{i,j}^1(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_{i,j}^1(\theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &< \sum_{j=1}^{n_i} \int_{A_{i,j}} \nu \lambda_{\mathcal{Y}}(dy) \\
 &= \nu \lambda_{\mathcal{Y}}(K_{x_i}).
 \end{aligned}$$

Por lo tanto,

$$\begin{aligned}
 \int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) &\leq I_{i,1} + I_{i,2} + I_{i,3} \\
 &< \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + (\nu + \rho) + M_i(\nu + \rho) \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &\quad + \nu \lambda_{\mathcal{Y}}(K_{x_i}) + \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + \rho + M_i \rho \lambda_{\mathcal{Y}}(K_{x_i}).
 \end{aligned}$$

Así,

$$\begin{aligned}
 \int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) &= \int_{K_{x_i}^c} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) \\
 &\quad + \int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) \\
 &< \nu + \frac{\epsilon}{4} + \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + (\nu + \rho) \\
 &\quad + M_i(\nu + \rho) \lambda_{\mathcal{Y}}(K_{x_i}) + \nu \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &\quad + \eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + \rho + M_i \rho \lambda_{\mathcal{Y}}(K_{x_i}) \\
 &= \frac{\epsilon}{4} + \nu [2 + \lambda_{\mathcal{Y}}(K_{x_i})(1 + M_i)] \\
 &\quad + 2\eta_i \lambda_{\mathcal{Y}}(K_{x_i}) + 2\rho [1 + M_i \lambda_{\mathcal{Y}}(K_{x_i})].
 \end{aligned}$$

Finalmente, tomando

$$\nu = \frac{\epsilon}{4 \max\{2 + \lambda_{\mathcal{Y}}(K_{x_i})(1 + M_i) : i = 1, \dots, T\}},$$

$$\eta_i = \eta = \frac{\epsilon}{8 \max\{\lambda_{\mathcal{Y}}(K_{x_i}) : i = 1, \dots, T\}},$$

y

$$\rho = \frac{\epsilon}{8 \max\{1 + M_i \lambda_{\mathcal{Y}}(K_{x_i}) : i = 1, \dots, T\}},$$

se tiene que

$$\int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f_{x_i}^0(y) \right| \lambda_{\mathcal{Y}}(dy) < \epsilon,$$

y como $\epsilon > 0$ es arbitrario, se concluye la demostración del teorema. ■

B.10 Demostración del Teorema 3.9

Sean $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^{\mathcal{X}}$, $\sigma_0 > 0$, $\epsilon > 0$ y $x_1, \dots, x_T \in \mathcal{X}$. Hay que demostrar que el conjunto dado por

$$\left\{ (P_{x_1}, \dots, P_{x_T}, \sigma) : \int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) < \epsilon, \right. \\ \left. P_{x_i} \in \mathcal{P}(\Theta), \sigma > 0 \right\},$$

tiene probabilidad estrictamente positiva. Notar que

$$\begin{aligned} & \int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\ & \leq \int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\ & + \int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy). \end{aligned}$$

Así, para concluir que

$$\int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) < \epsilon,$$

basta demostrar que

$$\int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy) < \frac{\epsilon}{2}, \quad (\text{B.6})$$

y

$$\int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) < \frac{\epsilon}{2}. \quad (\text{B.7})$$

En relación al primer término, se tiene que

$$\begin{aligned} & \int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\ &= \int_{\mathcal{Y}} \left| \int_{\Theta} [\psi(y, \theta, \sigma) - \psi(y, \theta, \sigma_0)] P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\ &\leq \int_{\mathcal{Y}} \int_{\Theta} |\psi(y, \theta, \sigma) - \psi(y, \theta, \sigma_0)| P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\ &\leq \int_{\mathcal{Y}} \left(\sup_{\theta \in \Theta} |\psi(y, \theta, \sigma) - \psi(y, \theta, \sigma_0)| \right) P_{x_i}(\Theta) \lambda_{\mathcal{Y}}(dy) \\ &\leq \left(\sup_{y \in \mathcal{Y}} \sup_{\theta \in \Theta} |\psi(y, \theta, \sigma) - \psi(y, \theta, \sigma_0)| \right) \lambda_{\mathcal{Y}}(\mathcal{Y}). \end{aligned}$$

Gracias al supuesto (ii) y, a la compacidad de \mathcal{Y} y Θ , se tiene que la familia de funciones $\{\sigma \rightarrow \psi(y, \theta, \sigma) : (y, \theta) \in \mathcal{Y} \times \Theta\}$ es uniformemente equicontinua para todo σ en una vecindad compacta de σ_0 . Así, dado $\delta > 0$, existe un compacto V_{σ_0} conteniendo a σ_0 en su interior tal que

$$\sup_{\sigma \in V_{\sigma_0}} \sup_{y \in \mathcal{Y}} \sup_{\theta \in \Theta} |\psi(y, \theta, \sigma) - \psi(y, \theta, \sigma_0)| < \delta.$$

Dado que $\lambda_{\mathcal{Y}}(\mathcal{Y}) < \infty$ (por ser $\lambda_{\mathcal{Y}}$ una medida finita), tomando $\delta = \frac{\epsilon}{2\lambda_{\mathcal{Y}}(\mathcal{Y})}$ se concluye (B.6).

El supuesto (ii) y la compacidad de \mathcal{Y} permiten concluir que la familia de funciones $\{\theta \rightarrow \psi(y, \theta, \sigma_0) : y \in \mathcal{Y}\}$ es uniformemente equicontinua. Dado que $\sup_{y \in \mathcal{Y}} \psi(y, \theta, \sigma_0) < \infty$, se deduce por el teorema de Arzela-Ascoli que existe una partición A_1, \dots, A_n de \mathcal{Y} y puntos $z_1 \in A_1, \dots, z_n \in A_n$, tales que

$$\sup_{y \in A_j} \sup_{\theta \in \Theta} |\psi(y, \theta, \sigma_0) - \psi(y_j, \theta, \sigma_0)| < \eta,$$

para $j = 1, \dots, n$ y $\eta > 0$ por determinar.

Se definen las funciones $h_j(\theta) = \psi(y_j, \theta, \sigma_0)$, $j = 1, \dots, n$. Claramente dichas funciones son continuas y acotadas. Esto permite construir vecindades $V_{P_{x_1}^0}, \dots, V_{P_{x_T}^0}$ (con respecto a la topología débil de medidas de probabilidad) de $P_{x_1}^0, \dots, P_{x_T}^0$ respectivamente. En efecto, sean

$$V_{P_{x_i}^0} := \left\{ P_{x_i} \in \mathcal{P}(\Theta) : \left| \int_{\Theta} h_j(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_j(\theta) P_{x_i}^0(d\theta) \right| < \nu, j = 1, \dots, n \right\},$$

para $i = 1, \dots, T$, con $\nu > 0$ por determinar.

Por otra parte, notar que para $i = 1, \dots, T$ se tiene

$$\int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \leq I_{i,1} + I_{i,2} + I_{i,3},$$

donde

$$I_{i,1} = \sum_{j=1}^n \int_{A_j} \left| \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}(d\theta) - \int_{\Theta} \psi(y_j, \theta, \sigma_0) P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy),$$

$$I_{i,2} = \sum_{j=1}^n \int_{A_j} \left| \int_{\Theta} \psi(y_j, \theta, \sigma_0) P_{x_i}(d\theta) - \int_{\Theta} \psi(y_j, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy),$$

e

$$I_{i,3} = \sum_{j=1}^n \int_{A_j} \left| \int_{\Theta} \psi(y_j, \theta, \sigma_0) P_{x_i}^0(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy).$$

Acotando cada término por separado se tiene que

$$\begin{aligned}
 I_{i,1} &= \sum_{j=1}^n \int_{A_j} \left| \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}(d\theta) - \int_{\Theta} \psi(y_j, \theta, \sigma_0) P_{x_i}(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^n \int_{A_j} \int_{\Theta} |\psi(y, \theta, \sigma_0) - \psi(y_j, \theta, \sigma_0)| P_{x_i}(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^n \int_{A_j} \left(\sup_{\theta \in \Theta} |\psi(y, \theta, \sigma_0) - \psi(y_j, \theta, \sigma_0)| \right) P_{x_i}(\Theta) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^n \left(\sup_{y \in A_j} \sup_{\theta \in \Theta} |\psi(y, \theta, \sigma_0) - \psi(y_j, \theta, \sigma_0)| \right) \lambda_{\mathcal{Y}}(A_j) \\
 &< \sum_{j=1}^n \eta \lambda_{\mathcal{Y}}(A_j) = \eta \lambda_{\mathcal{Y}}(\mathcal{Y}).
 \end{aligned}$$

Similarmente,

$$\begin{aligned}
 I_{i,3} &= \sum_{j=1}^n \int_{A_j} \left| \int_{\Theta} \psi(y_j, \theta, \sigma_0) P_{x_i}^0(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^n \int_{A_j} \int_{\Theta} |\psi(y, \theta, \sigma_0) - \psi(y_j, \theta, \sigma_0)| P_{x_i}^0(d\theta) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^n \int_{A_j} \left(\sup_{\theta \in \Theta} |\psi(y, \theta, \sigma_0) - \psi(y_j, \theta, \sigma_0)| \right) P_{x_i}^0(\Theta) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \sum_{j=1}^n \left(\sup_{y \in A_j} \sup_{\theta \in \Theta} |\psi(y, \theta, \sigma_0) - \psi(y_j, \theta, \sigma_0)| \right) \lambda_{\mathcal{Y}}(A_j) \\
 &< \sum_{j=1}^n \eta \lambda_{\mathcal{Y}}(A_j) = \eta \lambda_{\mathcal{Y}}(\mathcal{Y}).
 \end{aligned}$$

Por otra parte,

$$\begin{aligned}
 I_{i,2} &= \sum_{j=1}^n \int_{A_j} \left| \int_{\Theta} \psi(y_j, \theta, \sigma_0) P_{x_i}(d\theta) - \int_{\Theta} \psi(y_j, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &= \sum_{j=1}^n \int_{A_j} \left| \int_{\Theta} h_j(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_j(\theta) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) \\
 &< \sum_{j=1}^n \int_{A_j} \nu \lambda_{\mathcal{Y}}(dy) = \nu \sum_{j=1}^n \lambda_{\mathcal{Y}}(A_j) = \nu \lambda_{\mathcal{Y}}(\mathcal{Y}).
 \end{aligned}$$

Entonces,

$$\int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) < 2\eta \lambda_{\mathcal{Y}}(\mathcal{Y}) + \nu \lambda_{\mathcal{Y}}(\mathcal{Y}).$$

Imponiendo que

$$\eta = \frac{\epsilon}{8\lambda_{\mathcal{Y}}(\mathcal{Y})} \quad \text{y} \quad \nu = \frac{\epsilon}{4\lambda_{\mathcal{Y}}(\mathcal{Y})},$$

se deduce (B.7).

Por lo tanto, para todo

$$(P_{x_1}, \dots, P_{x_T}, \sigma) \in V_{P_{x_1}^0} \times \dots \times V_{P_{x_T}^0} \times V_{\sigma_0},$$

se verifica que

$$\int_{\mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \lambda_{\mathcal{Y}}(dy) < \epsilon.$$

Debido a que el DDP tiene soporte débil completo, entonces cada $V_{P_{x_i}^0}$ tiene probabilidad positiva. Lo mismo ocurre con V_{σ_0} , gracias a que π_{σ} tiene soporte completo en \mathbb{R}_+ . Así,

B.11. DEMOSTRACIÓN DEL TEOREMA 3.10

$V_{P_{x_1}^0} \times \dots \times V_{P_{x_T}^0} \times V_{\sigma_0}$ tiene probabilidad positiva. Todo lo anterior, más la arbitrariedad de $\epsilon > 0$, permiten concluir que el soporte de Hellinger del proceso

$$\left\{ \int_{\Theta} \psi(\cdot, \theta, \sigma(\omega)) F_x(\omega)(d\theta) : x \in \mathcal{X} \right\},$$

está dado por

$$\prod_{x \in \mathcal{X}} \left\{ \int_{\Theta} \psi(\cdot, \theta, \sigma) P_x(d\theta) : P_x \in \mathcal{P}(\Theta), \sigma > 0 \right\}.$$

■

B.11 Demostración del Teorema 3.10

Sean $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^{\mathcal{X}}$, $\sigma_0 > 0$, $\epsilon > 0$ y $x_1, \dots, x_T \in \mathcal{X}$. Hay que demostrar que el conjunto dado por

$$\left\{ (P_{x_1}, \dots, P_{x_T}, \sigma) : \sup_{y \in \mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| < \epsilon, \right. \\ \left. P_{x_i} \in \mathcal{P}(\Theta), \sigma > 0 \right\},$$

tiene probabilidad estrictamente positiva. Notar que

$$\begin{aligned} & \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \\ & \leq \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}^0(d\theta) \right| \\ & \quad + \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}^0(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right|. \end{aligned}$$

Gracias al supuesto (ii) y la compacidad de $\mathcal{Y} \times \Theta$ se obtiene que la familia de funciones $\{\sigma \rightarrow \psi(y, \theta, \sigma) : (y, \theta) \in \mathcal{Y} \times \Theta\}$ es uniformemente equicontinua para todo σ en una vecindad compacta de σ_0 . Así, dado $\delta > 0$, existe un compacto $V_{\sigma_0} \subseteq \mathbb{R}_+$ conteniendo a σ_0 en su interior

B.11. DEMOSTRACIÓN DEL TEOREMA 3.10

tal que

$$\sup_{\sigma \in V_{\sigma_0}} \sup_{y \in \mathcal{Y}} \sup_{\theta \in \Theta} |\psi(y, \theta, \sigma) - \psi(y, \theta, \sigma_0)| < \delta.$$

Entonces,

$$\begin{aligned} & \sup_{y \in \mathcal{Y}, \sigma \in V_{\sigma_0}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}^0(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| \\ & \leq \left(\sup_{\sigma \in V_{\sigma_0}} \sup_{y \in \mathcal{Y}} \sup_{\theta \in \Theta} |\psi(y, \theta, \sigma) - \psi(y, \theta, \sigma_0)| \right) P_{x_i}^0(\Theta) \\ & < \delta. \end{aligned}$$

La propiedad (ii) y la compacidad de $\mathcal{Y} \times V_{\sigma_0}$ implican que la familia de funciones $\{\theta \rightarrow \psi(y, \theta, \sigma) : (y, \sigma) \in \mathcal{Y} \times V_{\sigma_0}\}$ es uniformemente equicontinua. Dado que $\sup_{(y, \sigma) \in \mathcal{Y} \times V_{\sigma_0}} < \infty$, se deduce por el teorema de Arzela-Ascoli que existe una partición A_1, \dots, A_n de $\mathcal{Y} \times V_{\sigma_0}$ y puntos $(y_1, \sigma_1) \in A_1, \dots, (y_n, \sigma_n) \in A_n$ tales que

$$\sup_{(y, \sigma) \in A_j} \sup_{\theta \in \Theta} |\psi(y, \theta, \sigma) - \psi(y_j, \theta, \sigma_j)| < \eta,$$

para $j = 1, \dots, n$ y $\eta > 0$ por determinar.

Se definen las funciones $h_j(\theta) = \psi(y_j, \theta, \sigma_j)$, $j = 1, \dots, n$. Claramente dichas funciones son continuas y acotadas. Esto permite construir vecindades $V_{P_{x_1}^0}, \dots, V_{P_{x_T}^0}$ (con respecto a la topología débil de medidas de probabilidad) de $P_{x_1}^0, \dots, P_{x_T}^0$ respectivamente. En efecto, sean

$$V_{P_{x_i}^0} := \left\{ P_{x_i} \in \mathcal{P}(\Theta) : \left| \int_{\Theta} h_j(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_j(\theta) P_{x_i}^0(d\theta) \right| < \nu, j = 1, \dots, n \right\},$$

para $i = 1, \dots, T$, con $\nu > 0$ por determinar.

Tomar $P_{x_i} \in V_{P_{x_i}^0}$, $i = 1, \dots, T$, $(y, \sigma) \in \mathcal{Y} \times V_{\sigma_0}$. Sin pérdida de generalidad se puede

suponer que $(y, \sigma) \in A_j$ para algún $j \in \{1, \dots, n\}$. Luego,

$$\begin{aligned}
 & \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}^0(d\theta) \right| \\
 & \leq \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y_j, \theta, \sigma_j) P_{x_i}(d\theta) \right| \\
 & + \left| \int_{\Theta} \psi(y_j, \theta, \sigma_j) P_{x_i}(d\theta) - \int_{\Theta} \psi(y_j, \theta, \sigma_j) P_{x_i}^0(d\theta) \right| \\
 & + \left| \int_{\Theta} \psi(y_j, \theta, \sigma_j) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}^0(d\theta) \right|,
 \end{aligned}$$

donde

$$\begin{aligned}
 & \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y_j, \theta, \sigma_j) P_{x_i}(d\theta) \right| \\
 & \leq \left(\sup_{(y, \sigma) \in A_j} \sup_{\theta \in \Theta} \left| \psi(y, \theta, \sigma) - \psi(y_j, \theta, \sigma_j) \right| \right) P_{x_i}(\Theta) \\
 & < \eta.
 \end{aligned}$$

Similarmente,

$$\begin{aligned}
 & \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}^0(d\theta) - \int_{\Theta} \psi(y_j, \theta, \sigma_j) P_{x_i}^0(d\theta) \right| \\
 & \leq \left(\sup_{(y, \sigma) \in A_j} \sup_{\theta \in \Theta} \left| \psi(y, \theta, \sigma) - \psi(y_j, \theta, \sigma_j) \right| \right) P_{x_i}^0(\Theta) \\
 & < \eta.
 \end{aligned}$$

Además,

$$\begin{aligned} & \left| \int_{\Theta} \psi(y_j, \theta, \sigma_j) P_{x_i}(d\theta) - \int_{\Theta} \psi(y_j, \theta, \sigma_j) P_{x_i}^0(d\theta) \right| \\ &= \left| \int_{\Theta} h_j(\theta) P_{x_i}(d\theta) - \int_{\Theta} h_j(\theta) P_{x_i}^0(d\theta) \right| \\ &< \nu. \end{aligned}$$

De lo anterior se concluye que

$$\left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}^0(d\theta) \right| < 2\eta + \nu,$$

para todo $(y, \sigma) \in \mathcal{Y} \times V_{\sigma_0}$ y $P_{x_i} \in V_{P_{x_i}^0}$, $i = 1, \dots, T$. Así,

$$\sup_{y \in \mathcal{Y}, \sigma \in V_{\sigma_0}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}^0(d\theta) \right| < 2\eta + \nu.$$

Finalmente, para todo $\sigma \in V_{\sigma_0}$ se tiene que

$$\sup_{y \in \mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| < \delta + 2\eta + \nu.$$

Tomando

$$\delta = \frac{\epsilon}{3}, \quad \eta = \frac{\epsilon}{6} \quad \text{y} \quad \nu = \frac{\epsilon}{3},$$

se concluye que

$$\sup_{y \in \mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| < \epsilon,$$

para todo $\sigma \in V_{\sigma_0}$ y $P_{x_i} \in V_{P_{x_i}^0}$, $i = 1, \dots, n$.

Finalmente, dado que tanto π_σ como el DDP tienen soporte completo y $\epsilon > 0$ es arbitrario, se concluye que $V_{P_{x_1}^0} \times \dots \times V_{P_{x_T}^0} \times V_{\sigma_0}$ tiene probabilidad positiva y por lo tanto el soporte producto \mathbb{L}^∞ del proceso

$$\left\{ \int_{\Theta} \psi(\cdot, \theta, \sigma(\omega)) F_x(\omega)(d\theta) : x \in \mathcal{X} \right\},$$

está dado por el conjunto

$$\prod_{x \in \mathcal{X}} \left\{ \int_{\Theta} \psi(\cdot, \theta, \sigma) P_x(d\theta) : P_x \in \mathcal{P}(\Theta), \sigma > 0 \right\}.$$

■

B.12 Demostración del Teorema 3.11

Sean $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^{\mathcal{X}}$, $\sigma_0 > 0$, $\epsilon > 0$ y $x_1, \dots, x_T \in \mathcal{X}$. Hay que demostrar que el conjunto dado por

$$\{(P_{x_1}, \dots, P_{x_T}, \sigma) : f(\cdot, P_{x_i}, \sigma) \in KL(f(\cdot, P_{x_i}^0, \sigma_0), \epsilon), P_{x_i} \in \mathcal{P}(\Theta), \sigma \in \mathbb{R}_+\},$$

tiene probabilidad estrictamente positiva, donde

$$f(y, P_{x_i}, \sigma) = \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta),$$

y

$$KL(f_0, \epsilon) = \left\{ f : \int_{\mathcal{Y}} f_0(y) \ln \left(\frac{f_0(y)}{f(y)} \right) \lambda_{\mathcal{Y}}(dy) < \epsilon \right\}.$$

B.12. DEMOSTRACIÓN DEL TEOREMA 3.11

Sea

$$W_\delta = \left\{ (P_{x_1}, \dots, P_{x_T}, \sigma) : \sup_{y \in \mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) - \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) \right| < \delta, \right. \\ \left. P_{x_i} \in \mathcal{P}(\Theta), \sigma \in \mathbb{R}_+ \right\},$$

con $\delta > 0$ por determinar.

Gracias al supuesto (iii) se tiene que

$$\int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) > 0,$$

para todo $y \in \mathcal{Y}, i = 1, \dots, T$. Luego, por compacidad de \mathcal{Y} se tiene que

$$c_{0,i} := \inf_{y \in \mathcal{Y}} \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) > 0,$$

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

Notar que si $(P_{x_1}, \dots, P_{x_T}, \sigma) \in W_\delta$, entonces

$$\inf_{y \in \mathcal{Y}} \int_{\Theta} \psi(y, \theta, \sigma) P_{x_i}(d\theta) > \inf_{y \in \mathcal{Y}} \int_{\Theta} \psi(y, \theta, \sigma_0) P_{x_i}^0(d\theta) - \delta \geq \frac{c_{0,i}}{2}, \quad i = 1, \dots, T,$$

para $\delta \leq c_1 := \min \left\{ \frac{c_{0,1}}{2}, \dots, \frac{c_{0,T}}{2} \right\}$.

B.12. DEMOSTRACIÓN DEL TEOREMA 3.11

Luego, para $i = 1, \dots, T$, se tiene que

$$\begin{aligned} \int_{\mathcal{Y}} f(y, P_{x_i}^0, \sigma_0) \ln \left(\frac{f(y, P_{x_i}^0, \sigma_0)}{f(y, P_{x_i}, \sigma)} \right) \lambda_{\mathcal{Y}}(dy) &\leq \sup_{y \in \mathcal{Y}} \left| \frac{f(y, P_{x_i}^0, \sigma_0)}{f(y, P_{x_i}, \sigma)} - 1 \right| \\ &= \frac{\sup_{y \in \mathcal{Y}} |f(y, P_{x_i}^0, \sigma_0) - f(y, P_{x_i}, \sigma)|}{\inf_{y \in \mathcal{Y}} f(y, P_{x_i}, \sigma)} \\ &< \frac{2\delta}{c_{0,i}}. \end{aligned}$$

Definiendo $c_2 := \epsilon c_1$ y tomando $\delta \leq \min\{c_1, c_2\}$ se concluye que

$$\int_{\mathcal{Y}} f(y, P_{x_i}^0, \sigma_0) \ln \left(\frac{f(y, P_{x_i}^0, \sigma_0)}{f(y, P_{x_i}, \sigma)} \right) \lambda_{\mathcal{Y}}(dy) < \epsilon,$$

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

Así,

$$W_\delta \subseteq \{(P_{x_1}, \dots, P_{x_T}, \sigma) : f(\cdot, P_{x_i}, \sigma) \in KL(f(\cdot, P_{x_i}^0, \sigma_0), \epsilon), P_{x_i} \in \mathcal{P}(\Theta), \sigma \in \mathbb{R}_+\},$$

y por el teorema 3.10 se concluye que W_δ tiene probabilidad estrictamente positiva, y por lo tanto, también

$$\{(P_{x_1}, \dots, P_{x_T}, \sigma) : f(\cdot, P_{x_i}, \sigma) \in KL(f(\cdot, P_{x_i}^0, \sigma_0), \epsilon), P_{x_i} \in \mathcal{P}(\Theta), \sigma \in \mathbb{R}_+\}.$$

Dada la arbitrariedad de $\epsilon > 0$, se concluye que el soporte producto KL del proceso

$$\left\{ \int_{\Theta} \psi(\cdot, \theta, \sigma(\omega)) F_x(\omega)(d\theta) : x \in \mathcal{X} \right\},$$

está dado por el conjunto

$$\prod_{x \in \mathcal{X}} \left\{ \int_{\Theta} \psi(\cdot, \theta, \sigma) P_x(d\theta) : P_x \in \mathcal{P}(\Theta), \sigma > 0 \right\}.$$



B.13 Demostración del Teorema 3.12

Sean $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{G}_{\mathcal{X}}^*$ y $\epsilon > 0$. Dado que la función $(y, \theta) \rightarrow \psi(y, \theta)$ es acotada, entonces existe una constante positiva $C > 1$ tal que $\psi(y, \theta) \leq C$ para todo $y \in \mathcal{Y}$ y $\theta \in \Theta$. Se define $\epsilon^* = \frac{\epsilon}{C}$. Entonces $\epsilon^* < \epsilon$.

Notar que la familia de funciones $\{\Theta \ni \theta \rightarrow \psi(y, \theta) : y \in \mathcal{Y}\}$ es uniformemente continua (por continuidad de $(y, \theta) \rightarrow \psi(y, \theta)$) y $\sup_{y \in \mathcal{Y}} \psi(y, \theta) < \infty$, para todo $\theta \in \Theta$ (por compacidad de \mathcal{Y}). Luego, por el teorema de Arzelá-Ascoli, dado $\epsilon^* > 0$, existe una partición A_1, \dots, A_n de \mathcal{Y} y puntos $y_1 \in A_1, \dots, y_n \in A_n$ tal que

$$\sup_{y \in A_i} \sup_{\theta \in \Theta} |\psi(y, \theta) - \psi(y_i, \theta)| < \epsilon^*.$$

Notar que

$$\begin{aligned} & \sup_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right| \\ & \leq \max \left\{ \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right| : i = 1, \dots, n \right\}. \end{aligned}$$

Entonces

$$\begin{aligned} & \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right| < \epsilon, i = 1, \dots, n \right\} \\ & \subseteq \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right| < \epsilon \right\}. \end{aligned}$$

B.13. DEMOSTRACIÓN DEL TEOREMA 3.12

Luego basta demostrar que

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right| < \epsilon, i = 1, \dots, n \right\} > 0.$$

Notar que

$$\begin{aligned} & \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right| \\ \leq & \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y_i, \theta) G_x(\omega)(d\theta) \right| \\ & + \left| \int_{\Theta} \psi(y_i, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y_i, \theta) P_x^0(d\theta) \right| \\ & + \left| \int_{\Theta} \psi(y_i, \theta) P_x^0(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right|. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} & \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right| \\ \leq & \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y_i, \theta) G_x(\omega)(d\theta) \right| \\ & + \sup_{x \in \mathcal{X}} \left| \int_{\Theta} \psi(y_i, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y_i, \theta) P_x^0(d\theta) \right| \\ & + \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \left| \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) - \int_{\Theta} \psi(y_i, \theta) P_x^0(d\theta) \right|. \end{aligned}$$

Del teorema 3.6 se desprende que

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \left| \int_{\Theta} \psi(y_i, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y_i, \theta) P_x^0(d\theta) \right| < \epsilon, i = 1, \dots, n \right\} > 0,$$

ya que la función $\theta \rightarrow \psi(y_i, \theta)$ pertenece a $\mathcal{C}_0(\Theta)$, para $i = 1, \dots, n$. Se denomina $\Omega_0 \subseteq \Omega$ al conjunto de probabilidad positiva anterior.

Por otra parte, para todo $i = 1, \dots, n$

$$\sup_{x \in \mathcal{X}} \sup_{y \in A_i} \left| \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) - \int_{\Theta} \psi(y_i, \theta) P_x^0(d\theta) \right| \leq \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \int_{\Theta} |\psi(y, \theta) - \psi(y_i, \theta)| P_x^0(d\theta).$$

Notar que

$$\begin{aligned} \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \int_{\Theta} |\psi(y, \theta) - \psi(y_i, \theta)| P_x^0(d\theta) &\leq \sup_{x \in \mathcal{X}} \int_{\Theta} \sup_{y \in A_i} |\psi(y, \theta) - \psi(y_i, \theta)| P_x^0(d\theta) \\ &\leq \left(\sup_{y \in A_i} \sup_{\theta \in \Theta} |\psi(y, \theta) - \psi(y_i, \theta)| \right) \left(\sup_{x \in \mathcal{X}} P_x^0(\Theta) \right) \\ &< \epsilon^* \\ &< \epsilon. \end{aligned}$$

Luego, para todo $i = 1, \dots, n$

$$\sup_{x \in \mathcal{X}} \sup_{y \in A_i} \left| \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) - \int_{\Theta} \psi(y_i, \theta) P_x^0(d\theta) \right| < \epsilon.$$

Falta analizar el término

$$\begin{aligned} &\sup_{x \in \mathcal{X}} \sup_{y \in A_i} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y_i, \theta) G_x(\omega)(d\theta) \right| \\ &\leq \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \sum_{j=1}^{\infty} \pi_j(x, \omega) |\psi(y, \theta_j(\omega)) - \psi(y_i, \theta_j(\omega))|. \end{aligned}$$

Dado que $\sum_{j=1}^{\infty} \pi_j(x, \omega) = 1$ c.s., existe $\Omega_2 \subseteq \Omega$ con $P(\Omega_2) = 1$ tal que para todo $\omega \in \Omega_0 \cap \Omega_2$, $h_n(x, \omega) := \sum_{j=1}^n \pi_j(x, \omega) \rightarrow 1$ cuando $n \rightarrow \infty$, para todo $x \in \mathcal{X}$. La convergencia anterior es uniforme gracias al teorema de Dini, pues $\{h_n\}_{n \in \mathbb{N}}$ es una sucesión

B.13. DEMOSTRACIÓN DEL TEOREMA 3.12

monótona creciente de funciones continuas sobre un compacto convergiendo a una función continua. Adicionalmente, gracias al teorema de Egoroff existe $\Omega_3 \subseteq \Omega_0 \cap \Omega_2$ con $\mathbb{P}(\Omega_3) > 0$ tal que $h_n(x, \omega) \rightarrow 1$ uniformemente en $x \in \mathcal{X}$ y en $\omega \in \Omega_3$. Así, existe $n_1 \in \mathbb{N}$ tal que $\sum_{j=n_1+1}^{\infty} \pi_j(x, \omega) < \epsilon^*$, para todo $\omega \in \Omega_3$ y $x \in \mathcal{X}$.

Se define

$$\Omega_4 := \Omega_3 \cap \{\omega \in \Omega : \theta_1(\omega), \dots, \theta_{n_1}(\omega) \in \Theta\}.$$

Notar que $\mathbb{P}(\Omega_4) > 0$ por independencia. Además, para todo $\omega \in \Omega_4$ se cumple que

$$\begin{aligned} & \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \sum_{j=1}^{\infty} \pi_j(x, \omega) |\psi(y, \theta_j(\omega)) - \psi(y_i, \theta_j(\omega))| \\ & \leq \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \sum_{j=1}^{n_1} \pi_j(x, \omega) |\psi(y, \theta_j(\omega)) - \psi(y_i, \theta_j(\omega))| \\ & \quad + \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \sum_{j=n_1+1}^{\infty} \pi_j(x, \omega) \psi(y, \theta_j(\omega)) \\ & \quad + \sup_{x \in \mathcal{X}} \sum_{j=n_1+1}^{\infty} \pi_j(x, \omega) \psi(y, \theta_j(\omega)) \\ & \leq \sup_{x \in \mathcal{X}} \sum_{j=1}^{n_1} \pi_j(x, \omega) \sup_{y \in A_i} |\psi(y, \theta_j(\omega)) - \psi(y_i, \theta_j(\omega))| \\ & \quad + \sup_{x \in \mathcal{X}} \sum_{j=n_1+1}^{\infty} \left(\sup_{y \in A_i} \psi(y, \theta_j(\omega)) \right) \pi_j(x, \omega) \\ & \quad + C \sup_{x \in \mathcal{X}} \sum_{j=n_1+1}^{\infty} \pi_j(x, \omega) \\ & < \epsilon^* + C\epsilon^* + C\epsilon^* = \epsilon^* + 2C\epsilon^* < 3\epsilon, \end{aligned}$$

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

Luego, para todo $\omega \in \Omega_4$ se tiene que

$$\begin{aligned} & \sup_{x \in \mathcal{X}} \sup_{y \in A_i} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right| < 5\epsilon, \quad i = 1, \dots, n, \\ \Rightarrow & \sup_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \left| \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta) - \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \right| < 5\epsilon, \end{aligned}$$

y se concluye la demostración del teorema. ■

B.14 Demostración del Teorema 3.13

Sean $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{G}_{\mathcal{X}}^*$ y $\epsilon > 0$. Por el teorema 3.12 existe $\Omega_0 \subseteq \Omega$ con $\mathbb{P}(\Omega_0) > 0$ tal que para todo $y \in \mathcal{Y}$, $x \in \mathcal{X}$, $\omega \in \Omega_0$

$$|f(y|x, G_x(\omega)) - f^0(y|x)| < \epsilon.$$

Suponiendo que $f^0(y|x) - \epsilon > 0$ (esto se logra tomando $\epsilon > 0$, lo suficientemente pequeño) se concluye que

$$\frac{f^0(y|x)}{f^0(y|x) + \epsilon} \leq \frac{f^0(y|x)}{f(y|x, G_x(\omega))} \leq \frac{f^0(y|x)}{f^0(y|x) - \epsilon},$$

que es equivalente a

$$\frac{1}{1 + \frac{\epsilon}{f^0(y|x)}} \leq \frac{f^0(y|x)}{f(y|x, G_x(\omega))} \leq \frac{1}{1 - \frac{\epsilon}{f^0(y|x)}},$$

siempre y cuando $f^0(y|x) > 0$.

Notar que

$$\begin{aligned}
 \sup_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} f^0(y|x) &= \sup_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \\
 &\leq \sup_{x \in \mathcal{X}} \int_{\Theta} \left(\sup_{y \in \mathcal{Y}} \psi(y, \theta) \right) P_x^0(d\theta) \\
 &\leq C \sup_{x \in \mathcal{X}} P_x^0(\Theta) = C,
 \end{aligned}$$

donde $C > 1$ es, tal que $\psi(y, \theta) \leq C$, para todo $y \in \mathcal{Y}$, $\theta \in \Theta$. Así,

$$\sup_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} f^0(y|x) < \infty.$$

En lo que sigue se supone que

$$\inf_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} f^0(y|x) > 0.$$

Esto se logra, por ejemplo, si $\psi(y, \theta) \geq \delta > 0$, para todo $y \in \mathcal{Y}$, $\theta \in \Theta$. En efecto,

$$\begin{aligned}
 \inf_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} f^0(y|x) &= \inf_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} \int_{\Theta} \psi(y, \theta) P_x^0(d\theta) \\
 &\geq \inf_{x \in \mathcal{X}} \int_{\Theta} \left(\inf_{y \in \mathcal{Y}} \psi(y, \theta) \right) P_x^0(d\theta) \\
 &\geq \delta \inf_{x \in \mathcal{X}} P_x^0(\Theta) = \delta > 0.
 \end{aligned}$$

Luego, existen $\xi_1, \xi_2 > 0$, tales que

$$\xi_1 \leq f^0(y|x) \leq \xi_2,$$

B.14. DEMOSTRACIÓN DEL TEOREMA 3.13

para todo $y \in \mathcal{Y}$, $x \in \mathcal{X}$. De lo anterior se concluye que

$$M_1(\epsilon) \equiv \frac{\xi_1}{\xi_2 + \epsilon} \leq \frac{f^0(y|x)}{f^0(y|x) + \epsilon} \leq \frac{f^0(y|x)}{f(y|x, G_x(\omega))}.$$

Similarmente, suponiendo que $\xi_1 - \epsilon > 0$ (lo cual es cierto para $\epsilon > 0$, lo suficientemente pequeño) se cumple que

$$\frac{f^0(y|x)}{f(y|x, G_x(\omega))} \leq \frac{f^0(y|x)}{f^0(y|x) - \epsilon} \leq \frac{\xi_2}{\xi_1 - \epsilon} \equiv M_2(\epsilon).$$

Por lo tanto, existen $M_1(\epsilon) > 0$ y $M_2(\epsilon) < \infty$ tales que

$$M_1(\epsilon) \leq \frac{f^0(y|x)}{f(y|x, G_x(\omega))} \leq M_2(\epsilon),$$

para todo $y \in \mathcal{Y}$, $x \in \mathcal{X}$. Notar que $M_1(\epsilon)$ es decreciente en ϵ y $M_2(\epsilon)$ es creciente en ϵ . Dado que $\ln(\cdot)$ es uniformemente continua y acotada en el intervalo $[M_1(\epsilon), M_2(\epsilon)]$, y estrictamente monótona, entonces para todo $\epsilon^* > 0$, existe $\epsilon > 0$, tal que

$$\begin{aligned} & \mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \int_{\mathcal{Y}} f^0(y|x) \ln \left(\frac{f^0(y|x)}{f(y|x, G_x(\omega))} \right) \lambda_{\mathcal{Y}}(dy) < \epsilon^* \right\} \\ & \geq \mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} |f(y|x, G_x(\omega)) - f^0(y|x)| < \epsilon \right\} > 0. \end{aligned}$$

B.14. DEMOSTRACIÓN DEL TEOREMA 3.13

En efecto,

$$\begin{aligned}
 \frac{f^0(y|x)}{f(y|x, G_x(\omega))} \leq M_2(\omega) &\Rightarrow \ln \left(\frac{f^0(y|x)}{f(y|x, G_x(\omega))} \right) \leq \ln(M_2(\omega)) \\
 &\Rightarrow \int_{\mathcal{Y}} f^0(y|x) \ln \left(\frac{f^0(y|x)}{f(y|x, G_x(\omega))} \right) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \ln(M_2(\omega)) \int_{\mathcal{Y}} f^0(y|x) \lambda_{\mathcal{Y}}(dy) = \ln(M_2(\omega)) \\
 &\Rightarrow \sup_{x \in \mathcal{X}} \int_{\mathcal{Y}} f^0(y|x) \ln \left(\frac{f^0(y|x)}{f(y|x, G_x(\omega))} \right) \lambda_{\mathcal{Y}}(dy) \\
 &\leq \ln(M_2(\omega)) \equiv \epsilon^*.
 \end{aligned}$$

Así, para $\epsilon^* > 0$, basta tomar $\epsilon = \frac{\xi_1 \exp(\epsilon^*) - \xi_2}{\exp(\epsilon^*)}$. Notar que $\ln(M_2(\epsilon)) > 0$, pues $\frac{\xi_2}{\xi_1 - \epsilon} > 1$.

En el escenario que

$$\inf_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} f^0(y|x) = 0,$$

tomar $\delta > 0$ (por definir) y considerar

$$\tilde{f}^0(y|x) = \frac{\max\{f^0(y|x), \delta\}}{\int_{\mathcal{Y}} \max\{f^0(y|x), \delta\} \lambda_{\mathcal{Y}}(dy)}.$$

Entonces,

$$\inf_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} \tilde{f}^0(y|x) > 0,$$

y

$$f^0(y|x) \leq L \tilde{f}^0(y|x),$$

con

$$L = \int_{\mathcal{Y}} \max\{f^0(y|x), \delta\} \lambda_{\mathcal{Y}}(dy).$$

Siguiendo las ideas de la demostración del teorema 2 de [Petrone & Wasserman \(2002\)](#) se tiene

que

$$\begin{aligned} & KL(f^0(y|x), f(y|x, G_x(\omega))) \\ & \leq (L+1) \ln(L) \\ & + L \left[KL(\tilde{f}^0(y|x), f(y|x, G_x(\omega))) + \sqrt{KL(\tilde{f}^0(y|x), f(y|x, G_x(\omega)))} \right]. \end{aligned}$$

Luego,

$$KL(\tilde{f}^0(y|x), f(y|x, G_x(\omega))) < \epsilon' \Rightarrow KL(f^0(y|x), f(y|x, G_x(\omega))) \leq \epsilon'',$$

con

$$\epsilon'' = (L+1) \ln(L) + L[\epsilon' + \sqrt{\epsilon'}].$$

Dado que $L \geq 1$, se tiene que $\ln(L) \geq 0$ y entonces $\epsilon'' > 0$.

Finalmente, se concluye que para todo $\epsilon'' > 0$, existen $\delta > 0$, $\epsilon' > 0$ tal que

$$\begin{aligned} & \mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} KL(f^0(y|x), f(y|x, G_x(\omega))) < \epsilon'' \right\} \\ & \geq \mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} KL(\tilde{f}^0(y|x), f(y|x, G_x(\omega))) < \epsilon' \right\} > 0, \end{aligned}$$

y se concluye el teorema. ■

B.15 Demostración del Teorema 3.14

Sean $\{P_x^0 : x \in \mathcal{X}\} \in \mathcal{G}_{\mathcal{X}}^*$ y $\epsilon > 0$. Definiendo

$$f^0(y|x) := \int_{\Theta} \psi(y, \theta) P_x^0(d\theta),$$

$$m_0(y, x) := q(x) f^0(y|x),$$

y

$$m^{(\omega)}(y, x) := q(x) f(y|x, G_x(\omega)),$$

entonces

$$\begin{aligned} KL(m_0, m^{(\omega)}) &= \int_{\mathcal{X}} \int_{\mathcal{Y}} m_0(y, x) \ln \left(\frac{m_0(y, x)}{m^{(\omega)}(y, x)} \right) \lambda_{\mathcal{Y}}(dy) dx \\ &= \int_{\mathcal{X}} q(x) \left[\int_{\mathcal{Y}} f^0(y|x) \ln \left(\frac{f^0(y|x)}{f(y|x, G_x(\omega))} \right) \lambda_{\mathcal{Y}}(dy) \right] dx \\ &\leq \sup_{x \in \mathcal{X}} \int_{\mathcal{Y}} f^0(y|x) \ln \left(\frac{f^0(y|x)}{f(y|x, G_x(\omega))} \right) \lambda_{\mathcal{Y}}(dy). \end{aligned}$$

Por el teorema 3.13 se concluye que

$$\begin{aligned} 0 &< \mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} \int_{\mathcal{Y}} f^0 \left(\frac{f^0(y|x)}{f(y|x, G_x(\omega))} \right) \lambda_{\mathcal{Y}}(dy) < \epsilon \right\} \\ &\leq \mathbb{P} \{ \omega \in \Omega : KL(m_0, m^{(\omega)}) < \epsilon \}. \end{aligned}$$

Finalmente, por un teorema de [Schwartz \(1965\)](#) se concluye la demostración del teorema. ■

B.16 Demostración del Teorema B.1

Teorema B.1. Sea $(\tilde{\Theta}, \tilde{d})$ un espacio métrico polaco, $\Theta \subseteq \tilde{\Theta}$ un conjunto cerrado y $\mathcal{P}(\Theta)$ el espacio de todas las medidas de probabilidad sobre el espacio medible $(\Theta, \mathcal{B}(\Theta))$, con $\mathcal{B}(\Theta)$ la σ -álgebra de los Borelianos asociada a la topología traza subyacente. Sea λ una medida Boreliana localmente finita estrictamente positiva definida sobre el espacio medible $(\tilde{\Theta}, \mathcal{B}(\tilde{\Theta}))$, con $\mathcal{B}(\tilde{\Theta})$ la σ -álgebra de los Borelianos respectiva. Sea $F_0 \in \mathcal{P}(\Theta)$ tal que $\text{supp}(F_0) = \Theta$ y $F_0 \ll \lambda$. Para $P_0 \in \mathcal{P}(\Theta)$, definamos

$$U(P_0, f_1, \dots, f_k, \epsilon) = \left\{ P \in \mathcal{P}(\Theta) : \left| \int_{\Theta} f_i(\theta) dP(\theta) - \int_{\Theta} f_i(\theta) dP_0(\theta) \right| < \epsilon, i = 1, \dots, k \right\},$$

donde $\epsilon > 0$, $k \in \mathbb{N} \setminus \{0\}$ y $f_i \in \mathcal{C}_b(\Theta)$, $i = 1, \dots, k$. Entonces $\exists Q \in \mathcal{P}(\Theta)$ tal que $Q \in U(P_0, f_1, \dots, f_k, \epsilon)$ y $Q \ll F_0$.

DEMOSTRACIÓN: Dado que Θ es un conjunto cerrado en $\tilde{\Theta}$, entonces (Θ, d) es un espacio métrico polaco donde d es la restricción de la métrica \tilde{d} al espacio Θ (ver capítulo 3 de [Aliprantis & Border \(2006\)](#)). Luego, gracias al teorema 6.3 de [Parthasarathy \(1967\)](#) se tiene que el conjunto de medidas de probabilidad cuyos soportes son subconjuntos finitos de un subconjunto denso de Θ , es denso en $\mathcal{P}(\Theta)$. Así, dado $\epsilon > 0$ existe $Q^* \in \mathcal{P}(\Theta)$ de la forma

$$Q^*(\cdot) := \sum_{j=1}^N W_j \delta_{\theta_j}(\cdot),$$

con $N \in \mathbb{N} \setminus \{0\}$,

$$(W_1, \dots, W_N) \in \Delta_N := \left\{ (w_1, \dots, w_N) : w_i \geq 0, i = 1, \dots, N, \sum_{j=1}^N w_j = 1 \right\},$$

y $\theta_1, \dots, \theta_N \in \Theta$, tales que

$$\left| \int_{\Theta} f_i(\theta) dQ^*(\theta) - \int_{\Theta} f_i(\theta) dP_0(\theta) \right| < \frac{\epsilon}{2}, \quad i = 1, \dots, k.$$

B.16. DEMOSTRACIÓN DEL TEOREMA B.1

Sea

$$B_{\tilde{d}}(\theta, r) = \{\vartheta \in \tilde{\Theta} : \tilde{d}(\theta, \vartheta) < r\}.$$

Es posible determinar $\delta > 0$ tal que

$$B_{\tilde{d}}(\theta_{j_1}, \delta) \cap B_{\tilde{d}}(\theta_{j_2}, \delta) = \emptyset, \quad \forall j_1 \neq j_2,$$

y

$$f_i(\theta_j) - \frac{\epsilon}{2N} < f_i(\theta) < f_i(\theta_j) + \frac{\epsilon}{2N}, \quad \forall \theta \in B_{\tilde{d}}(\theta_j, \delta) \cap \Theta, \quad j = 1, \dots, N, \quad i = 1, \dots, k.$$

En efecto, basta escoger $\delta = \min\{\delta_1, \delta_2, \delta_3\}$ donde δ_1 garantiza la separación de las bolas $B_{\tilde{d}}(\theta_1, \delta_1), \dots, B_{\tilde{d}}(\theta_N, \delta_1)$, δ_2 surge de la continuidad de f_1, \dots, f_k y δ_3 asegura que las bolas $B_{\tilde{d}}(\theta_1, \delta_3), \dots, B_{\tilde{d}}(\theta_N, \delta_3)$ tienen una medida finita con respecto a λ (gracias a su propiedad de finitud local). Para lo que sigue, definamos

$$q(\theta) := \sum_{j=1}^N \left(\frac{W_j}{c_{\theta_j, \delta}} \right) \mathbb{1}_{B_{\tilde{d}}(\theta_j, \delta) \cap \Theta}(\theta),$$

con $c_{\theta_j, \delta} = \lambda(B_{\tilde{d}}(\theta_j, \delta) \cap \Theta)$. Notemos que estos números están bien definidos gracias a que $B_{\tilde{d}}(\theta_j, \delta) \in \mathcal{B}(\tilde{\Theta})$, $\forall j = 1, \dots, N$, y λ es una medida localmente finita estrictamente positiva. Por construcción sigue que

$$\int_{\Theta} q(\theta) \lambda(d\theta) = 1.$$

Así, para

$$Q(A) := \int_A q(\theta) \lambda(d\theta), \quad A \in \mathcal{B}(\Theta),$$

se tiene que $Q \in \mathcal{P}(\Theta)$ y $Q \ll \lambda$. De lo anterior y la monotonía de la integral se tiene que

$$\begin{aligned} W_j f_i(\theta_j) - W_j \left(f_i(\theta_j) + \frac{\epsilon}{2N} \right) &< W_j f_i(\theta_j) - \int_{B_{\tilde{d}}(\theta_j, \delta) \cap \Theta} f_i(\theta) q(\theta) \lambda(d\theta) \\ &< W_j f_i(\theta_j) - W_j \left(f_i(\theta_j) - \frac{\epsilon}{2N} \right). \end{aligned}$$

Luego,

$$\left| W_j f_i(\theta_j) - \int_{B_{\tilde{d}}(\theta_j, \delta) \cap \Theta} f_i(\theta) q(\theta) \lambda(d\theta) \right| < \frac{W_j \epsilon}{2N} < \frac{\epsilon}{2N},$$

y entonces

$$\left| \int_{\Theta} f_i(\theta) dQ^*(\theta) - \int_{\Theta} f_i(\theta) q(\theta) \lambda(d\theta) \right| \leq \sum_{j=1}^N \left| W_j f_i(\theta_j) - \int_{B(\theta_j, \delta)} f_i(\theta) q(\theta) \lambda(d\theta) \right| < \frac{\epsilon}{2}.$$

Por lo tanto,

$$\begin{aligned} \left| \int_{\Theta} f_i(\theta) dQ(\theta) - \int_{\Theta} f_i(\theta) dP_0(\theta) \right| &\leq \left| \int_{\Theta} f_i(\theta) dQ(\theta) - \int_{\Theta} f_i(\theta) dQ^*(\theta) \right| \\ &+ \left| \int_{\Theta} f_i(\theta) dQ^*(\theta) - \int_{\Theta} f_i(\theta) dP_0(\theta) \right| \\ &< \epsilon, \end{aligned}$$

y se deduce que $Q \in U(P_0, f_1, \dots, f_k, \epsilon)$.

Notemos que $\mathcal{B}(\Theta) \subseteq \mathcal{B}(\tilde{\Theta})$, pues Θ es cerrado en $\tilde{\Theta}$. Dado que $\text{supp}(F_0) = \Theta$ y λ es una medida estrictamente positiva, entonces $\forall B \in \mathcal{B}(\Theta)$ tal que $F_0(B) = 0$, se cumple que $\lambda(B) = 0$. Como $Q \ll \lambda$, se concluye que $Q(B) = 0$, y por lo tanto, $Q \ll F_0$. ■

B.17 Demostración del Lema B.1

Lema B.1. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, $\mathcal{X} \subseteq \mathbb{R}^q$ un conjunto compacto, Θ un espacio Polaco y $\{\pi_i(x) : i = 1, 2, \dots\}$ definido como

$$\pi_i(x) = V_i(x, \omega) \prod_{j < i} (1 - V_j(x, \omega)),$$

tal que

(a) $V_i \in \mathcal{C}(\mathcal{X}, \mathbb{R}) := \{V : \mathcal{X} \rightarrow \mathbb{R} : \text{fn. continua}\}$, c.s., $\forall i \in \mathbb{N} \setminus \{0\}$.

(b) Para toda función continua $g : \mathcal{X} \rightarrow [0, 1]$,

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |V_i(x, \omega) - g(x)| < \epsilon \right\} > 0,$$

para todo $i = 1, 2, \dots$ y $\epsilon > 0$.

Entonces

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |\pi_1(x, \omega) - F_x(A_1)| < \epsilon_1, \dots, \sup_{x \in \mathcal{X}} |\pi_k(x, \omega) - F_x(A_k)| < \epsilon_k \right\} > 0,$$

para una partición medible $\{A_i : i = 1, \dots, k\}$ de Θ , $\epsilon_i > 0$ y $\{F_x : x \in \mathcal{X}\} \in \mathcal{P}(\Theta)^{\mathcal{X}}$ tal que $x \rightarrow F_x(B)$ es una función continua, $\forall B \in \mathcal{B}(\Theta)$.

DEMOSTRACIÓN: Gracias a la compacidad de \mathcal{X} se tiene que la aplicación $x \rightarrow F_x(B)$ es uniformemente continua, para todo $B \in \mathcal{B}(\Theta)$. Luego, sin pérdida de generalidad, podemos tomar una partición $\{A_i : i = 1, \dots, k\}$ de Θ tal que $0 < F_x(A_i) < 1$, $i = 1, \dots, k$, para todo $x \in \mathcal{X}$. Por demostrar que

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |\pi_1(x, \omega) - F_x(A_1)| < \epsilon_1, \dots, \sup_{x \in \mathcal{X}} |\pi_k(x, \omega) - F_x(A_k)| < \epsilon_k \right\} > 0.$$

Construyamos funciones $g_i : \mathcal{X} \rightarrow [0, 1]$, $i = 1, \dots, k$, tales que

$$\begin{aligned} g_1(x) &= F_x(A_1), \\ g_i(x) \prod_{j < i} (1 - g_j(x)) &= F_x(A_i), \quad 2 \leq i \leq k-1, \\ g_k(x) &\equiv 1. \end{aligned}$$

Notemos que g_1, \dots, g_k están bien definidas y son funciones continuas. Más aún, $Im(g_i) \subseteq$

(0, 1) para todo $i = 1, \dots, k - 1$ y se puede demostrar que

$$g_i(x) = \frac{F_x(A_i)}{\sum_{j=i}^k F_x(A_j)}, \quad i = 2, \dots, k - 1.$$

Por hipótesis acerca de los V_i 's se tiene que

$$\mathbb{P} \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |V_1(x, \omega) - g_1(x)| < \epsilon_1^*, \dots, \sup_{x \in \mathcal{X}} |V_k(x, \omega) - 1| < \epsilon_k^* \right\} > 0,$$

para $\epsilon_1^*, \dots, \epsilon_k^* > 0$. En lo que sigue se demuestra que para determinada elección de $\epsilon_i^* > 0$, $i = 1, \dots, k$, se verifica que

$$\begin{aligned} & \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |V_1(x, \omega) - g_1(x)| < \epsilon_1^*, \dots, \sup_{x \in \mathcal{X}} |V_k(x, \omega) - 1| < \epsilon_k^* \right\} \\ & \subseteq \left\{ \omega \in \Omega : \sup_{x \in \mathcal{X}} |\pi_1(x, \omega) - F_x(A_1)| < \epsilon_1, \dots, \sup_{x \in \mathcal{X}} |\pi_k(x, \omega) - F_x(A_k)| < \epsilon_k \right\}, \end{aligned} \quad (\text{B.8})$$

lo cual demuestra el lema B.1. En efecto, siguiendo el argumento de [Pati et al. \(2013\)](#), definamos las funciones

$$\begin{aligned} f_1 : [0, 1]^k & \rightarrow \mathbb{R}_+, \\ \vec{p} = (p_1, \dots, p_k) & \rightarrow f_1(\vec{p}) = p_1, \end{aligned}$$

y para $i = 2, \dots, k$,

$$\begin{aligned} f_i : [0, 1]^k & \rightarrow \mathbb{R}_+, \\ \vec{p} = (p_1, \dots, p_k) & \rightarrow f_i(\vec{p}) = p_i \prod_{j < i} (1 - p_j). \end{aligned}$$

Notemos que para $i = 2, \dots, k$ se tiene que

$$\begin{aligned} \sup_{x \in \mathcal{X}} |f_i(\vec{p}(x)) - f_i(\vec{q}(x))| & \leq (i - 1) \sup_{x \in \mathcal{X}} |p_i(x) - q_i(x)| \\ & + \sum_{j < i} \sup_{x \in \mathcal{X}} |f_j(\vec{p}(x)) - f_j(\vec{q}(x))|. \end{aligned}$$

Luego, tomando

$$\begin{aligned}\vec{p}(x) &= (V_1(x), \dots, V_k(x)), \\ \vec{q}(x) &= (g_1(x), \dots, g_{k-1}(x), 1),\end{aligned}$$

se concluye la existencia de $\epsilon_i^* > 0$, $i = 1, \dots, k$, tales que se satisfaga (B.8), y se concluye la demostración del lema. ■

B.18 Demostración del Lema B.2

Lema B.2. *Bajo las condiciones del teorema 3.1 se tiene que*

$$\lim_{n \rightarrow \infty} \int_{\Theta} f_j(\theta) G_{x_n}(\omega)(d\theta) = \int_{\Theta} f_j(\theta) G_{x_0}(\omega)(d\theta), \quad \forall j = 1, \dots, k.$$

DEMOSTRACIÓN: Las integrales anteriores se pueden escribir, $\forall j = 1, \dots, k$, como

$$\int_{\Theta} f_j(\theta) G_{x_n}(\omega)(d\theta) = \sum_{i=1}^{\infty} \left\{ V_i(x_n, \omega) \prod_{l < i} [1 - V_l(x_n, \omega)] \right\} f_j(\theta_i(x_n, \omega)),$$

y

$$\int_{\Theta} f_j(\theta) G_{x_0}(\omega)(d\theta) = \sum_{i=1}^{\infty} \left\{ V_i(x_0, \omega) \prod_{l < i} [1 - V_l(x_0, \omega)] \right\} f_j(\theta_i(x_0, \omega)).$$

Dado que $\omega \in W$ y $j = 1, \dots, k$ son arbitrarios, dichas variables no serán escritas explícitamente en lo sigue. Definamos las funciones $h_n, h : \mathbb{N} \setminus \{0\} \rightarrow \mathbb{R}$ como

$$h_n(i) := \left(V_i(x_n, \omega) \prod_{l < i} [1 - V_l(x_n, \omega)] \right) f_j(\theta_i(x_n, \omega)),$$

y

$$h(i) := \left(V_i(x_0, \omega) \prod_{l < i} [1 - V_l(x_0, \omega)] \right) f_j(\theta_i(x_0, \omega)),$$

B.18. DEMOSTRACIÓN DEL LEMA B.2

$\forall i \in \mathbb{N} \setminus \{0\}$. Dotemos al conjunto $\mathbb{N} \setminus \{0\}$ de la σ -álgebra potencia \mathcal{F} (esto es, el conjunto potencia de $\mathbb{N} \setminus \{0\}$) y la medida cuenta puntos $m(\cdot)$. Así, $(\mathbb{N} \setminus \{0\}, \mathcal{F}, m)$ es un espacio de medida y las funciones h_n, h son \mathcal{F} -medibles. Por álgebra y composición de funciones continuas se tiene que

$$\lim_{n \rightarrow \infty} h_n(i) = h(i), \quad \forall i \geq 1.$$

Gracias a que

$$\sum_{i=1}^{\infty} \left\{ V_i(x, \omega) \prod_{l < i} [1 - V_l(x, \omega)] \right\} = 1, \quad \forall x \in \mathcal{X}, \omega \in W,$$

y f_1, \dots, f_k son funciones acotadas, entonces h_n, h son funciones m -integrables. Así, podemos construir la función $g : \mathbb{N} \setminus \{0\} \rightarrow \mathbb{R}$, tal que

$$g(i) := \left(\frac{1}{2}\right)^{i+1} M_j, \quad \forall i \geq 1,$$

donde $M_j > 0$ es una constante finita que acota uniformemente a $|f_j|$. Se observa que

$$\sum_{i \geq 1} g(i) = 2M_j < \infty,$$

y que $|h_n(i)| \leq g(i)$, $\forall i, n \in \mathbb{N} \setminus \{0\}$. Luego, por el teorema de convergencia dominada, se tiene que

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{\mathbb{N} \setminus \{0\}} h_n(u) m(du) &= \int_{\mathbb{N} \setminus \{0\}} \lim_{n \rightarrow \infty} h_n(u) m(du) \\ &= \int_{\mathbb{N} \setminus \{0\}} h(u) m(du). \end{aligned}$$

B.18. DEMOSTRACIÓN DEL LEMA B.2

Dado que $\omega \in W$ y $j = 1, \dots, k$ son arbitrarios, se concluye que

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sum_{i=1}^{\infty} \left\{ V_i(x_n, \omega) \prod_{l < i} [1 - V_l(x_n, \omega)] \right\} f_j(\theta_i(x_n, \omega)) \\ &= \sum_{i=1}^{\infty} \left\{ V_i(x_0, \omega) \prod_{l < i} [1 - V_l(x_0, \omega)] \right\} f_j(\theta_i(x_0, \omega)), \end{aligned}$$

y se concluye el lema. ■

Bibliografía

- AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society* 139 655–673.
- ALIPRANTIS, C. & BORDER, K. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Berlin Heidelberg.
- ANDERSEN, H. H. (1995). *Linear and Graphical Models: For the Multivariate Complex Normal Distribution*, vol. 101. Springer Science & Business Media.
- ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian non-parametric problems. *The Annals of Statistics* 2 1152–1174.
- BALL, F. G., DRYDEN, I. L. & GOLALIZADEH, M. (2008). Brownian motion and Ornstein-Uhlenbeck processes in planar shape space. *Methodology and Computing in Applied Probability* 10 1–22.
- BARRIENTOS, A. F., JARA, A. & QUINTANA, F. A. (2012). On the support of MacEachern's dependent Dirichlet processes and extensions. *Bayesian Analysis* 7 277–310.
- BHATTACHARYA, A. & DUNSON, D. B. (2010). Nonparametric Bayesian density estimation on manifolds with applications to planar shapes. *Biometrika* 97 851–865.

BIBLIOGRAFÍA

- BHATTACHARYA, A. & DUNSON, D. B. (2012a). Nonparametric Bayes classification and hypothesis testing on manifolds. *Journal of Multivariate Analysis* 111 1–19.
- BHATTACHARYA, A. & DUNSON, D. B. (2012b). Strong consistency of nonparametric Bayes density estimation on compact metric spaces with applications to specific manifolds. *Annals of the Institute of Statistical Mathematics* 64 687–714.
- BHATTACHARYA, R. & PATRANGENARU, V. (2002). Nonparametric estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference* 108 23–35.
- BHATTACHARYA, R. & PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *The Annals of Statistics* 31 1–29.
- BHATTACHARYA, R. & PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *The Annals of statistics* 33 1225–1259.
- BILLERA, L. J., HOLMES, S. P. & VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics* 27 733–767.
- BILLINGSLEY, P. (1995). *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley.
- BILLINGSLEY, P. (1999). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley.
- BLACKWELL, D. & MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* 1 353–355.
- BLEI, D. M., JORDAN, M. I. ET AL. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis* 1 121–144.
- BOURBAKI, N. (1998). *General Topology: Chapters 5–10*. Springer Berlin Heidelberg.

BIBLIOGRAFÍA

- BREZIS, H. (2010). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer New York.
- BRIGNELL, C. J., BROWNE, W. J. & DRYDEN, I. L. (2005). Covariance weighted Procrustes analysis. In *In*. University Press, 107–110.
- CHUNG, Y. & DUNSON, D. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *Journal of the American Statistical Association* 104 1646–1660.
- CHUNG, Y. & DUNSON, D. B. (2011). The local Dirichlet process. *Annals of the Institute of Statistical Mathematics* 63 59–80.
- CIFARELLI, D. & REGAZZINI, E. (1978). Problemi statistici non parametrici in condizioni di scambiabilità parziale e impiego di medie associative. Tech. rep., Quaderni Istituto Matematica Finanziaria dell'Università di Torino.
- CIFARELLI, D. M. & REGAZZINI, E. (1990). Distribution functions of means of a Dirichlet process. *The Annals of Statistics* 18 429–442.
- CLAUDE, J. (2008). *Morphometrics with R. Use R!* Springer New York.
- COHN, D. L. (2013). *Measure Theory*. Springer Science & Business Media.
- DE IORIO, M., MÜLLER, P., ROSNER, G. L. & MACEACHERN, S. N. (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association* 99 205–215.
- DO CARMO, M. (1992). *Riemannian Geometry*. Mathematics (Boston, Mass.). Birkhäuser.
- DRYDEN, I. & MARDIA, K. (2016). *Statistical Shape Analysis: With Applications in R*. Wiley Series in Probability and Statistics. Wiley.
- DRYDEN, I. L. (2016). *shapes package*. R Foundation for Statistical Computing, Vienna, Austria. Contributed package, Version 1.1-13., URL <http://www.R-project.org>.
- DRYDEN, I. L. & MARDIA, K. V. (1998). *Statistical Shape Analysis*. John Wiley.

BIBLIOGRAFÍA

- DUNSON, D. B. & PARK, J.-H. (2008). Kernel stick-breaking processes. *Biometrika* 95 307–323.
- DUNSON, D. B., PILLAI, N. & PARK, J.-H. (2007). Bayesian density regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69 163–183.
- DUNSON, D. B., XUE, Y. & CARIN, L. (2008). The matrix stick-breaking process: flexible Bayes meta-analysis. *Journal of the American Statistical Association* 103 317–327.
- EPIFANI, I. & LIJOI, A. (2010). Nonparametric priors for vectors of survival functions. *Statistica Sinica* 20 1455–1484.
- FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1 209–230.
- FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics* 2 615–629.
- FERGUSON, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. *Recent Advances in Statistics* 24 287–302.
- FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 222 309–368.
- FLORENS, J.-P., MOUCHART, M., ROLIN, J.-M. ET AL. (1990). *Elements of Bayesian Statistics*. Marcel Dekker (Basel).
- FOX, N. I., MICHEAS, A. C. & PENG, Y. (2016). Applications of Bayesian Procrustes shape analysis to ensemble radar reflectivity nowcast verification. *Atmospheric Research* 176 75–86.
- FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré* 10 215–310.

BIBLIOGRAFÍA

- GELFAND, A. E., KOTTAS, A. & MACEACHERN, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* 100 1021–1035.
- GEMAN, S. & GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6 721–741.
- GHOSAL, S., GHOSH, J. K., RAMAMOORTHY, R. ET AL. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics* 27 143–158.
- GHOSAL, S., VAN DER VAART, A. ET AL. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *The Annals of Statistics* 35 697–723.
- GHOSH, J. & RAMAMOORTHY, R. (2003). *Bayesian Nonparametrics*. Springer Series in Statistics. Springer.
- GIKHMAN, I. & SKOROKHOD, A. (1969). *Introduction to the Theory of Random Processes*. W. B. Saunders Company.
- GOLALIZADEH, M. (2010). A useful family of stochastic processes for modeling shape diffusions. *Journal of Statistical Research of Iran* 7 21–36.
- GOODALL, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)* 53 285–339.
- GOWER, J. C. (1975). Generalized procrustes analysis. *Psychometrika* 40 33–51.
- GRIFFIN, J. E. & STEEL, M. J. (2006). Order-based dependent Dirichlet processes. *Journal of the American Statistical Association* 101 179–194.
- GUSTAFSON, P. (2015). *Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited Data*. CRC Press.

BIBLIOGRAFÍA

- GUTIÉRREZ, L., MENA, R. H. & RUGGIERO, M. (2016). A time dependent Bayesian nonparametric model for air quality analysis. *Computational Statistics & Data Analysis* 95 161–175.
- HANSON, T. E., BRANSCUM, A. J. & JOHNSON, W. O. (2005). Bayesian Nonparametric Modeling and Data Analysis: An Introduction. In D. Dey & C. Rao, eds., *Bayesian Thinking - Modeling and Computation*, vol. 25 of *Handbook of Statistics*. Elsevier, 245–278.
- HJORT, N., HOLMES, C., MÜLLER, P. & WALKER, S. (2010). *Bayesian Nonparametrics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- HJORT, N. L. & ONGARO, A. (2005). Exact inference for random Dirichlet means. *Statistical Inference for Stochastic Processes* 8 227–254.
- HSIAO, C. (1983). Chapter 4 - Identification. In *Handbook of Econometrics*, vol. 1. Elsevier, 223–283.
- HSU, E. (2002). *Stochastic Analysis on Manifolds*. Contemporary Mathematics. American Mathematical Society.
- ISHWARAN, H. & JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* 96 161–173.
- ISHWARAN, H. & ZAREPOUR, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* 87 371–390.
- JARA, A. & HANSON, T. E. (2011). A class of mixtures of dependent tail-free processes. *Biometrika* 98 553–566.
- JARA, A., LESAFFRE, E., DE IORIO, M. & QUINTANA, F. (2010). Bayesian semiparametric inference for multivariate doubly-interval-censored data. *The Annals of Applied Statistics* 4 2126–2149.
- JOST, J. (2008). *Riemannian Geometry and Geometric Analysis*. Universitext. Springer Berlin Heidelberg.

BIBLIOGRAFÍA

- KALLENBERG, O. (1997). *Foundations of Modern Probability*. Probability and its Applications A Series of the Applied Probability Trust. Springer.
- KARATZAS, I. & SHREVE, S. (1991). *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York.
- KENDALL, D. G. (1977). The diffusion of shape. *Advances in Applied Probability* 9 428–430.
- KENDALL, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society* 16 81–121.
- KENDALL, D. G., BARDEN, D., CARNE, T. K. & LE, H. (1999). *Shape and Shape Theory*. Wiley Series in Probability and Statistics. Wiley.
- KENDALL, W. S. (1988). Symbolic computation and the diffusion of shapes of triads. *Advances in Applied Probability* 20 775–797.
- KENDALL, W. S. (1990). The diffusion of Euclidean shape. In *In Disorder in Physical Systems*. Oxford Univ. Press, 428–430.
- KENDALL, W. S. (1998). A diffusion model for Bookstein triangle shape. *Advances in Applied Probability* 30 317–334.
- KENT, J. T. (1994). The complex Bingham distribution and shape analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* 56 285–299.
- KENT, J. T. & MARDIA, K. V. (1997). Consistency of Procrustes estimators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59 281–290.
- KHOSHNEVISAN, D. (2002). *Multiparameter Processes: An Introduction to Random Fields*. Monographs in Mathematics. Springer.
- KLINGENBERG, W. (1995). *Riemannian Geometry*. De Gruyter Studies in Mathematics. De Gruyter.

BIBLIOGRAFÍA

- KONG, A., LIU, J. S. & WONG, W. H. (1994). Sequential imputations and Bayesian missing data problems. *Journal of the American statistical association* 89 278–288.
- KOOPMANS, T. C. & REIERSOL, O. (1950). The identification of structural characteristics. *The Annals of Mathematical Statistics* 21 165–181.
- KORWAR, R. M. & HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *The Annals of Probability* 1 705–711.
- KOSCHAT, M. A. & SWAYNE, D. F. (1991). A weighted Procrustes criterion. *Psychometrika* 56 229–239.
- LANG, S. (1995). *Differential and Riemannian Manifolds*. Graduate Texts in Mathematics. Springer.
- LAWRENCE, E., BINGHAM, D., LIU, C. & NAIR, V. N. (2008). Bayesian inference for multivariate ordinal data using parameter expansion. *Technometrics* 50 182–191.
- LE, H. (1991). A stochastic calculus approach to the shape distribution induced by a complex normal model. *Mathematical Proceedings of the Cambridge Philosophical Society* 109 221–228.
- LE, H. (1994). Brownian motions on shape and size-and-shape spaces. *Journal of Applied Probability* 31 101–113.
- LEISEN, F. & LIJOI, A. (2011). Vectors of two-parameter Poisson-Dirichlet processes. *Journal of Multivariate Analysis* 102 482–495.
- LIJOI, A., NIPOTI, B., PRÜNSTER, I. ET AL. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* 20 1260–1291.
- LIJOI, A., PRÜNSTER, I. & WALKER, S. G. (2005). On consistency of nonparametric normal mixtures for Bayesian density estimation. *Journal of the American Statistical Association* 100 1292–1296.

BIBLIOGRAFÍA

- LINDLEY, D. V. (1972). *Bayesian Statistics: A Review*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics.
- LISSITZ, R. W., SCHÖNEMANN, P. H. & LINGOES, J. C. (1976). A solution to the weighted Procrustes problem in which the transformation is in agreement with the loss function. *Psychometrika* 41 547–550.
- LIU, C., RUBIN, D. B. & WU, Y. N. (1998). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika* 85 755–770.
- LIU, J. S. (1996). Nonparametric hierarchical Bayes via sequential imputations. *The Annals of Statistics* 24 911–930.
- LO, A. Y. ET AL. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics* 12 351–357.
- MACEACHERN, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science, Alexandria, VA*. American Statistical Association, 50–55.
- MACEACHERN, S. N. (2000). Dependent dirichlet processes. Tech. rep., Department of Statistics, The Ohio State University.
- MACEACHERN, S. N., CLYDE, M. & LIU, J. S. (1999). Sequential importance sampling for nonparametric Bayes models: the next generation. *Canadian Journal of Statistics* 27 251–267.
- MARDIA, K. & DRYDEN, I. (1999). The complex Watson distribution and shape analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61 913–926.
- MARDIA, K. & JUPP, P. (2000). *Directional Statistics*. Wiley Series in Probability and Statistics. Wiley.

BIBLIOGRAFÍA

- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21 1087–1092.
- MICHEAS, A. C., DEY, D. K. & MARDIA, K. V. (2006). Complex elliptical distributions with application to shape analysis. *Journal of Statistical Planning and Inference* 136 2961–2982.
- MICHEAS, A. C., FOX, N. I., LACK, S. A. & WIKLE, C. K. (2007). Cell identification and verification of QPF ensembles using shape analysis techniques. *Journal of Hydrology* 343 105–116.
- MICHEAS, A. C. & PENG, Y. (2010). Bayesian Procrustes analysis with applications to hydrology. *Journal of Applied Statistics* 37 41–55.
- MITRA, R. & MÜLLER, P. (2015). *Nonparametric Bayesian Inference in Biostatistics*. Frontiers in Probability and the Statistical Sciences. Springer International Publishing.
- MULIERE, P. & SECCHI, P. (1995). A note on a proper Bayesian Bootstrap. Tech. rep., Università degli Studi di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi.
- MULIERE, P. & TARDELLA, L. (1998). Approximating distributions of random functionals of Ferguson-Dirichlet priors. *Canadian Journal of Statistics* 26 283–297.
- MÜLLER, P., ERKANLI, A. & WEST, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* 83 67–79.
- MÜLLER, P. & QUINTANA, F. (2010). Random partition models with regression on covariates. *Journal of Statistical Planning and Inference* 140 2801–2808.
- MÜLLER, P., QUINTANA, F., JARA, A. & HANSON, T. (2015). *Bayesian Nonparametric Data Analysis*. Springer Series in Statistics. Springer International Publishing.
- MÜLLER, P., QUINTANA, F. & ROSNER, G. (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66 735–749.

BIBLIOGRAFÍA

- MÜLLER, P., QUINTANA, F. & ROSNER, G. L. (2011). A product partition model with regression on covariates. *Journal of Computational and Graphical Statistics* 20 260–278.
- MUNKRES, J. R. (2000). *Topology*. Prentice Hall.
- NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden-Day.
- NEWTON, M. A., QUINTANA, F. A. & ZHANG, Y. (1998). Nonparametric Bayes methods using predictive updating. In D. Dey, P. Müller & D. Sinha, eds., *Practical Nonparametric and Semiparametric Bayesian Statistics*. Springer, 45–62.
- NEWTON, M. A. & ZHANG, Y. (1999). A recursive algorithm for nonparametric analysis with missing data. *Biometrika* 86 15–26.
- NIETO-BARAJAS, L. E., PRÜNSTER, I., WALKER, S. G. ET AL. (2004). Normalized random measures driven by increasing additive processes. *The Annals of Statistics* 32 2343–2360.
- NIPOTI, B. (2011). *Dependent completely random measures and statistical applications*. Unpublished doctoral thesis, Department of Mathematics, University of Pavia.
- OKSENDAL, B. (2013). *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer Berlin Heidelberg.
- OPPER, M. & SAAD, D. (2001). *Advanced Mean Field Methods: Theory and Practice*. Neural Information Processing Series. MIT Press.
- PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Providence, USA: AMS Chelsea Publishing.
- PATI, D., DUNSON, D. B. & TOKDAR, S. T. (2013). Posterior consistency in conditional distribution estimation. *Journal of Multivariate Analysis* 116 456–472.
- PATRANGENARU, V. & ELLINGSON, L. (2015). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press.

BIBLIOGRAFÍA

- PELLETIER, B. (2005). Kernel density estimation on Riemannian manifolds. *Statistics & Probability Letters* 73 297–304.
- PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *Journal of Nonparametric Statistics* 18 57–67.
- PETERSEN, P. (2016). *Riemannian Geometry*. Graduate Texts in Mathematics. Springer International Publishing.
- PETRONE, S. & WASSERMAN, L. (2002). Consistency of Bernstein polynomial posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 79–100.
- PITMAN, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In T. S. Ferguson, L. S. Shapley & J. B. MacQueen, eds., *Statistics, Probability and Game Theory. Papers in Honor of David Blackwell*. IMS Lecture Notes - Monograph Series, Hayward, California, 245–268.
- PITMAN, J. & YOR, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability* 25 855–900.
- QUINTANA, F. A. (2010). Linear regression with a dependent skewed Dirichlet process. *Chilean Journal of Statistics* 1 35–49.
- R CORE TEAM (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RAO, M. (2005). *Conditional Measures and Applications*. Monographs and Textbooks in Pure and Applied Mathematics. CRC Press.
- REN, L., DU, L., CARIN, L. & DUNSON, D. (2011). Logistic stick-breaking process. *Journal of Machine Learning Research* 12 203–239.

BIBLIOGRAFÍA

- RODRIGUEZ, A. & DUNSON, D. B. (2009). Nonparametric Bayesian models through probit stick-breaking processes. Tech. rep., University of California, Santa Cruz, Department of Applied Mathematics and Statistics.
- RODRIGUEZ, A., DUNSON, D. B. & GELFAND, A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association* 103 1131–1154.
- ROTHENBERG, T. J. (1971). Identification in parametric models. *Econometrica: Journal of the Econometric Society* 39 577–591.
- ROYDEN, H. L., FITZPATRICK, P. & HALL, P. (1988). *Real Analysis*. Macmillan New York.
- SAKAI, T. (1996). *Riemannian Geometry*. Fields Institute Communications. American Mathematical Society.
- SALMASO, L. & BROMBIN, C. (2013). *Permutation Tests in Shape Analysis*. SpringerBriefs in Statistics. Springer New York.
- SAN MARTÍN, E. & GONZÁLEZ, J. (2010). Bayesian identifiability: contributions to an inconclusive debate. *Chilean Journal of Statistics* 1 69–91.
- SAN MARTÍN, E. & QUINTANA, F. (2002). Consistency and identifiability revisited. *Brazilian Journal of Probability and Statistics* 16 99–106.
- SCHERVISH, M. (1995). *Theory of Statistics*. Springer Series in Statistics. Springer New York.
- SCHWARTZ, L. (1965). On Bayes procedures. *Probability Theory and Related Fields* 4 10–26.
- SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* 2 639–650.
- SHAKERI, M., LOMBAERT, H., DATTA, A. N., OSER, N., LÉTOURNEAU-GUILLON, L., LAPOINTE, L. V., MARTIN, F., MALFAIT, D., TUCHOLKA, A., LIPPÉ, S. ET AL. (2016). Statistical shape analysis of subcortical structures using spectral matching. *Computerized Medical Imaging and Graphics* 52 58–71.

BIBLIOGRAFÍA

- SHIRYAYEV, A. N. (1992). *On Analytical Methods In Probability Theory*, chap. 9. Dordrecht: Springer Netherlands, 62–108.
- SMALL, C. (1996). *The Statistical Theory of Shape*. Springer Series in Statistics. Springer New York.
- TAO, T. (2011). *An Introduction to Measure Theory*. Graduate Studies in Mathematics. American Mathematical Society.
- TEH, Y. W., JORDAN, M. I., BEAL, M. J. & BLEI, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association* 101 1566–1581.
- THEOBALD, D. L. (2009). A nonisotropic Bayesian approach for superpositioning multiple macromolecules. In *Statistical Tools for Challenges in Bioinformatics, Proceedings of the 28th Leeds Annual Statistical Research (LASR) Workshop*. 55–59.
- THEOBALD, D. L. (2012). *Likelihood and Empirical Bayes superposition of Multiple Macromolecular Structures*, chap. 8. Springer Berlin Heidelberg, 191–208.
- TOKDAR, S. T., ZHU, Y. M. & GHOSH, J. K. (2010). Bayesian density regression with logistic Gaussian process and subspace projection. *Bayesian Analysis* 5 1–26.
- WAINWRIGHT, M. J. & JORDAN, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning* 1 1–305.
- WEIL, A. (1967). *Basic Number Theory*. Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete. Berlin: Springer.
- WOODING, R. A. (1956). The multivariate distribution of complex normal variables. *Biometrika* 43 212–215.
- ZIEZOLD, H. (1977). *On expected figures and a strong law of large numbers for random elements in quasi-metric spaces*, chap. 62. Dordrecht: Springer Netherlands, 591–602.